Babies and the Blackout: A Time Series Analysis of New York City Birth Data

Alan J. Izenman and Sandy L. Zabell, University of Chicago

1. Babies and the Blackout. At exactly 5:27 p.m., November 9, 1965, most of New York City was plunged into darkness because of a massive power failure affecting much of the Northeastern United States. On Wednesday, August 10, 1966, the New York Times carried a front page article with the headline "Births Up 9 Months After the Blackout," which began: "A sharp increase in births has been reported by several large hospitals here, nine months after the 1965 blackout." Above average numbers of births the previous Monday were said to have occurred at Mount Sinai, Bellevue, Columbia-Presbyterian, St. Vincent's, Brookdale and Coney Island hospitals, while births at New York and Brooklyn Jewish Hospitals were reported to be normal. The article added that "there were 16 births at Mount Sinai yesterday [Tuesday], 13 at Columbia Presbyterian and 10 at St. Vincent's, all above average;" this was contrasted with Nassau and Suffolk counties, "many of whose commuters were stranded in the city Nov. 9," and where "the number of births was reported normal," as well as "hospitals in Albany, Rochester, New Haven and Providence," where "the lights went on in midevening."

Next day (Thursday, August 11), a follow-up article appeared (buried on page 35) with the somewhat more cautious lead "Theories Abound on Birth Increase - Possible Link With Blackout Will Not Be Determined for Two More Weeks." By Friday readers were informed that "The birth rate began returning to normal in several leading hospitals here yesterday [Wednesday] following a sharp rise nine months after the 1965 blackout," and the case was closed on Saturday with a short article on page 50 entitled "Birth Rate in City Returns to Normal."

A week later the British magazine New Scientist reported the "Apparent sharp rise in births in New York City" [7]; a year later The Lancet, a respected medical journal, stated unequivocally "the last time New Yorkers demonstrated an unexpectedly vigorous procreative urge they were stimulated... by the stygian darkness of electric-power cuts." [2] At present the story of the 'blackout babies' appears to be an accepted part of American folklore. The episode seems plausible, the story carried by a respected and usually reliable newspaper. But just how good is the evidence for an increase in births nine months after the blackout? Is it really credible that a one-day increase in conceptions would result in a one- or two-day elevation in births 271 days later with virtually no variability or spread? Considerations such as these suggested that, 10 years after the New York Times articles had appeared, an assessment of the published evidence was in order.

2. The Times' Evidence. The first article carried by the Times cited six hospitals as having experienced a sharp increase in births on Monday, August 8. Of these six, Mount Sinai hospital certainly experienced a sharp rise in deliveries (28 compared to a daily average of 11). But one hospital does not a baby boom make. Of the five other hospitals mentioned, four (see Table 2.1) reported increases of no more than four over their daily average, hardly convincing evidence given that two other hospitals are said to have had normal numbers of births and the absence of any information in the article about the variability in these numbers or how the hospitals cited were chosen (there are over 100 hospitals in New York). Finally Bellevue, the last hospital for which data is given, presents somewhat different problems. On Wednesday the Times had reported that "At Bellevue there were 29 new babies in the nursery yesterday, compared with 11 a week ago and an average of 20." This statement is ambiguous as to whether the new babies referred to were born on Monday or Tuesday. Far worse, it is simply wrong. Without remarking on the inconsistency with the Wednesday article, both the Friday and Saturday reports in the Times state the average number of births per day at Bellevue to be 6. Data we present later show, in fact, that there were only 4 deliveries at Bellevue on Monday, 7 on Tuesday. The 'baby boom' has begun to burst.

The three subsequent articles in the Times series describe a pattern of continued increase in births on Tuesday, followed by a decline and return to normal on Wednesday and Thursday. (To facilitate the discussion we shall refer to the four articles in the series as T1, T2, T3, and T4.) The data given in T1-T4 are summarized below in Table 2.1. There are a number of inconsistencies, none serious. (It is interesting to note that St. Vincent's, whose 10 births on Monday were cited as evidence for a "sharp increase in births," is listed in T1 as having 10 births on Tuesday but in T2 as only having an "average" number of births that day.) All in all, the data seem inconclusive and one inclines to adopt the opinion of Dr. Christopher Tietze (quoted in T1), that "I am skeptical until I see data from the entire city. There can be daily fluctuations in individual hospitals that can be misleading."

Such data, giving the number of live births in New York City occurring by day from 1961 to 1966, was obtained by us from the New York City Department of Health. Detailed information about the series is given in Section 4. In Table 2.2 we list a portion of the data, the number of births for each day in August 1966; these numbers are graphed in Figure 2.1. As Figure 2.1 clearly shows, although an increase in births did indeed take place on Monday and Tuesday, August 8 and 9, 1966, similar increases took place on every other Monday and Tuesday of that August! In fact, the fluctuation in births throughout the week from a low on the weekends to a high in the early part of the week is a characteristic feature of the entire series of birth data throughout all six years. (Such weekday-weekend variation is attributed in [9] to a preference for performing elective deliveries on weekdays when the patient is delivered by her personal physician, while [5] opines that it is "probably caused by induced or delayed labor through conscious intent of the mother with or without medical assistance.") Figure 2.1 also shows that births on August 8th and 9th were not appreciably different from those on any other Monday and Tuesday in August. In fact, as seen in Table 2.2, births on those two days were, if anything, slightly lower than usual: 449 births on August 8 (compared to 452, 453, 470 and 451 births on other Mondays in August) and 440 on August 9 (compared to 470, 499, 519 and 468 births on other Tuesdays in August). The 'baby boom' has vanished.

3. A Review of the Literature. Despite such (to us) unequivocal evidence against a one- or two-day surge in New York City's birth rate nine months after the blackout, an article has appeared in the professional literature claiming precisely such an effect. In 1968, Professor L. B. Borst reported in the American Journal of Obstetrics and Gynecology that "daily birth records in New York City disclose a 30% increase in live births at five Manhattan hospitals on August 7 [(!!)], 1966, 270 days after the blackout of Nov. 9-10, 1965." [4] Noting that while power had not been "restored until the following day in Manhattan and parts of the Bronx, whereas in Brooklyn and Queens power was restored at various times during the evening and, in Richmond, almost immediately," Borst reasoned that computing the ratio of Manhattan births to total New York City births would simultaneously correct for the weekday-weekend effect discussed above and detect a blackout effect on the birthrate in the form of a percentage increase in the number of NYC births occurring in Manhattan. Using statistics for the number of live births in five (unspecified) Manhattan hospitals from August 1 to 13 and dividing the sum of these by total NYC births, Borst observed a distinct peak on August 7 which he concluded was a "very special day" (the percentage for August 7 differing from the mean percentage excluding August 7 by 7 average deviations from the mean).

Professor Borst omits from his article two pieces of information necessary to assess the validity of his conclusions. On the one hand, there is the disturbing issue of data selection: no mention is made of how the five hospitals studied were chosen. On the other, although the aggregate percentage of NYC births which occurred in the five hospitals under study can be approximately read off from a bar graph, the raw data for the individual hospitals is not given. Upon request, Professor Borst provided us with a copy of his data which is given in Table 3.1. (Note that data for Mt. Sinai was collected but not used by Borst in his article.)

Several interesting points emerge from inspection of Table 3.1. First, as mentioned earlier, the data for Bellevue Hospital show that births there were not unusually high on August 8-9, 1966, and in any case, were not as high as 29 on either day. Second, the total births in the five hospitals studied by Professor Borst (last row of Table 3.1) do not display noticeable nonrandom variation throughout the 13-day period for which statistics are provided. Certainly nothing exceptional appears to have happened on Sunday, August 7. The effect reported is entirely due to the seemingly innocent "normalization" of dividing these totals by total NYC births (which decrease on Sundays). If the daily trend in the five hospitals under study were the same as that for New York as a whole, this would seem a reasonable procedure. If, however, the trend in these five hospitals differs from that of the city as a whole, then the computed birth ratio of the two will exhibit variations unrelated to hypothesized blackout effects. We suggest that this is the case here. If the weekday-weekend variation exhibited in total NYC births is due to induction and/or delay of labor at some hospitals to avoid weekend deliveries, scheduling of elective deliveries primarily on weekdays, etc., this would be an effect more likely to occur in private hospitals where patients are frequently delivered by their own personal physician or a specialist than in large municipal hospitals with a large charity caseload and interns on duty at fixed hours. Indeed, such a difference has been reported by Menaker and Menaker [9], who state that "considerably less variation occurred in this regard at the municipal hospitals as compared with the "private" hospitals, which showed a weekend decline, most marked on Sunday." Three of the five hospitals used by Professor Borst fall into the former category (Bellevue, Harlem, and Metropolitan); the other two (Sloan and New York) are "private voluntary" (as opposed to "proprietary"). All but New York handle a large volume of so-called "service" cases. (It is perhaps not insignificant that Mount Sinai, the one hospital not used, alone displays a sharp increase in births on Monday.) Taking a ratio with a roughly stable numerator and a denominator which is minimized on Sunday, Professor Borst has observed a percentage increase in births which is an artifact of his methodology.

If the above explanation is correct, we should expect to see similar peaks in this birth ratio the Sundays before and after August 7. Unfortunately it is not possible to check this from Professor Borst's data as his statistics range only from the Monday before until the Saturday after August 7. However, in 1970 Dr. Walter Menaker [8] obtained statistics allowing him to compute the ratio of total Manhattan births to NYC births for the three Sundays in question; the results-- 98/356 (or 27.5%) on July 31, 97/344 (or 28.2%) on August 7 and 110/377 (or 29.2%) on August 14, --show that August 7 was in no way exceptional.

While a 1 or 2 day effect on births seems clearly ruled out, it is still possible that an effect on the birth rate took place over a longer period of time. Indeed, going on to note that 800,000 people were caught in the subways during the blackout and citing newspaper headlines such as "30% of Labor Force Too Weary To Work," Dr. Menaker felt it far more likely that the blackout would depress rather than increase the city's birthrate. Looking at births one week before and one week after August 9, Menaker noted that total births for this period were lower than the combined total for the week immediately prior and the week immediately following.

In Figure 3.1 we have graphed a smoothed version of the NYC birth data for 1961-1966 (see Figure 4.1); details of the method of smoothing appear in Section 4. Notice the regular seasonal pattern, namely two peaks, the first of which is smaller in magnitude and also of shorter duration than the second; the second peak occurs during the summer and is typically bifurcated with a single dip whose extent varies from year to year. (Such seasonal birth patterns for a number of countries have been extensively studied in [10].) In 1966, however, the summer peak contains two distinct dips, the first (indicated by an arrow in Figure 3.1) occurring in late July-early August and corresponding to the decrease noted by Menaker. The yearly variations in the summer peak make it impossible to conclude from simple inspection of Figure 3.1 whether or not this decrease in births is "significant". In any case the effect is quite small (a decline of at most several hundred births during a one-month interval in which over 12,000 births occurred). Indeed, Dr. Menaker himself concluded that "the evidence presented here for a decrease in conceptions during the Blackout cannot be considered direct or conclusive. 'Statistical significance' would have little or no meaning here. It should be emphasized that those who have postulated an increase in conceptions during the Blackout have failed to produce satisfactory evidence for such an increase. The evidence presented here suggests a decrease."

An attempt to give "statistical significance" to such aggregate birth statistics was later undertaken by Professor J. Richard Udry of the School of Public Health at the University of North Carolina, Chapel Hill [12]. Udry reasoned that "if there were an unusual number of conceptions on November 10th, then the period between June 27 and August 14, 1966, would contain a greater percentage of the year's births than that contained by the same period in other years." Udry's calculations (which we have confirmed) are given in Table 3.2. The results appear to support Udry's conclusion that "1966 is not an unusual year...we therefore cannot conclude from the data presented here that the great blackout of 1965 produced any significant increase (or decrease) in the number of conceptions."

Professor Udry's article, however, contains several "loose ends." Little attempt is made to contrast the seasonal pattern for 1968 with that of previous years nor is there any mention of the downward trend in NYC births that Figure 4.1 exhibits. (The existence of this trend makes the comparison of yearly percentages such as those in Table 3.2 somewhat dubious.) More troubling is the lack of attention to considerations of statistical power. A simple order of magnitude calculation will make the problem clear. Assume that on the night of the Blackout the incidence of intercouse in NYC rose 25%. If such an increase resulted in a corresponding increase in conceptions, approximately 110 extra births would occur nine months later, spread over a two month interval. (There were approximately 446 births per day during the 1961-1966 period.) Professor Udry's test attempts to detect this increase of 110 during a seven-week interval in which 21,290 births occurred. In terms of the percentages given in Table 3.2, an increase of 0.06% is in question, although the percentages involved are only calculated to the nearest tenth! If a (still sizeable) increase in conception of 10% occurred, the possibility of detection is even worse. At a very minimum, a power calculation to determine an optimal test interval is in order.

This last point highlights the real fallacy of a 'baby boom'. Even if a sizeable increase in intercourse took place on the night of the Blackout, the intervention of natural and human agencies would result in few additional conceptions. (E.g. contrast the average of 446 births per day with any reasonable estimate of the number of acts of intercourse taking place in New York City on any given night.) These additional births, at most several hundred in number, would occur over an eight week period nine months later. Engulfed in a sea of variability resulting from long-term trends, seasonal effects, weekend-weekday effects and random fluctuation, even the most sophisticated of statistical techniques will be hard put to detect any effect actually present.

4. Analysis of the Birth Data. As mentioned in Section 2, the data obtained from the NYC Department of Health consists of the total number of births per day in NYC over the 6 year period, 1961-1966, a total of 2191 days of birth data. These daily birth totals are graphed in Figure 4.1. They range from a low of 303 to a high of 563, with a mean and standard-deviation of 446.74 and 40.84 respectively. There is a clear decline in births for the last three years, 1964-6, whereas for the first three years the overall level is relatively stable. In addition, the series exhibits a regular seasonal pattern.

The spectrum of the data was estimated by first employing a cosine taper extending over the first and last 10% of the data (see Section 5.2 of [3]), adding a sufficient number of zeroes to the end of the tapered data to make the total length 4096, computing the raw periodogram (using the Fast Fourier Transform), and then smoothing the result by taking moving averages of 7 adjacent periodogram values. This is graphed (log-scaled) in Figure 4.2(a); an even smoother version obtained by successively applying moving averages of 15, 31 and 63 is graphed in Figure 4.2(b). Features of these graphs include a trend-seasonality component at low frequencies and also two large peaks, the second of which is clearly a harmonic of the first, which in turn occurs at a frequency of 0.143, or a 7-day cycle. This 7-day cycle is the "day-of-the-week" effect noted by Menaker [8] and suggests that there are significant differences in the number of births between different days of the week.

The data was first smoothed to provide an estimate of the sum of the trend and seasonal components. This was done by employing a suitable low-pass filter (see Section 6.4 in [3]) with cutoff frequency $\omega_c = 0.1122$, corresponding to picking up a 28-day cycle. This smoothed version of the data is graphed in Figure 3.1. The resulting residuals after smoothing are graphed in Figure 4.3; these range from a low of -127 to a high of 111, and have a mean and standard-deviation of zero and 35.10 respectively.

Next, averages for each of the seven days of the week over the length of the data set (which is now the residuals from smoothing the data) were computed; a graph of them is given in Figure 4.4. Notice the relatively high level of 5 consecutive week-days compared with the 2 weekend-days. These 7 values were successively subtracted from the series of residuals to give a new series ("filtered-deweeked" residuals). These are graphed in Figure 4.5. They range from a low of -106 to a high of 95, and have a mean and standard-deviation of zero and 25.51 respectively.

The spectrum of these second-stage residuals was then estimated as before. The trend-seasonality component has been removed, leaving an expected "dip" in the spectrum at the low frequencies and the two spikes of Figure 4.2 have now been reduced to the general level of spectrum values. A full-normal probability plot of these ordered residuals is graphed in Figure 4.6. The bulk of the points fall on a straight-line through the origin, while the lower-tail of the data is somewhat heavier (fatter) than that of the normal indicating an excess of large negative residuals. A graph of the residuals for the period in question, namely May-September 1966, is given in Figure 4.7; clearly nothing unusual is happening.

This last set of residuals was finally analyzed for outliers. This was done by setting up bands at the yearly mean plus-and minus two standard-deviations, the standard-deviation being estimated for each year separately; hence, residuals were termed "extreme" relative only to other residuals of the same year. Table 4.1 lists the estimated standard-deviations for each of the six years, together with their extreme-residual day numbers (a minus sign next to the number indicates a negative residual). Five days that appear more-or-less regularly in the table are the fixed holidays of New Year (day 1), Memorial Day (day 150/151), and Christmas Day (day 359/360), and the variable holidays of Labor Day (days 247, 246, 245, 251, 248, 247) and Thanksgiving (days 327, 326, 332, 331, 329, 328). In several instances, the day preceding or succeeding a holiday will also occur in the table. Independence day (day 185/186) is a curious omission from this list (it only appears once). Since these days may be regarded as somewhat "special", there is a valid case for omitting them in any further consideration of the set of residuals; a revised full-normal probability plot of the residuals (leaving out the 28 holiday-days) looked much better, the peculiarities in the left-hand tail having been removed. This effect has also been noted in [5].

Looking at the data now after estimating a trend-seasonality component, a day-of-the-week component, and computing residuals, we arrive at the following conclusions.

(a) There was no increase in births due to the Blackout;

(b) If there was any change in the general level of births, it was both small in magnitude and in a downward direction.

A more careful and detailed analysis of this data set will appear elsewhere. However, preliminary studies using simple ARIMA models (see [6]) indicate that the decline mentioned above in (b) is not significant.

5. Discussion. The episode of the missing baby boom highlights a number of issues only partly statistical in nature. The first of these concerns journalistic responsibility. Despite the paucity of evidence adduced in the New York Times articles, there is at present a widespread belief in a baby boom 9 months after the Blackout, an impressive testament to the power of the press as an opinion-maker. Although in this case the issue is hardly one of national importance, it does highlight the problems both journalists and the public face as issues of increasing statistical complexity become common in public affairs.

A second issue is the statistical refereeing of articles appearing in nonstatistical scientific journals. Most of the objections to the published literature that we have raised are of an elementary nature that surely would have been picked up had adequate refereeing taken place. Its absence is particularly hard to understand given an already unmanageable flood of scientific literature and the inability of many non-statisticians to independently evaluate evidence of a statistical nature.

Finally, there is an issue of data availability. A number of sources were contacted for data related to our study. What was originally thought of as a simple request for data ended up lasting 4 months and requiring 5 letters, 2 phone-calls, and a personal visit! In our case we were fortunate--all persons contacted were quite cooperative and the major difficulty involved was the problem of recovering records often 5 to 7 years old. Others have been less fortunate. As illustrated by our study, it is essential that data analysed in the scientific literature be given in full (if possible) or at least be readily available for scrutiny.

Acknowledgements. We should like to thank Frieda Nelson of the New York City Department of Health for providing us with tabulations of births by day for the 6 years of the study, and also Earl Westfall for carrying out the computations referred to in Section 4.

This research was carried out in the Dept. of Statistics, University of Chicago, under partial support of NSF Research Grants No. MCS72-04364 A04 and SOC72-05228 A04 and by U.S. Energy Research and Development Administration Contract No. EY-76-S-02-2751.*000. By acceptance of this article, the publisher and/or recipient acknowledges the U.S. Government's right to retain a non-exclusive, royalty-free license in and to any copyright covering this paper.

References

[1] Anon. (1966). "Birth Rate in City Returns to Normal," New York Times, August 13, p. 50.

[2] Anon. (1967). "Tides and Babies," Lancet (Annotations) 2, 505.

[3] Bloomfield, P. (1976). Fourier Analysis of Time Series: An Introduction. J. Wiley and Sons, New York.

[4] Borst, L. B. (1968). "Natality and the Blackout," Amer. J. Obstet. Gynec. 101, 422-423.

[5] Borst, L. B. and Osley, M. (1975). "Holiday Effects upon Natality," Amer. J. Obstet. Gynec. 122, 902-903.

[6] Box, G. E. P. and Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control. Revised Edition. Holden-Day, San Francisco.

[7] Low, I. (1966). "Blackout Babies and Breeding Seasons," New Scientist 31, 378-379.

[8] Menaker, W. (1970). "Evidence for a Drop in Conceptions in the Big Blackout." Amer. J. Obstet. Gynec. 107, 804-806.

[9] Menaker, W. and Menaker, A. (1959), "Lunar Periodicity in Human Reproduction: A Likely Unit of Biological Time," Amer. J. Obstet. Gynec. 77, 905-914.

[10] Rosenberg, H. M. (1966) "Seasonal Variation of Births, United States, 1933-1963", Vital and Health Statistics, National Center for Health Statistics, PHS Publ. No. 1000, Series 21, No. 9. Public Health Service, Washington.

[11] Tolchin, M. (1966). (1) "Births Up 9 Months After the Blackout," New York Times, August 10, p. 1; (2) "Theories Abound on Birth Increase," NYT, August 11, p. 35; (3) "Hospitals Report Birth Rates Gradually Returning to Normal," NYT, August 12, p. 33.

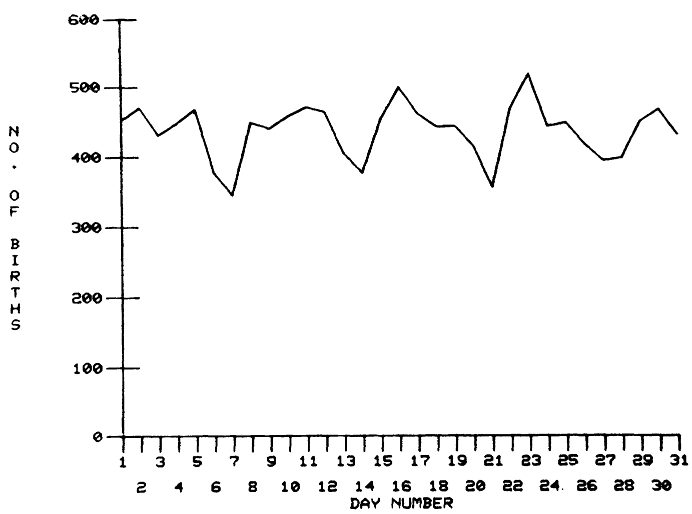
[12] Udry, J. R. (1970). "The Effect of the Great Blackout of 1965 on Births in New York City," Demography 7, 325-327.

	Average	(Aug. 8) M	(Aug. 9) Tu	(Aug. 10) W	(Aug. 11) Th
BELLEVUE	20[a] or 6[c,d]	29	slightly above average	1	2
BRONX MUNICIPAL	7	16	16	9	8
BROOKDALE	10	13[a] or 15[b,c]	15	14	13
BROOKLYN JEWISH	15	normal	slightly above average	18	8
COLUMBIA-PRESBYTERIAN	11[a] or 12[d]	15	13	average	18
CONEY ISLAND	4[b] or 5[a]	8	7	average	average
FRENCH	3		5	average	10
MOUNT SINAI	11	28	16	17	15
NEW YORK	13	normal	slightly above average	average	5
ST. LUKE'S	5	14-15	14-15	9	7
ST. VINCENT'S	7	10	10; average	average	average

Table 2.1. Daily birth data as reported in four New York Times articles.

a. In T1
b. In T2
c. In T3
d. In T4

Figure 2.1
TOTAL BIRTHS IN NYC: AUGUST, 1966



August	1	Mon	452		17	Wed	461
	2	Tues	470		18	Thurs	442
	3	Wed	431		19	Fri	444
	4	Thurs	448		20	Sat	415
	5	Fri	467		21	Sun	356
	6	Sat	377		22	Mon	470
	7	Sun	344		23	Tues	519
	8	Mon	449		24	Wed	443
	9	Tues	440		25	Thurs	449
	10	Wed	457		26	Fri	418
	11	Thurs	471		27	Sat	394
	12	Fri	463		28	Sun	399
	13	Sat	405		29	Mon	451
	14	Sun	377		30	Tues	468
	15	Mon	453		31	Wed	432
	16	Tues	499				

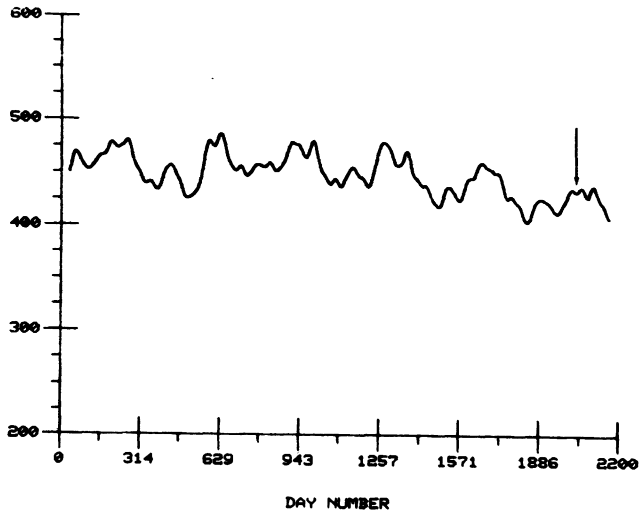
Table 2.2. Total live births occurring in August, 1966, for New York City.

August	1	2	3	4	5	6	7	8	9	10	11	12	13
Bellevue	2	1	5	7	7	6	10	4	7	2	2	1	2
Harlem	13	6	7	11	6	8	3	5	3	8	7	10	9
Metropolitan	11	14	5	10	11	6	6	8	7	10	8	9	6
Mt. Sinai	6	14	14	11	17	12	9	28	16	20	16	20	11
New York	8	11	6	13	10	10	11	12	10	13	6	11	16
Sloan	10	14	14	8	11	5	13	15	13	12	18	11	12
Total*	44	46	37	49	45	35	43	44	40	45	49	42	45

Table 3.1. Daily birth data for 6 individual hospitals in New York City, August 1966. (We should like to thank Professor L. B. Borst for providing us with these numbers, which originate with the New York Department of Health.)

*omitting Mt. Sinai births.

Figure 3.1
NYC BIRTHS (FILTERED: OMEGA=.1122,S=30) 1961-66



Year	1961	1962	1963	1964	1965	1966
% of year's total births	13.9	13.9	13.9	13.9	14.1	13.9

Table 3.2. Percentage of year's total births occurring in New York City, June 29-August 16 (1964: June 28-August 15), during 1961-1966 (Table 1 in [12]).

Figure 4.1
NYC BIRTHS -- 1961-1966

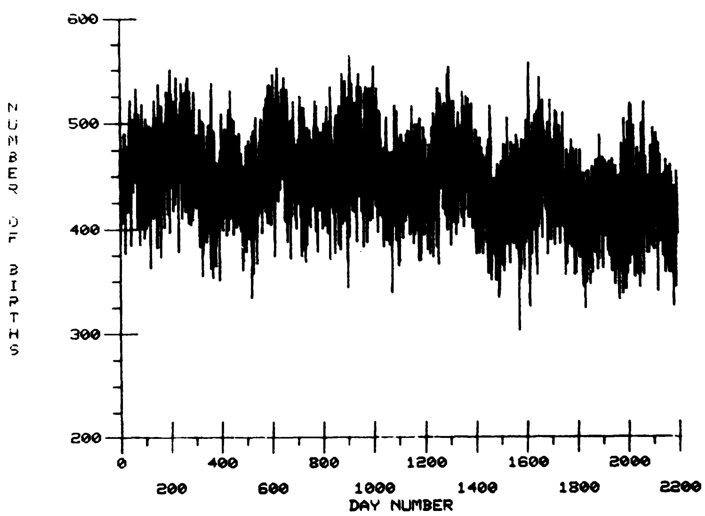


Figure 4.2a
LOG SPECTRUM: RAW;MA 7

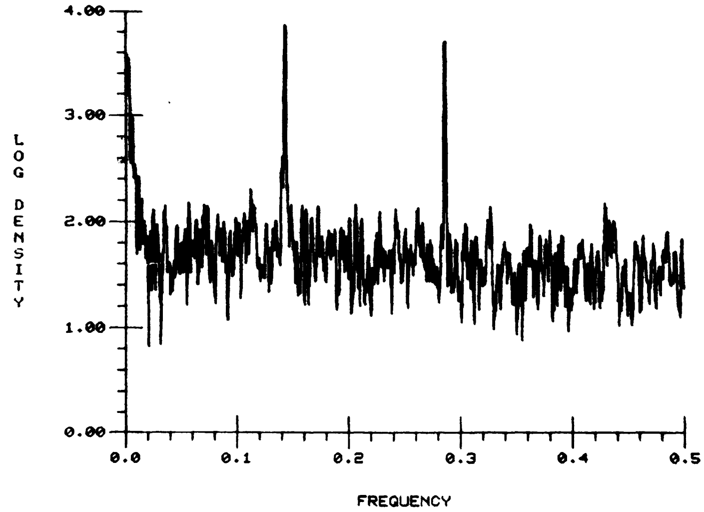


Figure 4.2b
LOG SPECTRUM: RAW;MA7,15,31,63

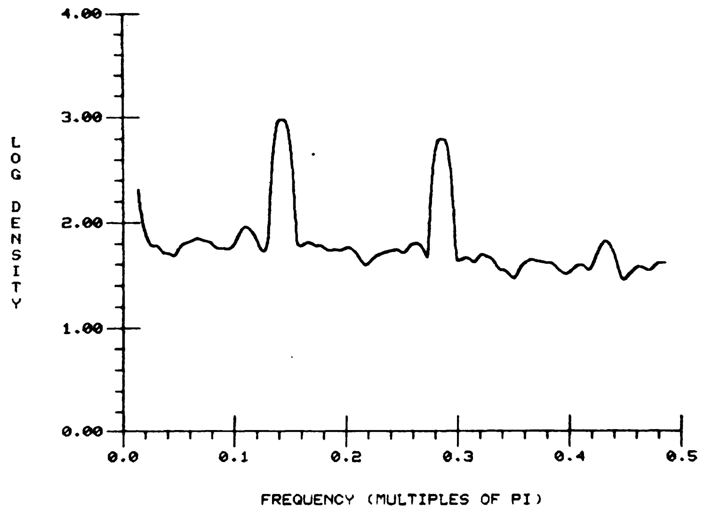


Figure 4.3
RES: FILTERED

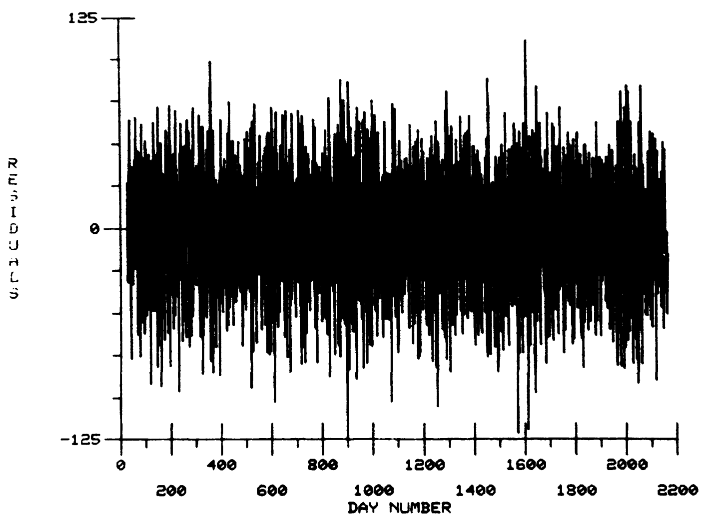


Figure 4.6
NORM P PLOT: RES(FILT,DWK)

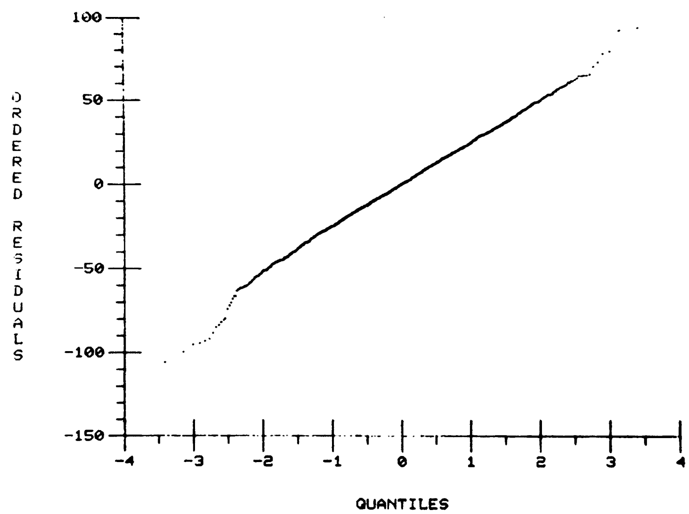
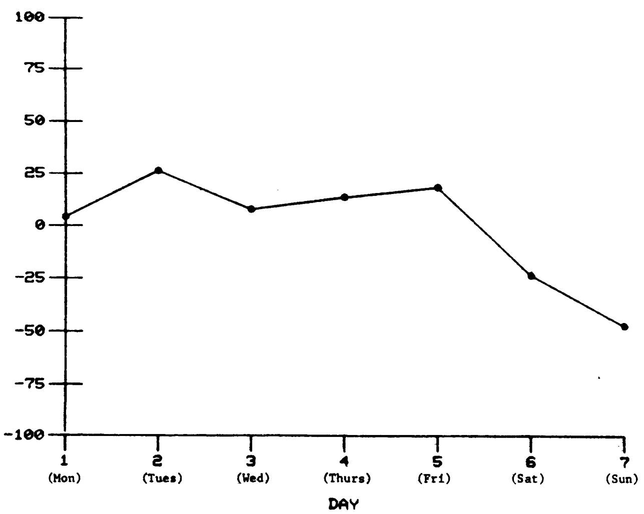


Figure 4.4
DAY-OF-WEEK EFFECT



	1961	1962	1963	1964	1965	1966
1.	31^{-}	1^{-}	1^{-}	1^{-}	1^{-}	1^{-}
2.	84	94^{-}	101	29^{-}	7^{-}	146
3.	111^{-}	122^{-}	113^{-}	77^{-}	30^{-}	150^{-}
4.	150^{-}	162	148	88^{-}	108^{-}	151^{-}
5.	154	171	150^{-}	130	137	158
6.	196	176^{-}	158	145	148	174^{-}
7.	247^{-}	245	167^{-}	159^{-}	150^{-}	181
8.	292	246^{-}	177	203	151^{-}	185^{-}
9.	327^{-}	300	180	252^{-}	168^{-}	187
10.	359^{-}	315	186^{-}	289	190	195
11.	362	319	245^{-}	359^{-}	205	235
12.	363	326^{-}	249^{-}	360^{-}	251	248^{-}
13.		359^{-}	269	364	281	255^{-}
14.		361	316^{-}	365	329^{-}	256^{-}
15.			332^{-}			283
16.			342^{-}			303
17.			347^{-}			328^{-}
18.			352			332^{-}
19.			359^{-}			
s_i	27.07	25.28	27.34	23.17	25.59	24.46,

Table 4.1. Residuals after filtering and removing "day-of-week" effec having |residual value| > $2s_i$, where i denotes year. Entry is day number.

Figure 4.5
RES: FILT,DWK

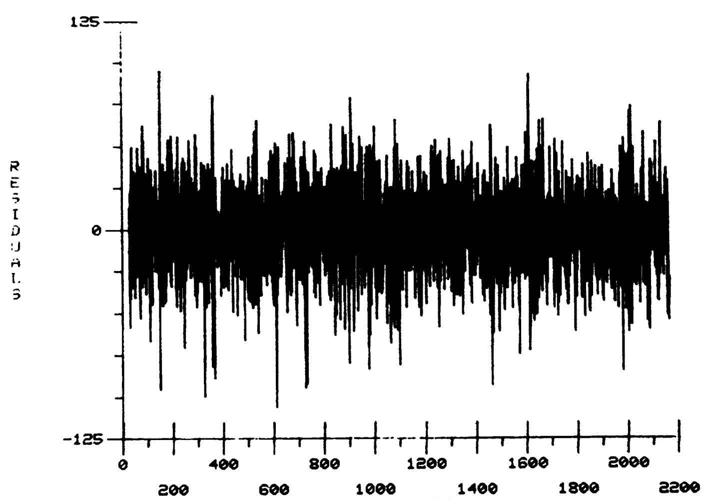
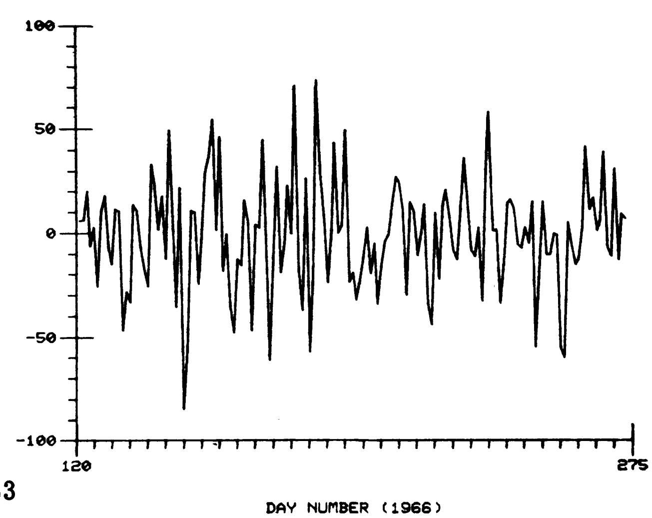


Figure 4.7
RES: FILT,DWK (MAY 1 - SEPT 30, 1966)



EDUCATIONAL ATTAINMENT AND THE LABOR FORCE PARTICIPATION OF WHITE WOMEN IN THE UNITED STATES, 1930-1970

A. J. Jaffe, Columbia University

Jeanne Clare Ridley, Georgetown University

A number of studies have documented the large increases that have occurred in the educational attainment and labor force participation rates of American women in the 20th century, Cf. (Bancroft, 1958, Durand, 1948, Folger and Nam, 1967, Jaffe and Steward, 1951, Oppenheimer, 1970, Ridley, 1971, Sweet, 1968). Also, well documented is the fact that the more education a woman has the greater her propensity to work (Bowen and Finegan, 1969, Cain, 1966). Indeed, one common explanation for the increased labor force participation of American women in recent decades has been the improvements in educational attainment women have experienced.

This paper investigates the question of how much of the observed increase in labor force participation of American women can be attributed to their increased educational attainment. Our analysis of this question has been confined to the 40 year period between 1930 and 1970 and to only white women.

In the first part of the paper we attempt to measure the direct influence of increased educational attainment on labor force participation. We then turn to a more detailed consideration of the changes that have occurred in the labor force participation rates of the various educational groups. Finally, as a possible explanation for our findings regarding educational attainment and labor force participation, we examine changes in fertility. Obviously, many factors have in the past and are at present influencing the extent to which American women participate in the labor force. We are not attempting to explore or measure all of these factors. Rather we are concerned with the interrelationships of educational attainment and labor force participation.

THE DATA

For our analysis we have drawn on data from the decennial censuses of 1930 to 1970. Our analysis is restricted to this 40 year period for a number of reasons. First, census data on educational attainment and labor force participation are available only from 1970 onward. Prior to 1940, the decennial census did not ask a question on educational attainment. Also, in 1940 the labor force concept replaced the gainful worker concept. We could, however, derive for 1930 educational attainment and labor force participation rates and thus obtain labor force participation rates by level of education in 1930. We obtained the educational level of white women by "younging" the 1940 population. We then adjusted the rates in accordance with the known totals in the labor force in 1930 which we estimated for each age group (Deming 1943). To obtain the labor force numbers in 1930 the gainfully occupied numbers were converted to agree as closely as possible to the 1940 labor force definition (Edwards, 1943). We are aware that the data on educational attainment and labor force participation are not completely comparable from census to census. Since, however, we are concerned with overall trends we did not attempt to correct for what appeared to be slight incomparabilities.

Note should be made that an important change occurred in the 1970 census data on children ever born. Prior to 1970 the question of children ever born was asked only of women who had ever been married; in 1970 this question was asked of all women. This change added somewhere between 10 to 20 children per 100 women (U. S. Bureau of the Census, 1970). Such a change we judged did not markedly affect the fertility trends discussed in this paper and therefore we did not adjust for this change.

The analysis reported here is restricted to white women for the simple reason that white and black women differ greatly in their labor force behavior. We hope to extend our analysis to black women in the near future.

RESULTS

The Influence of Education

Table 1 presents the results of standardizing the labor force participation rates on the educational distributions of 1930 and 1970. By standardizing we are attempting to answer the question: what would have been the age specific labor force participation rates if the educational level of women had remained unchanged over the entire 40 year period? Since there is no uniquely correct educational distribution to use as a standard we used both dates. This provides a range within which the influence of education can be said to lie. In the upper panel of Table 1, the labor force participation rates standardized on the 1970 educational distribution are shown and in the lower panel the rates standardized on the 1930 distribution are shown. The last three columns in Table 1 show: the observed change in the labor force participation rates, column (a); the absolute change in the standardized rates, column (b); the change attributable to increased educational attainment, column (c). Thus, the total change in labor force rates has been subdivided into two components, that due to other factors and that due to increased educational attainment.

Table 1. Standardized Labor Force Participation Rates for White Women by Age, United States: 1930 to 1970

Age	1930	1940	1950	1960	1970	Change 1930 to 1970 (a)	(b)	(c)
	Standardized on 1970 distribution of educational attainment							
20 to 24 years	50.0	55.0	47.7	48.1	56.2	14.5	6.2	8.3
25 to 29 years	34.0	40.2	34.1	34.9	44.1	16.1	10.1	6.0
30 to 34 years	24.7	33.2	30.4	33.9	42.6	21.3	17.9	3.4
35 to 39 years	23.2	30.3	34.6	39.2	47.6	28.0	24.4	3.6
40 to 44 years	21.9	27.4	35.7	44.1	50.9	32.2	29.0	3.2
45 to 49 years	20.9	25.5	35.7	47.4	52.7	35.0	31.8	3.2
50 to 54 years	19.9	25.3	33.8	47.3	52.0	35.2	32.1	3.1
	Standardized on 1930 distribution of educational attainment							
20 to 24 years	41.7	42.5	38.7	39.8	45.8	14.5	4.1	10.4
25 to 29 years	28.0	32.9	29.2	31.6	39.2	16.1	11.2	4.9
30 to 34 years	21.3	27.7	28.1	33.0	40.6	21.3	19.3	2.0
35 to 39 years	19.7	25.3	31.9	37.8	45.6	28.0	25.9	2.1
40 to 44 years	18.7	23.1	32.2	41.6	48.5	32.2	29.8	2.4
45 to 49 years	17.7	21.5	31.2	43.7	49.6	35.0	31.9	3.1
50 to 54 years	16.7	22.3	29.2	42.8	48.1	35.2	31.4	3.8

(a) Absolute change in unstandardized labor force rates.

(b) Absolute change in standardized rates; i. e., change due to factors other than changes in educational attainment.

(c) Absolute change resulting from increased educational attainment.

The observed change in the total rates for each age group shown in column (a) is simply the observed rates for 1930 subtracted from those for 1970. The observed rates for these two years and the intervening years are shown in Table 2. The influence of all factors except increased education (column (b)) is the difference between the standardized rates for the two dates. Since the influence of education has been held constant by standardization, then any residual must be due to other factors. Finally, the changes in the observed rate (column (a) minus the changes in the standardized rates (column (b) represent the changes due to increased education.

An inspection of column (c) clearly indicates that the direct influence of increased educational attainment has had but little effect on the labor force participation rates of white women. This is true regardless of whether the labor force rates are standardized on the 1930 or 1970 educational distribution. For instance, although at ages 20 to 24 the increase in educational attainment may have accounted for over a half to almost three quarters of the observed increase in the labor force rates, and at ages 25 to 29 almost one third, at ages 30 to 54 the influence of education accounts for less than 20 percent and in most instances less than 10 percent. As can be seen, however, these are precisely the ages at which the largest observed increases in labor force participation occurred.

We also divided the 40 year period into two periods and carried out the same calculations for the period between 1930 and 1950 and between 1950 and 1970. The same pattern of changes were obtained with education having less influence than other factors. Because of limitations of space, these results are not shown.

Labor Force Rates By Level of Education

Table 2 presents the observed labor force participation rates by educational attainment. The most rapid increases in labor force participation between 1930 and 1970 occurred among women with less than a high school education. Only among 20 to 24 year old women did the rate for college graduates (16+ years of education) increase slightly more than those with lower levels of educational attainment. At all older ages, 25 and above, those not having finished high school (under 12 years of education) had the greatest increases. Moreover, with increasing age even larger increases in labor force participation rates are observed for women at the lower educational levels. For example, at ages 25 to 29 among women with less than a high school education, the percentage increases in labor force participation was 51, the comparable percentage increase for women 50 to 54 was 214. For college graduates, the increase was 16 percent for those 25 to 29 and 69 percent for those 50 to 54.

How may we account for the greater increases in labor force participation rates among women having the lower educational attainment? One obvious answer is purely statistical. Increases must be asymptotic; as the labor force rate approaches 100 percent further increases are not possible. An alternative way of measuring the changes between 1930 and 1970 is by expressing the increase in the participation rates as a percent of those not in the labor force in 1930. The results of this approach produced a pattern very similar to that previously observed. To summarize: at ages 20 to 24, the increase was greatest among those with the lowest level of education; at ages 25 to 29, there was little difference by level of education; at ages 30 to 34 up through ages 40 to 44, the increases were larger among those at lower levels of education; at ages 45 to 49 and 50 to 54, there was little difference by level of education.

Educational Attainment and Fertility

The changes in labor force participation rates by educational level tended to be inversely related to the changes that occurred in fertility by education. As may be seen in the following table, the fertility of women with less than a high school education increased less rapidly than the fertility of women who finished high school or went on to college.

Children Ever Born Per White Woman

Age and Educational Attainment	1940	1970
20 to 24 years	.48	.68
Under 12 years	.73	1.39
12 years	.28	.71
13 to 15 years	.12	.26
16+ years	.07	.17
25 to 29 years	1.09	1.74
Under 12 years	1.43	2.38
12 years	.76	1.77
13 to 15 years	.61	1.39
16+ years	.31	.78
30 to 34 years	1.64	2.60
Under 12 years	2.03	3.00
12 years	1.16	2.54
13 to 15 years	1.05	2.32
16+ years	.70	1.75
35 to 39 years	2.11	2.98
Under 12 years	2.50	3.14
12 years	1.46	2.87
13 to 15 years	1.37	2.74
16+ years	.92	2.27
40 to 44 years	2.42	2.93
Under 12 years	2.80	3.03
12 years	1.67	2.77
13 to 15 years	1.60	2.77
16+ years	1.07	2.42

Table 2. Labor Force Participation Rates for White Women by Age and Educational Attainment, United States: 1930 to 1970

Age and Educational Attainment	1930	1940	1950	1960	1970
20 to 24 years	41.7	45.5	43.4	44.8	56.2
Under 12 years	34.5	35.1	31.2	33.2	37.1
12 years	55.4	56.8	51.1	48.7	59.3
13 to 15 years	47.1	47.7	46.7	49.4	56.6
16+ years	63.8	64.8	67.5	71.8	79.6
25 to 29 years	28.0	34.1	31.3	33.3	44.1
Under 12 years	23.4	27.1	25.4	29.1	35.4
12 years	31.9	39.8	33.4	32.9	42.2
13 to 15 years	39.1	44.6	36.7	37.3	46.9
16+ years	53.4	58.9	48.3	48.5	61.7
30 to 34 years	21.3	28.9	29.2	33.5	42.6
Under 12 years	18.7	23.7	26.3	32.3	39.3
12 years	24.0	34.2	29.9	33.5	42.8
13 to 15 years	26.7	35.5	33.1	33.9	41.5
16+ years	40.4	50.4	40.1	39.6	51.1
35 to 39 years	19.7	26.2	33.0	38.9	47.7
Under 12 years	17.3	21.8	30.0	36.9	44.3
12 years	23.3	31.7	34.5	39.1	48.4
13 to 15 years	25.2	33.1	36.7	39.7	46.2
16+ years	39.5	48.6	47.0	46.4	56.3
40 to 44 years	18.7	23.7	33.7	43.2	50.9
Under 12 years	16.5	20.2	29.8	40.0	47.0
12 years	22.6	28.6	36.7	44.9	52.6
13 to 15 years	24.4	31.3	39.6	46.0	50.8
16+ years	38.8	46.9	51.1	55.0	58.4
45 to 49 years	17.7	22.0	33.0	46.0	52.7
Under 12 years	15.5	18.8	28.2	41.3	47.5
12 years	21.8	26.7	38.0	49.1	55.4
13 to 15 years	24.8	31.0	42.0	51.8	55.1
16+ years	39.1	46.5	54.7	63.9	64.5
50 to 54 years	16.7	20.6	30.6	45.6	51.9
Under 12 years	14.5	20.1	26.0	39.8	45.6
12 years	20.9	25.7	36.4	50.0	55.4
13 to 15 years	26.1	31.6	42.2	54.3	56.7
16+ years	40.0	46.4	55.9	69.3	67.6

Thus, at ages 20 to 24 the increase in fertility was 90 percent for those with less than 4 years of high school education while for college graduates the percentage increase was 143 percent. At ages 35 to 39 the increases were 26 percent and 147 percent. A result of these differential increases in fertility has been the convergence of family size among the various educational groups just as there has been a convergence in the labor force participation rates among the various educational groups.

Although neither the changes in the educational attainment or in fertility behavior of white women can be interpreted as the "cause" of the changes in the other, these two changes occurred simultaneously. Since the fertility of women with less than a high school education increased less rapidly than the fertility of women who finished high school or went on to college, it is not surprising that women with less than a high school education entered the labor force in greater numbers. The changes in fertility in fact, parallel the pattern by educational level of those entering the labor force since 1930 noted above. Generally, while fertility was once an important factor in whether or not women were in the labor force, it now appears that childbearing no longer is as an important factor as it once was.

DISCUSSION

Many factors have influenced the labor force participation of white women in the United States over the last 40 years. Increased education is but one. Other factors include the large increase in urbanization, the almost complete disappearance of the rural population, and the increasing availability of jobs for women through the growth in the "paper" work industry. This latter factor, of course, has contributed to the very large increases in the white collar clerical occupations. Additional factors influencing the increased participation of women in the labor force have been the rapid dimunition of the foreign born population, the increasing income of large proportions of the changing roles of men and women and of course the feminist movement.

We have concluded that increased educational attainment in and of itself has not had much direct effect on the labor force participation of women. We suspect, however, that it has been an integral part of the overall matrix of interrelated factors. Possibly without the increased educational attainment of women some of the other factors would have had a lesser impact upon the labor force activity of women.

In a 1956 article (Jaffe, 1956) stated: ". . .if women complete their family formation by the late twenties, then they can enter (or reenter) the working force in the early thirties at which age they are more acceptable to employers. Thus, even without any decrease in the ultimate size of the fmaily, a lowering of the age at which family formation is completed may result in an increase in the proportion of women who leave the home for outside jobs." We suspect that the rapid diminution after age 30 and the completion of childbearing by close to age 35 have been important factors in the entry of older women into the labor force since 1930. It is possible also that the increase in family size during the baby boom period may have propelled many women into the labor force since the differentials in fertility by labor force status decreased between 1940 and 1960 (Ridley, 1971). These fertility differentials have diminished even further in the decade 1960 to 1970. The particularly large increases in labor force participation among younger white women between 1960 and 1970 have, no doubt, further contributed to the declines in fertility since 1960. Generally, the more recent changes in labor force and fertility reflect the increased educational attainment of women together with the other vast socioeconomic changes which have occurred since 1930.

In terms of the future, we expect that further increases in educational attainment of women will not materially affect their future labor force participation rates. Rather, these rates will reflect future job opportunities to a significant degree. Such future job opportunities will depend not only on the state of the economy but the degree of success women achieve in their quest for equal employment opportunities.

Acknowledgements

This paper is part of a larger study supported by a grant from the National Institute of Child Health and Human Development (No. HD06130-02). We are indebted to Thomas T. Kane who served as a research assistant on this project.

REFERENCES

Bancroft, G. 1958. The American Labor Force: Its Growth and Changing Composition. 1950 Census Monograph Series. New York: John Wiley and Sons.

Bowen, W. G. and T. A. Finegan. 1969. The Economics of Labor Force Participation. Princeton: Princeton University Press.

Cain, G. 1966. Married Women in the Labor Force: An Economic Analysis. Chicago: University of Chicago Press.

Deming, William Edwards. 1943. Adjustment of Data. New York: John Wiley and Sons.

Durand, J. D. 1948. The Labor Force in the United States: 1890 to 1960. New York: Social Science Research Council.

Edwards, Alba M. 1943. Comparative Occupational Statistics for the United States: 1870 to 1940. U. S. Bureau of the Census. Washington, D. C.: U. S. Government Printing Office.

Folger, K., and C. B. Nam. 1967. Education of the American Population. Washington, D. C.: U. S. Government Printing Office.

Jaffe, A. J., and C. D. Stewart. 1951. Manpower Resources and Utilization. New York: John Wiley and Sons.

Jaffe, A. J. 1956. Trends in the Participation of Women in the Working Force, Monthly Labor Review, 76, 5: 559.

Oppenheimer, V. K. 1970. The Female Labor Force in the United States: Demographic and Economic Factors Governing the Growth and Changing Composition. Monograph Series No. 5. Berkeley: University of California Press.

Ridley, J. C. 1971. The Changing Position of American Women: Education, Labor Force Participation, and Fertility. In The Family in Transition. Fogarty International Center Proceedings No. 3. Washington, D. C.: U. S. Government Printing Office.

Sweet, J. A. 1968. Women in the Labor Force. New York: Seminar Press.

U. S. Bureau of the Census, 1970. Census of Population. Subject Report, PC(2)-3A. Table 65.

COMPARING DATA ANALYSIS TECHNIQUES FOR EVALUATION DESIGNS WITH NON-NORMAL POPULATIONS

Elaine S. Jeffers, University of Maryland, Eastern Shore*

The data collection phases for evaluation designs may involve analysis of data from distributions that are skewed, restricted in range, lacking in variability or otherwise possessing characteristics rendering it difficult to assume normality of the mean and variance of the parent population. When assumptions have been violated, the following are some examples of transformations that can be used to result in a closer approximation to the normal distribution:

1. The square-root transformation

This transformation is useable when data are in the form of frequency counts as in "yes" "no" responses. The transformed score (X') would be:

$$X' = \sqrt{X}$$

2. The arc sine transformation

This transformation is useable when scores are in proportions, as in "percentage correct responses." The transformed score (X') would be:

$$X' = \arcsin \sqrt{X}$$

3. The logarithmic transformation

This transformation is useable when the data are decidedly skewed. The transformed score (X') would be:

$$X' = \log X$$

If the measures are small, the transformation might be:

$$X' = \log (X + 1)$$

Transformations can be used for fairly complex designs. They are also useable when the researcher wants to present a more elaborate analysis which would extend beyond a classical analysis of variance designs. Transformed data, however, changes the relationships between the statistics observed from the original data. Hence, when generalizing results, caution may need to be used in relating the results of the converted measures to the original measures.

Non-parametric and other non-traditional tests are other alternatives that have been used when assumptions of a normal distribution have not been met. May and Konklin developed a simplified version of the nonparametric counterpart to the F test for trend (4). Cooper used a non-parametric test for increasing trend by making comparisons of all possible pairs of observations (1). Lewis and Johnson illustrated the use of Kendall's Coefficient of Concordance for the evaluation of the extent of agreement among a set of judges, each of whom ranks in entirety a set of objects (3). These authors demonstrated that critical values of "W" are given only by approximation and only for large numbers. It was shown that Kendall's Coefficient of Concordance can be used to test more specific hypothesis about agreement on the criteria of choice within a group of judges (3). Simms and Collons illustrated the use of an algorithm for the Coefficient of Concordance which allows for the identification of the particular combination of individuals in a group, and all possible combinations of such individuals who exhibit or display the greatest amount of consensus (7).

Reynolds has stressed that indiscriminate use of tests of significance contributes to theoretical inadequacies in terms of scope and explanatory power (5). Young has stressed that a large part of the data which social scientists must deal with is not distributed according to known mathematical functions; non-parametric techniques (order statistics) offer the opportunity to test hypothesis which require no assumption about the form of the distribution of the population (8). Siegel has offered a variety of such non-parametric techniques which are particularly useable at the nominal and ordinal level of measurement (6).

Non-parametric tests can be used for pilot testing of instruments within an evaluation design, where assumption of normality cannot be taken for granted. In an abbreviated pilot-tested needs analysis procedure, first and second year university students ranked their preferences for the following goals for an educational research center:

1) Develop skills in evaluation of Educational Programs.
2) Develop skills in behavioral objectives and criterion-referenced testing--applications and development.
3) Develop knowledge of Instructional Systems Development.
4) Improve personal proficiency with Instructional Systems Development.
5) Make contacts with happenings in Educational Research.
6) Earn a PHD.
7) Build an Academic Recommendation.
8) Obtain tools needed for Educational Research--math, statistics, computer operations and research design.
9) Become familiar with Professional Societies, and Journals of Educational Research.
10) Take courses in other centers to keep abreast of most recent trends in Education.
11) Understanding of knowledge of diffusion.
12) Developing Skills in Curriculum Development.

13) Further knowledge in concept of Creativity and Research.

14) Gain realistic and practical experience in Evaluation.

15) Relate Evaluation and Teaching Skills to Humanistic Education Field.

The sign-rank test was used to test the hypothesis of no differences in ranking of these goals by first and second year university students as follows:

Ho (Null Hypothesis)..There is no difference in the ranking of these goals as perceived by first and second year students; alpha level equal .05.

H_1 (Alternative Hypothesis)..There is a difference in the ranking of these goals as perceived by first and second year students.

Table I presents an application of the sign-rank test used to test differences for these purposes. The null hypothesis tested by the sign test is that:

$p(XA > XB) = p(XA < XB) = \frac{1}{2}$

Another way of stating the null hypothesis is that the median difference is zero. Focus is placed on the direction of the differences rather than the size of the differences. Reference is made to Table D in Siegel--Table of Probabilities Associated with Values as Small, as observed Values of x in the Binomial Test (B). The probability of getting this result for a two-tailed test is .180. This is outside the rejection region and the decision is to accept the null hypothesis.

Non-parametric tests can be used in the preliminary stages of validating an instrument that will later be used for evaluation purposes. Table II illustrates the use of the Coefficient of Concordance to test the degree of agreement among Freshman, Sophomore, Junior and Senior students, concerning the quality of library services. Comparable groups of students from each of these four levels were asked to rank the quality of library services in four different testing situations using different instruments. The results (W = .99) indicated a high degree of agreement among the four groups of students.

Table III lists a variety of non-parametric methods which can be used at various levels of measurement (nominal, ordinal and interval). A researcher who decides on the use of a non-parametric test must decide on the most appropriate test to use in terms of the level of measurement. This is an important consideration since the use of a non-parametric test yields less powerful results when a parametric tests can be used for the same purpose. The researcher must decide upon the relative consequence of the Type I or the Type II error. This is a decision that should be made based on the type of variables that are being investigated and the sampling environment. Hence, in researching the effects of a treatment to reduce the cause of cancer or in evaluating the effects of an educational program, it is important that Type II errors not be made.

The power of non-parametric methods relative to Type I and Type II errors have been summarized in detail by Festinger and Katz (2). A Type I error is made when the decision is made to reject the null hypothesis when it should be accepted; a Type II error is made when the decision is made to accept the null hypothesis when it should be rejected. Power = 1 minus the probability of a Type II error; stated alternatively, power is the probability of appropriately rejecting the null hypothesis.

Within an evaluation setting, generalizability of results is not necessarily a goal. However, it is very important that results be valid for a particular decision-making setting. Failure to adequately consider the importance of this factor results in the possibility of collecting data that lacks decision-maker validity. Hence, in an evaluation setting, the decision to use a parametric or a non-parametric test would depend on the relative consequences of the power of that test on decision-making accuracy.

*The author wishes to acknowledge appreciation of the University of Mass. and Dr. N. Nagarajan formerly of the University of Maryland, Eastern Shore, for data presented.

Table I

An Application of the Sign-Rank Test of Differences in
1st and 2nd Year Student Responses to Group Goals

2nd Year (a)	1st Year (B)	2nd yr-1st yr. (A-B)	Direction of Difference	Signs
6	5	+1	XA > XB	+
4	2	+2	XA > XB	+
4	2	+2	XA > XB	+
2	2	0	XA = XB	0
7	5	+2	XA > XB	+
6	2	+4	XA > XB	+
5	0	+5	XA > XB	+
7	8	-1	XA < XB	-
7	4	+3	XA > XB	+
7	2	+5	XA > XB	+
3	4	-1	XA < XB	-
2	3	-1	XA < XB	-
4	3	+1	XA > XB	+
5	6	-1	XA < XB	-
4	2	+2	XA > XB	+

x = number of minority signs = 4 N = Number of non-zero ranks = 14

Table II

Coefficient of Concordance

Groups	I	II	III	IV
	Conditions			
A	4	2	3	1
B	4	3	2	1
C	4	3	2	1
D	4	3	2	1
R_j	16	11	9	4

Sum or ranks = 40 Mean of R_j = 10

$$W = \frac{s}{\frac{1}{12} K^2 (N^3 - N)} \qquad W = 0.99$$

s = Sum of squares of observed deviations from mean of R_j

k = # of sets of rankings--# of judges (i.e. A,B,C,D)

n = # of entities (objects or individuals) ranked (i.e. Conditions I-IV)

$\frac{1}{12} K^2 (N^3 - n)$ = max. possible sum of squared deviations , i.e. sum "s" which would occur with perfect agreement among K ranks

TABLE III (From Siegel's Non-Parametric Statistics)

	NONPARAMETRIC STATISTICAL TEST**				NONPARAMETRIC MEASURE OF CORRELATION
	Two-sample case		k-sample case		
One-sample case (Chap. 4)	Related samples (Chap. 5)	Independent samples (Chap. 6)	Related Samples (Chap. 7)	Independent Samples (Chap. 8)	(Chap. 9)
Binomial test, pp. 36-42 χ^2 one-sample test, pp. 42-47 (Nominal measurement)	McNemar test for the significance of changes, pp. 63-67 (Nominal measurement)	Fisher exact probability test, pp. 96-104 χ^2 test for two independent samples, pp. 104-111 (Nominal measurement)	Cochran Q test, pp. 161-166 (Nominal measurement)	χ^2 test for k independent samples, pp. 175-179 (Nominal measurement)	Contingency coefficient C, pp. 196-202 (Nominal measurement)
Kolmogorov-Smirnov one-sample test, pp. 47-52 One-sample runs test, pp. 52-58 (Ordinal measurement)	Sign test, pp. 68-75 Wilcoxon matched-pairs signed-ranks tests,*** pp. 75-83 (Ordinal measurement)	Median test, pp. 111-116 Mann-Whitney U test, pp. 116-127 Kolmogorov-Smirnov two-sample test, pp. 127-136 Wald-Wolfowitz runs test, pp. 136-145 Moses test of extreme reactions, pp. 145-152 (Ordinal measurement)	Friedman two-way analysis of variance, pp. 166-127 (Ordinal measurement)	Extension of the median test, pp. 179-184 Kruskal-Wallis one-way analysis of variance, pp. 184-193 (Ordinal measurement)	Spearman rank correlation coefficient: r_s, pp. 202-213 Kendall rank correlation coefficient: r, pp.213-223 Kendall partial rank correlation coefficient: $r_{zy,z}$, pp. 223-229 Kendall coefficient of concordance: W, pp. 229-238 (Ordinal measurement
	Walsh test, pp. 83-87 Randomization test for matched pairs, pp. 88-92 (Interval measurement)	Randomization test for two independent samples, pp. 152-156 (Interval measurement)			

**Each column lists, cumulatively downward, the tests applicable to the given level of measurement. For example, in the case of k related samples, when ordinal measurement has been achieved both the Friedman two-way analysis of variance and the Cochran Q test are applicable.

***The Wilcoxon test requires ordinal measurement not only within pairs, as is required for the sign test, but also of the differences between pairs. See the discussion on pp. 75-76.

References

(1) Cooper, M. "A Non-Parametric Test for Increasing Trend". Educational and Psychological Measurement, Vol. 35, 1975, pp. 303-306.

(2) Festinger, L. and Katz, D. Research Methods in the Behavioral Sciences. New York: Holt, Rinehart and Winston, Inc., 1966, pp. 536-577.

(3) Lewis, G. H. and Johnson, R. G. "Kendall's Coefficient of Concordance for Sociometric Rankings with Self Excluded". Sociometry, 1971, Vol. 34, No. 4, December 1971, pp. 496-503.

(4) May, R. and Konkin, P. "A Non-Parametric Test of an Ordered Hypothesis for K Independent Samples." Educational and Psychological Measurement, Vol. 30, 1970, pp. 251-258.

(5) Reynolds, R. Jr. "Replication and Substantive Import: A Critique on the Use of Statistical Inference in Social Research". Sociology and Social Research, Vol. 53, No. 3, April 1969, pp. 299-310.

(6) Siegel, S. Non-Parametric Statistics for The Behavioral Sciences. New York: McGraw-Hill Book Co., 1956.

(7) Simms, C. and Collons, R. "Sociological Research Applications of an Algorithm for the Coefficient of Concordance". American Sociologist, Vol. 4, No. 4, November 1969, pp. 321-324.

(8) Young, P. Scientific Social Surveys and Research. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1966. pp. 340-341.

THE EFFECT OF TRANSFORMATIONS ON THE ANALYSIS OF VARIANCE

Paula Ann Johnson, Reynolds Metals

T. F. Higginbotham, Auburn University

ABSTRACT

Nine transformations were studied for effect on the analysis of variance. They were ARC TANGENT, $\sqrt{x}$, $\sqrt{x} + \sqrt{x+1}$, LOG_{10}, EXP(x), INVERSE, LOG_{ε}, SINE, and HYPERBOLIC TANGENT.

The factorial experiment used as a basis for this study was a three factor factorial with three levels per factor and three observations per cell. One effect was altered by the addition of a constant to provide one effect with a high percentage of significant records.

The transformations were applied to observations distributed according to the following distributions: Beta, Binomial, Erlang, Exponential, Gamma, Log-normal, Normal, Poisson, Uniform, and actual data from highway accident damage costs. For all distributions except the highway accident damage costs (8436), 10,000 experiments were studied.

When analysis of variance results were compared (before and after transformation), $\sqrt{x}$ and $\sqrt{x} + \sqrt{x+1}$ consistently changed the fewest results. All other transformations produced unstable results across the range of distributions, showing little effect to sometimes drastic change according to the distribution involved.

INTRODUCTION

It has been common practice for many years to transform raw data prior to performing an analysis of variance in order that the assumptions of the analysis of variance may be more accorately approximated; for example, see Davies (1954), or Fryer (1966).

In addition, it is well known that the analysis of variance tends to be robust, see Norton (1952), under some conditions. If robustness could be expected under all conditions, a study of the effects of transformations of raw data would be pointless.

In a computer aided review of pertinent literature, no single study was found concerning the impact of various transformations on the analysis of variance. Yet the practice of transforming of raw data is so common as to have been made available in various computer statistical packages; for example, Biomedical Statistical Package, (Dixon, 1973).

The purpose of this study was to examine the impact of various common statistical distributions at the .01 and .05 levels of significance of the analysis of variance. The transformations selected were ARC TANGENT, $\sqrt{x}$, $\sqrt{x} + \sqrt{x+1}$, LOG_{10}, EXP(x), INVERSE, LOG_{ε}, SINE, and HYPERBOLIC TANGENT, selected because of availability in BMD (Dixon, 1973). The statistical distributions were Beta, Binomial, Erlang, Exponential, Gamma, Log-normal, Normal, Poisson, and Uniform, selected on the basis of being available in SIMSCRIPT. In addition to add a practical flavor to the problem, Alabama Highway Department Accident Damage Cost[1] data was included.

The factorial experiment used as a basis for this study was a three factor (A, B, C) factorial with three levels per factor and three observations per cell. This experiment was selected for the following reasons:

(1) an ABC design would provide two and three way interaction effects,

(2) three observations per cell would tend to provide "stability" in the cell,

(3) eighty-one observations would provide 54 degrees of freedom for test purposes (all factors were assumed fixed),

(4) increasing the number of factors would materially increase the number of computations without materially affecting the outcome, and

(5) increasing the number of observations per cell would materially increase the number of computations without materially affecting the outcome.

It was decided to force real significance[2] on one of the effects in the experiment so that one effect would have a relatively large percentage of significant records. By adding a constant to one level of a factor, the population mean of that level would be changed, forcing real significance. The third level of factor A was thus changed, except in the case of Highway (10). The individual constants added are specified in the discussion of each distribution.

Because it is not possible to perform all of the transformations on a value of zero, all remaining zero values (after the addition of the constant) were changed to 1.0. The only two distributions affected by this were the Binomial (02) and Poisson (08). The exact number of values so changed is specified in the individual distribution discussion.

The value ranges for the distributions were purposefully kept small because of the magnitude of values generated by transformation 06 (EXP(x)).

[1]David B. Brown, Personal Commun., 1975
[2]William D. Spears, Personal Commun., 1975

This is a restriction imposed by IBM 370 hardware. Any value over 18 was changed to 18. Not all distributions had values requiring this change but all are noted in the distribution discussions.

BETA DISTRIBUTION RESULTS

A random variable X is said to be distributed according to the Beta distribution of its density function is given by

$$f(x) = \frac{\Gamma(\alpha + \beta + 2)}{\Gamma(\alpha + 1)\Gamma(\beta + 1)} X^{\alpha}(1-x)^{\beta} \quad 0<x<1$$

where α and β are parameters with $\alpha > -1$ and $\beta > -1$ (Beyer, 1966, p. 18). The parameters used were $\alpha = 2$ and $\beta = 2$. The constant added to force real significance of main effect A was 0.25.

<u>Adjusted Original Effect Significant</u>

.05 Level of Significance

Significance tests of main effect A showed that more than 99% of all significant records remained significant after all transformations except 07 (Inverse). Only 73.32% remained significant after transformation 07. This pattern continued for all main effects and interactions; 24-32% remained significant at the .05 level of significance. For effects other than A, transformations 05 (LOG_{10}) and 08 LOG_{ε}) caused the percentage of significant records to drop to between 59 and 64% while all other transformations left at least 78% significant.

.01 Level of Significance

Significance tests of main effect A showed that more than 95% of all significant records remained significant after all transformations except 07 (Inverse). Only 40.57% remained significant after transformation 07. This pattern continued for all other main effects and interactions; 8-20% remained significant at the .01 level of significance. For all effects other than A, transformations 05 (LOG_{10}) and 08 (LOG_{ε}) caused the percentage of significant records to drop to between 42 and 58% while all other transformations left at least 70% significant.

<u>Adjusted Original Effect Not Significant</u>

.05 Level of Significance

Effects other than A were changed to significant less than 4% of the time by all transformations. Main effect A was affected more than 4% of the time by all transformations except 09 (SINE).

.01 Level of Significance

Effects other than A were changed to significant less than 2% of the time by all transformations. Transformation 06 (EXP(x)) changed 12.36% to significant on main effect A. All other transformations on A affected less than 5% of the records.

BINOMIAL DISTRIBUTION RESULTS

A random variable X is said to be distributed according to the Binomial Distribution if the probability function is given by

$$f(x) = \binom{n}{x} p^x q^{n-x}$$

where $p + q = 1$ and $\binom{n}{x} = \frac{n!}{x!(n-x)!}$. The parameters used were Trials = n = 6 and Probability of Success = p = 0.5. The constant added to force real significance of main effect A was 1.50. To enable all transformations to be performed, 8578 observations (1.05% of the total number) that were found to be zero after the addition of the constant were changed to 1.0.

<u>Adjusted Original Effect Significant</u>

.05 Level of Significance

Significant tests of main effect A showed that more than 97% of all significant records remained significant after all transformations except 09 (SINE) and 10 (HYPERBOLIC TANGENT). Only 85.27% remained significant after transformation 09; 56.74% after transformation 10. This pattern continued through all other main effects and interactions with transformations 09 and 10 causing less than 41% and 30% respectively to remain significant. For all effects other than A, only transformations 03 ($\sqrt{x}$) and 04 ($\sqrt{x} + \sqrt{x+1}$) left more than 80% significant.

.01 Level of Significance

Significance tests on main effect A showed that four of the transformations caused less than 93% of the significant records to remain significant. They were 06 (EXP(x)) with 84.73%, 07 (INVERSE) with 87.85%, 09 (SINE) with 65.48%, and 10 (HYPERBOLIC TANGENT) with only 15.78% remaining significant. For effects other than A, only transformations 03 ($\sqrt{x}$) and 04 ($\sqrt{x} + \sqrt{x+1}$) left more than 75% significant.

<u>Adjusted Original Effect Not Significant</u>

.05 Level of Significance

Effects other than A were changed to significant less than 6% of the time by all transformations. Main effect A was changed to significant more than 6% of the time by all transformations except 03 ($\sqrt{x}$) and 04 ($\sqrt{x} + \sqrt{x+1}$).

.01 Level of Significance

Effects other than A were changed to significant less than 3% of the time by all transformations. Main effect A was changed to significant less than 6% of the time by all transformations except 06 (EXP(x)) which changed 12.37%.

ERLANG DISTRIBUTION RESULTS

The Erlangian distribution of order k has as its probability density function:

$$f(t) = \frac{\lambda}{\Gamma(k)} (\lambda t)^{k-1} \varepsilon^{-\lambda t}.$$ (Parzen, 1962, p. 199).

The parameters used were mean = 4 and k = 3. There were 139 observations (.017% of the total number) greater than 18.0 which were changed to 18.0. The constant added to force real significance of main effect A was 2.0.

Adjusted Original Effect Significant

.05 Level of Significance

Significance tests of main effect A showed that more than 91% of all significant records remained significant after six of the transformations. However, transformation 06 (EXP(x)) left only 6.58% significant. Transformation 09 (SINE) left 30.65% significant and 10 (HYPERBOLIC TANGENT) left 41.18% significant. For effects other than A, only transformations 03 ($\sqrt{x}$) and 04 ($\sqrt{x}$ + $\sqrt{x+1}$) left more than 78% of the records significant.

.01 Level of Significance

Significance tests of main effect A showed that more than 95% of all significant records remained significant after five of the transformations. Transformation 06 (EXP(x)) left only 1.76% significant; 07 (INVERSE) left 69.77% significant; 09 (SINE) left 12.29% significant; and 10 (HYPERBOLIC TANGENT) left 10.18% significant. For effects other than A, only transformations 03 ($\sqrt{x}$) and 04 ($\sqrt{x+1}$ left more than 75% significant (except for effect ABC where only 67% were left significant). Throughout all effects, transformations 06, 09, and 10 caused the greatest loss of significance.

Adjusted Original Effect Not Significant

.05 Level of Significance

Effects other than A were changed to significant less than 6% of the time by all transformations. Main effect A was changed more than 25% of the time by all transformations except 06 (EXP(x)) which only changed .41%.

.01 Level of Significance

Effects other than A were changed to significant less than 2% of the time by all transformations. Main effect A was changed more than 16% of the time by all transformations except 06 (EXP(x)) which changed none, and 10 (HYPERBOLIC TANGENT) which changed 5.09%.

EXPONENTIAL DISTRIBUTION RESULTS

A continuous random variable X assuming all non-negative values is said to have an Exponential distribution with parameter $\alpha > 0$ if its probability density function is given by

$$f(x) = \alpha\varepsilon^{-\alpha x},\ x>0$$
$$= 0 \quad \text{elsewhere.}$$

Meyer, 1965, p. 173. The parameter used was mean = 2. There were 108 observations (.013% of the total number) greater than 18.0 which were changed to 18.0. The constant added to force real significance of main effect A was 2.0.

Adjusted Original Effect Significant

.05 Level of Significance

Significance tests of main effect A showed that only three transformations changed any significant values. Transformation 06 (EXP(x)) left only 5.21% significant; 07 (INVERSE) left 35.67% significant; and 09 (SINE) left 66.25% significant. For effects other than A, only transformations 03 ($\sqrt{x}$) and 04 ($\sqrt{x}$ + $\sqrt{x+1}$) left at least 70% significant. Transformations 06 and 07 left less than 8% significant for all effects other than A.

.01 Level of Significance

Significance tests of main effect A showed that only three transformations left less than 99% significant. They were transformation 06 (EXP(x)) which left only 1.39% significant; 07 (INVERSE) which left 13.54% significant; and 09 (SINE) which left 43.96% significant. For effects other than A, only transformation 04 ($\sqrt{x}$ + $\sqrt{x+1}$) left at least 70% significant (except on effect ABC which was only 57%). As in the .05 level of significance above, transformations 06 and 07 left the least percentage of significant records.

Adjusted Original Effect Not Significant

.05 Level of Significance

Effects other than A were changed to significant less than 7% of the time by all transformations. Main effect A was changed more than 31% of the time by all transformations except 06 (EXP(x)) which changed none.

.01 Level of Significance

Effects other than A were changed to significant less than 2% of the time by all transformations. Main effect A was changed more than 18% of the time by all transformations except 06 (EXP(x)) which changed none and 07 (INVERSE) which changed 9.21%.

GAMMA DISTRIBUTION RESULTS

A random variable X is said to be distributed according to the Gamma distribution if its density function is given by

$$f(x) = \frac{1}{\Gamma(\alpha+1)\beta^{\alpha+1}} \quad X^{\alpha}\varepsilon^{-X/\beta} \quad 0<X<\infty$$

where α and β are parameters with $\alpha > -1$ and $\beta>0$. (Beyer, 1966, p. 18). The parameters used were mean = 4 and k = 3. There were 130 observations (.016% of the total number) greater than 18.0 which were changed to 18.0. The constant added to force real significance of main effect A was 2.0.

Adjusted Original Effect Significant

.05 Level of Significance

Significance tests of main effect A showed that more than 91% of all significant records remained significant after six of the transformations. However, transformation 06 (EXP(x)) left only 6.26% significant. Transformation 09 (SINE) left 30.55% significant and 10 (HYPERBOLIC TANGENT) left 40.26% significant. For effects other than A, only transformations 03 ($\sqrt{x}$) and 04 ($\sqrt{x}$

+ $\sqrt{x+1}$)left more than 75% of the records significant.

.01 Level of Significance

Significance tests of main effect A showed that more than 95% of all significant records remained significant after five of the transformations. Transformation 06 (EXP(x)) left only 1.58% significant; 07 (INVERSE) left 69.29% significant; 09 (SINE) left 11.71% significant; and 10 (HYPERBOLIC TANGENT) left 8.90% significant. For effects other than A, only transformations 03 ($\sqrt{x}$) and 04 ($\sqrt{x} + \sqrt{x+1}$) left at least 70% significant. Throughout all effects, transformations 06, 09, and 10 caused the greatest loss of significance.

Adjusted Original Effect Not Significant

.05 Level of Significance

Effects other than A were changed to significant less than 6% of the time by all transformations. Main effect A was changed more than 21% of the time by all transformations except 06 (EXP(x)) which only changed .10%.

.01 Level of Significance

Effects other than A were changed to significant less than 2% of the time by all transformations. Main effect A was changed more than 15% of the time by all transformations except 06 (EXP(x)) which changed none and 10 (HYPERBOLIC TANGENT) which changed 5.06%.

LOG NORMAL RESULTS

A random variable X is said to be distributed according to the Log-normal distribution if its probability density function is given by

$$f(x) = \frac{1}{\sqrt{2\pi}} x^{-1} \varepsilon^{-(\ln x - \alpha)^2/2\beta^2} \quad x>0,\ \beta>0,\ \alpha \geq 0$$
$$= 0 \quad \text{elsehwere}$$

(Miller, 1965, p. 77). The parameters used were mean = 4 and standard deviation = 1. The constant added to force real significance of main effect A was 2.00.

Adjusted Original Effect Significant

.05 Level of Significance

Significance tests of main effect A showed that more than 90% of all significant records remained significant after all transformations except 09 (SINE). Transformation 09 left 20.59% significant. For all other effects, transformation 09 caused the greatest drop in significance. For effects other than A, only transformations 03 ($\sqrt{x}$) and 04 ($\sqrt{x} + \sqrt{x+1}$) left at least 85% significant.

.01 Level of Significance

Significance tests of main effect A showed that six of the transformations had no effect. Transformation 06 (EXP(x)) left 78.74% significant; 09 (SINE) left only 7.76% significant; and 10 (HYPERBOLIC TANGENT) left 70.12% significant. For effects other than A, only transformations 03 ($\sqrt{x}$) and 04 ($\sqrt{x} + \sqrt{x+1}$) left at least 80% significant. Transformation 09 caused the greatest drop in significance.

Adjusted Original Effect Not Significant

.05 Level of Significance

Effects other than A were changed to significant less than 6% of the time by all transformations except 10 (HYPERBOLIC TANGENT). Transformation 10 changed approximately 28% of all records on effects other than ABC.

.01 Level of Significance

Effects other than A were changed to significant less than 2% of the time by all transformations except 10 (HYPERBOLIC TANGENT). Transformation 10 changed approximately 11% of all records on effects other than ABC.

NORMAL DISTRIBUTION RESULTS

A random variable X is said to be distributed according to the Normal distribution if its density function is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \varepsilon^{-(x-\mu)^2/2\sigma^2} \quad -\infty < x < \infty.$$

(Beyer, 1966, p. 18). The parameters used were mean = 6 and standard deviation = 1. The constant added to force real significance of main effect A was 3.0.

Adjusted Original Effect Significant

.05 Level of Significance

Significance tests on main effect A showed that more than 97% of all significant records remained significant after all transformations 09 (SINE). Only 67.53% remained significant after transformation 09. For effects other than A, transformations 03 ($\sqrt{x}$), 04 ($\sqrt{x} + \sqrt{x+1}$), 05 (LOG_{10}), and 08 (LOG_{ε}) left at least 76% of the records significant. Transformation 09 continued to cause the greatest drop in significance for all effects (except ABC which was reduced to zero significant records by transformation 10 (HYPERBOLIC TANGENT)).

.01 Level of Significance

Significance tests on main effect A showed that more than 96% of all significant records remained significant after all transformations except 09 (SINE). Only 42.36% remained significant after transformation 09. For effects other than A transformations 03 ($\sqrt{x}$) and 04 ($\sqrt{x} + \sqrt{x+1}$) left at least 76% of the records significant. Effect ABC had all significant records changed by transformation 10 (HYPERBOLIC TANGENT) which left 67-97%.

Adjusted Original Effect Not Significant

.05 Level of Significance

Effects other than A were changed to significant less than 8% of the time by all transformations except 10 (HYPERBOLIC TANGENT). Transfortion 10 changed no records on effect ABC but changed 69-99% of records for other effects.

.01 Level of Significance

Effects other than A were changed to significant less than 4% of the time by all transformations except 10 (HYPERBOLIC TANGENT). Transformation 10 changed no records on effect ABC but changed 67-98% of records for other effects.

POISSON DISTRIBUTION RESULTS

A random variable X is said to be distributed according to the Poisson distribution if its probability function is given by

$$f(x) = \frac{\varepsilon^{-m} m^{x}}{x!} \quad m > 0, 1, 2, \cdots$$

(Beyer, 1966, p. 19). The parameter used was mean = 4. The constant added to force real significance of main effect A was 2.0. To enable all transformations to be performed, 9972 observations (1.23% of the total number) that were found to be zero after the addition of the constant were changed to 1.0.

Adjusted Original Effect Significant

.05 Level of Significance

Significance tests of main effect A showed that more than 97% of all significant records remained significant after six of the transformations. However, transformation 06 (EXP(x)) left only 26.31% significant; 09 (SINE) left 8.55% significant; and 10 (HYPERBOLIC TANGENT) left 36.39% significant. For effects other than A, the above three transformations caused the greatest loss of significance. Transformations 03 ($\sqrt{x}$) and 04 ($\sqrt{x} + \sqrt{x+1}$) were the only ones to leave more than 80% significant on all effects.

.01 Level of Significance

Significance tests of main effect A showed that at least 90% of all significant records remained significant after five of the transformations. Transformation 06 (EXP(x)) left only 9.78% significant; 07 (INVERSE) left only 79.91% significant; 09 (SINE) left 2.04% significant; and 10 (HYPERBOLIC TANGENT) left 9.30% significant. The above four transformations caused the greatest loss of significance for all effects. Transformations 03 ($\sqrt{x}$) and 04 ($\sqrt{x} + \sqrt{x+1}$) were the only transformations to leave at least 68% of all effects significant.

Adjusted Original Effect Not Significant

.05 Level of Significance

For all effects other than A, less than 7% of the records were changed by all transformations. For main effect A, all transformations changed more than 15% of the records except 06 (EXP(x)).

.01 Level of Significance

For all effects other than A, less than 3% of the records were changed by all transformations. For main effect A, all transformations changed more than 12% of the records except 06 (EXP(x)) which changed none; 09 (SINE) changed 3.55% of the records and 10 (HYPERBOLIC TANGENT) changed 3.95% of the records.

UNIFORM DISTRIBUTION RESULTS

A random variable is said to be distributed uniformly over the interval [a,b], where both a and b are finite, if its probability density function is given by

$$f(x) = \frac{1}{b-a}, \quad a \le x \le b$$
$$= 0, \quad \text{elsewhere}$$

(Meyer, 1965, p. 64). The parameters used were lower bound = a = 1 and upper bound = b = 5. The constant added to force real significance of main effect A was 1.0.

Adjusted Original Effect Significant

.05 Level of Significance

Significance tests of main effect A showed that more than 97% of all significant records remained significant after all transformations except 09 (SINE) and 10 (HYPERBOLIC TANGENT). Transformation 09 left 61.26% significant; 10 left 80.68% significant. For effects other than A, the above two transformations caused the greatest drop in significance. For all effects only transformations 03 ($\sqrt{x}$) and 04 ($\sqrt{x} + \sqrt{x+1}$) consistently left at least 85% of the records significant.

.01 Level of Significance

Significance tests of main effect A showed that five of the transformations left more than 95% of the records. However, transformation 02 (ARC TANGENT) left 88.39% significant; 07 (INVERSE) left 84.24%; 09 (SINE) left 40.70%; and 10 (HYPERBOLIC TANGENT) left 36.52%. For effect other than A, only transformations 03 ($\sqrt{x}$) and 04 ($\sqrt{x} + \sqrt{x+1}$) left more than 79% of the records significant.

Adjusted Original Effect Not Significant

.05 Level of Significance

For effects other than A, less than 6% of the records were changed by all transformations. Main effect A was changed more than 6% of the time by all transformations except 09 (SINE) which only changed .10%.

.01 Level of Significance

For effects other than A, less than 3% of the records were changed by all transformations. Main effect A was changed more than 6% of the time by all transformations 02 (ARC TANGENT), 06 (EXP(x)), 07 (INVERSE), and 10 (HYPERBOLIC TANGENT).

HIGHWAY DISTRIBUTION RESULTS

The distribution was generated from actual highway accident damage costs as estimated by investigating officers. The observations were obtained by reading the Alabama Highway Department accident file. Each observation was divided by 1000 so amounts could be transformed by 06 (EXP(x)). A frequency table of observation (expressed in thousands of dollars) is found in Table 1. After the above coding, 353 observations (.05% of the total number) were found to be greater than 18.0 and were changed to 18.0. There was no constant added in this distribution to force significance.

Adjusted Original Effect Significant

.05 Level of Significance

Significance tests on all main effects and interactions showed that only transformations 02 (ARC TANGENT), 03 ($\sqrt{x}$), 04 ($\sqrt{x+1}$) and 10 (HYPERBOLIC TANGENT) left at least 78% of the records significant. The other five transformations caused a much greater drop in significance. Transformation 07 (INVERSE) caused the greatest drop in significance.

.01 Level of Significance

Significance tests of all main effects and interactions showed that only transformations 02 (ARC TANGENT), 03 ($\sqrt{x}$), 04 ($\sqrt{x} + \sqrt{x+1}$) and 10 (HYPERBOLIC TANGENT) left at least 74% of the records significant. Transformation 07 (INVERSE) left less than 6% of the records significant.

Adjusted Original Effect Not Significant

.05 Level of Significance

All effects were changed less than 11% of the time by all transformations. Transformation 05 (LOG_{10}) and 08 (LOG_{ε}) caused the greatest changes.

.01 Level of Significance

All effects were changed less than 5% of the time by all transformations.

SUMMARY

Nine transformations were studied for effect on the analysis of variance. They were ARC TANGENT, $\sqrt{x}$, $\sqrt{x} + \sqrt{x+1}$, LOG_{10}, EXP(x), INVERSE, LOG_{ε}, SINE AND HYPERBOLIC TANGENT.

The factorial experiment used as a basis for this study was a three factor factorial with three levels per factor and three observations per cell. One effect was altered by the addition of a constant to provide one effect with a high percentage of significant records.

The transformations were applied to observations distributed according to the following distributions: Beta, Binomial, Erlang, Exponential, Gamma, Log-normal, Normal, Poisson, Uniform and actual data from highway damage costs. For all distributions except the highway accident damage costs (8436), 10,000 experiments were studied.

When analysis of variance results were compared (before and after transformation), $\sqrt{x}$ and $\sqrt{x} + \sqrt{x+1}$ consistently changed the fewest results. All other transformations produced unstable results across the range of distributions, showing little effect to sometimes drastic change according to the distribution involved.

BIBLIOGRAPHY

1. Beyer, William H., Ed. CRC Handbook of Tables for Probability and Statistics. Cleveland, Ohio: Chemical Rubber Co., 1966.

2. Davies, Owen L., Ed. Design and Analysis of Industrial Experiments. London: Oliver and Boyd, 1954.

3. Dixon, W. J., Ed. Biomedical Computer Programs. Los Angeles: University of California Press, 1973.

4. Freyer, H. C. Concepts and Methods of Experimental Statistics. Boston: Allyn & Bacon, 1966.

5. International Business Machines, IBM OS Full American National Standard Cobol. White Plains, New York: IBM, 1972.

6. Kiviat, P. J., Villaneuva, R.; and Markowity, H. M. SIMSCRIPT II.5 Programming Language. Edited by E. C. Russell. Arlington, Virginia: Consolidated Analysis Centers, 1973.

7. Meyer, Paul L. Introductory Probability and Statistical Applications. Reading, Mass.: Addison-Wesley Publishing Co., 1965.

8. Miller, Irwin, and Freund, John E. Probability and Statistics for Engineers. Englewood Cliffs, New Jersey: Prentice Hall, 1965.

9. Norton, Dee Wiswall. "An Empirical Investigation of the Effects of Normality and Heterogenity Upon the F-Test of Analysis of Variance." Ph.D. Dissertation, State University of Iowa, 1952.

10. Parzen, Emanuel. Stochastic Processes. San Francisco: Holden-Day, 1962.

11. Struble, George. Assembler Language Programming: The IBM System/360. Reading, Massachusetts: Addison-Wesley Publishing Co., 1969.

CORRELATIONS BETWEEN SUMS OF SQUARES AND F STATISTICS IN THE FIXED MODEL FOR UNBALANCED DESIGNS

John M. Jordan & G. Rex Bryce, Brigham Young University

For any analysis of variance model it is well known that the sums of squares of the main effects and the interactions are independent if the design is balanced. However, this independence does not necessarily follow for all pairs of sums of squares when the design is unbalanced, i.e., unequal cell frequencies. Since the latter is more likely to occur in practice, it would be useful to know the consequences of various degrees of imbalance on correlations between the sums of squares.

In the fixed model all the F ratios are correlated to some degree by virtue of a common denominator, the mean square for error. If the design is unbalanced and the sums of squares are fully adjusted, i.e., each sum of squares is obtained when its term is ordered last in the model, then the F ratios are also correlated through their numerators. As a result the tests are not independent and the joint probability of one or more Type I errors is a function of the strength of this correlation.

The Theory

This study depends on the representation of the analysis of variance model as a linear regression model by the use of the dummy variable technique within the cell means format. Define the cell means model in matrix format as

$$\underline{y} = X\underline{\mu} + \underline{e}$$

where $\underline{y}$ is an $n \times 1$ vector of responses, X is an $n \times p$ matrix of ones and zeroes indicating the location of the responses by cell, $\underline{\mu}$ is a $p \times 1$ vector of cell means and $\underline{e}$ is the $n \times 1$ vector of residuals.

We reparameterize this cell means model by the introduction of a full rank $p \times p$ matrix of contrasts, C, chosen to be of interest to the researcher. Write this new model as

$$\underline{y} = XC^{-1}C\underline{\mu} + \underline{e} = Z\underline{\beta} + \underline{e}$$

so that $\underline{\beta} = C\underline{\mu}$ is the set of coefficients dictated by the analyst's chosen set of linear combinations of the cell means and $Z = XC^{-1}$ is a "stretched" version of the inverse of the matrix of contrasts. We will assume throughout this study that all elements of $\underline{\beta}$ are known or specified as measures of the strength of particular contrasts defined on the terms in the model. For a 2×2 model in the usual ordering

$$\underline{\beta}' = \left[\beta_\mu \ \beta_A \ \ \beta_B \ \beta_{AB}\right] .$$

The purpose of this move from one full rank formulation to another should be clear. The parameters in both models are uniquely estimable; however, the original cell means model just estimates the cell means whereas the reparameterized model provides information which allows the kind of comparison of treatment effects needed by the researcher.

Now, if we were interested in estimating $\underline{\beta}$, we could obtain a vector of solutions to this system as

$$\hat{\underline{\beta}} = (Z'Z)^{-1}Z'\underline{y}$$

in the usual manner of multiple linear regression. However, for this study we are interested instead in a factored version of the $Z'Z$ matrix augmented by $Z'\underline{y}$ as a means of obtaining expressions for the sums of squares of the terms in the analysis of variance model. A Cholesky factorization of this augmented matrix can be obtained by premultiplication by a Cholesky operator in the form

$$R\left[Z'Z \mid Z'\underline{y}\right] = R\left[Z'Z \mid \underline{z}\right] = \left[T \mid \underline{t}\right]$$

and it will be useful to us to partition these matrices in order to write exact expressions for the sums of squares. If, for example, the interaction is ordered last in a model with two main effects, then the last partition in the $\underline{t}$ vector, $\underline{t}_{AB}$, will be the product of the elements in the partitioned last row of the R operator matrix with the elements of the $\underline{z}$ vector. The number of rows in $\underline{t}_{AB}$ will depend on the number of degrees of freedom associated with the interaction term.

If we symbolize this partitioned last row of R as R'_{AB} in this ordering and the matrix Z in the same ordering as Z^{AB}, then we have

$$\underline{t}_{AB} = R'_{AB}Z'^{AB}\underline{y}$$

therefore

$$\underline{t}'_{AB}I\underline{t}_{AB} = \underline{y}'Z^{AB}R_{AB}R'_{AB}Z'^{AB}\underline{y} .$$

This last formula is a quadratic form for the fully adjusted sum of squares for this interaction (SSAB) for any design, whether balanced or not. The same procedure could be used to find a general expression for the fully adjusted sum of squares for any term in the model by a simple reordering of the matrices. Having these, we can use well known theorems on the variance and covariance of quadratic forms, as in Searle (1971), to obtain the correlation between sums of squares for any design.

If we have $\underline{x} \sim N(\underline{\eta},V)$, then

$$\mathrm{Cov}(\underline{x}'P\underline{x},\underline{x}'Q\underline{x}) = 2\mathrm{tr}(PVQV) + 4\underline{\eta}'PVQ\underline{\eta} .$$

Since our response vector is normally and independently distributed with $\underline{\eta} = Z\underline{\beta}$ and $V = \sigma^2 I$, we can use matrix P as defined in the quadratic expression for SSAB and a corresponding Q defined for main effect B ordered last to write

$$\mathrm{Cov}(SSB,SSAB) = 2\sigma^4\mathrm{tr}(Z^{AB}R_{AB}R'_{AB}Z'^{AB}Z^{B}R_{B}R'_{B}Z'^{B})$$

$$+ 4\sigma^2\underline{\beta}'^{AB}Z'^{AB}Z^{AB}R_{AB}R'_{AB}Z'^{AB}Z^{B}R_{B}R'_{B}Z'^{B}Z^{B}\underline{\beta}^{B}$$

where the superscripts on $\underline{\beta}$, as usual, refer to a particular ordering of the elements in this vector. Observe that the second term in the covariance is the non-centrality portion of the expression since the whole term becomes zero if $\underline{\beta} = \underline{0}$.

Recall that the trace is invariant when cyclic operations are performed on its argument. With this in mind, we have found it convenient to rewrite this portion of the first term in the covariance as

$$\mathrm{tr}(R_{AB}R'_{AB}Z'^{AB}Z^{B}R_{B}R'_{B}Z'^{B}Z^{AB})$$

for which the dimensions are $p \times p$ rather than $n \times n$, where p is the number of cells in the design and n is the number of data points in the sample.

Before we can compute the correlation between SSAB and SSB, we need expressions in a similar form for the variance of each sum of squares. Searle (1971, p. 57) shows the general form for the variance of a quadratic form as

$$\mathrm{Var}(\underline{x}'P\underline{x}) = 2\mathrm{tr}(PV)^2 + 4\underline{\eta}'PVP\underline{\eta}$$

where, once again, $\underline{x} \sim N(\underline{\eta},V)$. Proceed as before to write the variance of (say) the interaction sum of squares as

$$\mathrm{Var}(SSAB) = 2\sigma^4\mathrm{tr}(Z^{AB}R_{AB}R'_{AB}Z'^{AB})^2$$

$$+ 4\sigma^2\underline{\beta}'^{AB}Z'^{AB}(Z^{AB}R_{AB}R'_{AB}Z'^{AB})^2Z^{AB}\underline{\beta}^{AB} \quad .$$

Again, the argument of the trace can be cycled for a more manageable matrix product.

The Application to 2 x 2 ECART Designs

Interested readers can verify, using a simple balanced design, that the formulas described above give a well defined zero correlation between any pair of sums of squares, i.e., zero covariance and non-zero variance.

However, our main interest is in unbalanced designs. In order to systematically study the effect of imbalance on correlation between sums of squares, we chose the ECART pattern (equal column and row totals) as a point of departure. This design is attributed to Powers and Herr (1975), who used it to study the partial confounding of row and column effects which occurs in unbalanced designs because of a lack of orthogonality in the design. Using this pattern we describe the cell frequencies in a 2×2 model as

	B_1	B_2
A_1	n - k	n + k
A_2	n + k	n - k

Then we can obtain the matrix product $Z = XC^{-1}$ for our reparameterized model, with X a $4n \times 4$ array of ones and zeroes corresponding to data location and C identified by the term ordered last in this matrix of contrasts.

For example, when interaction is ordered last, we obtain

$$Z^{AB} = X(C^{AB})^{-1} = X \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

from which $\underline{R}'_{AB} = \begin{bmatrix} a & 0 & 0 & b \end{bmatrix}$

where $a = \dfrac{k}{2n(n^2-k^2)}$ and $b = \dfrac{n}{2n(n^2-k^2)}$.

The Z matrix for main effect B ordered last is obtained by a reordering of the columns of the Z^{AB} matrix. If this is done, then the inverse of the transpose of the factored matrix yields a last row

$$\underline{R}'_{B} = \begin{bmatrix} 0 & a & 0 & b \end{bmatrix} \quad .$$

When these results are used to assemble the matrix product needed in the expression for the covariance between SSAB and SSB, the value of this product is zero. Since Var(SSAB) and Var(SSB) can be shown to be non-zero, we conclude that the correlation between SSAB and SSB is zero in the ECART design. The same result is obtained for the correlation between SSAB and SSA for this design.

However, the correlation between the sums of squares for the main effects in the ECART pattern is non-zero. We proceed directly to write the trace in the first term of the covariance between SSA and SSB as

$$2\sigma^4\mathrm{tr}(\underline{R}_A\underline{R}'_A Z'^{A}Z^{B}\underline{R}_B\underline{R}'_B Z'^{B}Z^{A}) = 2\sigma^4 k/n \; .$$

For the second covariance term, the noncentrality portion, we can reduce the matrix product to

$$16k\sigma^2\beta_A\beta_B(1-k^2/n^2)$$

so, as expected, the non-centrality term in the covariance between SSA and SSB is a function, in part, of the contrasts written on these terms. It also depends, as does the first term in the covariance, on the amount of imbalance and on σ^2.

Before correlation can be computed the variances of the sums of squares must be obtained. The matrix products needed here are easily deduced from our earlier work on the covariance terms. Since the algebra is very much the same as it was in finding the covariance, we will just assert that

$$\mathrm{Var}(SSA) = 2\sigma^4 + 16n\sigma^2\beta^2_A(1-k^2/n^2)$$

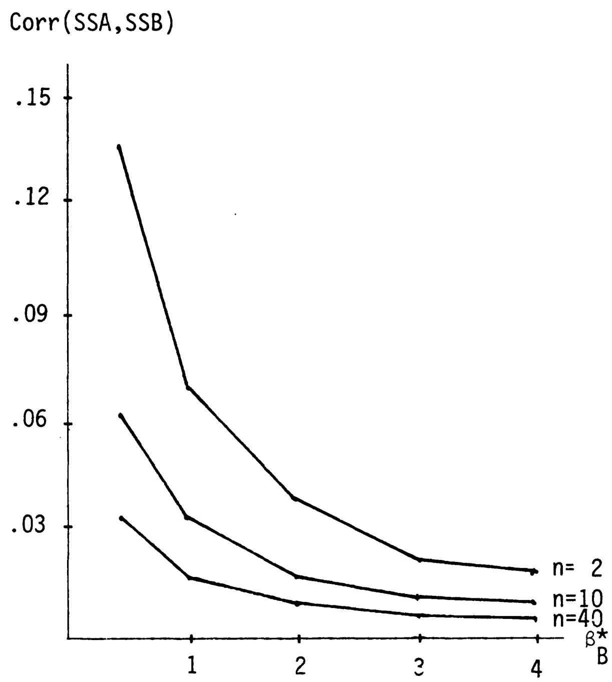
and the parallel expression can be obtained for Var(SSB).

We now have all the ingredients needed to write a general expression for the correlation between SSA and SSB for the ECART design. Before doing so, we will simplify the terms by defining $w = k/n$ as a measure of imbalance in the design. Also, for later ease in handling the contrasts on the main effects, convert them to standardized form as $\beta^*_A = \beta_A/\sigma$ and $\beta^*_B = \beta_B/\sigma$. Then the correlation becomes

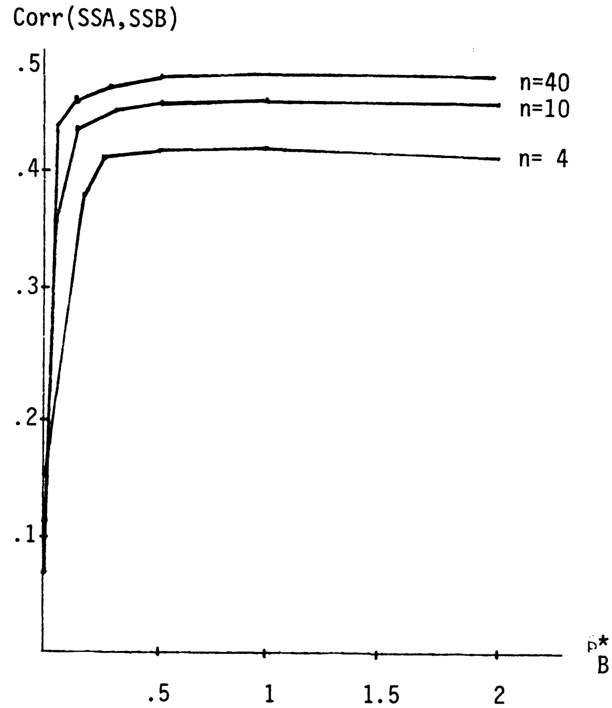
$$\mathrm{Corr}(SSA,SSB) = \frac{w^2 + 8k\beta^*_A\beta^*_B(1-w^2)}{\sqrt{\left[1+8n\beta^{*2}_A(1-w^2)\right]\left[1+8n\beta^{*2}_B(1-w^2)\right]}}$$

Under the null hypothesis, $\beta^*_A = \beta^*_B = 0$, the correlation between the main effects is equal to w^2 and so depends just on the measure of imbalance. Remember that n here is the average number of observations per cell; the total sample size is 4n . We conclude that, under this assumption, a small amount of imbalance in the ECART pattern, perhaps occurring fortuitously in a large sample, causes little correlation between SSA and SSB, whereas any amount of imbalance in a smaller sample will result in enough correlation that we would conclude important dependence exists between the main effects sums of squares and, therefore, between the F ratios used to test the main effects.

When one contrast, say β_A^* , is zero and the other is positive, the correlation between SSA and SSB decreases for increased sample size and for increased strength of the non-zero contrast, given a certain level of imbalance in the design. The values of this correlation are shown for selected cases in the figure below, when w = .5.



For both main effects non-zero, inspection of the general form for correlation between SSA and SSB in the ECART design will confirm that the second terms in the covariance and each of the variances will dominate the expression for most values of the parameters. Written with just these second terms, Corr(SSA,SSB) simplifies to approximately w = k/n , the measure of imbalance for a wide range of values of the parameters. The figure below shows this behavior for $\beta_A^* = .25$ over various levels of β_B^* for several sample sizes given w = .5 . The correlation begins to drop away from w as β_B^* increases but this decrease is small. We note in passing that the correlations are not affected if the interaction is ignored in obtaining main effects sums of squares rather than our fully adjusted approach.



The Application to Other 2 x 2 Designs

If the cell frequencies are identified only by location, i.e.,

	B_1	B_2
A_1	n_{11}	n_{12}
A_2	n_{21}	n_{22}

Then the Z'Z matrix can still be written in a simple form but the factored matrix is unworkable so it is not possible to write an expression for the correlation between sums of squares as a function of the useful parameters. Instead we must work directly with the matrix products which define covariances and variances for quadratic forms.

To preserve order in our consideration of the generalized 2 x 2 design, we will view new patterns as departures from the ECART design, whose characteristics are known to us. For an example, we will explain the case where $n_{..} = n_{11} + n_{12} + n_{21} + n_{22} = 12$. We will not consider patterns in which missing cells occur so the only two ECART designs are 2,4,4,2 and 1,5,5,1.

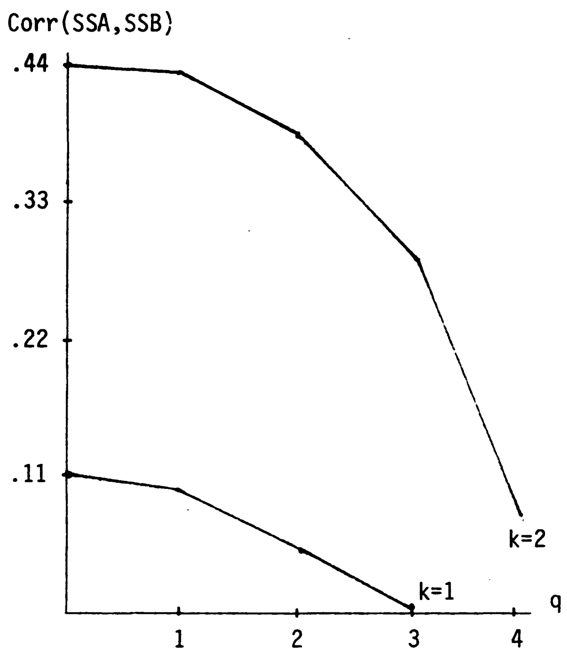
We know that the 2,4,4,2 ECART has n = 3 and k = 1 , therefore Corr(SSA,SSB) is .1111 when we assume the null hypotheses are true. Under the same null assumption, Corr(SSA,SSB) is .4444 for the 1,5,5,1 ECART.

Now, view departures from these patterns as functions of a new parameter q where, in the ECART design, q is used to modify the frequencies in the "n + k" cells:

n-k	n+k-q
n+k+q	n-k

Observe that $q = 1$ changes a 2,4,4,2 ECART to a 2,3,5,2 pattern but it changes a 1,5,5,1 ECART to a 1,4,6,1 pattern.

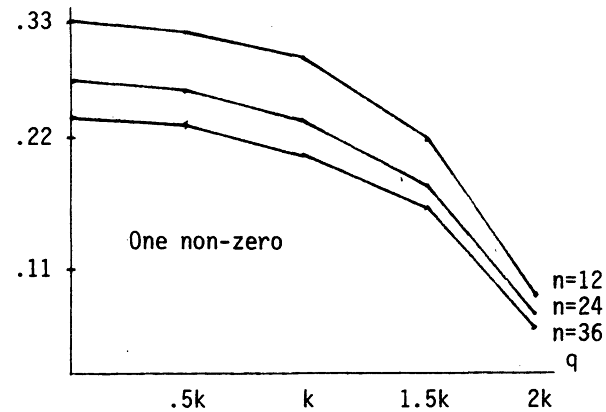
We will consider only the case in which the null hypotheses are true. Under this assumption, i.e., $\underline{\beta}^{*A} = [0\ 0\ 0\ 0]$ and $\underline{\beta}^{*B} = [0\ 0\ 0\ 0]$, Corr(SSA,SSB) decreases for departures from the ECART design and increases for departures from our proportional designs, as seen in the figure below. This agrees with the Powers and Herr (1975) conclusion that the ECART design is the most non-orthogonal design. From these starting points, the correlations for the departures from this design move in predictable directions for increasing values of q .

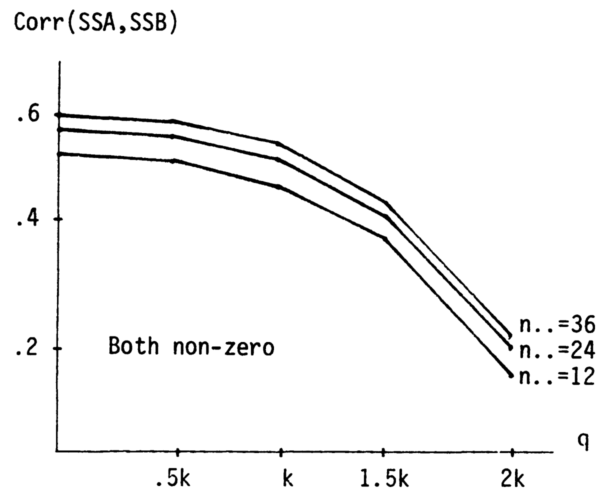


Finally, for the 2 x 2 design, we will briefly consider the change in Corr(SSA,SSB) which occurs as total sample size increases from n.. = 12 to n.. = 24 and n.. = 36 . The number of possible patterns increases greatly with larger sample sizes; we will study only the effect of the larger sample by discussing those patterns in which the ratios of k to n and q to k are the same as they were for the n.. = 12 cases where n = 3 and k = 2 .

Given these conditions on the larger sample sizes, Corr(SSA,SSB) remains the same when the null hypotheses are assumed true. This is the case where the contrasts are all zero so the second terms in the covariance and the variances are also zero. The unchanged correlation means that the change in the variances of the sums of squares as sample size increases is exactly offset by the change in covariance between the sums of squares. Given the fixed k to n ratio, this confirms the result found for the ECART design and extends it to departures caused by non-zero values of q.

However, the behavior of Corr(SSA,SSB) when sample size is increased and one or both of the standardized contrasts on the main effects is non-zero is not as easy to explain, since the rise in n.. causes the correlation to drop in one case and rise in the other. Refer to the figures below to see that the correlation falls for increased n.. over the range of q when one contrast is zero and the other is positive.

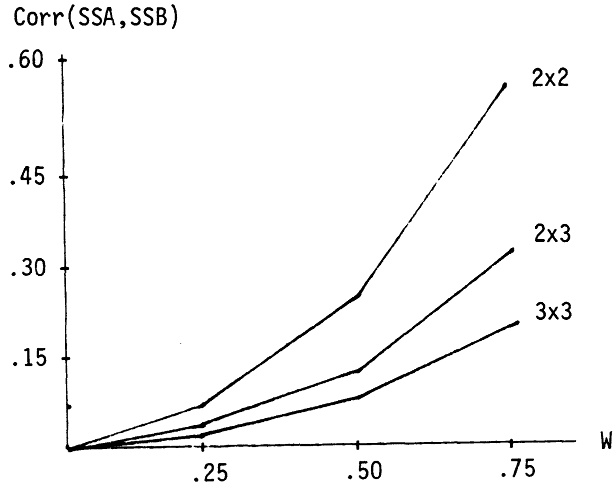




When both contrasts are positive the correlation rises with increased sample size. It would seem that, for both contrasts positive, we are denied the reward of lower correlation for a larger sample. But this assumes that orthogonality is an unmitigated virtue where a little reflection will confirm that it is not. If both null hypotheses about the main effects are, in fact, false then it is better to have high correlation between SSA and SSB so that the chance of rejecting both null hypotheses is enhanced and thus there is a pay-off from the larger sample,just as there is in the other case, where one null is true and the other is not, so the payoff from the larger sample comes in the form of lower correlation between the sums of squares of the main effects.

Application to Larger Designs

Since the determination of the number of levels for a categorical treatment is sometimes arbitrary, e.g., low, medium and high versus just low and high, we have studied to what extent the choice of number of levels affects the correlation between sums of squares. To isolate this influence we studied a fixed sample size, n.. = 36, for fixed amounts of imbalance in the ECART design for three different assumptions about the number of levels per treatment: the 2×2, the 2×3 and the 3×3 designs. As is shown in the figure below, the level of correlation decreases as the size of design increases, other things being equal, given the null hypotheses true.



This indicates that the larger versions of the ECART design are more orthogonal than the smaller versions with the same nominal level of imbalance. This is probably due to the presence in the larger designs of some cells with frequency of "n" which are unaffected by the parameter k .

Correlation Between F Ratios

When simultaneous F tests are done on the same data set in a fixed analysis of variance model, the F ratios are dependent because they have the same denominator, the error mean square. For a balanced design, the numerator sums of squares are independent, so all the dependence in the F ratios is caused by their common denominator. Hurlburt and Spiegel (1976) evaluated the conditional probability that both F ratios are significant given that one F ratio is significant under the assumption of independent numerator sums of squares. For unbalanced designs, the F ratios are also correlated through their numerators in many cases and our results so far enable us to measure this correlation since it depends, in part, on the covariance between sums of squares. This was done, under the assumption of true null hypotheses, for pairs of F ratios with independent numerators and for pairs with dependent numerators. These correlations will be used to approximate the joint probability that one or other or both F ratios are significant, i.e., that at least one Type I error is committed in the simultaneous tests.

Since the F statistic is a ratio of independent χ^2 statistics divided by their degrees of freedom, let x be distributed as a χ^2 with "a" degrees of freedom, let y be distributed as a χ^2 with "b" degrees of freedom and let z be distributed as a χ^2 with "c" degrees of freedom. We will use z as the denominator χ^2 distribution in defining a pair of F ratios, so that x and y will both be assumed independent of z but not independent of each other. Then, by definition, the general expression for the covariance between two F ratios is

$$\mathrm{Cov}(F_x,F_y) = \mathrm{Cov}\left[\frac{x/a}{z/c}, \frac{y/b}{z/c}\right]$$

$$= E\left[\frac{xy/ab}{(z/c)^2}\right] - E\left[\frac{x/a}{z/c}\right] E\left[\frac{y/b}{z/c}\right]$$

$$= \frac{c^2}{ab}\left[E(xy)E(z^{-2}) - E(x)E(y)E^2(z^{-1})\right]$$

but we can obtain

$$E(xy) = \mathrm{Cov}(x,y) + E(x)E(y)$$

and we know that the expected values of these random variables are just equal to their degrees of freedom if their distributions are centrally χ^2 . This will always be the case for the denominator variable z in our F statistic; we will be concerned with the joint probability of Type I errors made when the null hypotheses are true, so the numerator random variables will also have central χ^2 distributions.

Expected values of powers of these random variables can be obtained using the expectation of the gamma distribution. In this way, it can be shown that

$$E(z^{-2}) = \frac{1}{(c-2)(c-4)}$$

for $c > 4$, and also that

$$E(z^{-1}) = \frac{1}{c-2}$$

for $c > 2$ by the same approach.

Given all of this we have been able to show that

$$\mathrm{Corr}(F_x,F_y) = \frac{(c-2)\mathrm{Cov}(x,y) + 2ab}{2\sqrt{ab(a+c-2)(b+c-2)}}$$

for $c > 4$ and $a,b > -4$. The only operable restriction on this expression is that the number of degrees of freedom for the error sum of squares must be greater than 4 .

In the application to a 2×2 design this expression becomes, in terms of the covariance between main effect sums of squares, for the F ratios used to test these main effects:

$$\mathrm{Corr}(F_A,F_B) = \frac{(4n-6)\mathrm{Cov}(SSA,SSB) + 2\sigma^4}{2(4n-5)\sigma^4}$$

for a design with $4(n-1)$ degrees of freedom for error, such as the 2×2 ECART. When the Cov(SSA,SSB) for this size ECART design is substituted into this formula, under the assumption that the null hypotheses are true, we obtain

$$\mathrm{Corr}(F_A,F_B) = w^2 + \frac{1-w^2}{4n-5}$$

which approaches w^2 assymptotically as sample size becomes large, the same value found earlier for Corr(SSA,SSB) for the 2×2 ECART design under the same assumption.

For the generalized 2×2 design, the w^2 term is replaced by

$$\mathrm{tr}(Z^A \underline{R}_A \underline{R}'_A Z'^A Z^B \underline{R}_B \underline{R}'_B Z'^B) \quad .$$

The argument of the trace is one that we have seen before in calculations for the correlation between sums of squares.

Equivalent expressions for larger designs can be written as well. For a 2×3 size, the error sum of squares will have $6(n-1)$ degrees of freedom, again using n in our context as an average cell frequency, and the correlation becomes

$$\mathrm{Corr}(F_A,F_B) = \frac{(3n-4)\mathrm{tr}(M) + 1}{\sqrt{3(6n-7)(n-1)}}$$

where the argument M of the trace is the same as used above. If the same is done for the 3×3 design, the formula is

$$\mathrm{Corr}(F_A,F_B) = \frac{(9n-11)\mathrm{tr}(M) + 3}{18(n-1)}$$

The Overall Alpha Level for a Pair of F Tests

If the null hypotheses are both true in a pair of F tests, then an alpha error is committed whenever either one or other or both of the null hypotheses are rejected. We are interested in the joint probability that an alpha error will occur in some way, which will be the complement of the event that both decisions are correct, written for the main effects in a 2×2 model as

$$1 - \Pr(\text{accept } \beta_A = 0 \text{ and accept } \beta_B = 0) \quad .$$

This probability is described by a central bivariate F distribution for which we have been unable to find a closed form in the literature in the general class when the numerator mean squares are correlated. Hurlburt and Spiegel (1976) evaluated the integral of the bivariate F distribution when the mean squares are independent but extending their results is not straightforward. We have worked with an approximation to probabilities from the bivariate F distribution suggested by Johnson and Kotz (1972, p.242). The results of these calculations, using as input the correlations between pairs of F ratios for tests on the main effects found earlier in this study, indicate that the overall alpha level reacts strongly to imbalance in the design on the order of $w = .5$ or higher by our measure of imbalance.

We are now attempting to evaluate the bivariate F distribution directly for cases of correlated numerator mean squares in order to obtain precise measures of the overall alpha level. If the series expression for this distribution converges rapidly then accurate values for this probability should be possible.

The foregoing is a part of a much more detailed study found in Jordan (1976).

References

Hurlburt, Russel T. and Douglas K. Spiegel. "Dependence of F Ratios Sharing a Common Denominator Mean Square," The American Statistician, 30 (1976), 74-78.

Johnson, N.L. and S. Kotz. Distribution in Statistics: Continuous Multivariate Distributions. Volume IV. New York City: John Wiley and Sons, Inc., 1972.

Jordan, J.M. (1976). "Correlations Between Sums of Squares and F-Statistics in the Fixed Model for Unbalanced Designs." Unpublished Masters Thesis, Brigham Young University, Provo, Utah.

Powers, William A. and David G. Herr. "A Monte-Carlo Comparison of Error Probabilities of Five ANOVA Methods in Unbalanced, Two Way Designs." Unpublished Paper, University of North Carolina (Greensboro), 1975.

Searle, S.R. Linear Models. New York City: John Wiley & Sons, Inc., 1971.

SUCCESS OR FAILURE: THE PREVENTION OF UNWANTED PREGNANCIES

Gloria Kapantais and Evelyn Mathis
National Center for Health Statistics

Population control is not an idea new to the twentieth century. Although natural calamities historically have acted as checks on the world's population, nevertheless, people have sought artificial ways to limit their reproduction since primitive times. In fact, birth control in some form has been attempted for at least several thousand years.(1)

Despite its historical background, the birth control movement was slow in gaining momentum in the United States. While several individuals crusaded for population control in the early 1900's, there was little government interest in the field of family planning until the mid-1960's, since both Federal and State governments feared their involvement in such programs would antagonize religious and political groups. While the Kennedy Administration did sponsor research on human reproduction and contraception, it was not until the Johnson Administration that there was full support for family planning, including the allotment of Federal funds for the first national scope family planning programs. This dramatic change in the government's support of family planning programs is illustrated by President Nixon's message to the Congress on Population, July 18, 1969, in which he stated that:

> "It is my view that no American woman should be denied access to family planning assistance because of her economic condition; I believe, therefore, that we should establish as a national goal the provision of adequate family planning services within the next five years to all those who want them but cannot afford them. This we have the capacity to do."

During the last decade, therefore, public opinion in the United States has grown to recognize that family planning is vital to the individual and national health and well-being.

The primary concern of the family planning movement as it has evolved has been to stress the welfare of the family and the advantages of well-spaced and limited numbers of children. The movement is above all familistic, stressing the rights of parents--and of women, especially--to have the number of children they want.

While the emphasis on voluntary family planning as a health measure of considerable significance for both the individual family and the community is readily accepted, the interest of public welfare departments has also focused attention on the added dimension of family planning as a social measure, since the availability of family planning services is a crucial part of community efforts to reduce poverty and dependency.

An unwanted first birth is often a precipitating factor in becoming a welfare recipient, especially for young unmarried women. In addition, unwanted births can also perpetuate welfare dependency and mitigate the effectiveness of other social service programs designed to reduce or eliminate the need for public assistance.(2)

Problems stemming from improper child-spacing, lack of prenatal and postnatal care and misuse of contraceptives are all areas of concern to health care professionals, in addition to the health and social consequences of unwanted children.(3)

What then, has been the Federal government's role in promoting family planning services? In December 1970, Congress enacted the Family Planning Services and Population Research Act which legislated specific authority to assist in making comprehensive voluntary family planning services readily available to all persons desiring such services.

To help accomplish these goals, either new Federal programs were created or those already in existence extended their commitment to the field of family planning--among them at that time were the Children's Bureau; the Office of Economic Opportunity; Maternal and Child Health Services; and the Social and Rehabilitation Service. The National Center for Family Planning Services, established in October 1969, was the first Federal agency devoted exclusively to supporting family planning services. During its existence, the Center provided leadership in the extension of family planning services through special service project grants and other supporting programs.

Since these early beginnings of the Federal government's direct involvement in making family planning services available for all citizens, many changes have taken place, some of them, organizational and program changes.

This rapid expansion of family planning programs brought to the forefront the need for, and lack of, accurate and current information on the nature and extent of family planning services provided by public and private programs and the extent to which the total need for subsidized services was being met.

It was in such an atmosphere that the National Family Planning Services Data Collection System was developed as an answer to the legislative requirements and responsibilities imposed by the newly created Federal programs in this area. Primary responsibility for the methodological development and actual implementation of the data collection was assigned to the National Center for Health Statistics (NCHS), part of the Department of Health, Education, and Welfare.

This paper is based primarily upon data from two of the survey mechanisms developed under this system: the National Reporting System for Family Planning Services and the National Inventory of Family Planning Service Sites. The National Reporting System collects data on the numbers and characteristics of patients, number and types of visits made to clinics, and the number and types of services provided. In addition, the system also collects data about clinics and projects through which the patients are served. The projects that participate in the National Reporting System include both Federally-funded and non-Federally funded family planning projects. However, the majority of the clinics are funded by Titles V & X. The National Inventory provides location and characteristics data on all family planning clinics in the U.S., regardless of their funding. It is estimated that between 80 and 90 percent of the facilities in the National Reporting System are also in the National Inventory.

Prior to examining the data from these two survey mechanisms, background information will be given on recent trends in the total fertility rate and age specific birth rates for the United States. These show that the total fertility rate, which provides the implications of current levels of fertility for completed family size, has declined from a postwar peak of 3,724 births per 1,000 women in 1957 to the record low level of 1,895 in 1973. Age specific birth rates have declined with little interruption since 1957 among women in almost every age group. However, since 1957, the decline in the birth rate and in the fertility rate has been a gradual one. Table 1 contains preliminary data from 1975 that indicate that the birth rate was 14.8 and the fertility rate 66.7. Not much change has occurred in the birth rate between 1975, 1974, and 1973, and only a change of 5% occurred from 1972's birth rate of 15.6. In addition, there was only a 9% change in the fertility rate between 1972 and 1975. The fertility rate per 1,000 women aged 15-44 years dropped from 86.5 in 1969 to 66.7 in 1975, a decrease of close to 25%.(4),(5)

In spite of this constant decline in the birth and fertility rates, there still existed a substantial need for subsidized family planning services, such as those provided by the Federal government. This was due to the fact that, as late as 1974, it was estimated that of the 3.4 million women served by organized family planning programs, 1.8 million were in families with income below the poverty level; 2.2 million were in families at or below 125 percent of the poverty level, and 2.5 million were in families at or below 150 percent of the poverty level.(6)

According to data from the 1974 Inventory of Family Planning Service Sites, most of the organized family planning services being provided in this country are from public funds, with the Department of Health, Education, and Welfare providing the bulk of them. Table 2 shows that sixty-nine percent of the clinics in the country received some funding from the Department of Health, Education, and Welfare. State and county governments also were contributors to a substantial number of clinics. It is believed that the data from the 1975 Inventory will indicate similar findings.

As Table 3 indicates, the number of clinics providing family planning services increased by 18.2% from 1973 to 1974, and by only 2.2% from 1974 to 1975. However, the number of patients served by these clinics increased by 22.0% from 1973 to 1974, and by 24.5% from 1974 to 1975. Forty-three point seven percent of these total patients in 1975 were new patients. These figures indicate that organized family planning programs have grown during a period of overall U.S. fertility decline.

However, most of the fertility decline has been attributed to a reduction in unwanted and mistimed births, which has resulted from the more consistent use of contraception and a greater use of the most effective contraceptive methods. Of the 4,500 medical clinics that responded to the 1974 National Inventory of Family Planning Service Sites, Table 4 indicates that 99.6% of them employed oral contraceptives; 94.2% employed the intrauterine device and another 90% employed foam. (Obviously most clinics employ several different methods).

Data from the National Reporting System indicates that a majority of the female patients in family planning clinics employ one of the more effective methods of contraception--the pill or the IUD. For each year between 1969 and 1975, Table 5 shows that a minimum of 79.3% of the female patients used either the pill or the IUD--usually considered two of the most effective methods--with an average use of 83.5% for these two methods over these years. For all four of the methods listed in Table 5 (the pill, IUD, diaphragm and foam), the average use over this same time span was 90.7%. These figures tend to indicate that patients being served by organized family planning programs select and use the more efficient contraceptive methods.

The data in Table 6 also support this assumption, for, as this table indicates, 571,527 new patients (41.1%) used no method <u>before</u> enrolling in the clinic, while only 92,505 patients (6.7%) used no method <u>after</u> enrolling in the clinic (as reported at their last visit).

Before enrolling in the clinic 562,905 new patients (40.5%) used the pill, while as of their last visit 933,534 new patients (71.4%) used the pill. This increase from 562,905 to 993,534 represented an increase of more than 76% use of the pill. (7)

The number of patients being served in family planning clinics has continued to grow, the numbers being limited (in the opinion of the authors) only by the availability of services. Table 7 indicates that from the first year for which estimates are available (1969), the number of patients enrolled in family planning clinics has grown from less than 80,000 to more than 3.2 million women at the end of calendar year 1975. More than half, or 56.3%, of these women were not newly enrolled but had been in the program prior to 1975. In fact, the yearly increases in the proportion of continuation patients are all significant increases at the 95% level.

In addition to these data from the National Reporting System and the National Inventory, findings from other sources also support this premise that family planning programs in the United States are having a substantial impact in preventing births that are unwanted or mistimed.

Data from the 1970 National Fertility Study(8) were used by Cutright and Jaffee in a study to determine the effects of family planning programs on the fertility of low-income U.S. women. The authors concluded that the U.S. family planning program has--independent of other sociodemographic factors--reduced the fertility of low-income women by helping them to prevent unwanted and mistimed births. According to the authors, the program works because it gives women of lower socio-economic status access to modern and effective methods of contraception that they would not otherwise have. As a result, the rates of unwanted and mistimed pregnancy of patients are lower than those of comparable women who lack access to organized clinic programs.(9)

Charles F. Westoff, codirector of the 1970 National Fertility Study has stated that the results of this study indicate that the sharp marital fertility decline in the United States in the 1960's is entirely attributable to a reduction in unplanned births. This was accomplished by more consistent use of contraception, greater use of the most effective contraceptive methods and improved efficacy in their use.(10)

The National Survey of Family Growth, which provides data on fertility, fertility planning, and related aspects of maternal and child health, is a relatively new survey of the National Center for Health Statistics. Table 8, which presents 1973 data from the first cycle of this continuing survey, also demonstrates the impact of family planning programs in the U.S. By 1973, almost 7 in 10 married couples of reproductive age were using some type of contraception. According to Mr. Westoff, who has analyzed these data in conjunction with those from the National Fertility Studies, "It seems highly probable that by the end of the 1970's, almost all married couples at risk of unintended pregnancy in the United States will be using contraception, and almost all contraceptors will be protected by the most effective medical methods. We are rapidly approaching universal, highly effective contraception practice."(11)

Another source of data is the National Natality surveys conducted by the National Center for Health Statistics, which include a probability sample of women who have given birth during the year. According to these survey data, the proportion of births that were unwanted has declined substantially after 1968, when 13 percent of them were unwanted. By 1972, the percent of unwanted births had dropped to 8 percent. The decline in unwantedness was more pronounced among black wives than white wives so that, by 1972, only about one-fifth more black than white births were classified as unwanted.(12)

These data from selected studies on the availability and utilization of family planning services illustrate the growing acceptance in recent years of the idea that effective family planning practices offer undeniable health and welfare benefits to all individuals, and especially to low-income families. The provision of family planning services, especially to those who cannot afford private care, has been shown to be one of the most cost-effective preventive health care programs.(2) More important than economic considerations, however, are the medical and social benefits to be derived from successful fertility management. By providing family planning services which enable individuals to freely determine the number and spacing of their children, the quality of life for all individuals, but particularly the economically disadvantaged, will be improved. And this, despite all the various stated and unstated objectives, has been the ultimate goal of the family planning services program in this country--to raise the quality of life for all persons.

Thus, from a time just twenty years ago when it was common to provoke acrimonious debate over the socioeconomic as well as the moral issues involved in family planning, the concept today is almost universally accepted, tacitly or officially, as a sensible approach to improving the quality of life for all.

Tables are available upon request from the author.

References and Footnotes

(1) Westoff, Leslie Aldridge and Westoff, Charles F., From Now to Zero: Fertility, Contraception and Abortion in America. Boston: Little, Brown, and Co., 1971.

(2) Carter, Reginald and Nell, Cathy, Family Planning Services Provided to AFDC Recipients in Michigan. Studies in Welfare Policy No. 6, State of Micigan Department of Social Services, 1975.

(3) Thomas, Richard K., Family Planning in Memphis, Tennessee; Knowledge, Practice and Attitudes, Memphis Regional Medical Program, 1975.

(4) Provisional Statistics, Annual Summary for the U.S., 1975, NCHS Monthly Vital Statistics Report, Vol. 24, No. 13, June 30, 1976.

(5) Health: United States 1975, DHEW Publication No. (HRA) 76-1232.

(6) The poverty threshold for a nonfarm family of four persons was $5,038 in 1974, about 11 percent higher than the 1973 cutoff of $4,540. The poverty thresholds are updated each year to reflect changes in the annual average Consumer Price Index.

(7) It should also be noted that only 458,469 (81.4%) of the original 562,905 patients who used the pill prior to enrollment were still using the pill at the last visit. That is a decrease of more than 104,000 patients using the pill. Looking at it from this angle, the number of women who actually began taking the pill after enrolling in the clinic was 535,065. Therefore, the net increase in women taking the pill was 430,629 while the actual number of new "starts" was 535,065.

(8) Mr. Charles F. Westoff and Ms. Norman B. Ryder were codirectors of the 1965, 1970, and 1975 National Fertility Studies, conducted under contract by the Office of Population Research, Princeton University for the

Center for Population Research of the National Institute of Child Health and Human Development.

(9) Cutright, Phillips and Jaffe, Frederick S., "Family Planning Program Effects on the Fertility of Low-Income U.S. Women," Family Planning Perspectives, Vol. 8, No. 3, May/June 1976.

(10) Westoff, Charles F., "The Yield of the Imperfect: the 1970 National Fertility Study", Demography, Vol. 12, No. 4, Nov. 1975.

(11) Westoff, Charles F., "Trends in Contraceptive Practice: 1965-1973" Family Planning Perspectives, Vol. 8, No. 2, March/April 1976 p. 57.

(12) A report on the findings of the 1968, 1969, and 1972 surveys has been written by Robert H. Weller, Professor of Sociology at the Center for the Study of Population, Institute for Social Research, Florida State University. This NCHS report, entitled Wanted and Unwanted Childbearing in the United States, 1968, 1969, and 1972, is scheduled for publication in 1977.

EVALUATING SSA'S CURRENT PROCEDURES FOR ESTIMATING UNTAXED WAGES

Bertram Kestenbaum, Social Security Administration*

This paper presents an evaluation of the bias and precision in current methodology at the Social Security Administration (SSA) for estimating annual wages from truncated quarterly wage series. We find that SSA's methodology underestimates total wages when the truncation quarter is other than the first quarter and overestimates total wages for first quarter truncation, but that the net bias is small. We describe and summarize the error associated with individual estimates made with the current and competing methodologies. We suggest that a small modification of SSA's methodology should slightly improve performance.

NEED FOR ESTIMATES

Employers are required to report to the Social Security Administration wages and salaries paid to their employees which are subject to FICA taxes. When earnings exceed the maximum taxable by law, the employer truncates the earnings amount at that point. In 1972, there were almost 18.5 million wage reports truncated at the taxable ceiling of $9,000.

There are both program-related and non-program needs which require that these untaxed wages be estimated. Historically, the need was first felt when the Social Security Administration required a determination of the percentage of total wages that were taxed. Other program uses are the estimation of tax revenues that would accrue for different taxable maxima; and the modeling of lifetime earnings streams to test suggested revisions of the social security benefit formula[1].

In addition, the need to extrapolate for untaxed wages arises in various non-program research efforts using the Social Security Administration's files of statistical samples. Two examples are Gallaway's enquiry into the pecuniary returns to geographic migration [2], and McCall's application of Markovian stayer-mover models to income dynamics [3].

THE DATA BASE

The 1973 Exact Match project linked together survey data from the March 1973 Current Population Survey and its supplement on work experience and income of the population during the prior year with administrative records of the Social Security Administration and the Internal Revenue Service [4]. One general thrust of this endeavor was towards improvement of the quality of each agency's statistical output in the area of income distribution. A specific objective under this heading was the enabling of an evaluation of SSA's current extrapolative method to estimate untaxed wages.

Though the 1973 Exact Match was effected with the Social Security Administration's computerized Summary Earnings Record file, the level of summarization in that file rendered it unsuitable for anything more than the identification of potential cases with truncated earnings. We could determine which persons had wages reaching $9,000 but could not distinguish those who earned this amount from a single employer. Nor did we have available the quarterly earnings amounts on which the estimates are based.

Accordingly, for a 1-in-4 (approximately) subsample of candidates for our evaluation study--3,155 of them in all--a manual search of microfilmed records was undertaken. We determined that 259 persons had no truncation, having achieved the taxable maximum only from combined earnings with 2 or more employers. Also, six farmwork records were dropped, given the special nature of annual--rather than quarterly--reporting of farm employment, leaving a viable sample of 2,890 cases.

A validation study involving comparable variables y and z differs from a comparison of y and z in that we consider one of the two, say y, to be "true" or a close approximation of the "truth"; then, it may be used as a yardstick against which the accuracy of z is measured. The three-way Exact Match presented two potential y's: the higher-quality IRS Form 1040 entry for wages, based on attached Forms W-2, and the lower-quality CPS response, often based on recall and on one household member's perception of another's income. While it was our original intention to accept the CPS report when there was no usable Form 1040 amount, rather than give up the case, preliminary tabulations indicating substantial CPS understatement made us change our minds as to the suitability of the CPS to assume a role as a yardstick.

The joint filing of income tax returns, the common practice among married persons, presented some special problems to our evaluation efforts. Consider the situation from the following perspective. For a person with truncated wages who filed a nonjoint return, the Form 1040 wage entry is the sum of three amounts:

1. the wage that had been truncated--$9,000 or more;
2. wages from other covered employment, if any--which are known to us unless they, too, were truncated; and
3. wages from noncovered employment, if any--which we wish to assume are zero. An argument can be made that the likelihood of secondary jobholding in noncovered employment for someone with high earnings on his primary job is rather small.

For a joint filer, however, the Form 1040 entry is the sum of five amounts: these same three, plus the spouse's covered wages, if any, plus the spouse's noncovered wages, if any. Now numerous difficulties present themselves:

1. Did the March CPS find a spouse present (who did not file his/her own tax report)?

2. Have we secured the spouse's social security number, to use in obtaining the spouse's covered wages from the Summary Earnings Record file? We considered this a problem unless the CPS stated that the spouse had not worked at all in the prior year or was a self-employed or unpaid worker.

3. Does the Summary Earnings Record indicate that the spouse had both wages and self-employment, in which case the Summary Earnings Record figure is the combined amount?

4. Does the spouse also have a truncated wage?

5. Could the spouse perhaps have noncovered wages? We decided to be wary of noncovered wages if the spouse had no covered wages and the CPS classified the spouse in government, farm, or household employment, or in industries dominated by nonprofit organizations. (These are the four areas of employment where social security coverage is not complete.)

Table 1 details the erosion of the sample due to these various factors to an effective size of 2,470 cases. 1/

Table 1.--Number of Discarded Cases by Type

Basic Sample		2,890
Discarded Cases		420
Truncation with Two Employers	16	
Form 1040 Wages Less than $9,000	36	
No Form 1040 Found	38	
Spouse Not Found	13	
No Usuable Social Security Number for Spouse	60	
Spouse Has Both Wages and Self-employment Reported to SSA	6	
Spouse May Have Truncated Wages	161	
Spouse May Have Noncovered Wages	90	
Remaining Sample		2,470

Among the remaining cases there are, no doubt, a small fraction for which the assumptions necessary to equate the adjusted Form 1040 amount with the true truncated wage do not hold. Corrections to individual records are not feasible, but, in specifying the bias in our procedures, we will make allowance for this inaccuracy in our yardstick.

Now, what should be the products of our evaluation study? Certainly, to adjust the estimate of untaxed wages on the macro level, we need to specify the extent of bias in the SSA procedure. Also, a distribution of forecast errors by size and sign for each of several types of earnings patterns would be helpful to users in judging what confidence to place in point or interval estimates of truncated wage amounts; such a distribution is appended to the paper (as table A). In the balance of the paper, however, we would like to summarize the magnitude, type, and structure of the forecast errors and compare SSA's present method with other models.

METHOD II

Up to this point, we have not described "Method II" 2/, the technique used most often at the Social Security Administration to forecast untaxed wages. Its scheme is to determine the calendar quarter closest to the truncation quarter with wages greater than the reported truncation quarter wages; then, substitute this amount for the truncation quarter and any subsequent quarters. To illustrate: if reported wages in the first three calendar quarters of 1972 are $2,500, $3,500, and $3,000, respectively, Method II assigns $3,500 to the third and fourth calendar quarters. For reported wages of $3,200, $2,800, and $3,000, Method II passes over the second quarter's amount and substitutes $3,200 for the last two quarters.

When wages reported for the truncation quarter exceed all prior quarters' 3/--in effect, a contradiction of the Method II premise--Method II uses these reported wages for the truncation quarter and any subsequent quarters. This is a biased procedure, because the true ceiling quarter wages are at least equal to, but probably somewhat greater than, the reported amount.

A special case occurs when the first calendar quarter is the truncation quarter. Then, there is no series to extrapolate. Method II assumes, in such cases, that wages are at least four times the taxable maximum but does not attempt to make estimates on an individual basis. Instead, it computes an average for the group from the often-used procedure [5] of fitting a Pareto curve to the open end of the income distribution. The Pareto scheme expresses the "or more" cumulative frequency distribution as a simple two-parameter exponential function of the level of the variable and, hence, is completely determined by two points on the curve. To calculate the "or more" frequencies at two high dollar levels, Method II presumes that "or more" frequencies at four times and twice the maximum can be approximated by the frequency of first quarter truncation and the frequencies of first and second quarter truncation combined, respectively.

The bias in Method II, that is, the average value of the forecast error y-z, where y is the "true" wage obtained from the Form 1040, and z is its Method II estimate, is given in table 2. Negative biases represent overestimates.

The overall understatement of $203 is 3.5 percent of the average Form 1040 untaxed wage of $5,797. The Form 1040 average, however, is, itself, biased upward, as mentioned before, so that the actual

Method II understatement bias, we conjecture, may be closer to 1 percent or 2 percent.

Table 2.--Method II Biases by Earnings Pattern

Earnings Pattern	Percent of Population	Under-statement Bias (in dollars)
Total..................	100.0	203
First Quarter Truncation..	2.5	-11,292
Second Quarter Truncation:		
Ceiling Quarter Wages:		
Highest.................	2.6	3,072
Not Highest.............	11.0	179
Third Quarter Truncation:		
Ceiling Quarter Wages:		
Highest.................	2.8	1,730
Not Highest.............	34.0	319
Fourth Quarter Truncation:		
Ceiling Quarter Wages:		
Highest.................	2.6	978
Not Highest.............	44.5	444

Our very first reaction to the large overstatement for first quarter truncation was disapproval of the Pareto technique--and confusion, since we were aware of Gastwirth's short essay in the Review of Economics and Statistics criticizing the Pareto technique for producing estimates that were too low [6]. We, then, observed, however, that, for 21 of the 55 sample cases with first quarter truncation, the Form 1040 comparison amount was actually less than four times the taxable ceiling. The fault for the overstatement error lay, not with the Pareto method for estimating wages beyond four times the maximum, but, with the assumption that the employment relationship observed in the first quarter of the year continues throughout the year.

Indeed, we believe there exist two major sources of error--acting in opposite directions. One is the uncertainty of the worker's future employment status. Perhaps, he will stop working during the year or change to another employer. This error is greater for longer periods of uncertainty. The other--which we will return to later--is the failure to recognize the secular trend of rising wages over time.

We would like to digress a moment to consider Gastwirth's results. Analyzing the distribution of adjusted gross income (AGI) published in the annual IRS Statistics of Income series, Gastwirth concluded that the Pareto estimate for the open-ended interval substantially underestimates the true average AGI. Carrying out the calculations for 1972 [7], fitting the Pareto curve at $15,000 and $25,000, we obtain an estimate of average AGI in the $25,000-and-over interval of $36,633, which indeed falls 15 percent short of the actual average of $42,932.

We believe, however, that the performance of the Pareto approximation may be better for certain kinds of income than for others. For AGI, which includes interest, dividends, rents, and other unearned income, many high incomes may be very high incomes, so that the extrapolation of a Pareto curve fit at lower values could pass below the true locus for the group. Wage and salary income, on the other hand, is, for the most part, constrained to reasonable levels, so that the extrapolated Pareto curve may provide an adequate fit. In fact, tabulating our Exact Match file, we obtained $35,038, as the true average IRS wage for the $25,000-and-over group, and an estimate of $34,292--only 2 percent lower--from a Pareto fit at $15,000 and $25,000.

Except for first quarter truncation, Method II makes wage estimates on an individual basis, and we can summarize the precision of these estimates, following Theil [8] and Mincer and Zarnowitz [9], with the root mean square error (RMSE) measure, i.e., the square root of the average value of $(y-z)^2$. 4/ The root mean square error for Method II, as estimated from the sample, is $3,180; and RMSE's for various earnings patterns are given below in table 3. The RMSE is larger for third quarter truncation than for fourth quarter truncation and largest for second quarter truncation. The RMSE is about twice as large when the truncation quarter wages are highest than otherwise.

Table 3.--Root Mean Square Errors for Earnings Patterns

Ceiling Quarter	Are Ceiling Quarter Wages Highest?	
	Yes	No
Second	$10,371	$5,316
Third	5,301	2,491
Fourth	3,094	1,341

A very useful measure of the performance of Method II on the micro level is the proportion of wage amounts predicted within x percent of the true value. We obtain a proportion of 57 at x = 5% and a proportion of 77 at x = 10%.

ALTERNATIVE METHODS WHEN THE TRUNCATION QUARTER IS OTHER THAN THE FIRST CALENDAR QUARTER

Method II is, of course, only one of several extrapolative methods that could be used to estimate untaxed wages. Each carries a different theoretical model for the stochastic nature of quarterly wage earnings: Is the extrapolation of levels or changes (trends) more appropriate?

How many terms should be included? Should these terms be weighted identically or differentially? The heterogeneity in the movement of persons' earnings over time makes it difficult to decide a priori which model is best, but we can compare the empirical results, as in the top half of table 4.

The first line of results is not for an extrapolative model, but for a set of six regressions designed to obtain the best least-squares fit of the true untaxed wage as a linear function of the wage levels in quarters prior to the ceiling quarter. A separate regression was run to fit each of six earnings patterns, defined by the number of prior quarters (1, 2, or 3) and whether or not the ceiling quarter wages were highest.

The resultant bias is, of course, zero, and the root mean square error is minimized; so, the performance of the various extrapolative methods may be judged relative to this first line. It should be noted, however, that while this regression average has certain desirable properties, it does not do well with respect to the proportion of estimates correct to within 5 percent of the true value.

The second line of results is for Method II. Six other methods follow, some of which need to be elaborated on. For the "weighted average of prior levels" (line 6), we wanted more recent levels to be weighted more heavily, but the precise weights we used--2/3 and 1/3 for two prior quarters; 2/3, 2/9, and 1/9 for three prior quarters--were arbitrarily chosen. "Extrapolating prior trend" means carrying forward the most recent (line 7) or the average (line 8) signed difference between the wages of one quarter and the preceding quarter.

We would judge the extrapolation of highest prior quarter method (line 3) best, because of its smaller bias and the high proportion of wage estimates it predicts within 5 percent of the true wage. We think that it is best, not because its theory is most "correct," but because, by searching out the highest prior quarter wage, it somewhat offsets the problem identified earlier, namely, the tendency towards understatement which is characteristic of extrapolations that ignore the rising trend of wages.

It seems appropriate, therefore, to include an explicit trend component in our models, such as, for example, Milton Friedman did for his calculation of "permanent income" [10]. For this, we, first, adjusted wages downward to eliminate the presumed trend; then, recalculated the extrapolative estimates; and, finally, brought the trend component back in. A comparison of the lower and upper halves of table 4 reveals that both Method

Table 4.--Empirical Results with Different Models

Description of Model	Understatement Bias (in dollars)	Percent Correct within 5% of True Total Wage	Percent Correct within 10% of True Total Wage	Root Mean Square Error (in dollars)
NO OVERALL TREND ASSUMED				
1. Regression................	0	49.5	77.8	3,022
2. "Method II"...............	493	56.6	76.6	3,180
3. Extrapolating Maximum Prior Level.....................	271	59.6	77.7	3,160
4. Extrapolating Immediately Prior Level..............	507	55.5	75.8	3,179
5. Extrapolating Simple Average of Prior Levels.......	588	57.8	77.2	3,163
6. Extrapolating Weighted Average of Prior Levels*......	578	56.2	76.5	3,163
7. Extrapolating Immediate Prior Trend..............	264	39.5	61.4	3,656
8. Extrapolating Average Prior Trend....................	246	44.4	65.0	3,639
OVERALL TREND OF 1% ASSUMED				
2a. "Method II"...............	377	57.3	76.7	3,168
3a. Extrapolating Maximum Prior Level....................	148	59.9	78.1	3,157
5a. Extrapolating Simple Average of Prior Levels.......	472	59.8	78.0	3,149
OVERALL TREND OF 2% ASSUMED				
2b. "Method II"...............	281	57.0	76.8	3,167
3b. Extrapolating Maximum Prior Level....................	22	58.6	77.6	3,162
5b. Extrapolating Simple Average of Prior Levels.......	354	61.0	78.2	3,143

* More recent levels are weighted more heavily.

II and a simple average of prior quarters perform better with an assumed rising secular trend of 1 percent or 2 percent per quarter. We would tentatively suggest that this type of modification be implemented.

FOOTNOTES

*The author would like to acknowledge his appreciation to Fritz Scheuren and Beth Kilss for their gracious guidance and assistance. Thanks are also extended to the typists, Kathy Wetzel and Joan Reynolds.

1/ On top of the weights and initial raking adjustments (for survey undercoverage and failures to match to administrative records) already on the Exact Match file [11], we applied further minor adjustments to force the subsample to better reflect the known distribution of certain variables that are correlated with the performance of extrapolative technique in the population. Specifically, these adjustments were based on the assumptions that performance should depend on: (a) whether or not the recorded earnings amount for the truncation quarter is greater than in all preceding quarters; and (b) the range of, and (c) the number of quarters with prior earnings, since extrapolation error should be larger for series with greater variability and longer forecast spans.

2/ Its predecessor, Method I, could operate, at the macro level only, to produce estimates of total untaxed wages and related information, derived from curve fitting and calculus techniques [12].

3/ Our situation is different from the usual forecasting context in this respect, i.e., in that we have partial information for the forecast period.

4/ It must be noted that a small number of "outliers," i.e., large values for (y-z), can have a large effect on the magnitude of the RMSE measure.

REFERENCES

[1] Report of the Consultant Panel on Social Security to the United States Congressional Research Service, August 1976.

[2] Gallaway, L. E. Geographic Labor Mobility in the United States, 1957 to 1960, Social Security Administration, Office of Research and Statistics, Research Report No. 28, 1969.

[3] McCall, J. J. Income Mobility, Racial Discrimination, and Economic Growth, The Rand Corporation, 1973.

[4] Scheuren, F., et. al. Studies from Interagency Data Linkages, Social Security Administration, Report No. 4, 1975.

[5] Miller, H. P. Income Distribution in the United States, 1960 Census Monograph series, 1966, pp. 215-217.

[6] Gastwirth, J. L. appendix to "The Estimation of the Lorenz Curve and Gini Index," Review of Economics and Statistics, August 1972, p. 315.

[7] Internal Revenue Service. Statistics of Income, 1972--Individual Income Tax Returns, publication 79, January 1975, table 1.1.

[8] Theil, H. Applied Economic Forecasting, American Elsevier, 1971.

[9] Mincer, J. and Zarnowitz V. "The Evaluation of Economic Forecasts," in Economic Forecasts and Expectations, edited by Jacob Mincer, National Bureau of Economic Research, 1969.

[10] Friedman, M. A Theory of the Consumption Function, National Bureau of Economic Research, 1957.

[11] Studies from Interagency Data Linkages, Social Security Administration, Report No. 8, (in preparation).

[12] Resnick, M. "Annual Earnings and the Taxable Maximum for OASDHI," Social Security Bulletin, November 1966, pp. 42-43.

APPENDIX

TABLE A.--PERCENTAGE DISTRIBUTION BY SIZE AND SIGN OF METHOD II FORECAST ERROR: DISTRIBUTION OF DIFFERENCE BETWEEN THE WAGE REPORTED TO IRS AND ITS METHOD II ESTIMATE, WHEN THE CEILING QUARTER IS OTHER THAN THE FIRST CALENDAR QUARTER

EARNINGS PATTERN BY RANGE OF 1/ OF PRIOR QUARTER WAGES	TOTAL	DOLLAR SIZE OF FORECAST ERROR																	
		ABSOLUTE						OVERESTIMATES OF 51 DOLLARS OR MORE						UNDERESTIMATES OF 51 DOLLARS OR MORE					
		LESS THAN 51	51 TO 250	251 TO 500	501 TO 1,000	1,001 TO 5,000	5,001 OR MORE	TOTAL	51 TO 250	251 TO 500	501 TO 1,000	1,001 TO 5,000	5,001 OR MORE	TOTAL	51 TO 250	251 TO 500	501 TO 1,000	1,001 TO 5,000	5,001 OR MORE
TOTAL	100.00	10.60	19.79	19.69	18.85	25.13	5.93	31.00	7.53	6.99	5.33	9.03	2.12	58.39	12.26	12.70	13.52	16.10	3.81
ONE PRIOR QUARTER, TOTAL	100.00	8.55	4.77	10.88	11.58	38.74	25.49	40.76	1.73	4.50	4.71	18.15	11.67	50.70	3.04	6.38	6.87	20.59	13.82
CEILING QUARTER HIGHEST, TOTAL	100.00	3.38	3.64	14.88	4.86	40.70	32.53	38.57	1.48	8.55	1.93	19.98	6.63	58.04	2.16	6.33	2.93	20.72	25.90
CEILING QUARTER NOT HIGHEST, TOTAL	100.00	9.77	5.04	9.92	13.19	38.26	23.82	41.28	1.79	3.53	5.38	17.71	12.87	48.95	3.25	6.39	7.81	20.55	10.95
TWO PRIOR QUARTERS, TOTAL	100.00	9.09	13.65	18.24	20.70	33.96	4.37	34.85	5.73	6.52	7.25	14.04	1.31	56.07	7.92	11.72	13.45	19.92	3.06
CEILING QUARTER HIGHEST, TOTAL	100.00	14.44	17.98	21.35	20.66	17.13	8.42	19.39	7.25	7.43	2.68	2.03	0.00	66.15	10.73	13.92	17.98	15.10	8.42
LESS THAN 100	100.00	18.61	14.29	21.53	19.29	14.81	11.44	9.37	4.17	5.20	0.00	0.00	0.00	71.99	10.12	16.33	19.29	14.81	11.44
100 TO 249	100.00	6.29	22.95	26.42	33.47	10.85	0.00	34.44	16.56	6.28	6.35	5.25	0.00	59.25	6.39	20.14	27.12	5.60	0.00
250 TO 499	100.00	7.91	42.70	16.95	23.96	8.48	0.00	29.83	13.68	8.04	8.11	0.00	0.00	62.26	29.02	8.91	15.85	8.48	0.00
500 TO 999	100.00	10.48	0.00	27.92	21.95	31.00	8.65	8.79	0.00	8.79	0.00	0.00	0.00	80.73	0.00	19.13	21.95	31.00	8.65
1,000 TO 1,999	100.00	27.36	0.00	27.73	0.00	23.52	21.39	27.73	0.00	27.73	0.00	0.00	0.00	44.91	0.00	0.00	0.00	23.52	21.39
2,000 TO 4,999	100.00	25.45	0.00	0.00	0.00	48.09	26.46	18.52	0.00	0.00	0.00	18.52	0.00	56.03	0.00	0.00	0.00	29.57	26.46
CEILING QUARTER NOT HIGHEST, TOTAL	100.00	8.65	13.29	17.98	20.69	35.35	4.04	36.12	5.60	6.45	7.62	15.03	1.42	55.23	7.69	11.53	13.07	20.32	2.62
LESS THAN 100	100.00	15.95	13.37	19.47	19.18	29.09	2.95	16.07	2.75	4.38	2.26	6.68	0.00	67.99	10.62	15.09	16.92	22.41	2.95
100 TO 249	100.00	9.76	15.40	21.04	25.48	27.11	1.21	29.27	5.75	4.50	9.12	9.40	0.50	60.97	9.65	16.54	16.36	17.71	0.71
250 TO 499	100.00	3.77	16.17	23.07	22.19	31.41	3.39	39.53	5.91	11.38	9.12	11.95	1.17	56.70	10.26	11.69	13.07	19.46	2.22
500 TO 999	100.00	5.95	15.19	16.90	21.47	36.07	4.41	52.13	9.69	9.67	11.76	20.63	0.38	41.91	5.50	7.23	9.71	15.44	4.03
1,000 TO 1,999	100.00	2.30	5.11	9.36	16.50	63.24	3.49	54.40	5.11	2.23	11.64	33.21	2.21	43.30	0.00	7.13	4.86	30.03	1.28
2,000 TO 4,999	100.00	9.56	4.47	4.76	13.68	48.41	19.12	42.93	2.41	0.00	0.00	26.18	14.34	47.51	2.06	4.76	13.68	22.23	4.78
THREE PRIOR QUARTERS, TOTAL	100.00	12.37	28.93	23.36	19.50	14.31	1.51	25.18	10.62	8.06	4.01	2.49	0.00	62.43	18.31	15.30	15.49	11.82	1.51
CEILING QUARTER HIGHEST, TOTAL	100.00	14.01	52.34	17.60	7.89	2.11	6.05	2.86	1.37	0.00	1.49	0.00	0.00	83.13	50.97	17.60	6.40	2.11	6.05
LESS THAN 100	100.00	28.91	11.58	34.08	0.00	0.00	25.43	0.00	0.00	0.00	0.00	0.00	0.00	71.09	11.58	34.08	0.00	0.00	25.43
100 TO 249	100.00	26.32	65.41	8.28	0.00	0.00	0.00	6.90	6.90	0.00	0.00	0.00	0.00	66.79	58.51	8.28	0.00	0.00	0.00
250 TO 499	100.00	0.00	88.56	0.00	11.44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	88.56	0.00	11.44	0.00	0.00
500 TO 999	100.00	13.04	12.71	27.88	32.01	14.35	0.00	10.14	0.00	0.00	10.14	0.00	0.00	76.81	12.71	27.88	21.87	14.35	0.00
1,000 TO 1,999	100.00	0.00	74.65	25.35	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	74.65	25.35	0.00	0.00	0.00
2,000 TO 4,999	100.00	0.00	69.42	30.58	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	69.42	30.58	0.00	0.00	0.00
CEILING QUARTER NOT HIGHEST, TOTAL	100.00	12.28	27.56	23.70	20.19	15.03	1.25	26.50	11.17	8.54	4.16	2.63	0.00	61.23	16.39	15.16	16.03	12.40	1.25
LESS THAN 100	100.00	33.78	30.73	11.58	15.86	6.32	1.73	9.40	7.91	0.75	0.74	0.00	0.00	56.82	22.82	10.83	15.12	6.32	1.73
100 TO 249	100.00	17.08	38.04	15.82	13.81	14.69	0.56	20.05	11.88	2.94	1.62	3.61	0.00	62.87	26.16	12.88	12.19	11.08	0.56
250 TO 499	100.00	11.28	28.68	30.41	18.67	9.42	1.55	30.30	14.05	12.96	2.02	1.27	0.00	58.43	14.63	17.45	16.65	8.15	1.55
500 TO 999	100.00	5.29	22.90	30.98	24.74	15.01	1.08	30.69	10.35	13.12	5.24	1.98	0.00	64.02	12.55	17.86	19.50	13.03	1.08
1,000 TO 1,999	100.00	4.34	21.86	21.48	26.08	25.35	0.87	27.98	8.53	6.10	8.79	4.56	0.00	67.66	13.33	15.38	17.29	20.79	0.87
2,000 TO 4,999	100.00	7.90	19.57	11.36	22.22	36.66	2.29	39.64	11.37	3.83	13.99	10.45	0.00	52.46	8.20	7.53	8.23	26.21	2.29
5,000 OR MORE	100.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	100.00	0.00

Source: Derived from the 1973 CPS-IRS-SSA Exact Match Study conducted by the Census Bureau and Social Security Administration with the assistance of the Internal Revenue Service.

1/ The difference between the largest and smallest wage amounts in quarters prior to the ceiling quarter.

MEASURING THE WORLD MIDDLE CLASS

Nathan Keyfitz, Harvard University

Each discipline has its own way of looking at a social phenomenon and needs its own means of measurement. For economics production is central, and Gross National Product, using market prices as weights, usually divided by population, measures the amount of development. Sociologists see development as people changing their way of life, ceasing to be peasants and becoming farmers, or moving to the city and becoming middle class. When the people undergoing development themselves contemplate development, they see the object of their strivings as becoming literate, moving to the city, getting a job in a factory or, best of all, in an office. If they are beyond the age where change of status is feasible, they have these ambitions for their children. The middle-class way of life at which they aim involves using more and different kinds of food, clothing, housing, transport and recreation, and so money income and expenditures are the means by which the individual attains and maintains the middle-class life.

It is thus useful for some purposes to consider income, production, and expenditures as the means to development, and the modern way of life as the goal. That still leaves income as an indicator of development, but it is one among many. Broadening the statistical base, as this paper proposes, does not make empirical work immediately easier. Problems of finding data, of aggregating it, and of interpreting the resultant aggregation all abound.

The subject is in the condition that national income was in before Kuznets, Stone, and others did the conceptual work required for measurement of GNP and before Keynes developed a theory of the economic cycle that made it important. Yet if concern with attaining the middle-class style of life, for countries as different in other respects as Brazil and the USSR, is the objective of those involved, then we should try to find out how many individuals have attained it, and how many are attaining it each year. This despite difficulties of definition and measurement.

The GNP measures what passes through the market and makes only rough allowance for, or omits altogether, what is produced by the family for its own use. When peasants cease to make their own soap and buy it instead, a transition that has taken place within this century in French Canada and many other parts of the world, then use of soap becomes economically visible and can be entered into the national income accounts; when eating in restaurants increases, the labor of meal preparation acquires economic significance, which it does not have when meals are prepared at home. The GNP notes the raising of children as performed by professionals in day-care centers, but not when done at home. Despite heroic efforts to include all activities that may be called productive, the national accounts have a difficult time with those that do not pass through the market. The work of women in child care and housekeeping must constitute a quarter of total labor in most countries; its more exact valuation would depend on the birth rate and on the quality that could be imputed to such labor. But its concentration on the expanding area of exchange, which after all is an important facet of development, should be seen as the strength rather than the weakness of the national accounts.

A theory of employment that supposes distribution will take care of itself makes its measurement a subordinate objective of the national accounts. But distribution may well be central for development. If an average income of $4000 per year is the resultant of 1/100 of the population living at $380,000 per year and the other 99/100 living at $200 per year, the prognosis for development is very different from everyone having $4000. As evaluated by individuals in poor countries, some differentiations are more important than others--the jump from $100 to $10,000 can be incomparably more important than that from $10,000 to $100,000. This sense of a diminishing marginal utility of income is interpretable in terms of diminishing changes in way of life with successive increments of income. Not only does income as perceived by those concerned and their peers apparently have a break at the point where poverty shades into middle class life, when the individual most clearly changes his way of living, but the demand for energy may well have a discontinuity here also.

There is no society in which the individual is the economic unit (the need to raise children prevents this if nothing else does). People everywhere live in families, pool their incomes, and make no record of exchanges of money or of work within the family. Hence individual income is not a measurable concept. The distribution of income among the members of particular families being impossible to ascertain, the distribution of income among the individual inhabitants of a country can have little meaning. It is family incomes and their distribution that count.

But families are of different sizes. How do we interpret distribution when a family can have one or ten members--may consist of a widow, or a couple, or a father and mother and eight children? Simon Kuznets has investigated this and has developed methods for coping with it, always making use of money incomes.

Alternatively we will seek the point where poor shades into middle class, and describe the distribution in terms of this one cut. A man may be the sole earner in a family, but the whole family is poor or else the whole family is middle class. The Social Security Administration makes such a cut at the upper edge of poverty, and it classifies individual families according to where they stand in many different cross classifications. These include family size, age and sex of family head, farm-nonfarm residence, and income. Physical criteria are kept in view; it was judged, for instance, that in the base year 1970 a minimum income of $3743 was needed in urban areas to meet the nutritional and other needs of a family of four.

Matters would be much easier for the statistician trying to mark the upper bound of the poverty group if agreement could be reached on what constitutes minimum housing, minimum clothing, minimum nourishment. Not only have experts in the several fields little to say about such minima, but the expenditures of the poor themselves imply that their priorities are very different from those of any experts who would be hardy enough to prescribe for them: they may sacrifice nourishing food in favor of tickets to a baseball game; they may trade the family car for a newer model when what they "need" is medical services.

Yet when all that is said, a degree of uniformity appears in the purchases made at a given income. The couple that begins in middle-class life, or that climbs out of poverty, in any part of the world acquires as soon as possible a standard package of equipment that includes a dwelling with electricity and central heating or air conditioning; a refrigerator; a television set; an automobile. Until this basic equipment is in its possession the couple borrows up to the limits of its capacity, and only when it has obtained these artifacts does it think about saving.

In short, we need to recognize two components of growth. One is higher income in situ--peasants doing better but remaining peasants, laborers receiving higher wages but remaining laborers, rich people increasing their incomes. The other is transition across the poverty line, people going from poor to middle class.

For the United States we know from official sources (U.S. Statistical Abstract 1975, p. 400) that the fraction of households below the poverty line was 18.4 per cent in 1959 and 10.7 in 1969. That the mean income of all families rose by 75 per cent in current dollars, or 39 per cent in real dollars, during the same ten years needs to be supplemented by the fact that the fraction poor fell by 40 per cent.

Another measure of welfare is the proportion of income spent on food. According to successive family budget surveys in the United States (Monthly Labor Review, July 1974, 97: 8) this has gone down from 35.4 in 1935-39 to 29.6 in 1952 to 22.4 in 1963. The non-food expenditure for each unit of food has correspondingly risen from 1.82 to 2.38 to 3.46 in the same three surveys.

Corresponding to this on the production side is the number of non-farm workers for each farm worker. That ratio has gone from 51,760,000/7,160,000 = 7.2 in the United States in 1950 to 80,377,000/3,171,000 = 25.3 in April of 1975 (Statistical Abstract 1975, p. 343). In a present day poor country it may be as low as 0.5. The number of minutes of factory work required to buy a one-pound loaf of bread is a related indicator.

A conspicuous change in style of life occurs with the acquisition of an automobile. In the very first years of development, before any automobiles are scrapped, production figures indicate the number of individuals making that change; after a few years we want users of automobiles and not buyers. We could take sales less automobiles scrapped, except that scrapping is difficult to measure. We could take first purchases by families that do not have a car, but this again is not easy to ascertain. Most available is private automobile registrations, although one would like to avoid counting families twice if they have two cars. Unfortunately, the number of families with any cars registered is not widely available in national statistics.

Yet if geographical movement is what the automobile is good for, we must take into account that alternatives exist. A middle-class family in Europe is less likely to have a car than one of the same status in the United States, more likely to travel by streetcar, bus, and train. Intertemporal comparisons, say between the United States in 1920 and today, involve the same difficulty as international comparisons. We cannot stop with automobiles in use.

One supplement is homes served with electricity. Where this cannot be obtained, a proxy is the amount of electricity in use, if possible subtracting electricity that goes into industry. Running water and indoor flush toilets constitute measures of the middle class, with characteristic distortions. So also do television sets and school attendance.

Physical appurtenances and activities are not only indexes of a way of life, but they are themselves active agents in the change of mentality that is a part of development. Watching television affects a family's view of the world. Schooling that brings effective literacy is an indicator of the wish to be socially mobile,

TABLE 1 Per cent of family expenditure on three groups of items, United States, 1935-39 to 1963

	1935-39	1952	1963
Food	35.4	29.6	22.4
Transport	8.2	11.3	13.9
Health and recreation	11.7	17.4	19.5

Source: Monthly Labor Review, 97, 7: 8

and in addition the practice of reading generates habits of thought and behavior that conduce to mobility. Thus the number of persons who have completed some level of schooling, say 10 years, might correspond to the number above the poverty line.

Access to medical services is required to support the middle-class attitude towards sickness and death. We have data on physicians country by country that might serve as an indicator of the number of families that enjoy medical services. Unfortunately we have no guide to the quality of services provided, nor to the distribution of services among families. On the one hand the services might be dispersed among the entire population, so that all get some but no one gets enough. More commonly they are available only to people in cities, and especially to those who are well above the poverty line.

Medical services provide an example of the difficulty of comparing money incomes. In the Soviet Union a doctor is said to be paid something on the order of $135 per month, and the citizen gets medical services without paying for the doctor's time even at this rate. In the United States a physician who failed to earn 20 times as much as his opposite number in the Soviet Union would be badly off, and the patient pays for the doctor's time out of his own pocket. To compare medical services in the two countries by their cost seems less satisfactory than using numbers of physicians, and that is not very good either.

The several middle-class facilities do not come simultaneously but in a sequence; most families on the rise will acquire electricity, then perhaps piped water, then a television set, then an automobile, then a telephone, then write and receive mail at the middle-class average of something like one letter per person per day. The concept of a standard package is not to be taken so literally as to preclude the items being purchased in a sequence. We should be able to use the numbers of the several facilities to see roughly the order in which the appurtenances of middle-class status are acquired. Pending more appropriate data, I have converted total consumption into persons by multiplying by the United States ratio of consumption per person. For Mexico about the year 1970, we have the following twelve indicators (U.S. Statistical Abstract 1975, pp. 840-860):

	Estimated number of persons (millions)
Homes with electric lighting at 4 persons per home	19
Middle, secondary, or high school (at the rate of the new generation)	15
Homes with piped water, at 4 persons per home	13
Meat consumption at U.S. standard of 75 kg per person per year	13
People provided with hospital beds, at U.S. standard of 135 people per bed	9
Television sets at U.S. standard of 1.9 persons per set	8
Steel consumption at U.S. standard of 0.7 metric tons per person per year	7
Total energy consumption at U.S. standard of 12 tons coal equivalent per person per year	6
Consumption of electricity at U.S. standard of 9300 kwh per person per year	4
Automobiles at U.S. standard of 2.1 persons per automobile	4
Telephones at U.S. standard of 1.5 persons per telephone	3
Domestic mail sent at U.S. standard of 400 pieces per person per year	3

This display makes no pretence of providing definitive figures, but is rather an occasion for comment on the meaning of the 12 crude indicators shown. Automobiles and telephones may be the least unsatisfactory, though if there is a difference between the United States and Mexico in the proportion for business use, or the proportion of families having two or more, etc., that would make even these figures wrong. Electric lighting and piped water in the home and meat consumption do not by themselves qualify for middle-class status on an intuitive definition; some poor do have access to hospitals, perhaps in Mexico more than in other countries. Steel, total energy consumption and electricity are too high insofar as a larger proportion of these goes into investment and other collective, non-personal uses in Mexico than in the United States. (Investment will produce middle-class people in due course, but it does not directly measure their presence.) School attendance tells something about the status of the younger generation; the way the above number was calculated, middle-class status was also imputed to parents and grandparents of the pupils. Mail gives too low a number, because few other countries are as thoroughly saturated with commercial mail as the United States.

But the main difficulty of the table is its attempt to make a dichotomy--poor versus middle class--out of what is in most instances a substantially continuous distribution. Some poor people do receive mail, do have television sets, do use automobiles; on the other hand, some rich may not have these things or, more commonly, have more than one. The right way to proceed is to classify individuals according to combinations of items. One would take a sample survey covering the matters of the above table and others, and then try to see to what extent they are scalable, which is to say, come everywhere in the same sequence; and whatever the degree of scalability, one would try to find the point that would effectively discriminate the two groups into which the population could most meaningfully be divided.

The same criterion applies to this display as to early work on the national accounts--the test of success will be its arousing enough criticism, and enough efforts at improvement, to produce better figures.

Having defined the middle-class style of life in physical terms and then estimated its number, one would seek data to draw its energy and other implications. If the person drives 8000 miles per year he consumes about 400 gallons of gasoline, extracted from 20 barrels of oil. If he has a 1500-watt airconditioning unit and uses it 2000 hours during the year, a modest amount of cooling, then he draws 3000 kwh of electricity. His electric lighting may draw 200 watts for an average of three hours per day, or a little over 200 kwh. His refrigerator may draw 1000 kwh. Suppose in all he uses 5000 kwh per year and that 25 kwh are produced by a gallon of oil. Then he adds 200 gallons of diesel or other oil to the 400 gallons of high-test consumed by his car.

In contrast with these expenditures the peasant uses no electricity and perhaps a gallon or two of kerosene for lighting. The increase of income as one goes from peasant to middle-class status could well involve a discontinuity in resource use. Of course peasants of different incomes do have different amounts of consumption, and these have different materials and energy components, but such variations could well be less per dollar of income difference than differences across the poverty line.

FURTHER EXPLORATION OF CPS-IRS-SSA WAGE REPORTING DIFFERENCES FOR 1972

by Beth Kilss and Wendy Alvey*
Social Security Administration

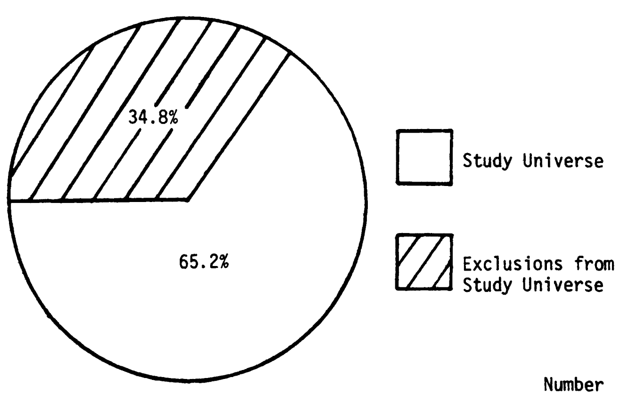
This paper compares wage data from the Census Bureau's Current Population Survey (CPS) with Social Security Administration (SSA) and Internal Revenue Service (IRS) administrative data on wages and salaries for persons eligible for interview in the CPS who filed a nonjoint tax return for 1972. Three-way wage comparisons are made to examine the extent of wage class agreement and evaluate CPS wage reporting. The data base that is used in this analysis was obtained by matching CPS, IRS, and SSA information for the same individual, as described elsewhere in these (and earlier) Proceedings.1/

In the first part of this paper, the study universe and the nature of the exclusions from it are defined. Next, some results obtained by comparing wages from the three sources are presented; and, finally, the study universe is be examined by several variables affecting the quality of wage reporting in the CPS.

RECONCILIATION OF CPS, IRS, AND SSA POPULATIONS

The Current Population Survey is essentially a sample of the civilian noninstitutional population of the 50 States and the District of Columbia. Thus, the population of those who file nonjoint returns includes certain individuals not found

Figure 1. --Derivation of Study Universe



	Number (in Millions)
All 1972 Nonjoint Returns (Estimated from 1972 Statistics of Income)	34.8
Less:	
CPS-IRS Noncomparable Nonjoint Returns	2.0
Equals:	
CPS-IRS Comparable Nonjoint Returns	32.8
Less:	
Earners with Imputed CPS Wages	3.2
Earners with SSA Self-Employment Income	0.9
Earners with No Wages from One or More Sources	6.0
Equals:	
Study Universe of CPS-IRS-SSA Comparable Nonjoint Returns	22.7

among persons eligible for interview in the CPS. These returns must be excluded from the study universe. In addition, because only nonjoint returns with wages from all three sources are being analyzed, further deletions are necessary. Figure 1 briefly summarizes the adjustments made.

As can be seen from the pie chart at the top of figure 1, the study universe is 65.2% of the 34.8 million nonjoint tax returns for 1972. The lower half of this figure provides a list of those that were excluded, the order in which they were excluded, and the size of each. First, approximately two million CPS-IRS noncomparable returns were deleted. These included returns for Armed Forces members, institutionalized individuals, and persons living overseas. Also included were persons filing tax returns for 1972 but who died prior to the March 1973 interview date and persons under 14 years of age. The exclusion of the latter group was necessary, since the March Supplement to the CPS does not ask income questions for them. Also, only 1972 tax-year returns were matched to the CPS; thus, any prior year delinquent returns had to be subtracted out, as well.2/ After these exclusions, 32.8 million CPS-IRS comparable nonjoint returns remained.

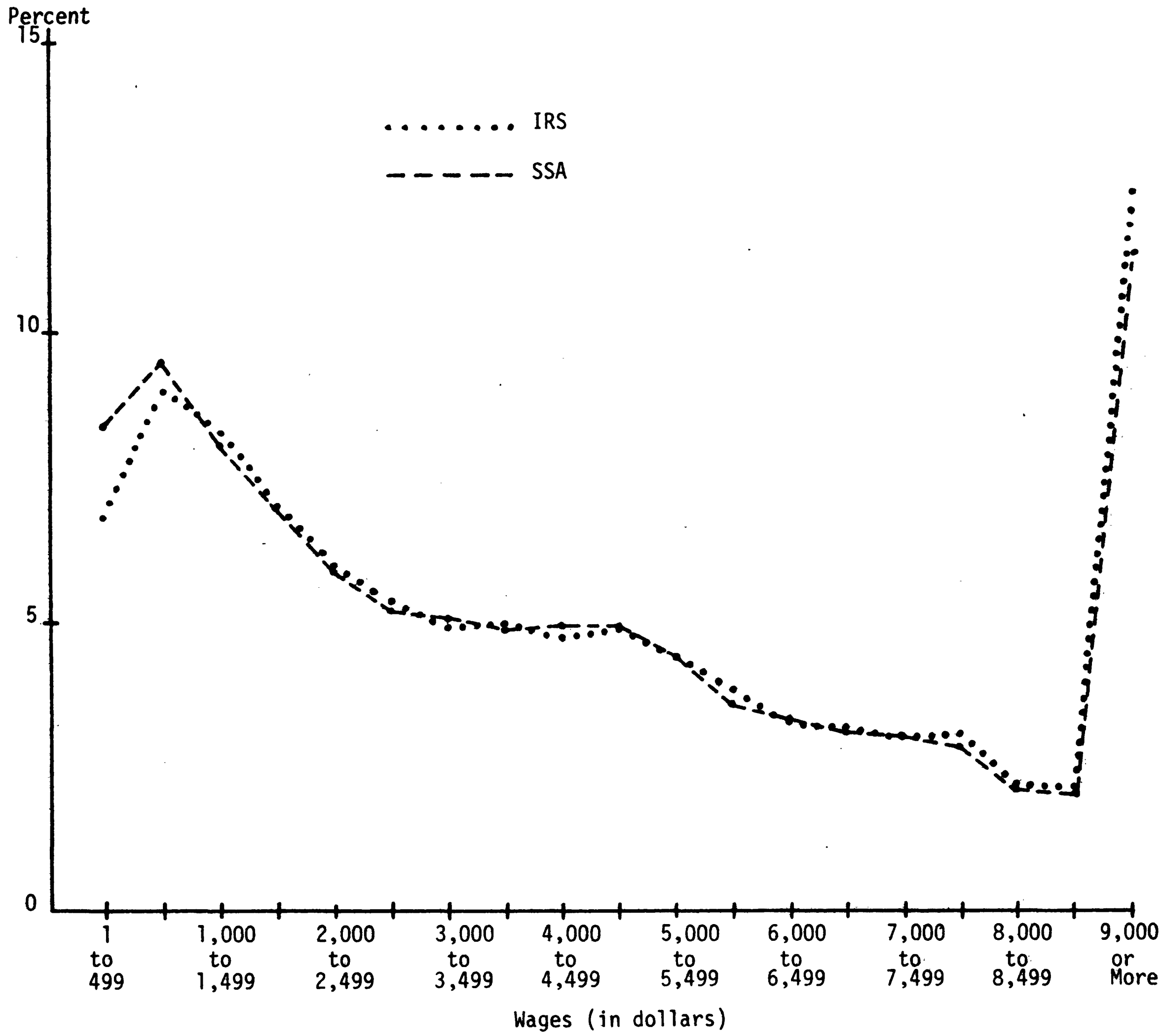
Three other groups were excluded from our analysis.3/ Persons not reporting their wages to the CPS enumerator (and for whom wages were imputed) accounted for the deletion of 3.2 million returns from the study universe. Then, since SSA's computerized administrative files do not distinguish between an individual's wages and self-employment income (this distinction is not required in calculating benefits), about 0.9 million earners with SSA self-employment income were subtracted out.4/ Finally, about six million earners with no reported wages from one or more of the three sources were deleted. This reconciliation yielded a study universe consisting of 22.7 million CPS-IRS-SSA wage earners who filed nonjoint tax returns for calender year 1972. It is this group which is dealt with below.

COMPARISON OF WAGE DISTRIBUTIONS FROM THE STUDY UNIVERSE

Before examining CPS, IRS, and SSA wage class agreement for the same individual, it is interesting to compare wage distributions for the study universe as a whole.

Figure 2 shows the percentage distribution of wages from administrative sources -- IRS and SSA. The 1972 wages are plotted along the horizontal axis in intervals of $500, up to $9,000. Since SSA wages are only reported up to the taxable maximum (which was $9,000 in 1972), for purposes of comparisons among the three sources, it was decided to end the distribution at this point. Therefore, the last class is an open-ended

Figure 2. --IRS and SSA Wage Distributions



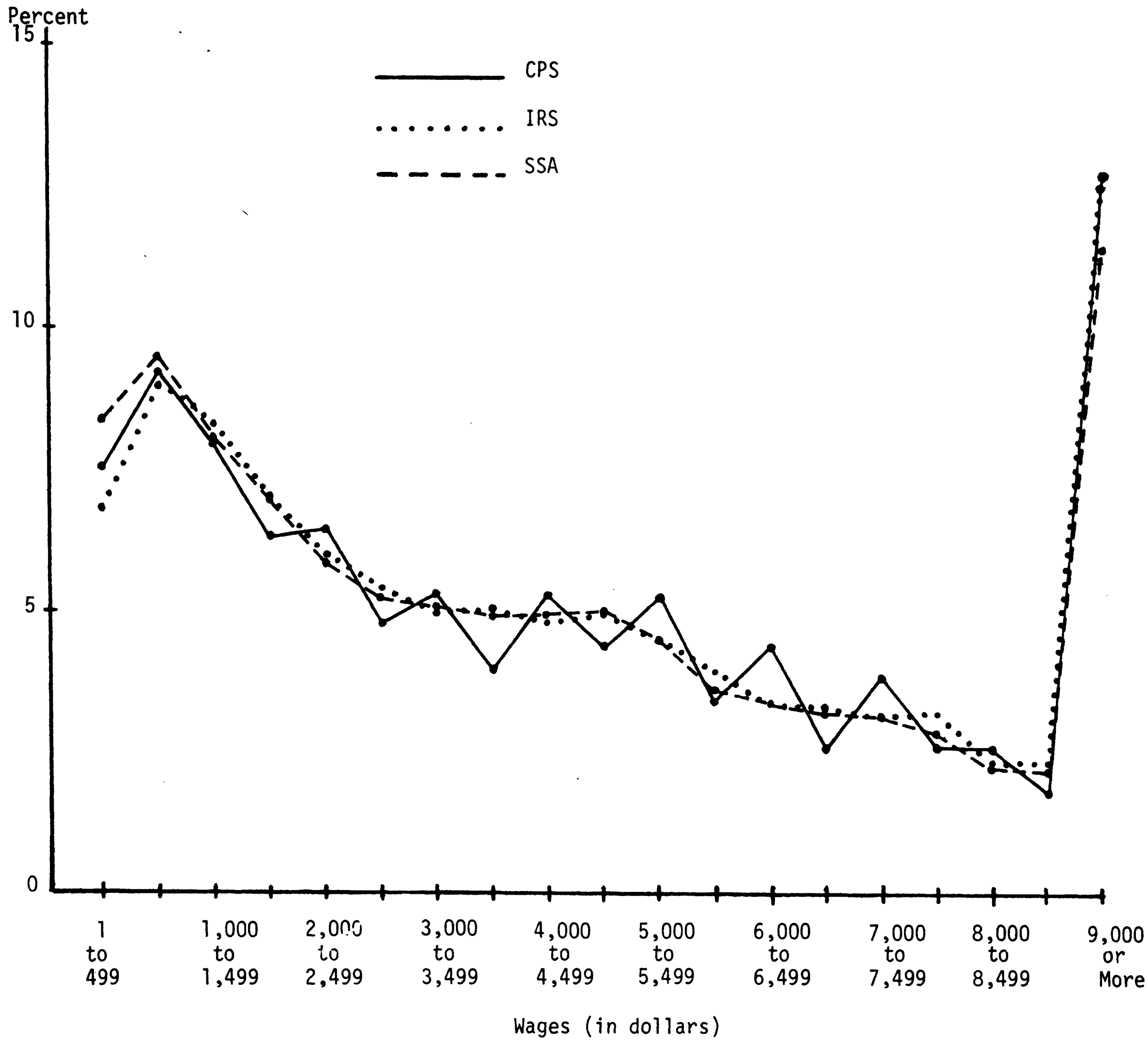
interval of $9,000 or more. The dotted line depicts the IRS distribution; the SSA distribution is shown by the broken line.

For the most part, the reporting of SSA wages appears to be nearly identical to that of wages obtained from IRS data. This was expected since wages from both sources are generally reported from carefully compiled payroll records, with legal penalties imposed on the taxfiler or his employer for fraudulent reporting. For a large part of both distributions -- from $1,000 to $8,999 -- the proportion of wage earners in each class declines as wages increase. However, two main differences in the IRS and SSA distributions are apparent: for persons with less than $1,000, the percent of those with SSA wages notably exceeds the proportion with IRS wages; while, for the $9,000 or more wage class, the reverse is true.

Definitional differences are probably the major factor in these two instances: SSA taxable wages include all wages received by an employee for services rendered in covered employment up to the annual taxable maximum for each employer. However, combinations of covered and noncovered employment are reported on Federal tax returns. It is this, no doubt, which accounts for most of the increased proportion of IRS wage earners beyond the taxable maximum, and, consequently, the smaller proportion at the lower end of the distribution.

Keeping in mind the similarities and differences between these two sources, consider the distribution of CPS wages shown by the addition of the solid line to the distributions presented in figure 2. (See figure 3.) Although the survey data follow the same overall pattern as the administrative data, the distribution is more jagged. This may be attributed to the fact that people often tend to report wages rounded to the nearest thousand dollars [9], thus leading to under- or overstatement of the CPS amount. Also, it is important to remember that wage classes, and not exact wage amounts, are being compared. Therefore, the placement of the upper and lower limits of each interval, as well as the length of each class, all affect what is being observed here.5/

Figure 3. --CPS, IRS, and SSA Wage Distributions



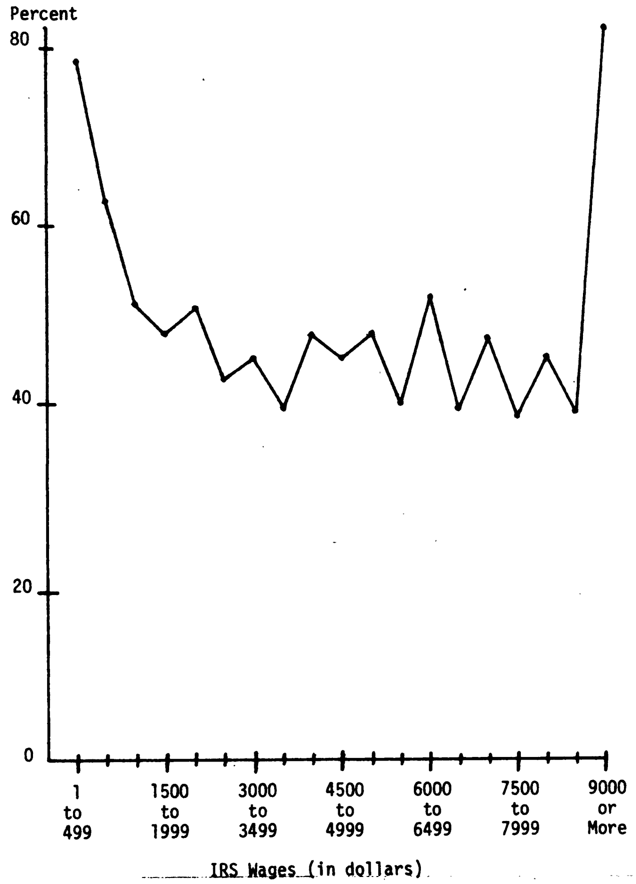
EXTENT OF CPS-IRS-SSA WAGE CLASS AGREEMENT

In the previous section, comparisons were made between marginal distributions of survey and administrative wage data. For purposes of further analysis, CPS, IRS, and SSA wage disparities will be discussed by examining their joint distribution. To do this, wages from all three sources were cross-classified in $500 classes. Agreement was said to exist if wages from more than one source fell into the same $500 interval. Several possible types of agreement resulted: three-way wage class agreement, various combinations of two-way agreement, and cases where none of the wages from any of the sources fell into the same wage class. While tables have been prepared presenting cross-tabulations for all of these agreement types,6/ the analysis in the remainder of this paper will be limited to an examination of three-way CPS-IRS-SSA wage class agreement.

As in figures 2 and 3, the horizontal axis in figure 4 is divided into $500 wage intervals. For the most part, three-way agreement fluctuates between 38 percent and 52 percent for the major part of the distribution. Agreement for the less-than-$1,000 or $9,000-or-greater categories is higher than for the middle-range size classes. This is partly because individuals at the tails of the distribution can only have wage disagreement in one direction. If you will notice, for the wage classes between $1,000 and $9,000, peaks are occurring for the intervals having lower limits which are multiples of $1,000, while the troughs in the graph fall in those classes with lower limits ending in "500." This same sawtoothed effect was observed for the CPS wage distribution in the previous chart.

What are some of the possible factors that may be contributing to these findings? Keep in mind that wages from different sources (administrative and survey) are being compared for the same individual. Because of the very nature of survey data, reporting differences occur. Some of the major "interview factors" which may be affecting CPS reporting are:

Figure 4. --Percent of Three-Way Agreement by IRS Wage Class



1. Type of interview.--manner in which survey information was obtained:

   a. Personal--face to face interview;
   b. Telephone--interviewer phoned individual for answers to survey questions.

2. Response status.--the relationship between the individual answering the survey questions and the individual about whom the survey questions were being asked:

   a. Self-response--the individual answered for himself;
   b. Proxy-response--answers to the survey questions were provided by someone else in the household.

3. Roundedness of CPS wages.--extent to which individual reported exact wage amount in survey interview:

   a. Nonrounded--wage amount ended in a number other than '000' or '500';
   b. Rounded--wage amount ends in '000' or '500'.

These factors are linked to the quality of CPS reporting, and, hence, to the agreement of CPS, IRS, and SSA wages for the same individual.7/ For example, it is expected that an individual who provides wage information from a Form W-2 during a personal interview will be much more likely to have his CPS, IRS, and SSA wages in agreement than a person whose spouse approximated the wage amount in a telephone interview.

In other words, a higher percent of three-way agreement was expected to occur for CPS wage data obtained in personal interviews rather than in telephone interviews, in interviews where the individual responded for himself rather than in cases where the responses were obtained from another household member, and in cases with nonrounded CPS wages rather than with rounded wages.

To test these suspicions, three-way agreement results were examined for each interview factor. The distribution was divided into three sections: less than $1,000, between $1,000 and $8,999, and $9,000 or more.

As shown in figure 5, the expected results occurred for each of the interview factors in the middle-range wage classes; that is, for those with wages of $1,000 - $8,999. However, some differences in the expected agreement pattern occur for wage classes at the extremes.

For example, for wages less than $1,000, three-way agreement is much higher in cases where the individual answered the survey questions for

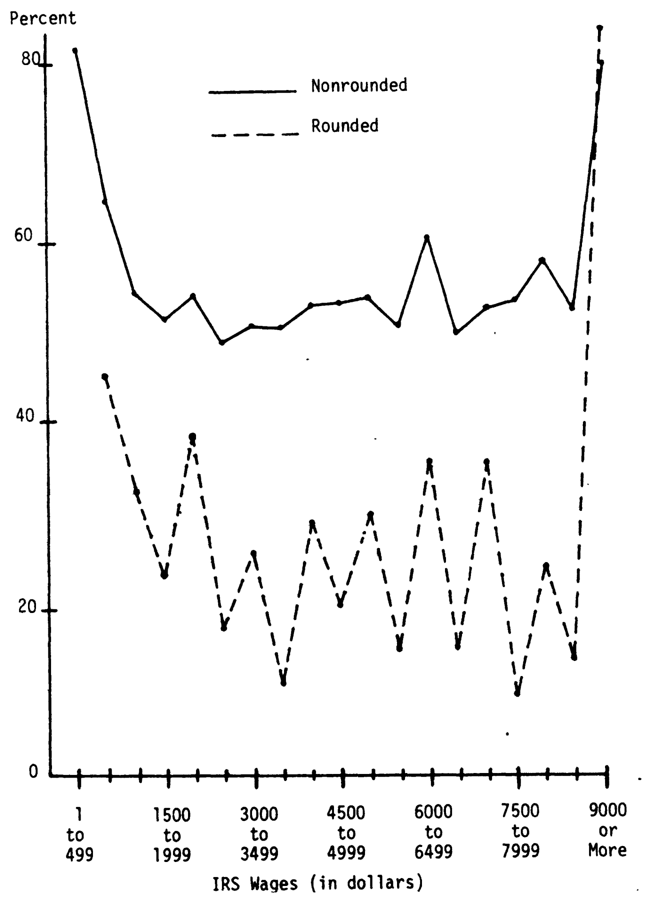
Figure 5.--Percent of Three-Way Agreement by Selected Interview Factors

Selected Interview Factors	IRS Wage Class		
	Less than $1,000	$1,000-$8,999	$9,000 or more
Total.........	61.2	46.4	89.7
PERSONAL INTERVIEW			
Total.........	61.6	47.0	89.8
Self-Response...	50.0	47.8	92.0
Nonrounded.....	55.7	55.2	89.3
Rounded........	7.5	21.6	95.0
Proxy-Response..	64.2	46.4	86.1
Nonrounded.....	66.4	53.4	88.1
Rounded........	29.5	24.0	84.2
TELEPHONE INTERVIEW			
Total........	60.4	45.8	89.5
Self-Response...	46.4	46.8	92.0
Nonrounded.....	48.6	57.0	90.8
Rounded........	33.0	21.3	93.2
Proxy-Response..	63.9	45.0	83.5
Nonrounded.....	67.6	51.6	82.7
Rounded........	30.8	27.9	84.1

someone else in the household rather than for himself. For wages of $9,000 or more, the behavior of response status is similar to that seen for the middle-range intervals; however, the expected trend for the roundedness of CPS wages is reversed. In other words, among persons in the highest wage class, three-way agreement for rounded CPS wages nearly always exceeds that of nonrounded cases. This is true for three out of the four combinations of interview factors shown in figure 5. Such an occurrence is not unexpected, since any individual reporting wages of $9,000 or more to all three sources will have three-way agreement, regardless of the roundedness of the actual amounts reported.

In order to further illustrate the effect of the interview factors on the levels of agreement, it is interesting to look at three-way agreement by one of the variables. Hence, it was decided to examine roundedness of CPS wages, the variable which appears to be most dominant in figure 5. In the next chart (figure 6), the solid line represents the percent distribution for nonrounded CPS wages and the broken line depicts that for rounded amounts.

Figure 6. --Percent of Three-Way Agreement by IRS Wage Class, Separately by Roundedness of CPS Wages



Three-way agreement for nonrounded CPS wages exceeds that of the rounded cases in every dollar size-class except the last. Furthermore, the sawtoothed effect noticed for the overall distribution in figure 4 is much less evident for persons with nonrounded survey wages. On the other hand, the reverse is true for the corresponding rounded cases: jaggedness of the distribution is far more pronounced than for the total study universe.

This, therefore, further supports the contention made earlier that at least one of the interview factors -- roundedness of CPS wages -- may be a useful predictor of survey wage agreement for the cases examined here. In the near future, regression analysis and other statistical techniques will be used to further study numerous variables, including the interview factors discussed in this paper, in an effort to determine their influence on the extent of agreement among wages from various sources.

FOOTNOTES

* The authors would like to thank H. Lock Oh and Penny Johnston for their assistance in preparing this paper. Helpful editorial comments were received from Ben Bridges, Fritz Scheuren, and Bertram Kestenbaum. Thanks are also due to Catherine Murphy and Deborah Dillard for their typing efforts.

1/ At the 1975 ASA meetings in Atlanta, several papers were presented which focused on the conceptual and reporting differences among the linked CPS, IRS, and SSA data sets. Some of the preliminary analyses presented included comparisons between matched CPS and IRS income data for 1972 [1], and similar comparisons for SSA and IRS wage data [2]. For more information on the basic study see [3] which appears elsewhere in these 1976 Proceedings. See also [4].

As mentioned in most of these referenced papers, all interagency data linkages were performed solely by Census Bureau personnel. Neither IRS nor SSA had access to identified records from each other's files. The tables of matched data used for this paper were produced by the Social Security Administration. However, the file used by SSA could not have been used for other than statistical purposes, since it was simply a random sample of unidentified records.

2/ The population total for CPS-IRS comparable nonjoint returns shown here is slightly smaller than the initial estimate given in [5]. Persons interested in full details should consult [6].

3/ Estimates for these groups and for the analysis of the study universe were obtained using sample weights which incorporated some adjustments for nonmatches and mismatches. These (initial raking) weights are described in detail in [7].

4/ Microfilm files are kept at SSA of the actual wage amounts received from each employer; however, these files were far too expensive to examine on a wholesale basis for the 1973 study. Use has been made of the microfilm files, though, on a subsample basis, as is discussed in [8].

5/ In order to take this factor into account, these distributions were examined by both $500 and $600 intervals. The pattern which emerged was similar for both cases.

6/ Copies of the tables from which these distributions were taken were available in handouts distributed after the session and can be obtained from the authors by writing to them at--Division of Economic and Long-Range Studies, Office of Research and Statistics, Social Security Administration, 1875 Connecticut Avenue, N.W., Washington, D.C. 20009.

7/ Two more papers which may be of interest, dealing with other factors affecting CPS reporting of income data appear in these Proceedings. See [3] and [10].

REFERENCES

[1] Herriot, R. and Spiers, E., "Measuring the Important on Income Statistics of Reporting Differences between the Current Population Survey and Administrative Sources," 1975 American Statistical Association Proceedings, Social Statistics Section, pp. 147-158.

[2] Millea, M.T. and Kilss, B., "Exploration of Differences between Linked Social Security and Internal Revenue Service Wage Data for 1972, "1975 American Statistical Association Proceedings, Social Statistics Section, 1976, pp. 138-146.

[3] Vaughan, D., and Yuskavage, R., "Investigating Discrepancies between Social Security Administration and Current Population Survey Benefit Data for 1972," 1976 American Statistical Association Proceedings, Social Statistics Section.

[4] Studies from Interagency Data Linkages, Report No. 4, Social Security Administration, 1975.

[5] Sailer, P., and Vogel, L. "Exploration of Differences between Linked Current Population Survey and Internal Revenue Service Income Data for 1972," 1975 American Statistical Association Proceedings, Social Statistics Section, 1976, pp. 129-137.

[6] Studies from Interagency Data Linkages, Report No. 9, Social Security Administration, (in preparation).

[7] Studies from Interagency Data Linkages, Report No. 8, Social Security Administration, (in preparation).

[8] Kestenbaum, B., "Evaluating SSA's Current Procedures for Estimating Untaxed Wages," 1976 American Statistical Association Proceedings, Social Statistics Section.

[9] Knott, J.J., "An Analysis of the Effect of Income Rounding in the Current Population Survey," 1971 American Statistical Association Proceedings, Social Statistics Section.

[10] Stevens, J. and Bailar, B., "The Relationship between Various Interviewer Characteristics and the Collection of Income Data," 1976 American Statistical Association Proceedings, Social Statistics Section.

A REVIEW OF RANDOMIZED RESPONSE MODELS AND SOME NEW RESULTS

Jong-IK Kim, Temple University
John A. Flueck, Temple University

1. INTRODUCTION

Randomized response techniques (RRT's) have been proposed and utilized for reducing response error problems when potentially "sensitive questions" are present in surveys of human populations. Warner (1965, 1971) introduced and pioneered in the development of the RRT and numerous others have sharpened and extended this approach (Table 1.).

The basic idea of the RRT is to create a stochastic (as opposed to a deterministic) relation between an individual's response and the question such that the characteristic of interest still is estimable. This stochastic relation provides some protection or "confidentiality" to the respondent and hence presumably increase his willingness to respond and to respond truthfully.

The paper will review briefly a number of RRT models and present a new multiple question RRT model. Some efficiency comparison and a few comments on applications will be offered.

2. REVIEW OF RRT MODELS

Table 1 presents a list of presently published RRT models with some identifying information. We will briefly review each of these sixteen models.

In initiating the work on RRT models, Warner (1965) presented a two related question single sample model for estimating the population proportion (π_A) of people who possessed a sensitive characteristic A. The model requires the respondent to use a randomization device in order to determine (unseen by the interviewer) which of the two related questions ($A,\bar{A}$) he will answer (Table 1). The two questions are of the form;

1. I am a member of group A,

2. I am a member of group Not-A,

and the randomization device selects question A with probability p and question $\bar{A}$ with probability (1-p). Thus, the unconditional probability of a "yes" answer under the assumption of truthful reporting is simply the sum of the two joint probabilities,

$$\begin{aligned}\Pr(\text{yes}) &= \Pr(\text{selected } A|A)\times\Pr(A) + \\ &\quad \Pr(\text{selected } \bar{A}|\bar{A})\times\Pr(\bar{A}) = p\pi_A+(1-p)(1-\pi_A). \quad (1)\end{aligned}$$

The unrestricted method of moments estimator of $\pi_A(\overset{m}{\pi}_A)$ is obtained by substituting n_1/n (the number of yes answers to the total number of answers) for the Pr(yes) in (1) and solving for $\overset{m}{\pi}_A$. This gives

$$\overset{m}{\pi}_A = \left(\frac{n_1}{n} - (1-p)\right) / (2p-1) \quad p \neq 1/2 \qquad (2)$$

which is also the maximum likelihood estimator ($\hat{\pi}_A$) of π_A as presented by Warner (1965). The variance of $\hat{\pi}_A$ is

$$\operatorname{Var}(\hat{\pi}_A) = \frac{\pi_A(1-\pi_A)}{n} + \frac{p(1-p)}{n(2p-1)^2} \quad p \neq \tfrac{1}{2}. \qquad (3)$$

Warner noted that the first term in (3) is the variance of a sample proportion under simple random sampling and the second term is the increase due to the RRT. He compared the mean square error (MSE) of this RRT model with the MSE of the direct question approach for various parameter values, including varying degrees of truthful reporting, and found that except for the cases of relatively small bias the MSE relative efficiency of the RRT could be substantial.

Warner also introduced the contamination design via the random determination of whether a direct sensitive question should be answered truthfully or falsely. This alternative RRT model (randomization of the truthfulness of the answer) produced estimators of π_A which were identical with those of equations (2) and (3). Lastly, he indicated that the RRT model could be easily extended from the dichotomous to the multichotomous population case.

Abul-Ela, <u>et al</u>. (1967) extended the RRT to the multichotomous population case by utilizing $Q(Q\geq 1)$ sensitive and Q' $(Q'>0)$ non-sensitive mutually exclusive related questions and n samples (Table 1). The MSE relative efficiency of this model compared with the direct question model was at times substantial, and the first field application of a trichotomous version apparently was not very successful. The field use of a RRT model appears to require careful training of the interviewers.

Greenberg, <u>et al</u>. (1969) modified the original Warner model by substituting an unrelated nonsensitive question (B) for the second question (Table 1). One or two samples (i=1 or i=1,2) are now required depending on whether the unrelated nonsensitive question character parameter value is known or unknown. This unrelated question model also was extended to the multichotomous case. Using the unrelated question RRT model, results from a field survey of "illegitimate births" were well supported by a record check. Also, Greenberg, <u>et al</u>. (1971) modified this unrelated question model to accommodate quantitative data.

Gould, <u>et al</u>. (1969) extended the unrelated question model to incorporate two RRT trials per respondent in an effort to reduce variability and improve the quality of the responses.

Moors (1971) increased the efficiency of the unrelated question model, when both π_A and π_B

Table 1. Published Randomized Response Models

Model Number	Reference	Question Structure	Number of Questions	Related Questions	Qualitative or Quantitative	Number of Questions Asked	Number of Trials
1	Warner JASA, 1965	1. (A_{11}, A_{12}); $A_{12} = \bar{A}_{11}$ 2. $(A_{11} \to T,\ A_{11} \to F)$	1. 2 2. 1	1. Y 2. -	QL	1	1
2	Abul-Ela, et al JASA, 1967	$(A_{ij}, B_{ij'})$; $i=1,\ldots n$, $j=1,..Q$, $j'=1,\ldots Q'$, $Q+Q'-1=n$,	Q+Q´	Y	QL	1	1
3	Greenberg, et al JASA, 1969	(A_{ij}, B_{ij}); $i=1$ or $i=1,2$	2	N	QL	1	1
4	Gould, et al Soc. Stat. Sec. Proc. of ASA, 1969	(A_{ij}, B_{ij}); $i=1,2$	2	N	QL	2	2
5	Moors JASA, 1971	$(A_{ij}, B_{ij'})$; $i=1,2$, $j=0,1$	1 or 2	N	QL	1	1
6	Greenberg, et al JASA, 1971	(A_{ij}, B_{ij}); $i=1,2$	2	N	QN	1	1
7	Warner JASA, 1971	$((A_{1j} \to T), (A_{1j} \to F)$ $j=1,\ldots,K$	K	-	QL	1	1
8	Boruch Soc. Sci. Res., 1972	$X \epsilon A \to (T_1, F_1)$ $X \notin A \to (T_2, F_2)$	1	-	QL	1	1
9	Folsom, et al JASA, 1973	$B_{ij} \cap (B_{ik}, A_i)$; $j \neq k$, $i=1,2$	3	N	QL	2	1
10	Eriksson ISR, 1973	$A \to (T_{1j}, F_{1k})$; $k=1,\ldots L$, $j \neq k$	1	-	QL, QN	≥ 1	≥ 1
11	Smith, et al Soc. Stat. Sec. Proc. of ASA, 1975	$(\{A_{ij}, B_{ij'}\}_1, \ldots, \{A_{ij}, B_{ij'}\}_L)$ $;i=1,\ldots M, j=1,\ldots,N$	N	-	QL	N	1
12	Liu, et al JASA, 1975	(A_{11})	1	-	QN	1	1
13	Hochberg Comm. in Stat., 1975	1. Stage 1 Stage 2 $\bigcap_{j=1}^{Q}(A_{ij}, Y) = \{$any Y... $\to$ Stop $\{$all N... $\to A_{ij}$	Stage 1: Q Stage 2: 1	Stage 1: N Stage 2: Y	QL, QN	Stage 1: Q Stage 2: 1	1
		2. $(A_{ij}) = \{X \notin A_{11} \ldots \to$ Stop $\{X \epsilon A_{11} \ldots \to (A_{1\ell}, Y, N)$ $\ell=1,\ldots r$ $j=1,\ t-r$	Stage 1: Q-r Stage 2: r	Stage 1: Y Stage 2: N	QL, QN	Stage 1: 1 Stage 2: r	1
14	Liu, et al JASA, 1976	(A_{11}, m)	1	N	QN	1	1

FOOTNOTES:

A refers to the sensitive question
B refers to the non-sensitive question
i indexes sample
j indexes question
k indexes answer
m is the number marked on ball

0 indexes null question
T refers to the truthful reporting
F refers to the nontruthful reporting
Y refers to yes
N refers to no
QL refers to qualitative and QN quantitative

were unknown, by asking only about the nonsensitive characteristic B in the second sample (P_2=0).

Warner (1971) showed how the RRT models could be represented by a general linear model, $\underline{Y} = \underline{T}\ \underline{\pi} + \underline{u}$. He also extended his original contamination model (Warner, 1965) to the multichotomous case (model 7, Table 1) and introduced the two stage randomized response technique (TSRRT).

Boruch (1972) indicated how the original Warner contamination model could be modified to include two randomization procedures. The first (T_1 or F_1 in Table 1) to be used if the respondent is a member of group A ($X \varepsilon A$) and the second (T_2,F2) to be used if $X \notin A$.

Folsom, et al. (Table 1) modified Moors' optimized unrelated question model to ask two questions in each sample, the first a direct question about a nonsensitive characteristic (B_{ij}) and the second a randomly selected sensitive (A_i) or nonsensitive question (B_{ik}). The new model consistently performed better than Moors' model and they reported on an application involving drinking and driving.

Eriksson (1973) presented a RRT multiple trial model which was applicable to quantitative or multichotomous data (Table 1). It appears to be a modification of the Warner (1971) contamination model. He also indicated that supplementary information can be utilized with the RRT estimator.

Smith, et al. (1975) reported on an application of the randomized version of Raghavarao and Federer's block total response method for obtaining information from a group of college students. The computed average variance for the seven questions was more than twice as large for the randomized block total response model (Table 1., 11) than for Warner's (1965) model.

Liu, et al. (1975) presented a single question discrete quantitative randomized location RRT model. It utilized a random location procedure and no comparisons were offered with possible competitors.

Hochberg (1976) introduced two particular members of Warner's (1971) class of TSRRT linear models (Table 1). No efficiency calculations or applications were mentioned, and neither model seemed to be very practical.

Liu and Chow (1976) presented a modification of Greenberg, et al.'s (1971) quantitative model (Table 1, 14) using a known distribution of nonsensitive characteristics. They suggest that this modification should considerably improve the efficiency of the estimator.

Clearly, numerous other RRT models are possible, and Greenberg, et al. (1974) have discussed and compared some of the above models in more detail.

3. Multiple Sensitive Question Model

Information on more than one sensitive characteristic can be obtained by the following multiple sensitive question (MSQ) model. The model's question structure is as follows:

Subsample i: $(A_{i1}, \bar{A}_{i2}, A_{i3}, \ldots, \bar{A}_{iQ}, B_{i1}, \bar{B}_{i2}, B_{i3}, \ldots, \bar{B}_{iQ'})$, $i=1,\ldots,Q+Q'$, and where

$\bar{A}_{ij}$ is the complement of question A_{ij}. We will assume for convenience of presentation that only the sensitive questions (A_{ij}) are of interest. Thus the total sample of respondents (N^*) will be randomly allocated into Q mutually exclusive subsamples and each respondent will select, via a randomization device, a single question and answer it "yes" or "no".

Let p_{ij} be the probability of the randomization device selecting the j th question in the i th subsample. Assume

$$\sum_{j=1}^{Q} p_{ij} = 1.$$

Define: π_j = Pr (Person in population has characteristic A_j), j=1,...Q,

$\sum_{j=1}^{Q} \pi_j \gtreqless 1$, and λ_i = Pr (respondent reports "yes" in subsample i).

A. Truthful Reporting.

For an even number of sensitive questions,

$$\lambda_i = p_{i1}\pi_1 + p_{i2}(1-\pi_2) + p_{i3}\pi_3 + \ldots + \left(1-\sum_{j=1}^{Q-1} p_{ij}\right)(1-\pi_Q),$$

and for an odd number the last term substitutes π_Q for $(1-\pi_Q)$.

Using matrix notation,

$\underline{\lambda} = \underline{P}\ \underline{\pi}$, where $\underline{\lambda}' = (\lambda_1, \lambda_2, \ldots, \lambda_Q)$,

$\underline{\pi}' = [\pi_1, (1-\pi_2), \pi_3, (1-\pi_4), \ldots, \pi_{Q-1}, (1-\pi_Q)]$,

$$\text{and } \underline{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & 1-\sum_{j=1}^{Q-1} p_{1j} \\ p_{21} & p_{22} & p_{23} & 1-\sum_{j=1}^{Q-1} p_{2j} \\ \vdots & \vdots & \vdots & \\ p_{Q1} & p_{Q2} & p_{Q3} & 1-\sum_{j=1}^{Q-1} p_{Qj} \end{bmatrix}$$

Thus, the observed values of $\underline{\lambda}, \hat{\underline{\lambda}}$, are linear combinations of $\underline{\pi}$ plus an error term $\underline{e}$. Therefore, $\hat{\underline{\pi}} = \underline{P}^{-1}\hat{\underline{\lambda}}$, given $|\underline{P}| \neq 0$, and we use Q subsamples to estimate Q sensitive characteristics.

The properties of $\hat{\underline{\pi}}$ are;

(1) $\hat{\underline{\pi}}$ is an unbiased estimator of $\underline{\pi}$.

$$E(\hat{\underline{\pi}}) = E(\underline{P}^{-1}\hat{\underline{\lambda}}) = \underline{P}^{-1}E(\hat{\underline{\lambda}}) = \underline{P}^{-1}\underline{P}\underline{\pi} = \underline{\pi},$$

(2) Variance and covariance matrix for related $\hat{\pi}_i$ and $\hat{\pi}_j$, $\forall$ i,j, $i \neq j$, is

$V(\hat{\underline{\pi}}) = \underline{P}^{-1}(\hat{\underline{\lambda}})(\underline{P}^{-1})'$, where $V(\hat{\underline{\lambda}})$ is a diagonal matrix with elements $\frac{\lambda_i(1-\lambda_i)}{n_i}$, i=1....,Q, and

(3) Variance matrix for unrelated $\hat{\pi}_i$ and $\hat{\pi}_j$, $\forall$ij, $i \neq j$, is achieved by replacing the $(ij)^{th}$ element of $V(\hat{\underline{\pi}})$ with 0.

Example: Two sensitive question case.

$$\hat{\underline{\pi}} = \begin{pmatrix} \hat{\pi}_1 \\ 1-\hat{\pi}_2 \end{pmatrix} = \begin{pmatrix} p_{11} & 1-p_{11} \\ p_{21} & 1-p_{21} \end{pmatrix}^{-1} \begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \end{pmatrix}$$

$$= \begin{pmatrix} p_1 & 1-p_1 \\ p_2 & 1-p_2 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \end{pmatrix},$$ dropping the second subscript on p_{ij}.

Then $\hat{\pi}_1 = \frac{\hat{\lambda}_1 (1-p_2) - \hat{\lambda}_2(1-p_1)}{p_1 - p_2}$ and $1-\hat{\pi}_2 = \frac{\hat{\lambda}_1 p_2 - \hat{\lambda}_2 p_1}{p_2 - p_1}$.

$$V(\hat{\underline{\pi}}) = V\begin{pmatrix} \hat{\pi}_1 \\ 1-\hat{\pi}_2 \end{pmatrix} = \begin{pmatrix} p_1 & 1-p_1 \\ p_2 & 1-p_2 \end{pmatrix}^{-1} \times$$

$$\begin{pmatrix} \frac{\lambda_1(1-\lambda_1)}{n_1} & 0 \\ 0 & \frac{\lambda_2(1-\lambda_1)}{n_2} \end{pmatrix} \begin{pmatrix} p_1 & 1-p_1 \\ p_2 & 1-p_2 \end{pmatrix}^{-1'}.$$

Thus, $$V(\hat{\pi}_1) = \frac{1}{(p_1-p_2)^2}\left[\frac{\lambda_1(1-\lambda_1)(1-p_2)^2}{n_1} + \frac{\lambda_2(1-\lambda_2)(1-p_1)^2}{n_2}\right],$$

and $$V(1-\hat{\pi}_2) = V(\hat{\pi}_2) = \frac{1}{(p_1 - p_2)^2} \times \left[\frac{\lambda_1(1-\lambda_1)\, p_2^2}{n_1} + \frac{\lambda_2(1-\lambda_2)p_1^2}{n_2}\right]$$ and $$Cov[\hat{\pi}_1, 1-\hat{\pi}_2] = -\frac{1}{(p_1-p_2)^2} \times \left[\frac{\lambda_1(1-\lambda_1)p_2(1-p_2)}{n_1} + \frac{\lambda_2(1-\lambda_2)p_1(1-p_1)}{n_2}\right]$$

for $\hat{\pi}_1$ and $\hat{\pi}_2$ related.

<u>Theorem 1</u>: If $\hat{\pi}_1$ and $\hat{\pi}_2$ are independent, $n_1 = n_2 = N^*/2$, $p_1+p_2 = 1$, and $p_1 \neq p_2$, then

$V(\hat{\pi}_1 + \hat{\pi}_2)_W \geq V(\hat{\pi}_1 + \hat{\pi}_2)_{MSQ}$, $\forall \pi_1, \pi_2$. The proof of this and the following theorem have been omitted due to the restriction on space.

<u>Theorem 2</u>: If $\hat{\pi}_1$ and $\hat{\pi}_2$ are dependent and all other conditions of Theorem 1 hold, then

$V(\hat{\pi}_1 + \hat{\pi}_2)_W \geq V(\hat{\pi}_1 + \hat{\pi}_2)_{MSQ}$, V $\hat{\pi}_1$, $\hat{\pi}_2$.

B. <u>Untruthful Reporting</u>.

Let T_i be the probability of telling the truth in the MSQ model given that the respondent has the i^{th} characteristic. Then,

$$\lambda_i = p_{i1}\pi_1 T_1 + p_{i2}(1-\pi_2) + p_{i2}\, \pi_2(1-T_2) + \ldots + (1 - \sum_{j=1}^{Q-1} p_{ij})(1-\pi_Q) + (1- \sum_{j=1}^{Q-1} p_{ij})\, \pi_Q(1-T_Q)$$

and $$\text{Bias}\,[\hat{\pi}_1 + (1-\hat{\pi}_2) + \ldots + (1-\hat{\pi}_Q)] = \pi_1(T_1-1) + \pi_2(1-T_2) + \ldots + \pi_Q(1-T_Q).$$

Now, let T_i' be the equivalent in Warner's (1965) model. Then Bias $(\hat{\pi}_i)_W = \pi_i(T_i'-1), \forall i$

<u>Corollary 1</u>. Under the conditions of Theorem 1, if $T_1 = T_2 = T_1' = T_2'$, then

$MSE(\hat{\pi}_1 + \hat{\pi}_2)_W > MSE(\hat{\pi}_1 + \hat{\pi}_2)_{MSQM}$, $\forall$ $\hat{\pi}_1$, $\hat{\pi}_2$.

<u>Corollary 2</u>. Under the conditions of Theorem 2, if $T_1 = T_2 = T_1' = T_2'$, then

$MSE(\hat{\pi}_1 + \hat{\pi}_2)_W > MSE(\hat{\pi}_1 + \hat{\pi}_2)_{MSQM}$, $\forall$ π_1, π_2.

In the interest of comparing the MSQ and Warner models under less restrictive conditions (i.e. $T_1 = T_2 \neq T_1' = T_2'$), Table 2 presents the MSE efficiency for the two question MSQ model compared with using the Warner model twice (where p_1 = .7, p_2 = .3 π_1= .2, π_2= .1, n_1+n_2= 1000, and n_i's optimally allocated). Provided it is reasonable to assume that $T_i = T_i'$ (i = 1,2), it is found that (see the diagonal elements in the Table 2) the efficiency of the MSQ model increases as T_i decreases. In other words, when the chance of untruthful reporting is high, the MSQ model is highly recommended.

(4). SOME CONCLUDING REMARKS

When information about more than one sensitive characteristic is of interest, one might use model 11 Table 1), repeated applications of the dichotomous models (table 1), or the MSQ model. In all MSE comparisions to date, the MSQ model generally has proved to be superior. It should be noted that the present MSQ model can handle discrete quantitative questions as well as questions of joint characteristics.

Lastly, the MSQ model can be easily extended to the multiple trial case, the continuous quantitative question case, and the double randomization (both questions and answers) case. Application of the MSQ model should be shortly forthcoming.

TABLE 2

MSE Efficiency for the Two Sensitive Question Case versus Warner's Model. $P_1=.7$, $P_2=.3$, $\pi_1=.2$, $\pi_2=.1$, $T_1=T_2$, $T_1'=T_2'$, $n_1+n_2 = 1000$ and Optimally Allocated. $\mathrm{Corr}(\hat{\pi}_1, \hat{\pi}_2) = -.72$.

$T_1'=T_2'$ \ $T_1=T_2$	1.00	.90	.80	.70	.60	.50
1.00	6.31	5.69	4.39	3.18	2.29	1.69
.90	7.30	6.58	5.07	3.67	2.65	1.95
.80	10.26	9.25	7.13	5.16	3.72	2.74
.70	15.21	13.70	10.57	7.65	5.52	4.06
.60	22.12	19.94	15.37	11.13	8.03	5.91
.50	31.02	27.95	21.55	15.60	11.25	8.28

REFERENCES

1. Abernathy, J. R., 1973, "The Two Alternate Questions Randomized Response Model for Human Surveys," JASA, 68, 525-530.

2. Abul-Ela, A. A., Greenberg, B. G., and Horvitz, D. G., 1967, "A Multi-Proportions Randomized Response Model," Journal of the American Statistical Association, 62, 99C-100.

3. Boruch, R. F., 1972, "Relations Among Statistical Methods for Assuring Confidentiality of Social Research Data," Social Science Research, 1 403-414.

4. Eriksson, Sven A., 1973, "A New Model for Randomized Response," International Statistical Review, 41, 1 101-113.

5. Folsom, R. E., Greenberg, B. G., Horvitz, D. G. and Abernathy, J. R., 1973, "The Two Alternative Questions Randomized Response Model for Human Surveys," JASA, 68, 525- 530.

6. Gould, A. L., Shah, B. V., Abernathy, J. R., 1969, "Unrelated Question Randomized Response Technique with Two Trials Per Respondent," Proceedings of ASA, Social Statistics Section.

7. Greenberg, B. G., Abul-Ela. A. A., Simmons, W. R. and Horvitz, D. G. 1969, "The Unrelated Question Randomized Response Model: Theoretical Framework," Journal of the American Statistical Association, 64, 520-539.

8. Greenberg, B. G., Kuebler, R. R. Jr., Abernathy, J. R. and Horvitz, D. G., 1971, "Application of the Randomized Response Technique in Obtaining Quantitative Data," Journal of the American Statistical Association, 66, 243-250.

9. Greenberg, B. G., Horvitz, D. G., and Abernathy, J. R., 1974," A Comparison of Randomized Response Designs. Reliability and Biometry," SIAM, Philadelphia, 787-815.

10. Hochberg, Y., 1975, "Two State Randomized Response Schemes for Estimating a Multinomial," Communications in Statistics, 4, 31, 1021-1032.

11. Horvitz, D. G., Greenberg, B. G., and Abernathy, J. R., 1975, "Recent development in randomized response designs," A Survey of Statistical Design and Linear Models. ed. J. N. Srivastava, North-Holland/American Elsevier, 271-285.

12. Liu, P. T., Chow L. P., and Mosley, H. W., 1975, "Use of the Randomized Response Technique with a New Randomizing Device," Journal of the American Statistical Association 70, 329-332.

13. Liu, P. T., and Chow, L. P., 1976, "A New Discrete Quantitative Randomized Response Model, Journal of the American Statistical Association, 71, 72-73.

14. Moors, J. J. A., 1971, "Optimization of the Unrelated Question Randomized Response Model," Journal of the American Statistical Association, 66, 627-629.

15. Raghavarao, D., and Federer, W. T., 1973, "Application of BIB Designs as an Alternative to the Randomized Response Method in Surveys." BU-490-M, Biometrics Unit, Cornell Univ.

16. Smith, L. L., Federer, W. T., and Raghararao, D., 1975, "A Comparison of Three Techniques for Eliciting Truthful Answers to Sensitive Questions," 1974 Proceedings of the ASA, Social Statistics Section, 447-452.

17. Warner, S. L., 1965, "Randomized Response: a Survey Technique for Eliminating Evasive Answer Bias," Journal of the American Statistical Association, 60, 63-69.

18. Warner, S. L., 1971, "The Linear Randomized Response Model, "Journal of the American Statistical Association, 66, 884-888.

A SURVEY OF CONSUMER PARTICIPANTS IN THE BIGGEST ANTITRUST REFUND EVER

Benjamin F. King, University of Washington

Antitrust litigation is an area of the law in which many diverse problems of statistical analysis arise. I have had the privilege of serving as the consulting statistician for the States of Washington, Oregon and Kansas, and subsequently for the U.S. District Court, in the Antibiotic Drug Case. This was a class action lawsuit against the five major manufacturers of tetracycline, a well-known broad spectrum antibiotic, in which it was alleged that purchasers suffered damages from improper pricing during the period 1954 through 1966.

Before the case was scheduled to go to trial, experts for the plaintiffs and the defendants locked horns on a number of challenging statistical issues -- the estimation of total retail purchases in the plaintiff states, the establishment of an index of prices, the breakdown of purchases into "just" and "unjust" components -- these are but some of the thorny problems that were encountered, few of which, if any, were amenable to classical textbook solutions. I echo my predecessors in this session when I mention the importance of statistics in questions of law, but I should like to emphasize that conventional methods are not always adequate - there is much room for innovation.

At any rate, the trial never took place. There was a $39.6 million settlement, with about $20 million of that to be distributed to members of the consumer class who would make claims for their amounts purchased. These purchasers resided in the States of California, Hawaii, Kansas, Oregon, Utah, and Washington. The story of the distribution of refunds is of the "they said it couldn't be done" variety. The largest previous class action refund involved about 70,000 businessmen with relatively good records of purchases (in re Gypsum Cases.) Hence there was great skepticism, especially on the part of the defendants, concerning the ability of the court to get the money back to the people who originally made purchases. An earlier distribution of $100 million to consumers in 43 states had resulted in only about 37,000 claims. Defendants' attorneys speculated that it would be unusual if that number were exceeded in the present six-state situation. It was, therefore, a great success to end up, as we did, with about one million claims and to mail checks to 885,000 claimants after validation procedures were complete. Because its magnitude is so much greater than previous class action refund operations, and because it demonstrated the feasibility of a number of things that were considered to be infeasible, this antitrust refund, referred to as Operation Money Back, has become a landmark in the history of antitrust litigation and a prototype for similar distributions in the future.

Yet, many important questions concerning consumer class actions remain unanswered: For example, antitrust lawyers are bothered by problems of giving proper notice in actions of this type. Because of a recent court ruling (Eisen vs. Carlisle & Jacquelin) it is necessary to notify members of a class of the fact of their membership by "the best means practicable", which has been determined to be first-class mail. Thus in the Antibiotic Drug Case, involving the six hold-out states, a notice of creation of the class was mailed to approximately 12 million households in 1971, and when the settlement was proposed in 1974, it was necessary to mail out another notice. When the settlement was approved by the court, claim forms were mailed, and to claimants who claimed purchases greater than $150, a request for further proof was sent out. You can imagine the effect of all of these mailings on a distribution involving rather modest settlement amounts, but covering even larger populations.

Of interest also to the designers of Opera-Money Back is the relative effectiveness of various types of communications in persuading consumers to participate in the refund.

There were serious questions raised about the incidence of fraud in the refund operation, although as described in other reports, considerable precautions against wide-scale cheating were taken.

In summary, it is desirable, for the future of class action litigation, to know a great deal more about the general characteristics and attitudes of the persons who made claims for damage refunds. In order to try to get some kind of a handle on these uncertainties, and in order to prevent a golden opportunity to study the phenomenon from fading away, we asked the court and the Attorneys General involved to allow us to use some of the money from uncashed refund checks to survey the population of claimants.* With the grant we designed a probability sample covering seven strata in the target population:

(1) Claimants with purchases of $150 or less -- for which no further proof was required.

(2) Those claiming over $150 but less than $1000 who submitted statements of proof or documentation from physicians, hospitals, drugstores, etc. Notarization of signature was also required.

(3) Those above $150 but less than $1000 who could not submit proof but who authorize the court to determine the amount of purchases. These were subsequently reduced to $150.

The remaining strata are for higher claim amounts, and for the nonrespondents to the request for proof of purchases after the first claim form was received. (The latter group received nothing in refund.)

The National Opinion Research Center of the University of Chicago (NORC) was given the contract for the field work which began about May 15. Since almost a year had passed since the mailing of the refund checks, there was difficulty in locating some of the selectees. The interview was by telephone -- about 55 minutes -- with nontelephone claimants given personal interviews. Respondents were mailed packets containing the essential documents involved in the class action in order to aid them in recall. I am happy to report that we obtained the targeted response rate, 85 per cent, but the survey took rather longer than expected, and we just came out of the field about August 10.

The few results that I will report today are from a lightning analysis (thanks to SPSS) of the first 727 cases out of a final 900 who either filed for $150 or less or responded to the request for proof after having made a larger claim. These figures are of course tentative and not for publication, but I shall be very surprised if they change very much. Incidentally, because of the stratified design with widely different weights, we selected six interpenetrating subsamples for ease of estimation of standard errors -- estimates of population percentages appear to have standard errors ranging from 1 to 3 percentage points, depending on the variable.

First a few remarks about the way in which our population compares with the general population of persons over 21 in the U.S. -- As some of you know, NORC sends out an annual survey, the General Social Survey, in which a number of standard demographic items are asked, as well as specific questions for the social science community. The GSS was executed this year in March and April -- hence we included a number of GSS items in our own survey for comparison purposes.

Our group has a different age distribution from the general population (I use California 1970 here.) We have fewer under 34 and more in the 35-54 range, corresponding to heads of families with small children back during the period in suit, 1954-66. Also there are slightly more elderly, but this may be a response effect.

Our group is definitely better educated. 47.7 per cent have one or more years of college as opposed to 29.9 per cent for California in 1970 (persons 25 or over). Since our population is older, the results are even more noteworthy, indicating that the people who get into these kinds of consumer refunds are probably more aware of what is going on in general.

Income is also higher than that picked up by NORC nationally in March, which is to be expected because of the pervasive correlation between income and education.

Only 2.9 per cent of our population is black. 3.5 per cent identified themselves as belonging to races other than white or black, (probably Orientals in California) and .7 per cent refused to answer that item. We suspected that we would be deficient in racial minorities. Most persons of Latin American background in the West are white, and in California a great deal of effort was expended by the Attorney General's office to try to assure that Spanish-speaking consumers participated. There was even a Spanish claim form distributed, and the regular claim form had a message in Spanish across the top. It was not very successful, however.

NORC asked a series of questions that are rather interesting -- the ANOMIA questions -- e.g., Agree/disagree? Next to health, money is the most important thing in life; it's hardly fair to bring a child into the world with the way things look for the future; these days a person doesn't really know whom he can count on; you sometimes can't help wondering whether anything is worthwhile anymore; etc. If you agree with many of these nine items you are in pretty bad shape psychologically. In the General Social Survey the results are rather depressing -- e.g., 33.4 of U.S. adults agree that money is the most important thing next to health; 40 per cent said that they wonder if anything is worthwhile anymore; 41.5 per cent think that it is hardly fair to bring a child into the world today; 59 per cent think that the lot of the average man is getting worse. I am happy to report, that averaging over the nine ANOMIA items, the mean proportion in our claimant group who agree with these dismal statements is about 18 percentage points lower than the U.S. adult population in general. We seem to have a happier group -- it will probably turn out to be a class distinction.

[To the extent that time permitted, Professor King reported additional marginal relative frequencies for some of the items in the questionnaire. Of particular interest to class action lawyers are questions about the willingness of the participants to get involved in another similar refund operation after their experience in Operation Money Back, and the minimum amount of refund that would be necessary to induce them to participate again. We emphasize that all reported figures are preliminary and subject to change in the final analysis. A complete history of the lawsuit and the refund operation including an analysis of the survey of participants will be published in the form of one or several monographs in the near future.]

* The "we" here refers to a term consisting of a Special Master of the Court, an economist, a data processing specialist, and a statistician.

A NEW LOOK AT MORTALITY INDEXES WITH EMPHASIS ON SMALL AREA ESTIMATION

Joel C. Kleinman,* National Center for Health Statistics

INTRODUCTION

One of the oldest and most widely used indicators of health status is the mortality rate. The limitations of the crude mortality rate for comparing specific areas or groups led to the use of various adjustments for factors such as age, race and sex in order to obtain an index which was free from the confounding effects of these factors (Kitagawa, 1966). Currently, the most commonly used adjustments are the direct and indirect methods. However, both these methods are determined to a large extent by death rates in the older age groups which limits their usefulness in health planning and certain epidemiological applications.

In addition, due to the historical use of mortality rates with large populations, relatively little work has been done concerning the standard errors of adjusted rates. However, with increasing need for small-area health status data (as mandated in the National Health Planning and Resource Development Act of 1974) the use of mortality indexes for small populations should become more common and thus there is a need for an assessment of the variability of these indexes.

This study compares four mortality indexes and their standard errors among United States counties using 1969-1971 mortality data.

DEFINITIONS

Indexes

For a particular race-sex group in each county, let

d_i = # deaths in age group i

p_i = population in age group i

$m_i = \frac{d_i}{p_i}$ = death rate in age group i

$d = \Sigma d_i$

$p = \Sigma p_i$

Replacing the small letters by capital letters will indicate the corresponding data for a standard population. The age-adjusted rate using the indirect method is usually expressed as the Standard Mortality Ratio (SMR):

$$SMR = \frac{d}{\Sigma M_i p_i} = \frac{\Sigma m_i p_i}{\Sigma M_i p_i} \qquad (1)$$

*The author wishes to thank Timothy Pierce for carrying out the data processing and computations upon which the results are based.

The adjusted rate using the direct method is usually expressed as

$$\Sigma \frac{P_i}{P} m_i$$

It can also be expressed as a ratio similar to the SMR, sometimes called the Comparative Mortality figure (CMF):

$$CMF = \frac{\Sigma \frac{P_i}{P} m_i}{\Sigma \frac{P_i}{P} M_i} = \frac{\Sigma m_i P_i}{\Sigma M_i P_i} \qquad (2)$$

Thus the only difference between the two is that the direct method applies the same standard age distribution to the county's age-specific death rates while the indirect method applies the county's age distribution to a standard set of rates. For this reason, the direct method is usually preferred since two counties with the same set of age-specific rates but different age distributions will have equal CMFs but unequal SMRs. However, in terms of identifying counties with "excess" deaths (an important application in planning) the use of the SMR is more appropriate since it weights the age-specific death rates as they apply to the county's age distribution. Thus, a relatively high death rate is weighted according to the population it actually affects. For this reason, we shall concentrate on indirect adjustment.

Yerushalmy (1951) has pointed out that both the direct and indirect methods of adjustment are greatly influenced by the numbers of deaths and thus death rates in the older age groups are emphasized. He proposed a weighted average of the ratios of each age-specific death rate to a standard age-specific death rate. An indirect method of adjustment which is analogous to Yerushalmy's index is the relative mortality index (RMI):

$$RMI = \frac{1}{p} \Sigma \frac{m_i}{M_i} p_i = \frac{1}{p} \Sigma \frac{d_i}{M_i}$$

This has the advantage of not requiring the county's population by age, an important property for inter-censal estimates. Unlike the two previous indexes, the RMI averages ratios of county age-specific death rates to standard rates and so it is not influenced by numbers of deaths. Indeed it may go too far in this direction. For example, based on United States white males in 1970 doubling of the death rate in the 5-14 age group would result in 8,481 additional deaths while a doubling of the age 35-44 death rate would result in 34,278 additional deaths. Yet the RMI would weight the first increase nearly twice as high as the second (based on the 1970 United States age distribution for white males).

A compromise between these extremes is the use of "productive years of life lost" as a way of counting deaths (Haenszel, 1950). The basic idea is to weight each death according to 70 minus the age of the decedent, the assumption being that each individual should be expected to live 70 productive years of life. When the deaths are grouped by age intervals, 70 minus the midpoint is used as the weight. Thus a years of life lost index can be defined as

$$YLL = \frac{\Sigma\, d_i(70-l_i)}{\Sigma\, M_i p_i (70-l_i)} = \frac{\Sigma m_i p_i (70-l_i)}{\Sigma M_i p_i (70-l_i)}$$

where l_i is the midpoint of the age interval and the sums are evaluated only for positive values of $70-l_i$. Thus, using the standard eleven age groups (under 1, 1-4, 5-14,..., 75-84, 85+), all deaths above age 65 are excluded from the calculation. Using the example from the previous paragraph, doubling the age 5-14 death rate results in an additional 508,860 years of life lost compared to 1,028,340 years of life lost in the 35-44 age group. The YLL index is the one which seems most appropriate for identifying areas with excess mortality in a health planning and resource allocation context.

Each index was computed separately for each color-sex group (white male, white female, other male, other female) using the 1969-1971 age-color-sex-specific rates for the total United States as the standards. For the CMF the total 1970 United States population age distribution was used as the standard. A combined index was formed for each of the "indirect" indexes (SMR, RMI, YLL) by extending the summation in each formula over 44 (= 4 color-sex groups x 11 age groups) categories.

Standard errors

Since these indexes will be based on small area data, it is important to compare their variances. With the exception of Chiang's (1961) work and an unpublished paper by Chiang and Linder (1969) little work has been done in this area. Any of the preceding indexes can be represented as

$$I = \frac{\Sigma w_i m_i}{w}, \quad w = \Sigma w_i$$

where the w_i and w are considered constants not subject to random variation. Thus the standard error of I is

$$s(I) = \sqrt{\frac{\Sigma w_i^2 \sigma_i^2}{w^2}}$$

where $\sigma_i^2 = \text{Var}(m_i)$. Following, Chiang (1961),

$$\sigma_i^2 = \frac{m_i\left[1-a_i n_i m_i\right]}{p_i\left[1+(1-a_i)n_i m_i\right]} \qquad i = 1,\ldots,10$$

where $\sigma_{11}^2 = 0$
a_i = fraction of last year of life
n_i = number of years in age interval

The standard life table values of $a_i (a_1 = .1, a_2 = .4, a_3 = \ldots = a_{10} = .5)$ were used (Chiang, 1968). For comparing the indexes, the coefficient of variation s(I)/I will be used.

In this study average annual death rates using 1969-1971 deaths and 1970 population were computed. That is,

$$m_i = \frac{d_i}{3p_i}$$

where d_i = total deaths for 3 years in age group i

p_i = 1970 county population in age group i

for each race-sex group. In Chiang's formula for σ_i^2, p_i was replaced by $3p_i$. Also when $d_i = 0$, m_i was set to zero in computing the index but m_i was set to M_i in computing σ_i^2.

RESULTS

The results are based on the 2,805 counties* with 5,000 or more total population in 1970. Five indexes were considered for each county: white male, white female, other male, other female, and combined. Each index was computed only if the population for the color group was at least 5,000. The combined indexes for the three indirect adjustment methods (SMR, RMI, YLL) were computed for all 2,805 counties. There were 2,701 counties with white population 5,000 or more, 695 with other population 5,000 or more, and 671 with both white and other population 5,000 or more.

There was substantial variation in the age distribution among the 2,805 counties: the percent over 65 varied from below 7 percent in the lowest decile to over 16 percent in the upper decile with a median of 11.4 percent. The crude death rate varied from below 7.4 per 1,000 in the lowest decile to over 14.0 per 1,000 in the highest decile with a median of 10.7 (the United States rate for 1969-1971 was 9.4 per 1,000). Incidentally the need for age adjustment is indicated by the fact that the correlation coefficient between the crude death rate and the percent over 65 was .88. The correlation coefficient between the crude death rate and each of the indexes discussed below was on the order of .3.

* Of the 3,140 United States counties, 13 were not considered due to different coding between FIPS and NCHS. These were the 5 boroughs of New York City, 2 counties in Hawaii, and 6 county-equivalents in Alaska.

The CMF and SMR gave nearly identical results. The correlation coefficients between the two were over .97 for each color-sex group. Their coefficients of variation (cv) were also similar although there was a tendency for the SMR to have a slightly smaller cv (it can be proven that SMR has a smaller variance than CMF). Thus, despite the theoretical differences between the two, the actual differences over U.S. counties are unimportant. For this reason, we will concentrate on comparing the three indirect indexes.

The correlation matrices for the indexes are shown in Table 1 based on the 671 counties with all color-sex indexes computed. For each index, the white males have the highest correlation with the combined index. The order of the remaining correlations are white female, other male and other female for SMR and RMI but white females have the lowest correlation with the combined YLL index. The correlations between the sex groups are moderate (.4-.7). The correlations between white males and other males (.39-.49) are higher than those between white females and other females (.20-.30).

Table 2 shows the correlation matrices for each color-sex group again based on the 671 counties. Except for other females, the RMI-YLL correlation is highest (.76-.88) and the SMR-RMI lowest (.45-.66). For other females SMR-YLL has the highest correlation (.77).

Results based on the 2,701 counties with white indexes computed and the 695 counties with other indexes computed are similar to those cited above.

Selected percentiles of the distributions of the indexes are presented in Table 3. The distributions of RMI and YLL have more variation than SMR, due in part to their larger random error component (see below). Thus, the RMI and YLL estimate greater "excess" mortality than does the SMR. For example, based on the combined indexes, the upper quintile of the RMI distribution consists of counties at least 26.4 percent above what would be expected based on U.S. rates while the SMR upper quintile estimates only 9.1 percent excess deaths; YLL is intermediate with its upper quintile at 21.4 percent. The RMI and YLL distributions are both skewed to the right while the SMR distributions are symmetric.

Simultaneous pictures of the counties classified in the extremes of the distributions are given in Table 4 which shows the areas in the highest quintile of each index. The percentage of counties in the highest quintile of at least one index was approximately 35 percent for each color-sex group. Of these fewer than one-fourth within each color-sex group were classified high by all three indexes. As was the case with the correlation coefficients, the RMI and SMR had the least agreement. However, the agreement between RMI and YLL was similar to that for SMR and YLL for white females, other males, and other females. For white males and combined, the RMI-YLL agreement was greater.

Table 5 illustrates the differences between the combined and color-sex specific YLL indexes in terms of counties classified as extreme (in the upper or lower quintile). Except for white males, the combined index identified only 56 percent-64 percent of the counties which had a high color-sex-specific index. These results are consistent with the correlations in Table 1 and point to the need for using race-sex-specific indexes whenever possible.

In terms of coefficients of variation (cv), the ordering of the indexes are as expected based on the age groups each emphasize: SMR has the smallest cv and RMI the largest. Table 6 shows selected percentiles of the cv's. Four-fifths of the counties had SMRs with cv's below 4 percent-9 percent. For YLL less than 40 percent had cv's in this range, and for RMI only 10-20 percent had cv's in this range. Eighty percent of the counties had YLL indexes with cv's below 12 percent-23 percent.

In order to facilitate a quick rule of thumb for anticipating cv's, Table 7 shows the range of cv's observed for selected county population size groups. For the SMR, cv's above 5 percent were virtually nonexistent in counties over 10,000 population. In the 5-10,000 group, however, more than 40 were over 5 percent. The RMI has larger cv's and it is not until the 100,000 + group that the majority of cv's are below 5 percent. Clearly, this index is not very useful for small populations. The YLL index is intermediate between the two. For populations above 25,000, virtually all cv's are below 10 percent and the majority are below 5 percent.

CONCLUSIONS

The results presented show that substantial differences exist among the indexes studied (except for Standardized Mortality Ratio and Comparative Mortality Figure). In addition, the counties with excess mortality vary for each color-sex group. The implications of these results for two types of applications are presented briefly below.

First, in terms of epidemiological investigations which use regression analysis of age-adjusted rates using counties or other ecological areas as the unit of analysis (e.g., Lave and Seskin, 1973), our results show that the results are influenced heavily by white male death rates in the older age groups. Even when separate color-sex groups are used, the death rates for white males in the three age groups between 55 and 84 account for nearly 80 percent of the variation in the white male SMR. Thus, more thoughtful specification of the model in such studies is called for. If a combined index is necessary and a relative risk model is appropriate, the relative mortality index (or Yerushalmy's amalagous direct method) seems a better alternative than the SMR or CMF. Of course, the ideal approach is to use age-specific rates but for small populations these rates have a large random error component.

For Health Planning applications, the use of the SMR or CMF is inappropriate since death rates among the elderly are probably least amenable to health planning intervention. Similarly the RMI seems to place too much emphasis on very small death rates in the young age groups. Thus, the years of life lost index seems a reasonable compromise. We have also shown that the SMR combined index which is the only one possible when published sources of county mortality data are used (e.g., National Center for Health Statistics, 1975), is a poor substitute for the color-sex specific YLL indexes. Thus, health planning agencies should attempt to go beyond the published data to obtain age-race-sex-specific death rates in order to be able to compute the YLL indexes.

REFERENCES

1. Chiang, C.L. (1961), Standard Error of the Age-Adjusted Death Rate. USDHEW, Vital Statistics-Special Reports 47:271-285.

2. Chiang, C.L. (1968), Introduction to Stochastic Processes in Biostatistics, Wiley, New York.

3. Chiang, C.L. & Linder, F.E. (1969), On the Standard Errors of Death Rates. Unpublished manuscript, Department of Biostatistics, Univ. of North Carolina, Chapel Hill.

4. Haenszel, W. (1950). A Standardized Rate for Mortality in Units of Lost Years of Life, Am. J. Pub. Health 40, 17-26.

5. Kitagawa, E.M. (1966), Theoretical Considerations in the Selection of a Mortality Index, and Some Empirical Comparisons, Human Biology 38:293-308.

6. Lave, L.B. & Seskin, E.P. (1973), An Analysis of the Association Between U.S. Mortality and Air Pollution, J. Am. Stat. Assn. 68:284-290.

7. National Center for Health Statistics (1975), Vital Statistics of the U.S., 1973, Vol. II, Part B. DHEW Pub. No. (HRA) 76-1102. Washington.

8. Yerushalmy, J. (1951), A Mortality Index for Use in Place of the Age-Adjusted Death Rate, A. J. Pub. Health 41:907-922.

Table 1. Correlation coefficients between color-sex groups for three mortality indexes: United States counties, 1969-1971[a]

Index		White Male (WM)	White Female (FM)	Other Male (OM)	Other Female (OF)	Combined (C)
	WM	1.0000				
	WF	.6173	1.0000			
SMR	OM	.4407	.2731	1.0000		
	OF	.4227	.2987	.7266	1.0000	
	C	.8708	.7699	.6330	.6041	1.0000
	WM	1.0000				
	WF	.4346	1.0000			
RMI	OM	.3892	.2214	1.0000		
	OF	.2795	.2038	.4152	1.0000	
	C	.8182	.6894	.5915	.5505	1.0000
	WM	1.0000				
	WF	.4925	1.0000			
YLL	OM	.4938	.2615	1.0000		
	OF	.3859	.2621	.6266	1.0000	
	C	.8336	.6017	.7438	.6633	1.0000

[a]Based on 671 counties with White and Other population 5,000 or more

SMR = Standardized Mortality Ratio
RMI = Relative Mortality Index
YLL = Years of Life Lost Index

Table 2. Correlation coefficients between indexes for each color-sex group: United States counties, 1969-1971[a]

Color-Sex		SMR	RMI	YLL
	SMR	1.0000		
WM	RMI	.6408	1.0000	
	YLL	.7297	.8841	1.0000
	SMR	1.0000		
WF	RMI	.4484	1.0000	
	YLL	.5100	.7603	1.0000
	SMR	1.0000		
OM	RMI	.6576	1.0000	
	YLL	.7574	.7942	1.0000
	SMR	1.0000		
OF	RMI	.5416	1.0000	
	YLL	.7667	.7023	1.0000
	SMR	1.0000		
C	RMI	.6444	1.0000	
	YLL	.7336	.8777	1.0000

[a]Based on 671 counties with White and Other population 5,000 or more

Table 3. Selected percentiles of mortality indexes: United States counties 1969-1971

Index	Color-Sex	Number of Counties[a]	10	20	Percentile 40	60	80	90
SMR	WM	2701	.867	.913	.978	1.035	1.106	1.164
	WF	2701	.845	.892	.960	1.018	1.087	1.141
	OM	695	.771	.880	.995	1.078	1.172	1.248
	OF	695	.792	.891	1.008	1.089	1.171	1.220
	C	2805	.871	.916	.978	1.029	1.090	1.139
RMI	WM	2701	.819	.901	1.036	1.159	1.347	1.505
	WF	2701	.735	.835	.964	1.081	1.245	1.404
	OM	695	.670	.799	.961	1.108	1.272	1.443
	OF	695	.689	.800	.984	1.100	1.293	1.471
	C	2805	.834	.913	1.019	1.122	1.264	1.375
YLL	WM	2701	.832	.917	1.024	1.127	1.271	1.404
	WF	2701	.766	.853	.967	1.066	1.191	1.313
	OM	695	.729	.848	.991	1.100	1.248	1.356
	OF	695	.749	.858	.994	1.119	1.255	1.376
	C	2805	.847	.919	1.015	1.098	1.215	1.316

[a]Indexes computed only when county population in the color group was 5,000 or more

Table 4. Percent of counties in highest quintile of each index by color and sex: United States counties 1969-1971

Color-Sex	All	Indexes for which county classified in highest quintile[a] SMR and RMI	SMR and YLL	RMI and YLL	SMR only	RMI only	YLL only	None	Total %	N
White male	7.8	1.8	2.9	6.0	7.6	5.4	3.0	65.5	100.0	2701
White female	5.8	2.0	2.9	6.1	9.2	6.0	5.3	62.8	100.0	2701
Other male	8.1	1.4	4.0	4.0	6.0	6.5	3.7	66.2	99.9	695
Other female	5.2	1.9	5.2	5.2	7.9	7.5	4.2	63.0	100.1	695
Combined	7.1	1.6	3.3	6.1	7.7	5.2	3.5	65.5	100.0	2805

[a]Numbers of counties in highest quintile of each index vary slightly due to ties and interpolation

Table 5. Relationship between counties classified by combined YLL and color-sex specific YLL indexes: United States counties 1969-1971

Color-Sex		Combined YLL Index Highest 20% (H)	Middle 60% (M)	Lowest 20% (L)	Total %	N[a]
WM	H	72.2	27.6	0.2	100.0	547
	M	8.1	84.7	7.2	100.0	1621
	L	0.9	19.4	79.7	100.0	533
	Total	20.5	60.2	19.3	100.0	2701
WF	H	57.7	41.6	0.7	100.0	542
	M	11.6	74.6	13.8	100.0	1618
	L	4.1	38.0	57.9	100.0	541
	Total	20.5	60.2	19.3	100.0	2701
OM	H	63.8	36.2	---	100.0	140
	M	8.4	80.8	10.8	100.0	417
	L	0.7	40.7	58.6	100.0	138
	Total	18.3	63.9	17.8	100.0	695
OF	H	56.2	43.1	0.7	100.0	139
	M	12.2	77.1	10.7	100.0	419
	L	1.4	44.6	54.0	100.0	137
	Total	18.3	63.9	17.8	100.0	695

[a]Indexes computed only when county population in the color group was 5,000 or more

Table 6. Selected Percentiles of coefficients of variation for each mortality index: United States counties 1969-1971

Index	Color-Sex	Number of Counties[a]	Percentile 10	20	40	60	80	90
	WM	2701	.018	.025	.034	.043	.053	.063
	WF	2701	.020	.028	.039	.050	.064	.078
SMR	OM	695	.030	.038	.052	.063	.075	.084
	OF	695	.034	.044	.060	.072	.087	.099
	C	2805	.013	.018	.025	.032	.040	.047
	WM	2701	.066	.091	.130	.167	.220	.258
	WF	2701	.082	.119	.175	.236	.320	.398
RMI	OM	695	.083	.120	.163	.201	.249	.290
	OF	695	.097	.143	.198	.251	.314	.386
	C	2805	.052	.073	.105	.138	.185	.222
	WM	2701	.045	.064	.092	.119	.151	.176
	WF	2701	.062	.090	.133	.175	.231	.280
YLL	OM	695	.051	.073	.101	.126	.147	.164
	OF	695	.062	.092	.126	.154	.176	.198
	C	2805	.035	.047	.071	.089	.122	.143

[a]Indexes computed only when population in the color group was 5,000 or more

Table 7. Coefficients of variation for combined mortality indexes by county population: United States counties, 1969-1971

Index	Population	Coefficient of Variation <1%	1%-	5%-	10%-	≥15%	Total %	N[a]
SMR	5,000	---	59.6	39.5	0.7	0.2	100.0	554
	10,000	---	98.0	1.9	---	0.1	100.0	1008
	25,000	---	99.6	0.2	0.2	---	100.0	570
	50,000	---	100.0	---	---	---	100.0	332
	100,000	27.8	71.9	0.4	---	---	100.1	270
	500,000	100.0	---	---	---	---	100.0	52
	1,000,000	100.0	---	---	---	---	100.0	19
	Total	5.2	86.0	8.6	0.2	0.1	100.0	2805
RMI	5,000	---	---	---	0.7	99.3	100.0	554
	10,000	---	---	0.5	60.6	38.9	100.0	1008
	25,000	---	---	61.9	37.4	0.7	100.0	570
	50,000	---	0.6	97.9	1.5	---	100.0	332
	100,000	---	68.1	31.5	---	0.4	100.0	270
	500,000	---	100.0	---	---	---	100.0	52
	1,000,000	10.5	89.5	---	---	---	100.0	19
	Total	0.1	9.1	27.4	29.7	33.8	100.1	2805
YLL	5,000	---	---	1.3	59.2	39.5	100.0	554
	10,000	---	---	60.0	39.0	1.0	100.0	1008
	25,000	---	4.4	95.3	0.2	0.2	100.1	570
	50,000	---	63.0	37.0	---	---	100.0	332
	100,000	---	99.6	---	0.4	---	100.0	270
	500,000	---	100.0	---	---	---	100.0	52
	1,000,000	36.8	63.2	---	---	---	100.0	19
	Total	0.2	20.2	45.6	25.8	8.2	100.0	2805

[a]Based on counties with population 5,000 or more

AN EXPERIMENT ON IMPROVING RESPONSE RATES AND REDUCING CALL BACKS IN HOUSEHOLD SURVEYS

Helen P. Koo, Research Triangle Institute
Jeanne Clare Ridley, Georgetown University
P. V. Piserchia, Research Triangle Institute
Deborah A. Dawson, Georgetown University
Christine A. Bachrach, Georgetown University
Mildrid I. Holt, Research Triangle Institute
Daniel G. Horvitz, Research Triangle Institute

INTRODUCTION

This paper reports the results of preliminary analyses of an experiment undertaken as part of a pretest for a national survey; more refined analyses are underway and will be reported in a subsequent paper. The national survey will cover white ever married women belonging to the birth cohorts of 1901-1910, i.e., women now 66 to 76 years of age. The field work for the pretest survey was carried out by Research Triangle Institute (RTI) from late June through early August 1976.

There were a number of reasons for undertaking the experiment. First, since response rates for older persons tend to be somewhat lower than among younger ones (Atchley, 1969; Benus et al., 1971; Kish, 1965), we decided to experiment with different methods of contacting respondents in hopes of finding a contact method which would improve our response rate. We also hoped that if one of the lower cost methods of contacting respondents proved to be equally or more effective than a higher cost method, we could take advantage of the less expensive method in the main survey. Also, we were concerned with the general problems encountered in surveys today, namely, increasing costs and declining response rates. Finally, a review of recent survey literature indicated that while the effects of contact procedures in mail and telephone surveys on response rates have been rather extensively studied, only a relatively few controlled studies have been carried out in which the methods of contacting respondents for personal interviews have been systematically varied (Brunner and Carroll, 1967; Cartwright and Tucher, 1967).

THE PRETEST SAMPLE

The sample for the pretest survey was designed to include households varying in geographic location, rural-urban residence, and socioeconomic status. From the 100 primary sampling units (PSU's) which make up RTI's national general purpose sample, 10 PSU's were selected in close proximity to or within the four metropolitan areas of Boston, Atlanta, Chicago, and Los Angeles. From each of the 10 PSU's, two secondary sampling units (SSU's), one urban and one rural, were purposively selected on the basis of 1970 census data to have a high proportion of white females belonging to the 1901-1910 birth cohorts and to include areas varying in socioeconomic status (SES), as indicated by median housing and rental values of the SSU. Within each SSU, all households containing a potentially eligible respondent as identified in a household screening were included in the sample for interviewing.

The screening for eligible respondents was carried out in March and April 1976. The response rate for the screening questionnaire was 89.9 percent. In the 2,342 housing units screened, 377 women were identified as potentially eligible for the pretest interview. Since many of the screening interviews were carried out with another member of a household or with a neighbor, we expected that some cases identified as eligible for an interview would not meet the eligibility requirements when actually contacted for the study. In addition, some of the more detailed eligibility requirements were not obtained in the screening interview.

THE EXPERIMENT

In the pretest survey, a 2 X 2 factorial experiment was carried out. The purposes of the experiment were to assess the main and interaction effects of two different types of procedures for contacting the potentially eligible respondents identified in the screening. The two factors tested were the use of a lead letter and the method of initial contact with the respondent. The first factor, at two levels, consisted of the central office sending (or not sending) a lead letter to the respondent prior to any attempt by the interviewer to contact her. The lead letter explained the sponsorship and purposes of the study and indicated that an interviewer would be contacting the respondent. The second factor, also at two levels, consisted of assigning (or not assigning) the interviewer to telephone the potential respondent before attempting a personal visit. In the prior telephone call, the interviewer gave information about the study and tried to make an appointment for an interview.

Within SSU's and within the four areas, the 377 potential respondents were randomly assigned into each of the four treatment combinations. The numbers of cases assigned to each treatment combination were: Lead Letter-Initial Telephone, 100; Lead Letter-Initial Visit, 91; No Letter-Initial Telephone, 96; and No Letter-Initial Visit, 90.

In the Initial Telephone treatment, interviewers were instructed to place up to four telephone calls to the respondent to set up an interview appointment. After no more than four unsuccessful attempts at telephoning, the interviewer made a personal visit to try to contact the respondent. In the Initial Visit treatment, up to two personal visit attempts were made prior to any telephone attempt to contact the respondent. Only after the second unproductive visit was the interviewer allowed to telephone to try to make an appointment.

During the course of the field work, 66 cases had to be removed from the experiment because the assigned treatment could not be followed. An additional 55 potential respondents were determined as ineligible for an interview because of age, race, marital status, or nativity eligibility requirements of the survey, or because they had died or moved. Thus a total of 256 cases, or 68 percent, were considered eligible for an interview and remained in the experiment. In the tables to be presented, only 255 of these 256 eligible, experimental cases are included because the data on one case was received too late to be included in the analysis.

RESULTS

This section describes the results of preliminary analysis of the effects of the experimental treatments on the response rate, the refusal rate, and the cost associated with obtaining completed interviews. Given the very limited time available for analysis, we felt that a reasonable technique for analyzing the results of the experiment was an unweighted analysis of variance of cell means (proportions). Because there were different numbers of individuals per cell and different cell proportions, the cell mean variances were not equal. However, because neither the cell sizes nor proportions varied greatly, it was felt that F-test procedures would be robust against the lack of strict

Table 1. Response Rates, Refusal Rates, Telephone and Visit Effort Ratios, Cost Ratios, and Number of Eligible Experimental Cases for High, Medium, and Low Socioeconomic Status (SES) Secondary Sampling Units (SSU's), by Treatment

	High SES (SSU's)			Medium SES (SSU's)			Low SES (SSU's)			All SSU's		
	Lead letter	No lead letter	Total	Lead letter	No lead letter	Total	Lead letter	No lead letter	Total	Lead letter	No lead letter	Total
A. Response Rates												
Initial Telephone	36.8	43.8	40.0	64.3	66.7	65.4	54.8	72.4	63.3	51,6	63.2	57.0
Initial Visit	35.3	33.3	34.4	50.0	47.4	48.6	52.9	63.6	58.2	47.8	52.2	50.0
Total	36.1	38.7	37.3	56.7	54.8	55.7	53.8	67.7	60.6	49.6	57.3	53.3
B. Refusal Rates												
Initial Telephone	57.9	56.3	57.1	28.6	25.0	26.9	16.1	24.1	20.0	31.3	33.3	32.2
Initial Visit	58.8	46.7	53.1	25.0	47.4	37.1	29.4	27.3	28.4	35.8	37.3	36.6
Total	58.3	51.6	55.2	26.7	38.7	32.8	23.1	25.8	24.4	33.6	35.5	34.5
C. Ratios of Numbers of Telephone (T) and Visit (V) Efforts for all Eligible Cases to Number of Interviews Completed												
Initial Telephone	5.2T 2.4V	8.2T 2.3V	6.8T 2.3V	3.3T 3.0V	2.5T 1.6V	2.9T 2.3V	4.9T 2.4V	3.5T 2.2V	4.1T 2.2V	4.5T 2.5V	4.1T 2.1V	4.4T 2.3V
Initial Visit	2.5T 7.1V	3.3T 7.8V	2.9T 7.3V	1.6T 6.2V	1.3T 5.9V	1.4T 6.2V	2.5T 4.2V	0.8T 3.8V	1.5T 4.0V	2.3T 5.2V	1.3T 5.0V	1.8T 5.0V
Total	3.9T 4.4V	6.2T 4.7V	5.1T 4.6V	2.5T 4.6V	1.8T 4.0V	2.2T 4.3V	3.7T 3.3V	2.2T 3.0V	2.8T 3.1V	3.4T 3.8V	2.8T 3.5V	3.0T 3.8V
D. Ratios of Total Cost of Efforts for all Eligible Cases to Number of Interviews Completed												
Initial Telephone	21.6	23.3	22.4	23.8	13.6	19.0	21.3	18.4	19.7	22.0	18.3	20.1
Initial Visit	51.2	57.4	54.0	44.9	42.9	43.9	31.6	26.6	28.9	38.6	35.2	36.8
Total	35.2	37.5	36.3	33.7	29.1	31.4	26.6	22.5	24.4	30.2	26.6	28.3
E. Number of Eligible Cases												
Initial Telephone	19	16	35	14	12	26	31	29	60	64	57	121
Initial Personal	17	15	32	16	19	35	34	33	67	67	67	134
Total	36	31	67	30	31	61	65	62	127	131	124	255

homogeneity of variances.

In the analyses of variance to be reported below, the SES of SSU's was included as a control variable because earlier regression analyses (results not shown) had indicated that it had important effects on whether a woman responded, refused, etc. The contrasts tested by the analyses of variance were: Initial Telephone versus Initial Visit; Lead Letter versus No Lead Letter; interactions between letter and initial contact treatments; and differences among the three levels of SES (of SSU's). Other interactions, such as those between SES and the treatments, were not included but will be examined in future analyses using other techniques. One disadvantage of analysis of variance of mean values is the inability to derive a "pure" error term directly; contrasts are tested by using, for error, interaction mean squares. In the analyses described below, the SES by initial contact methods, SES by letter treatments, and SES by initial contact methods by letter treatments sums of squares comprised the error sums of squares.

Response Rates. The response rate, defined as the proportion of the eligible experimental cases who completed interviews, was 53.3 percent for the pretest survey. The response rates for the four treatment combinations by the three levels of SES of SSU's are presented in Panel A of Table 1. In the analysis of variance, the simple averages of the response rates of these 12 cells were compared. These simple means for SES were: High SES, 37.3; Medium SES, 57.1; and Low SES, 60.9. The values tested for the treatment contrasts were:

	Lead Letter	No Lead Letter	Total
Initial Telephone	52.0	61.0	56.5
Initial Visit	46.1	48.1	47.1
Total	49.1	54.6	51.8

The results of the analysis of variance of the response rates are presented in Table 2.

Table 2. Analysis of Variance of Unweighted Response Rates by Treatment and Socioeconomic Status (SES) of Secondary Sampling Units (SSU's)

Source of variation	Degrees of freedom	Sums of squares	F	P-value
Assigned Initial Contact	1	.02634	7.90	0.031
Letter	1	.00906	2.72	0.150
Assigned Initial Contact x Letter	1	.00359	1.08	0.339
SES of SSU	2	.12874	19.31	0.003
Error	6	.02001		
Total	11	.18774		

The Initial Telephone versus Visit comparison was statistically significant, but the letter comparison was not. Nor were there significant interaction effects between letter and initial contact treatments. But the SES of SSU's had significant effects on response rates.

The Initial Telephone treatment produced a significantly larger response rate (57 percent) than the Initial Visit treatment (47 percent). The No Lead Letter treatment also resulted in a better response rate (55 percent) than the Lead Letter treatment (49 percent), but this was not statistically significant when the influence of initial contact and SES were adjusted for in the analysis of variance procedures used. Whether the Initial Telephone treatment occurred following a lead letter or not, it obtained higher response rates than the Initial Visit treatment. The SES of SSU's was important, the response rate in low and medium SES areas being markedly better (61 and 57 percent, respectively), than that of the high SES SSU's (37 percent). Although not examined in the analysis of variance, there appear to be some interesting interactions between the treatments and SES, as can be seen in Panel A of Table 1.

Refusal Rates. The refusal rate (proportion of eligible experimental cases refusing to participate) was 34.5 percent for the pretest survey. Besides refusal, there were other reasons for nonresponse, such as physical or mental incapacitation, and being away temporarily. But the present analysis examined only the rates of refusal because refusal accounted for about three-quarters of all nonrespondents.

Panel B of Table 1 presents the refusal rates by treatment and SES. Analysis of variance of the refusal rates found no significant treatment, treatment interaction, or SES effects. Thus while the Initial Telephone treatment seemed to result in better response rates, it did not reduce the incidence of refusal. This indicates that the other components of nonresponse require study.

Cost Effectiveness. Although our cost analysis is even more preliminary than those already discussed, the results are so striking that they are presented here. We do not expect that further analysis will change the conclusions drawn from these results.

For each eligible experimental case was recorded the total number of telephone calls placed and the number of attempts to make personal visits by all field personnel involved in the case to bring it to final resolution. The telephone and visit efforts were counted regardless of outcome (such as busy signal, no one at home, spoke with relative, etc.)

Panel C of Table 7 presents the ratios of the number of telephone and visit efforts expended on all eligible cases to the number of interviews completed. The Initial Telephone treatment required more telephone efforts (4.4 on the average) but fewer visit efforts (2.3) to obtain a completed interview than did the Initial Visit treatment (which took an average of 1.8 telephone efforts and 5.0 visit efforts). It appears that the prior telephone calls were successful in achieving their goal of setting up interview appointments and that generally the appointments were kept, thus reducing the number of personal visits required. Overall, whether a lead letter had been sent or not did not have much influence on the level of efforts expended, although some differences in the efforts needed by the two letter treatments appeared under the various SES conditions.

To study cost effectiveness, the telephone and visit efforts were converted to cost. The expense reports for as much of the interviewing period as was available were examined for an interviewer randomly selected from each of the four study areas. The data indicated that the average mileage cost per visit effort was \$2.75. Since detailed telephone expenses were not readily available, direct telephone costs were assumed to be \$.20 per call effort. The cost of labor (C) of each interviewer was assumed to be a function of the number of telephone and visit efforts she made, and the total cost then was found as a function of the direct costs and labor costs for telephone and visit efforts:

$$\text{Total cost} = (C_{\text{phone}} + \$.20)\ (\text{No. calls}) + (C_{\text{visit}} + \$2.75)\ (\text{No. visits})$$

Components of the labor costs for the four interviewers were estimated by least squares, yielding an estimate of the total cost as \$.90 per telephone effort and \$6.97 per visit effort. This cost function was used to convert the effort data to cost.

To assess cost effectiveness, the cost of pursuing all eligible women in a given treatment and SES condition was divided by the number of respondents in that condition who actually completed an interview. These cost ratios are presented in Panel D of Table 1.

Analysis of variance of the logarithm of the cost ratios was performed, with the results seen in Table 3.

Table 3. Analysis of Variance of the Logarithm of Cost Ratios by Treatment and Socioeconomic Status (SES) of Secondary Sampling Units (SSU's)

Source of variation	Degrees of freedom	Sums of squares	F	P-value
Assigned Initial Contact	1	1.55484	31.50	0.001
Letter	1	0.04460	0.90	0.379
Assigned Initial Contact x Letter	1	0.02331	0.47	0.518
SES of SSU	2	0.28349	2.87	0.133
Error	6	0.29618		
Total	11	2.20243		

The logarithm of cost ratios was analyzed since it was thought that the ratio of cost ratios between two treatment conditions was more meaningful than the difference between cost ratios.

The analysis of variance did not show significant differences in cost ratios among letter treatments or SES levels; nor were there significant treatment interactions. Only the initial contact treatments differed significantly in their cost ratios. The Initial Telephone treatment resulted in much smaller cost ratios than the Initial Visit treatment. Panel D of Table 1 shows that the Initial Visit treatment was superior in cost effectiveness following either of the letter treatments and in every SES condition. In the various letter treatment by SES conditions, the cost ratios of the Initial Telephone treatment ranged from 32 to 69 percent of those of the Initial Visit treatment. Thus, not only did the Initial Telephone treatment produce better response rates than the Initial Visit treatment, it also was a more cost effective method of obtaining the interviews.

The cost function used to convert the effort data to cost was estimated on the basis of incomplete data. When calculated on the basis of complete data, the ratio of cost associated with a visit effort to the cost of a telephone effort may be found to differ from the 7:1 ratio reported above. But even if the ratio should be only 2:1, the cost ratio of obtaining an interview by the Initial Telephone treatment would still be only 76 percent of that of the Initial Visit treatment (based on the telephone and visit efforts reported for the two treatments in Panel C of Table 1). Therefore, it seems safe to judge that when the cost function is recalculated based on complete data, the conclusion will remain that the Initial Telephone treatment was more cost effective than the Initial Visit treatment.

SUMMARY AND CONCLUSION

In the preliminary analyses presented, there is no evidence that sending a lead letter to a potential respondent affected the response or refusal rate, nor is there evidence that the lead letter influenced the cost effectiveness of obtaining interviews. If anything, it may have tended to depress response rates under certain SES by initial contact conditions.

The early analyses also showed that instructing interviewers to attempt to reach their respondents first by telephone to set up an appointment was an effective method. The Initial Telephone method produced better response rates, and at less relative cost, than trying to reach the respondent initially by a personal visit. It was cost effective because although more telephone calls had to be placed, fewer visits (the more expensive effort) needed to be made to obtain interviews. The Initial Telephone treatment did not reduce the refusal rate, however.

No significant interaction effects were found between the letter and initial contact treatments, either with respect to the response rate, refusal rate, or cost effectiveness.

Finally, the SES of SSU's affected response rates, but not refusal rates or cost effectiveness. High SES areas yielded much worse response rates than medium or low SES areas.

The alarmingly high refusal rate obtained in the pretest survey clearly requires further investigation. Until a detailed study has been performed, one can only speculate on the factors responsible for such a large proportion of the elderly women refusing to be interviewed.

The analyses reported are preliminary. They have yielded some interesting findings, but they should not be considered conclusive. In the near future, we plan to reexamine the main and interaction effects of the treatments, using other methods of analysis (including weighted methods), and also to investigate various interaction effects between SES and treatments. In particular, the categorical data methods of Grizzle, Starmer, and Koch (1969) will be used to yield a complete analysis of the response and refusal rates, as well as other components of the nonresponse rate. Cost effectiveness, based on more complete data, will also be further studied. For the cost ratios, Taylor series approximations to their variances will be used to perform additional analysis with respect to the various SES by treatment interactions.

REFERENCES

Atchley, R. C. 1969. "Respondents vs. refusers in an interview study of retired women: an analysis of selected characteristics," *Journal of Gerontology*, 24:42-47.

Benus, J. and J. C. Ackerman. 1971. "The problem of nonresponse in sample surveys," in *Working Papers on Survey Research in Poverty Areas* (Lansing, Witney, and Wolfe, eds.), Chapter 2, pp. 26-59.

Brunner, G. A. and S. J. Carroll, Jr. 1967. "The effect of prior telephone appointments on completion rates and response content," *Public Opinion Quarterly*, 31:652-654.

Cartwright, A. and W. Tucker. 1967. "An attempt to reduce the number of calls on an interview inquiry," *Public Opinion Quarterly*, 31:299-302.

Grizzle, J. E., C. F. Starmer, and G. G. Koch. 1969. "Analysis of categorical data by linear models," *Biometrics*, 25:489-504.

Kish, L. 1965. *Survey Sampling*. New York: John Wiley and Sons, Inc., pp. 532-562.

PERCEIVED HEALTH STATUS - HOW GOOD IS PROXY REPORTING?

Mary Grace Kovar and Ronald W. Wilson
National Center for Health Statistics

The measurement of health status has been a perennial problem which has plagued and intrigued workers in the health field for many years. Measures from vital statistics, primarily mortality measures based on death rates, were used as the best available method and are still used in the absence of other data, although it has long been recognized that the concept of health status includes the extent of disability and morbidity in the living population. Even sophisticated measures of death alone are not sufficient.

The attempt to incorporate measures of the health of the living population has led to an extensive literature on health indexes (the National Center for Health Statistics' Clearinghouse on Health Indexes is an excellent bibliographic reference for recent work). Many of the indexes for health status which have now been developed are extremely sophisticated and exhibit a high degree of ingenuity.[1, 2, 3, 4] Unfortunately, many of them depend on data which are not readily available. Some of the indexes require measures which are extremely difficult to obtain without long and costly household interview surveys with highly structured questionnaire design. These require careful interviewer training for consistency in data collection and extensive data processing to combine the responses into an index.

Since the need for a measure of health status has increased with the passage of legislation (P.L. 93-641 requires the measurement of health status for health planning), it would be extremely helpful to have a measure based on easily collectable data. One approach is to simply ask people their opinions about their own health. This has been done on a number of studies and the responses have been found to correlate highly with other measures of status, need, and utilization of health services.

The version of the question used in the Health Interview Survey conducted by the National Center for Health Statistics is "Compared to other persons ...'s age, would you say that his health is excellent, good, fair, or poor?" Some examples of the proportion of persons in each age-health status group who report specified measures of utilization or disability are shown in Table 1. It is obvious that with each decline in reported health status, the proportion for whom utilization of physicians' or hospital services is reported increases as does the proportion for whom limitation of activity is reported. Such relationships are consistent with the findings from other studies indicating that the simple question on health can be used as a predictor for other measures of interest. Perceived health status was also consistent with the reporting found on demographic measures. Old people, poor people, rural residents, e.g., had relatively high proportions for whom poor health was reported, corresponding with the generally high levels of disability among these groups.

It must be pointed out, however, that the question on the Health Interview Survey was not asked in isolation but was included as part of a questionnaire in which numerous other questions about health were asked. It followed questions on two-week disability days, physician visits, and limitation of activity. The relationships shown here may be influenced somewhat by the questionnaire context.

If perceived health status is indeed a useful measure as it appears to be, then a methodological question arises in the collection of the information. Is it necessary to ask each individual the question or can one family member report for the entire family living in the same household? This question is important, as all survey researchers know, because it is more costly and time consuming to collect data if each person must answer individually than it is if one respondent can report for the household.

An opportunity to evaluate the effect of self and proxy respondents arose from an experiment conducted by the Health Interview Survey in the Spring of 1972. The background and field experience of this study has been reported on in previous papers in the Proceedings of the Social Statistics Section and an extensive discussion of the methodology is contained in the technical appendix of a forthcoming NCHS publication.[5, 6] In brief, the independent weekly samples were assigned to experimental and control groups. During half of the quarter, the regular HIS respondent rules, where frequently only one adult responds for the household, were in effect. During the other half all adults were required to respond for themselves and additional visits were made as needed. Other than the change in the respondent rule, all regular survey procedures remained in effect so that the experiment was conducted in the context of an ongoing National survey and the quality of the regular data collection was maintained. National estimates were made separately for both the control and experimental periods so that aggregate data were available to compare the National estimate as derived with the usual respondent procedures with those under the self-respondent rule. Only 5 percent of adult males and 3 percent of females had a proxy respondent during the experimental weeks in contrast with 41 and 14 percent during the control or usual procedure weeks.

We will now look at the differences that occur in perceived health status using the experimental respondent rules and the standard respondent rules. Tables 2-7 present estimates of the percent of the population in a number of categories who would be classified as being in excellent, good, fair or poor health as derived by each procedure. Data are presented for both males and females by color, age, family income, marital status and education of the head of the family. In addition, the percent difference between the experimental and the control procedures is shown for each category. A negative difference indicates that the estimates based on the standard respondent rule yields a higher response, a positive difference indicates a higher level using the self-respondent rule. Table 2 shows that in general the standard respondent rules appear to yield somewhat higher estimates of the proportion of the persons, both male and female, whose health is rated as "excellent." For example, overall 49.2 percent of males are rated as "excellent" under the standard rule while 45.4 percent are rated as "excellent" under the self-response rule. None of the differences for females are statistically significant at the 10 percent level. Table 3 shows the estimates of persons rated as being in "good" health. The pattern here is reversed with the standard rule yielding lower rates than the self-rule. The results from these two tables seem to indicate that respondents tend to be more critical, i.e., more likely to rate themselves as "good" rather than "excellent," when evaluating their own status, than when evaluating someone else's health, usually that of a spouse. Some of the differences by socio-demographic categories are difficult to explain. The easiest explanation of these differences applies to the one found for married men who are least likely of all groups to be self respondents under the standard rule. Their wives, or whoever responds for them, are apparently more likely to report "excellent" health for them than they report for themselves. This results in a higher proportion of "excellent" ratings than under the self response rule.

The differences between the control and experimental groups on the reporting of "excellent" or "good" health status disappear when the two health status categories are combined (Table 4). There are no significant differences between the estimates based on the two respondent procedures.

Tables 5 and 6 show the impact of the respondent rules on the reporting of "fair" and "poor" health status. The pattern is less clear at this end of the health status spectrum. Only one statistically significant difference was found at the "fair" health status level. Only three significant differences were found at the "poor" health status level. Also, there is no pattern in the direction of the differences. However, as with the more positive end of the health status continuum, when the "poor" and "fair" categories are combined, (Table 7) the only significant difference occurred for females in families where the head had less than a high school education.

Conclusion

We have looked at the impact of two different respondent rule procedures, the use of a household respondent versus all self-respondents, on the reporting of perceived health status. While some differences occur in the "excellent" and "good" health categories, with the all self-respondent procedure giving lower estimates of "excellent" health and higher estimates of "good" health, the differences are minimal when the two categories are combined. By combining the "fair" and "poor" categories the differences also disappear. These findings would seem to indicate that if "excellent" plus "good" health can be interpreted as the positive end of a health status continuum and "fair" plus "poor" as the negative end, then the use of standard respondent rules provides a population estimate of health status comparable to a self-reported estimate. However, this is only a measure of positive or negative direction on a health status continuum. If the concern is for the strength or degree, that is, the differentation between "good" and "excellent" health status and to a lesser extent between "fair" and "poor" health status, then the two response rules provide somewhat different results. Even so, the differences that occur between the self and standard rules in the "excellent" and "good" categories are not large, between about 5 and 16 percent of the estimate under the standard rules.

Therefore, in conclusion, it appears that household respondents can be used to report the perceived health status of other household members.

REFERENCES

1. Berry, C.C., Bush, J.W., and Kaplan, R.M.: Testing for Variation in Social Group Preferences for Function Levels of a Health Index. Presented at the Annual Meeting of the American Statistical Association, August 25-28, 1975.

2. Chen, M.K.: "A Comprehensive Population Health Index Based on Mortality and Disability Data," Forthcoming in Social Indicators Research.

3. Pollard, W.E., Bobbitt, R.A, Bergner, M., Martin, D.P., and Gilson, B.S.: The Sickness Impact Profile: Reliability of a Health Status Measure. Medical Care 14(2):146-155, February 1976.

4. Sullivan, D.F.: A Single Index of Mortality and Morbidity. HSMHA Health Reports 86(4): 347-354, April 1971.

5. Haase, K.W. and Wilson, R.W.: "The Study Design of an Experiment to Measure the Effects of Using Proxy Responses in the National Health Interview Survey," 1972 Social Statistics Section, Proceedings of the American Statistical Association, 289-293.

6. Kovar, M.G. and Wright, R.A.: "An Experiment with Alternate Respondent Rules in the National Health Interview Survey," 1973 Social Statistics Section, Proceedings of the American Statistical Association, 311-316.

Table 1. Percent of persons in specified health status and age category with ten or more doctor visits in past 12 months, with one or more hospital episodes in past 12 months, and with limitation in major activity: United States, 1974

Health status	Age: Under 17	17-44	45-64	65 and over
Doctor visits:	Percent with 10 or more doctor visits in past 12 months			
Health status				
Excellent	3.7	6.0	4.6	8.5
Good	7.1	10.1	8.7	14.0
Fair	23.1	23.3	22.2	25.6
Poor	51.3	48.8	45.6	43.6
Hospital episodes:	Percent with one or more hospital episodes in past 12 months			
Health status				
Excellent	4.3	9.0	7.0	10.3
Good	6.5	12.7	11.1	13.6
Fair	14.9	22.4	19.0	21.7
Poor	31.5	40.2	35.6	35.2
Limitation in major activity:	Percent with limitation in major activity			
Health status				
Excellent	0.6	1.3	3.2	14.4
Good	2.1	5.2	11.2	31.2
Fair	13.7	23.0	42.8	60.2
Poor	48.4	68.1	86.0	88.4

NOTE: Data in this paper are based on household interviews of the civilian, noninstitutionalized population.

SOURCE: Unpublished data from the Health Interview Survey, National Center for Health Statistics.

Tables 2, 3, & 4. POSITIVE HEALTH STATUS: Rate per 100 population based on self-respondent and standard-respondent rules and percent difference between these rates by sex and selected population characteristics: United States, Spring 1972

Selected characteristics	Male			Female		
	Self respondent rule	Standard respondent rule	Percent difference	Self respondent rule	Standard respondent rule	Percent difference
Table 2. EXCELLENT						
Total - 17+ years	45.4	49.2	-7.7†	40.9	41.9	-2.4
Color						
White	45.9	50.7	-9.5†	41.9	43.7	-4.1
All other	40.4	37.3	8.3	32.2	28.3	13.8
Age						
17-24 years	54.8	58.1	-5.7	50.9	52.8	-3.6
25-44 years	56.8	59.5	-4.5	49.7	49.1	1.2
45-64 years	35.8	40.7	-12.0†	33.6	34.9	-3.7
65 years and over	27.6	28.6	-3.5	27.2	29.1	-6.5
Family income						
Less than $5,000	31.2	29.6	5.4	28.2	27.0	4.4
$5,000-$9,999	41.4	45.2	-8.4†	38.5	40.4	-4.7
$10,000-$14,999	51.1	55.7	-8.3	48.1	50.5	-4.8
$15,000 or more	57.9	64.0	-9.5†	53.8	56.5	-4.8
Marital status						
Married	45.8	50.7	-9.7†	42.3	43.5	-2.8
Widowed, separated, divorced	33.8	34.0	-0.6	32.7	33.9	-3.5
Never married	49.3	49.8	-1.0	47.6	47.8	-0.4
Education of head of family						
Less than 12 years	32.3	33.5	-3.6	29.5	28.5	3.5
12 years	49.6	55.8	-11.1†	44.3	45.9	-3.5
13 years or more	61.4	66.5	-7.7†	57.6	60.0	-4.0
Table 3. GOOD						
Total - 17+ years	38.2	35.3	8.2†	41.1	39.1	5.1†
Color						
White	38.1	34.6	10.1†	41.0	38.7	5.9†
All other	38.8	40.8	-4.9	42.2	42.4	-0.5
Age						
17-24 years	37.3	35.1	6.3	40.8	38.1	7.1
25-44 years	34.2	32.6	4.9	39.8	38.9	2.3
45-64 years	42.1	38.1	10.5†	43.8	40.4	8.4†
65 years and over	40.2	36.5	10.1	38.9	38.1	2.1
Family income						
Less than $5,000	36.3	36.2	0.3	40.0	39.2	2.0
$5,000-$9,999	40.9	38.2	7.1	44.7	40.2	11.2†
$10,000-$14,999	39.4	36.4	8.2	40.5	40.0	1.3
$15,000 or more	34.2	30.0	14.0	36.5	35.7	2.2
Marital status						
Married	38.0	34.0	11.8†	41.6	39.5	5.3
Widowed, separated, divorced	39.2	41.4	-5.3	39.8	37.3	6.7
Never married	38.3	37.8	1.3	40.4	40.4	0.0
Education of head of family						
Less than 12 years	41.1	40.8	0.7	43.8	42.4	3.3
12 years	39.6	34.3	15.5†	42.8	40.8	4.9
13 years or more	32.2	27.7	16.2†	34.3	31.5	8.5
Table 4. EXCELLENT + GOOD						
Total - 17+ years	83.5	84.5	-1.2	82.0	81.0	1.2
Color						
White	84.0	85.3	-1.5	82.9	82.4	0.6
All other	79.3	78.1	1.5	74.4	70.7	5.2
Age						
17-24 years	92.1	93.2	-1.2	91.8	90.9	1.0
25-44 years	91.0	92.1	-1.2	89.5	88.0	1.7
45-64 years	77.9	78.8	-1.1	77.4	75.3	2.8
65 years and over	67.8	65.1	4.1	66.1	67.2	-1.6
Family income						
Less than $5,000	67.5	65.8	2.6	68.2	66.2	3.0
$5,000-$9,999	82.3	83.4	-1.3	83.2	80.6	3.2
$10,000-$14,999	90.5	92.1	-1.7	88.7	90.6	-2.1
$15,000 or more	92.1	94.0	-2.0	90.3	92.2	-2.1
Marital status						
Married	83.9	84.8	-1.1	83.9	82.9	1.2
Widowed, separated, divorced	73.0	75.4	-3.2	72.5	71.2	1.8
Never married	87.6	87.5	0.1	88.0	88.3	-0.3
Education of head of family						
Less than 12 years	73.4	74.4	-1.3	73.3	70.8	3.5
12 years	89.1	90.1	-1.1	87.1	86.7	0.5
13 years or more	93.6	94.2	-0.6	91.9	91.5	0.4

† Difference is statistically significant at the 0.10 level.

SOURCE: Unpublished data from the Health Interview Survey, National Center for Health Statistics.

Tables 5, 6, & 7. NEGATIVE HEALTH STATUS: Rate per 100 population based on self-respondent and standard-respondent rules and percent difference between these rates by sex and selected population characteristics: United States, Spring 1972

Selected characteristics	Male			Female		
	Self respondent rule	Standard respondent rule	Percent difference	Self respondent rule	Standard respondent rule	Percent difference
Table 5. FAIR						
Total - 17+ years	11.8	11.3	4.4	13.9	14.2	-2.1
Color						
White	11.5	10.8	6.5	13.2	13.2	0.0
All other	14.2	15.3	-7.2	20.2	22.4	-9.8
Age						
17-24 years	6.0	5.3	13.2	6.7	8.4	-20.2
25-44 years	6.8	5.9	15.3	8.7	9.2	-5.4
45-64 years	15.1	15.4	-1.9	17.6	17.9	-1.7
65 years and over	23.4	24.6	-4.9	24.4	23.9	2.1
Family income						
Less than $5,000	22.3	22.3	0.0	24.0	24.7	-2.8
$5,000-$9,999	12.1	13.3	-9.0	13.2	15.1	-12.6
$10,000-$14,999	7.8	6.3	23.8	9.5	7.0	35.7†
$15,000 or more	6.0	4.9	22.4	7.3	6.3	15.9
Marital status						
Married	11.6	11.0	5.5	13.0	13.1	-0.8
Widowed, separated, divorced	18.4	18.1	1.7	19.4	20.7	-6.3
Never married	9.2	9.3	-1.1	9.3	8.6	8.1
Education of head of family						
Less than 12 years	18.7	18.3	2.2	20.3	21.8	-6.9
12 years	8.4	7.8	7.7	10.3	9.9	4.0
13 years or more	4.5	4.5	0.0	6.5	7.0	-7.1
Table 6. POOR						
Total - 17+ years	3.8	3.5	8.6	3.2	3.9	-17.9
Color						
White	3.5	3.2	9.4	3.1	3.6	-13.9
All other	6.0	6.1	-1.6	4.5	6.2	-27.4
Age						
17-24 years	0.8	0.8	0.0	0.5	0.0	0.0
25-44 years	1.3	1.5	-13.3	1.2	2.0	-40.0
45-64 years	6.2	4.9	26.5	4.2	6.0	-30.0†
65 years and over	7.6	9.7	-21.6	8.0	7.9	1.3
Family income						
Less than $5,000	9.4	11.2	-16.1	7.1	8.3	-14.5
$5,000-$9,999	4.7	2.9	62.1†	2.9	3.5	-17.1
$10,000-$14,999	1.0	1.2	-16.7	1.3	1.9	-31.6
$15,000 or more	0.8	0.8	0.0	1.2	1.3	-7.7
Marital status						
Married	3.6	3.5	2.9	2.2	3.2	-31.3†
Widowed, separated, divorced	7.9	5.8	36.2	7.4	7.3	1.4
Never married	2.5	2.6	-3.8	1.4	2.2	-36.4
Education of head of family						
Less than 12 years	6.9	6.6	4.5	5.3	6.7	-20.9
12 years	1.4	1.3	7.7	1.6	2.1	-23.8
13 years or more	1.4	1.2	16.7	1.1	1.3	-15.4
Table 7. FAIR + POOR						
Total - 17+ years	15.6	14.9	4.7	17.2	18.2	-5.5
Color						
White	15.0	14.1	6.4	16.2	16.8	-3.6
All other	20.3	21.4	-5.1	24.7	28.5	-13.3
Age						
17-24 years	6.8	6.1	11.5	7.2	8.4	-14.3
25-44 years	8.1	7.4	9.5	10.0	11.2	-10.7
45-64 years	21.3	20.3	4.9	21.7	23.9	-9.2
65 years and over	31.0	34.3	-9.6	32.4	31.8	1.9
Family income						
Less than $5,000	31.7	33.6	-5.7	31.1	33.0	-5.8
$5,000-$9,999	16.8	16.2	3.7	16.1	18.7	-13.9
$10,000-$14,999	8.7	7.5	16.0	10.8	9.0	20.0
$15,000 or more	6.8	5.7	19.3	8.4	7.5	12.0
Marital status						
Married	15.2	14.5	4.8	15.2	16.3	-6.7
Widowed, separated, divorced	26.3	24.0	9.6	26.8	28.0	-4.3
Never married	11.7	11.9	-1.7	10.7	10.8	-0.9
Education of head of family						
Less than 12 years	25.7	24.9	3.2	25.7	28.4	-9.5†
12 years	9.8	9.2	6.5	11.9	12.1	-1.7
13 years or more	5.9	5.7	3.5	7.6	8.4	-9.5

† Difference is statistically significant at the 0.10 level.

SOURCE: Unpublished data from the Health Interview Survey, National Center for Health Statistics.

Table 8. Estimated population by sex and selected population characteristics: United States, Spring 1972

Selected characteristic	Male	Female
	Population in thousands	
Total 19+ years	61,218	69,591
Age		
19-44 years	32,919	35,861
45-64 years	20,026	22,154
65 years and over	8,273	11,576
Color		
White	54,844	61,790
All other	6,374	7,801
Family income		
Less than $5,000	11,741	17,200
$5,000-$9,999	18,059	20,191
$10,000-$14,999	14,941	15,041
$15,000 or more	13,278	13,075
Education of head of family		
Less than 12 years	25,677	30,188
12 years	17,909	20,844
13 years or more	16,930	17,768
Marital status		
Married	46,170	47,396
Widowed, separated, divorced	5,103	14,674
Never married	9,945	7,521

NOTE: For official population estimates see Bureau of the Census reports on the civilian population of the United States, in Current Population Reports, Series P-20, P-25, and P-60.

SOURCE: Unpublished data from the Health Interview Survey, National Center for Health Statistics.

A MODEL OF POPULATION GROWTH INVOLVING MORTALITY-FERTILITY INTERACTIONS: SOME TENTATIVE RESULTS FOR INDIA *

G. K. Kripalani, Western Michigan University

In analyses of population growth in the low income countries, there have not been many attempts to consider in an integrated and comprehensive way the interactions between the major determinants of change. A reasonably adequate framework of analysis must take into account mortality-fertility interactions in low income societies as significant reductions in mortality are achieved and sustained. Only such an empirically realistic framework of analysis may be expected to provide meaningful insights into the role played by policy and non-policy parameters in the model and to provide a valid basis for making population projections.

There is general consensus that high rates of population growth witnessed in many parts of the low-income world in the last two decades were due, in large measure, to very significant declines in mortality rather than rises in fertility rates. What has remained a controversial issue, however, is the nature and extent of household fertility behaviour response to continuously improving mortality experience.

While still treating mortality changes as exogenous, the present research hypothesizes that birth rates may respond in downward fashion to declines in death rates. The main elements of the hypothesis pertain to household family formation behaviour and are: (a) the concept of Desired Family Size; (b) household response to past mortality changes via lagged adjustment in planned fertility; (c) 'myopic' expectations about future mortality improvements; (d) possible changes in (i) desired family size, (ii) preferred child-spacing pattern and (iii) household behaviour parameters reflecting degree of risk-aversion in response to mortality improvements and the historical consistency of this process. The expectations hypothesis involves distributed lags and myopic expectations. The expected gain in the force of mortality in the future c periods hence, expectations formed at time t is the product of the expected gain in the current period and a myopia factor. Theoretically speaking, the myopia factor may be handled as an Exponential process or a Poisson process (a) independent of or (b) dependent on the past history of the process. The past history of the process will include variations in past mortality gains and the time length of the process. The expectations for the current period involve distributed lags and incorporate the effect of the past history of the process. Mathematically the hypothesis is written as:

$$E\ y\ (t + c/t) = M\ (t + c/t).\ E\ y\ (t/t) \quad \ldots.(1)$$

where

$$E\ y\ (t/t) = L\ E\ y\ (t - 1/t - 1) + (1 - L)\ y\ (t - 1) \quad \ldots.(2)$$

where L = lag parameter lying between 0 and 1;

$E\ y\ (t + c/t)$ = expected change in the force of mortality in the period $(t + c)$, expectations formed at time t; and

$y\ (t - 1)$ = actual change in the force of mortality observed in the time period $(t - 1)$.

$M\ (t + c/t)$ = Myopia factor at time t for time period $(t + c)$ in the future.

Changes in mortality rates play an important role in this model on account of the concept of the Desired Completed Family Size and its fixity in the face of changes in mortality. Declining mortality rates and the taking into account of mortality improvements in the decision-making process for determining planned fertility rates imply that under the assumptions made, planned fertility rates respond to changes in mortality via number of currently living children and expected survival rates. Decline in mortality rates will induce declines in planned fertility rates in order to achieve the goal of a fixed DCFS.

In determining planned fertility or the desired number of children to be born over the remaining child-bearing ages of the mother expected survival rates play a critical role. These rates which are subjectively conceived by the family are influenced by the attitude towards past and future mortality changes, attitude toward risk, etc. In many poor traditional societies which are the main focus of this study, where parents may, in a sense, be considered as regarding children as capital on which they depend in their old age, there may be and possibly is a tendency to play "safe" within reasonable limits and these reasonable limits will depend upon family's tastes regarding risk-taking and risk-aversion. For a poor developing society, it may be assumed that people are generally prone to be more of risk-averters type, when the question of their basic livelihood and survival in old age is concerned, recognizing the almost total absence of Social Security programs in these societies. This implies the existence of a bias towards "over-saving" for the future in the form of children. An important relevant issue in this context is that of increasing uncertainties to which parents expose themselves by their dependence on capital in children in a situation of rapidly changing social and cultural values in these societies as economic development occurs. As a consequence, expected income streams in old age from investments in children will be realized with a lower probability than previously. When this holds true, the 'bias' towards over-savings in the form of children will be strengthened. A family's attitude towards risk-taking, 'bias' and other factors will determine the extent to which it will take into account past and expected future mortality improvements to determine its future fertility. These elements are reflected in and captured by the lag parameter L, and the myopia parameter M. Lagged response and myopic expectations are devices that capture these 'biases' and neutralize the reduced probability of realizing expected future income streams.

There may be other important reasons for lagged fertility response to mortality improvements. In realistic situations, households may not be aware of mortality improvements unless they have been underway for quite a while and even then they may not be able to form accurate quantitative judgments about the magnitude of these changes and their impact on family size. Further, in tradition-bound poor societies with a long history of unchanging environment, where traditional values and attitudes have long experience to back them, response to change may be expected to develop only slowly and cautiously. Safe response to these changes may also require the acquisition of knowledge of the use of contraceptives, etc., and of the availability of such devices at reasonable costs. For the present it is not our purpose to make a detailed investigation of the whole host of relevant forces or considerations relevant to low income developing countries that play a part in lagged and partial fertility response by households to declining mortality, but only to visualize the feasibility of such response.

A female population whose family formation behaviour has a goal of achieving a fixed Desired Completed Family Size (DCFS) will respond to mortality changes by appropriate adjustment in their planned fertility. Mortality improvements unaccompanied by any downward adjustments in actual fertility will result in an accelerated population growth of existing numbers and further the households will discover that the number of children surviving to adulthood exceeds the quantity aimed at. Even if instantaneous and 'full' adjustments in planned fertility are made immediately following mortality disturbance and are realized, in the early stages for a time, however, the population will grow at a rate faster than previously on account of the fact that more females would survive to adulthood and higher ages than would have been the case in the absence of any downward disturbances in mortality. Thus mortality improvements will lead in the immediate future to an accelerated rate of population growth even if instantaneous and 'full' fertility adjustments accompany mortality changes. In cases in which fertility responses to mortality declines are neither instantaneous nor 'full' additional sources contributing to accelerated rate of population growth will operate. Both the magnitude and the duration of this process will depend principally upon the lag parameter. In elaborate models in which family formation takes place over a life and time span, the myopia parameter representing expected mortality improvements in the future will also be relevant in determining the sequence of the rates of population growth.

A Simplified Version

A simplified formulation is developed for the purpose of gaining qualitative insights into the role played by model parameters and for throwing into sharp focus the relationship between fertility and mortality rates in determining age composition structure and rate of population growth. The population is divided into four equal age groups 0, 1, 2 and 3. Age group 0 relates to children and age group 1 consists of all adults in child-bearing period of life. Children are born to females in age group 1 only. Since all children are born in one time period, myopia is absent. The myopia parameter M (t + c/t + c) is equal to unity.

Family formation behaviour assumptions are: (i) The family is aiming at a Desired Completed Family Size (DCFS) which is assumed given and fixed and does not change as mortality rates change. DCFS is defined as the number of children born who are desired to survive to adulthood, say age 1. (ii) The family has a fixed preferred child-spacing pattern which does not change as mortality and fertility changes occur. (iii) Families respond to mortality improvements by lagged adjustments in planned fertility. Since a single period covers the whole child-bearing time span, it will be unrealistic to ignore completely mortality changes currently under way whose impact on emerging profile of children living at various ages of the mother's child-bearing span could easily be visible. To take into account the mortality disturbances during the current period, relation (2) has been modified as follows:

$$E\,y(t/t) = L\,E\,y(t-1/t-1) + (1-L)\,y(t) \qquad \ldots\ (3)$$

Derivation of Formulae

Let a(x, t+c) denote change in the force of mortality at age x during the time period t+c. The mortality disturbance starts at time t. Prior to time t, age-specific mortality and fertility schedules remained unchanged. Let u(x, t+c) denote the force of mortality at age x at time t+c. When the discussion is general and applies to all age groups, we will, for the sake of brevity, use the notation u(t+c) to refer to the force of mortality at any age at time t+c. We now have:

$$u(t+c+g) = u(t) - a(t) - a(t-1) - \ldots - a(t+c-1) - ga(t+c) \qquad \ldots. (4)$$

where a(t+c) refers to improvement in the force of mortality in the relevant age group during time period (t+c), c is an integer and g a fraction.

Let S(x, t-1) refer to before-disturbance one period survival rate schedule. When the discussion is general and applicable to all age groups we will use the notation S(t-1) or simply S as the predisturbance survival rate schedule. Let SR(t+c) refer to actual survival rate for any age during time period (t+c); that is from time (t+c) to (t+c+1), after disturbance. Now

$$u(t+g) = u(t) - ga(t) \qquad \ldots. (5)$$

where g lies between 0 and 1. Hence, we have:

$$SR(t) = \exp\left[-\int_0^1 u(t+g)\, dg\right] \qquad \ldots. (6)$$

$$= S. \exp\left[a(t)/2\right] \qquad \ldots. (7)$$

$$SR(t+1) = \exp\left[-\int_0^1 u(t+1+g)\, dg\right]$$

$$= S. \exp\left[a(t) + a(t+1)/2\right] \qquad \ldots. (8)$$

and so on. In general:

$$SR(t+c) = S. \exp\left[\sum_{c=o}^{c-1} a(t+c) + a(t+c)/2\right] \qquad \ldots. (9)$$

Let y(x, t+c) denote the periodic rate of decline in the force of mortality at age x during time period t+c, where y(x, t+c) is a function of x (age) and t+c (time). Let Ey(t+c) denote the expected change in the force of mortality over time period t+c. Using the lag relationship (3) and assuming that no mortality gains were expected prior to time period t, we derive:

$$Ey(t) = (1-L)\, a(t) \qquad \ldots. (10)$$

$$Ey(t+1) = (1-L)\, [La(t) + a(t+1)] \qquad \ldots. (11)$$

and in general:

$$Ey(t+c) = (1-L)\, [L^{c}a(t) + L^{c-1} a(t+1) + \ldots a(t+c)] \qquad \ldots. (12)$$

Expected survival rates denoted by ESR(t) can be expressed in terms of pre-disturbance values as follows:

$$ESR(t) = \exp\left[-\int_0^1 Eu(t+g)\, dg\right] \qquad \ldots. (13)$$

$$= S. \exp\left[(1-L)\, a(t)/2\right] \qquad \ldots. (14)$$

Similarly in general:

$$ESR\ (t+c) = S. \exp\left[\sum_{c=o}^{c-1} a(t+c).\right] \exp\left[(1-L) \sum_{j=o}^{c} L^{c-j}\, a(t+j)/2\right] \qquad \ldots. (15)$$

The households fertility response to mortality improvements is such that the goal of DCFS is to be attained. Let G be the DCFS. If D(t+c) is the planned fertility for period t+c, we have the general relationship:

$$D(t+c).\ ESR(t+c) = D.S. = G.$$

where D is the pre-disturbance total fertility. It can be shown that:

$$D(t) = D. \exp\left[-(1-L)\, a(t)/2\right] \qquad \ldots. (16)$$

$$D(t+1) = D \exp\left[-a(t)\right]. \exp\left[-(1-L). \{La(t) + a(t+1)\}/2\right] \qquad \ldots. (17)$$

$$D(t+c) = D \exp\left[-\sum_{c=o}^{c-1} a(t+c)\right]. \exp\left[-(1-L) \sum_{j=o}^{c} L^{c-j}. a(t+j)/2\right] \qquad \ldots. (18)$$

These relations show that the female children born per potential mother will continuously decline from the initial level of D before disturbance and asymptotically approach the full adjustment level of

$$D \exp\left[-\sum_{j=o}^{c-1} a(t+j) - a(t+c)/2\right] \ldots. (19)$$

Let G(t+c) refer to long-term stable population growth factor corresponding to age-specific mortality and fertility schedules of period (t+c). It can be shown that:

$$G(t+c) = D(t+c).\ SR\ (t+c) \qquad \ldots. (20)$$

Using expression for SR(t+c) in terms of pre-disturbance values, and substituting also for D(t+c) we have:

$$G(t) = G.\ \exp\ [La(t)/2] \qquad \ldots.(21)$$

$$G(t+1) = G.\ \exp\ [La(t+1)/2 - L(1-L)\ a(t)/2] \qquad \ldots.(22)$$

Similarly, we have, after simplification:

$$G(t+c) = G.\ \exp\ [La(t+c)/2 - L(1-L)\ a(t+c-1)/2 \ldots. - L^c\ (1-L)a(t)/2] \qquad \ldots.(23)$$

Empirical Results for India

For reasons of space, a detailed discussion of the choice of parameter values and of the assumptions underlying the projections is not given here. The following information based on results of 1951, 1961 and 1971 Population Censuses of India is, however, important in making judgments about these assumed values.

(a) The percent growth rates of India's population during 1941-50, 1951-60 and 1961-70 decades were 13.4%, 21.64% and 24.57%. Between 1951 and 1971, India's population increased by 51.1 percent.

(b) If it is assumed that no significant mortality improvements occurred in India in the few decades prior to 1951 so that stable population condition could be taken as a reasonably rough approximation, the long-run stable population one period (20 years) growth factor G may be assumed at $(1.134)^2 = 1.286$. This means that on average, in the absence of significant mortality improvements that actually occurred in India during the fifties and to a much lesser extent during the sixties, India's population between 1951 and 1971 would have increased by 28.6%. The difference of 22.5% may be attributed to mortality and fertility shifts that may have taken place during the 20-year period 1951-71.

(c) Analyses of India's census data suggests that there is little evidence of significant fertility declines occurring during 1951-70 in response to very significant mortality declines underway in that period. This means that the value of lag parameter L in relation (3) is very close to unity.

(d) Based on India's Official Life Tables, the survival rates from birth to age 20 are as follows:

Period	Male	Female
1941-50	.58	.57
1951-60	.72	.71
1961-70	.77	.75

Thus, between 1946 and 1956 (mid-points of the decades), the female's 20-year survival rate increased by 24.56 percent; the percentage for period between 1956 and 1966 was only 5.92 percent. For the 20-year period 1946 to 1966, the 20-year female survival rate increased by 31.93 percent. Evidence is very clear that mortality declines which were very significant during the fifties had considerably slowed down during the sixties. Mortality gains reflected in the above survival rate were of the order of 2.2 percent per year in fifties, but only of 0.6 percent per year in the sixties.

(e) Life expectancy at birth for females was 35 years based on 1941-50 Life Table, 40.0 on 1951-60 Life Table, and 45.6 years on 1961-70 Life Table. Thus, over the 20-years between 1951 and 1971 Censuses, female life expectancy at birth increased by over 10 years, or by nearly 30 percent.

The following assumptions have been made in making population projections:

(i) Calculations have been made for females only. It is assumed that similar orders of magnitude will emerge for males and total population. 50% of children born are assumed female.

(ii) Mortality disturbance is assumed to start at time t that is 1951. It is assumed that $a(t) = .30$; $a(t+1) = .10$, $a(t+2) = .05$ and $a(t+3) = .05$. This means that the forces of mortality between ages 0 and 20 declined on average by amount .30 during 1951-70, will decline by amount .10 during 1971-90 and by amount .05 during 1991-2010 and 2011-2030. In terms of life expectancy, these assumptions are equivalent to assuming that female life expectancy at birth will be 50.0 years in 1980, 55.0 years in 2000, and 57.5 in 2020. Future mortality gains are assumed to be smaller since existing cheap sources of mortality declines are assumed to have been, by and large, almost entirely used up, and further gains are likely to depend on improvements in diet, nutrition, etc., that is, factors which depend on gains in per capita income.

(iii) For making population projections Model Life Tables West-Females for Life Expectancy at Birth equal to 50 years, 55 years and 57.5 years given in Coale and Demeny have been used.

The Summary Table below gives the main results. Projections are made for 2 extreme values of L, the Lag Parameter viz. L=0 representing no lag or full fertility response and L=1 representing no fertility response.

Summary Table

Projections for Female Population for India for Years 1991 and 2010 and Estimation of Important Population Parameters.
(Initial 1971 census figure assumed at 1000)

	Actual 1971	Projected 1991		Projected 2011	
		L=0	L=1.0	L=0	L=1.0
A. Female Population by age-group					
0 (0-20)	506	582	611	718	783
1 (20-40)	286	456	456	537	565
2 (40-60)	148	238	238	393	393
3 (60-80)	60	86	86	147	147
4 Total	1000	1362	1391	1795	1888
B. Proportion age group 0.	.506	.427	.439	.400	.415
C. Proportion age group 1.	.286	.335	.328	.299	.299
D. Proportion "labor force" (=age groups 1 and 2)	.434	.510	.499	.518	.507
E. Dependency Ratio.	.566	.490	.501	.482	.493
F. (i) Estimated total fertility (female children only)	2.53 (1951)	1.78	1.87	1.65	1.70
(ii) Fertility as proportion of 1951 fertility	1	.705	.741	.653	.670
G. Projected Population given actual 1971 total population (millions)	458	746	762	984	1034
H. Rate of Population Increase (%)					
(i) over period	-	36.2%	39.1%	31.8%	35.7%
(ii) annualized rate	-	1.54%	1.65%	1.38%	1.53%
I. Projected Population for year 2001	-	-	-	856	887

Important results are:

(i) If the hypothesis that households' fertility behaviour takes no account of mortality improvements during the current period is true, then India's population is expected to be 762 million by 1991, 887 million by 2001, and 1034 million by 2011.

(ii) India's population increased by 51.6 percent during 1951-71; it is projected to grow by 36.2% under assumption of full fertility response L=0, and by 39.1% under assumption of no fertility response to current mortality improvements during the period 1971-91. The rates of growth for the period 1991-2011 are 31.8% (L=0) and 35.7% (L=1).

(iii) India's child population (age group 0-20) which formed 50.6% in 1971 is projected to fall to 42.7 in 1991 and 40.0% in 2011 under full-fertility response assumption and to 43.9% in 1991 and 41.5% in 2011 under no fertility response assumption.

(iv) Total fertility is expected to fall to 70.5% of its 1951 level by 1991 in full fertility response case (L=0) and to 74.1% in no fertility response case (L=1). These work out to nearly 30 percent and 25 percent declines in total fertility and represent very significant reductions in fertility.

(v) The proportion of female population in child-bearing ages is expected to increase from 28.6% in 1971 to 33.5% if L=0 and 32.8% if L=1 in 1991. After that in the next 20 years 1991-2011 it is expected to fall to about 30 percent on both assumptions.

(vi) The proportion of population in labor force age groups 1 and 2 is expected to rise from 43.4% in 1971 to 51.0% in 1991 and 51.8% in 2010 if full fertility response assumption holds, L=0 case; and to 49.9% in 1991 and 50.7% in 2011 if no fertility response assumption holds, L=1 case.

(vii) The dependency ratio is expected to decline over time on both sets of assumptions.

References & Bibliography

Coale, A. J.: The Growth and Structure of Human Populations, Princeton University Press, Princeton, N. J., 1972.

Coale, A. J. and Demeny, P.: Regional Model Life Tables and Stable Populations, Princeton University Press, Princeton, 1966.

Hoover, P. F. and Longley-Cook, L. H.: Life and Other Contingencies, University Press, Cambridge, England, 1953. Published for the Institute of Actuaries and the Faculty of Actuaries.

Milbank Memorial Fund: "The Interrelations of Demographic, Economic and Social Problems in Selected Underdeveloped Areas." New York, 1954.

Registrar General of India: Official Population Census Reports, 1951, 1961 and 1971.

Registrar General of India: Official Life Tables for India, 1941-50, 1951-60 and 1961-70.

World Bank: Population Planning-Sector Working Paper, World Bank, March 1972.

APPENDIX 1.1

Value of F(t+c), given a(t)=.30; a(t+1)=.10; a(t+2)=.05 and a(t+3)=.05

F(t+c)	L=0	L=.1	L=.25	L=.5	L=.75	L=1
F(t)	.30	.30	.30	.30	.30	.30
F(t+1)	.10	.130	.175	.250	.325	.40
F(t+2)	.05	.063	.094	.175	.294	.45
F(t+3)	.05	.056	.073	.138	.270	.50

Note: $F(t)=a(t)$

$F(t+1)=La(t)+a(t+1)$

$F(t+c)=La(t+c-1)+a(t+c) \quad c=1,\ 2,\ 3.$

APPENDIX 1.2

Value of J(t+c) where J(t+c) is defined by D(t+c)/D=J(t+c). [Ratio of Total Fertility in any period (t+c) to Total Fertility in the Base Period before disturbance].

J(t+c)	L=0	L=.1	L=.25	L=.5	L=.75	L=1
J(t)	.861	.874	.894	.928	.963	1.00
J(t+1)	.705	.699	.644	.696	.711	.741
J(t+2)	.653	.651	.647	.641	.646	.670
J(t+3)	.622	.622	.621	.616	.617	.638

Note: $J(t+c)=\exp\left[-\sum_{0}^{c-1} a(t+c)\right].$

$\exp\left[-(1-L)F(t+c)/2\right];$

$F(t)=a(t);$

$F(t+c)=L\ a(t+c-1)+a(t+c)$

APPENDIX 1.3

Projection of children born during time periods 1 (1971-1990) and 2 (1991-2010) and surviving to the end of each period.

	Time Period 1 (1971-1990)		Time Period 2 (1991-2010)	
	L=0	L=1.0	L=0	L=1.0
(1)	(2)	(3)	(4)	(5)
1. Female-periods	408	408	514	544
2. Pre-disturbance total fertility	2.53	2.53	2.53	2.53
3. J(t+c)	.705	.741	.653	.670
4. Estimated fertility	1.78	1.87	1.65	1.70
5. Total female births	726	763	848	925
6. Survival Rate	.801	.801	.847	.847
7. Female children living t=1	582	611	718	783

APPENDIX 1.4

Projected 1990 Female Population for India

X	1971	Survival Rate	Survivors 1991
0-1	506	.902	456
1-2	286	.832	238
2-3	148	.579	86

APPENDIX 1.5

Projected 2010 Female Population for India

X	1991	Survival Rate	Survivors 2010
0-1 L=0	582	.924	537
L=1	611	.924	565
1-2	456	.861	393
2-3	238	.616	147
3-4	86	.0	0

APPENDIX 1.6

Values of exp [K(t+c)] for different values of C and L

C	L=0	L=.1	L=.25	L=.5	L=.75	L=1.0
0	1	1.015	1.038	1.077	1.119	1.161
1	1	.991	.984	.987	.995	1.051
2	1	.997	.990	.981	.982	1.025
3	1	.999	.997	.990	.985	1.025

* This research is a joint effort of this author and Dr. George S. Tolley of the University of Chicago. For the drafting of this report, however, the present author bears complete responsibility for all deficiencies in the paper.

THE USE OF UNIQUE STATISTICAL WEIGHTS FOR ESTIMATING VARIANCES WITH THE BALANCED HALF-SAMPLE TECHNIQUE

Stanley Lemeshow, University of Massachusetts

This paper considers the use of the balanced half-sample method for estimating the variance of an estimated mean when each individual from a selected sample has been assigned a unique statistical weight. Such individual weighting is an attempt to bring the contribution of sample individuals falling into specified demographic classes into closer alignment with known population figures. Previous work by Kish and Frankel (1968, 1970) and Simmons and Baird (1968) concerned the use of such weights in complex sample surveys such as the Health Examination Survey of the National Center for Health Statistics.

1. Estimating the variance of $\hat{\mu}$ with unique statistical weights

For simplicity a population composed of L strata is considered. Let N be the total number of individuals in this population and let N_i be the number of individuals in the i^{th} stratum, $i = 1, \cdots, L$. Suppose each individual in this population can be classified into one of D demographic groups and let M_j be the number of individuals in the j^{th} demographic group, $j = 1, \cdots, D$. Let N_{ij} be the number of individuals in the i^{th} stratum belonging to the j^{th} demographic group. The population totals, N_i, M_j, $i = 1, \cdots, D$ are know.

Population

Strata \ Demo Group	1	2	···	D	
1	N_{11}	N_{12}	···	N_{1D}	N_1
2	N_{21}	N_{22}	···	N_{2D}	N_2
⋮	⋮	⋮		⋮	⋮
L	N_{L1}	N_{L2}	···	N_{LD}	N_L
	M_1	M_2	···	M_D	N

Clearly, $\sum_{i=1}^{L} N_i = \sum_{j=1}^{D} M_j = \sum_{i=1}^{L} \sum_{j=1}^{D} N_{ij} = N,$

$\sum_{j=1}^{D} N_{ij} = N_i, \; i = 1, \cdots, L, \; \sum_{i=1}^{L} N_{ij} = M_j,$

$j = 1, \cdots, D.$

Now suppose a random sample of n_i individuals is drawn with replacement from the i^{th} stratum, $i=1, \cdots, L$. Let $n = \sum_{i=1}^{L} n_i$ be the total number of individuals selected from the population. Further, suppose that upon selection each individual is placed, at random, into one of two groups and the demographic class to which the individual belongs is noted. Let n_{ijk} be the number of individuals in the j^{th} demographic class within the i^{th} stratum who were placed at random into the k^{th} group, $k = 1, 2$.

Let $m_{j1} = \sum_{i=1}^{L} n_{ij1}, \; m_{j2} = \sum_{i=1}^{L} n_{ij2}, \; m_j = \sum_{i=1}^{L} n_{ij},$

$j = 1, 2, \cdots, D, \; n_{ij} = \sum_{k=1}^{2} n_{ijk}, \; i = 1, \cdots, L,$

$j = 1, \cdots, D.$

Sample

Group 1

Strata \ Demo	1	2	···	D	
1	n_{111}	n_{121}	···	n_{1D1}	$\frac{n_1}{2}$
2	n_{211}	n_{221}	···	n_{2D1}	$\frac{n_2}{2}$
⋮	⋮	⋮		⋮	⋮
L	n_{L11}	n_{L21}	···	n_{LD1}	$\frac{n_L}{2}$
	m_{11}	m_{21}	···	m_{D1}	$\frac{n}{2}$

Group 2

Strata \ Demo	1	2	···	D	
1	n_{112}	n_{122}	···	n_{1D2}	$\frac{n_1}{2}$
2	n_{212}	n_{222}	···	n_{2D2}	$\frac{n_2}{2}$
⋮	⋮	⋮		⋮	⋮
L	n_{L12}	n_{L22}	···	n_{LD2}	$\frac{n_L}{2}$
	m_{12}	m_{22}	···	m_{D2}	$\frac{n}{2}$

Total Sample

Strata \ Demo	1	2	···	D	
1	n_{11}	n_{12}	···	n_{1D}	n_1
2	n_{21}	n_{22}	···	n_{2D}	n_2
⋮	⋮	⋮		⋮	⋮
L	n_{L1}	n_{L2}	···	n_{LD}	n_L
	m_1	m_2	···	m_D	n

Note that $\sum_{i=1}^{L} n_i = \sum_{j=1}^{D} m_j = \sum_{i=1}^{L}\sum_{j=1}^{D} n_{ij} = n$,

$\sum_{j=1}^{D} n_{ij} = n_i$, $i = 1, \ldots, L$. We also note that since the division into groups is done at random, m_{jk} is usually not equal to $m_j/2$, $k = 1, 2$.

Let $w_{ijk\ell}$ be a statistical weight assigned to the ℓ^{th} individual in the k^{th} group of the j^{th} demographic class in the i^{th} stratum. The $w_{ijk\ell}$ are selected so that following properties hold:

$$\text{(i)} \quad \sum_{i=1}^{L}\sum_{j=1}^{D}\sum_{k=1}^{2}\sum_{\ell=1}^{n_{ijk}} w_{ijk\ell} = N \tag{1}$$

$$\text{(ii)} \quad \sum_{j=1}^{D}\sum_{k=1}^{2}\sum_{\ell=1}^{n_{ijk}} w_{ijk\ell} = N_i\ , \ i=1, \ldots, L \tag{2}$$

$$\text{(iii)} \quad \sum_{i=1}^{L}\sum_{k=1}^{2}\sum_{\ell=1}^{n_{ijk}} w_{ijk\ell} = M_j\ , \ j=1, \ldots, D \tag{3}$$

Let us assume that the demographic breakdown in each stratum is the same. That is,

$$N_{ij} = \frac{N_i M_j}{N} .$$

Then, by taking

$$W_{ijk\ell} = \frac{N_i M_j}{N(n_{ij})} , \tag{4}$$

it is easily verified that (1), (2), and (3) hold. Hence, constructing statistical weights in this manner satisfies the desired properties.

Let $x_{ijk\ell}$ be a measurement taken on the ℓ^{th} individual in the k^{th} random group of the j^{th} demographic class in the i^{th} stratum and let μ_{ij} be the mean of the individuals in the j^{th} demographic class in the i^{th} stratum. Let $\mu = \frac{1}{N}\sum_{i=1}^{L}\sum_{j=1}^{D} N_{ij}\,\mu_{ij}$ be the mean in the entire population. Then, μ may be estimated using the statistical weights as follows:

$$\hat{\mu} = \frac{\sum_{i=1}^{L}\sum_{j=1}^{D}\sum_{k=1}^{2}\sum_{\ell=1}^{n_{ijk}} w_{ijk\ell}\ x_{ijk\ell}}{\sum_{i=1}^{L}\sum_{j=1}^{D}\sum_{k=1}^{2}\sum_{\ell=1}^{n_{ijk}} w_{ijk\ell}} . \tag{5}$$

We observe that since the n_{ij} are random variables, $\hat{\mu}$ is not a linear estimate but is, instead, a combined ratio estimate.

For this estimate, the denominator of $\hat{\mu}$ equals N. That is

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{L}\sum_{j=1}^{D}\sum_{k=1}^{2}\sum_{\ell=1}^{n_{ijk}} w_{ijk\ell}\ x_{ijk\ell}$$

Even here, $\hat{\mu}$ is not a linear estimate since the n_{ij} occur in each of the weights. In cases to be discussed later, where the estimation of the mean of a smaller subgroup in the population is considered, the denominator may not be constant.

Now, we note that

$$E\,(\hat{\mu}\,|\,n_{ij} \neq 0) = \mu.$$

Clearly, unbiased estimates of the mean of any collection of the demographic classes of individuals ("domain of interest") can be obtained, subject to the condition that the total of the n_{ijk} in this collection $\neq 0$, by making use of the indicator random variable $\gamma_{ijk\ell}$ as follows:

Let

$$\gamma_{ijk\ell} = \begin{cases} 1 & \text{if individual belongs to the domain of interest} \\ 0 & \text{if individual belongs to some other domain.} \end{cases}$$

Then, the combined ratio estimate

$$\hat{\mu}_D = \frac{\sum_{i=1}^{L}\sum_{j=1}^{D}\sum_{k=1}^{2}\sum_{\ell=1}^{n_{ijk}} \gamma_{ijk\ell}\ w_{ijk\ell}\ x_{ijk\ell}}{\sum_{i=1}^{L}\sum_{j=1}^{D}\sum_{k=1}^{2}\sum_{\ell=1}^{n_{ijk}} \gamma_{ijk\ell}\ w_{ijk\ell}} \tag{6}$$

can be shown to be an unbiased estimate of the mean (to the extent that the $n_{ij} \neq 0$) of the domain of interest.

Once statistical weights of this nature have been established and the corresponding estimate defined, the problem of estimating the variance remains.

Two ways of adjusting individual weights for the purpose of estimating the variance of $\hat{\mu}$ are considered. The first, which we will refer to as the "single-set method", uses the entire sample, as described above, to construct the individual weights. These weights are then used in all calculations. That is, the same individual weights are used for estimates based on only part of the sample individuals (e.g., a half-sample) as are used for estimates based on the entire sample. Clearly, since the weights were designed to bring the entire sample into closer alignment with the population, the resulting estimates based on a fraction of the sample will not necessarily reflect the population parameter. A second method, which we will refer to as the "multiple-set method", considers the subgroup of the sample from which a population estimate is to be computed and assigns weights to the individuals in the subgroup. There is then a closer alignment between specified demographic classes in the subgroup and the population.

a. The Single-Set Method

The single-set method always assigns the individuals the statistical weight

$$w_{ijk\ell} = \frac{N_i M_j}{N(n_{ij})} \quad .$$

Hence, for the balanced half-sample method, letting δ_{ij} be an element from the appropriate Plackett-Burman matrix, we compute m half-sample estimates $\hat{\mu}_{(p)}$, p = 1, $\cdots$, m, where

$$\hat{\mu}_{(p)} = \frac{\sum_{i=1}^{L}\{\delta_{pi}\sum_{j=1}^{D}\sum_{\ell=1}^{n_{ij1}} w_{ij1\ell}\, x_{ij1\ell} + (1-\delta_{pi})\sum_{j=1}^{D}\sum_{\ell=1}^{n_{ij2}} w_{ij2\ell}\, x_{ij2\ell}\}}{\sum_{i=1}^{L}\{\delta_{pi}\sum_{j=1}^{D}\sum_{\ell=1}^{n_{ij1}} w_{ij1\ell} + (1-\delta_{pi})\sum_{j=1}^{D}\sum_{\ell=1}^{n_{ij2}} w_{ij2\ell}\}} \qquad (7)$$

and

$$\delta_{pi} = \begin{cases} +1 \text{ if the first group is to be used from the } i^{th} \text{ stratum} \\ 0 \text{ if the second group is to be used from the } i^{th} \text{ stratum.} \end{cases}$$

Once the $\hat{\mu}_{(p)}$ are calculated, the variance of $\hat{\mu}$ may be estimated by using

$$\hat{V}_B(\hat{\mu}) = \frac{1}{m}\sum_{i=1}^{m}(\hat{\mu}_{(i)} - \hat{\mu})^2 \qquad (8)$$

where $\hat{\mu}$ is defined as in (5).

b. The Multiple Set Method

Consider a second set of individual statistical weights

$$V_{ijk\ell} = \frac{N_i M_j}{N(n_{ijk})} \quad .$$

Then, for the balanced half-sample method we compute m half sample estimates $\hat{\mu}_{(p)}$, p = 1, $\cdots$, m, where

$$\hat{\mu}_{(p)} = \frac{\sum_{i=1}^{L}\{\delta_{pi}\sum_{j=1}^{D}\sum_{\ell=1}^{n_{ij1}} v_{ij1\ell}\, x_{ij1\ell} + (1-\delta_{pi})\sum_{j=1}^{D}\sum_{\ell=1}^{n_{ij2}} v_{ij2\ell}\, x_{ij2\ell}\}}{\sum_{i=1}^{L}\{\delta_{pi}\sum_{j=1}^{D}\sum_{\ell=1}^{n_{ij1}} v_{ij1\ell} + (1-\delta_{pi})\sum_{j=1}^{D}\sum_{\ell=1}^{n_{ij2}} v_{ij2\ell}\}} \qquad (9)$$

We note that E $(\hat{\mu}_{(p)} | n_{ijk} \neq 0) = \mu$. Hence, $\hat{\mu}(p)$, p = 1, $\cdots$, m, is an unbiased estimate of μ provided all $n_{ijk} > 0$.

The balanced half-sample estimates of the variance of $\hat{\mu}$ are defined as in (8) with the $\hat{\mu}_{(i)}$, i = 1, $\cdots$, m, defined as in (9).

It is often of interest to estimate the variance of $\hat{\mu}$ for certain domains of interest in the population. That is, we might want to estimate the variance of the estimated mean, $\hat{\mu}_D$, of one of the demographic classes into which the population is divided. As before, an indicator variable, $\gamma_{ijk\ell}$, which takes on the value 1 if the individual belongs to the demographic class for which an estimate is desired and 0 otherwise, is used. Then, the p^{th} half-sample estimate of $\hat{\mu}_D$ is

$$\hat{\mu}_{D(p)} = \frac{\sum_{i=1}^{L}\{\delta_{pi}\sum_{j=1}^{D}\sum_{\ell=1}^{n_{ij1}} \gamma_{ij1\ell}\, w^*_{ij1\ell}\, x_{ij1\ell} + (1-\delta_{pi})\sum_{j=1}^{D}\sum_{\ell=1}^{n_{ij2}} \gamma_{ij2\ell}\, w^*_{ij2\ell}\, x_{ij2\ell}\}}{\sum_{i=1}^{L}\{\delta_{pi}\sum_{j=1}^{D}\sum_{\ell=1}^{n_{ij1}} \delta_{ij1\ell}\, w^*_{ij1\ell} + (1-\delta_{pi})\sum_{j=1}^{D}\sum_{\ell=1}^{n_{ij2}} \gamma_{ij2\ell}\, w^*_{ij2\ell}\}} \, ,$$

where

$$w^*_{ijk\ell} = \begin{cases} w_{ijk\ell} \text{ if the single-set method is used} \\ v_{ijk\ell} \text{ if the multiple-set method is used.} \end{cases}$$

Then, the balanced half-sample estimate of the variance of $\hat{\mu}_D$ is

$$\hat{V}_B(\hat{\mu}_D) = \frac{1}{m}\sum_{i=1}^{m}(\hat{\mu}_{D(i)} - \hat{\mu}_D)^2, \text{ where } \hat{\mu}_D \text{ is de-}$$

fined as in (6).

2. The Sampling Experiment

Sampling experiments were performed using both the single-set or multiple-set weights in conjunction with the balanced half-sample method to estimate the variance of $\hat{\mu}$.

In the sampling experiments, n_i observations were randomly selected from the i^{th} stratum, i = 1, $\cdots$, L. On the j^{th} draw from the i^{th} stratum we observe the random pair (x_{ij}, u_{ij}) where

$$X_{ij} \sim N(\mu^{(i)}, \sigma^{(i)}_{xx})$$

and

$$U_{ij} \sim U(0,1) \quad .$$

The population is categorized into D = 2 demographic classes, and the probability of falling into either of the two is specified. That is, let

$$P_1 = \frac{M_1}{N}$$

be the proportion of the population belonging to the first demographic class. This proportion is a specified constant for all strata.

Then, if

$$u_{ij} \begin{cases} < p_1, \text{ the individual is a member of the first demographic class} \\ > p_1, \text{ the individual is a member of the second demographic class.} \end{cases}$$

We label the first demographic class "domain 1" and the second "domain 2". Hence, for each of the n observations drawn from this population we have a value for x as well as the demographic class to which the observation belongs.

The means and variances of x are specified for each of the L strata. Also specific are N, $p_1=M_1/N$, N_i and n_i, i = 1, ..., L. We will be interested in obtaining estimates of the variance of the estimated mean of the entire population, of doman 1 alone and of domain 2 alone.

The following "situations" were considered:

Situation (i-1) L = 3, $\mu_x^{(1)}=5$, $\mu_x^{(2)}=10$, $\mu_x^{(3)}=15$, $\sigma_{xx}^{(i)}=1$, i=1,2,3, $p_1=.8$, $N_1=1000$, $N_2=2000$, $N_3=7000$, N=10000, $n_i=100$, i=1,2,3.

Situation (i-2): Situation (i-1) with $n_i=10$, i=1,2,3.

Situation (i-3): Situation (i-1) with $n_i=20$, i=1,2,3.

Situation (i-4): Situation (i-1) with $n_i=50$, i=1,2,3.

Situation (i-5): Situation (i-1) with $p_1=.9$

Situation (i-6): " " " $p_1=.7$

Situation (i-7): " " " $p_1=.6$

Situation (i-8): " " " $p_1=.5$

Situation (i-9): " " " $N_1=3333$, $N_2=3333$, $N_3=3334$.

Situation (i-10): Situation (i-1) with $N_1=7000$, $N_2=2000$, $N_3=1000$

Situation (i-11): Situation (i-1) with $\mu^{(i)}=10$, i=1,2,3

Situation (i-12): Situation (i-1) with $\mu^{(1)}=5$, $\mu^{(2)}=6$, $\mu^{(3)}=7$

Situation (ii) L=15, $\mu_x^{(1)}=5$, $\mu_x^{(i)}=\mu_x^{(1)} + 5(i-1)$, i=2,...,15, $\sigma_{xx}^{(i)}=1$, i=1,...,15, $p_1=.8$, $N_1=300$, $N_i=N_1 + 100(i-1)$, i=2,...,15, N=15000, $n_i=n=100$, i=1,...,L.

For situations (i-1) - (i-11), 1000 repetitions of the experiment were taken and for situation (ii), 200 repetitions were taken.

The variance of $\hat{\mu}$ was estimated with $V_B(\hat{\mu})$.

Table 1 presents the estimated absolute relative bias[1] and variance[2], of the estimated variance of $\hat{\mu}$ which resulted from each of the sampling experiments performed for the thirteen situations under consideration. Results are presented using both the single-set and multiple-set methods for the more prevalent domain 1 individuals, the less prevalent domain 2 individuals, and for all sample individuals. Considered in this table, as described earlier, is an L strata situation where we vary the number of observations selected in each stratum, the probability of selecting an individual from domain 1, the population sizes of the strata, the means of the strata, or the number of strata.

Situations (i-1) through (i-10) correspond to having $\mu^{(1)} = 5$, $\mu^{(2)} = 10$, $\mu^{(3)} = 15$. That is, these strata are rather spread out - particularly since $\sigma_{xx}^{(i)} = 1$, i = 1,2,3. As is seen in Table 1, when we estimate the variance of μ for specific domains of interest using the single-set method with the balanced half-sample technique, the resulting variance estimates show extremely high bias and variability when compared to the same estimates using the multiple-set method. In these situations, use of $v_{ijk\ell}$ is advantageous since they depend on the n_{ijk} instead of the n_{ij}. The latter were used to construct the $w_{ijk\ell}$. When calculations are performed on the entire sample, the number of individuals in each stratum is equally divided among the two established groups and there no longer appears to be an advantage to using the multiple-set method.

We note that when groups are established within each stratum, n_{ijk}, the number of individuals belonging to a specific demographic class within a group, is a random variable. That is, n_{ijk} is not necessarily $n_{ij}/2$. In fact, when a small number of observations are selected from each stratum it is possible, as is indicated by the "*", that $n_{ijk} = 0$ in some cases.

In situations (i-11) and (i-12) where the means in each stratum are either equal (all $\mu^{(i)}= 5$) or close, ($\mu^{(i)} = 5,6,7$, $\sigma_{xx}^{(i)} = 1$), there appears to be no advantage to using the multiple-set method.

Calculations with situation (ii) indicate that, with 15 strata with widely differing means, it is certainly advisable to use the multiple-set method. For domain 1, the absolute relative bias was reduced from 60.157 using the single-set method to .054 using the multiple-set method. For domain 2, the absolute relative bias was reduced from 221.663 using the single-set method to .045 using the multiple-set method. In fact, even for calculations involving all sample individuals, for these 15 highly spread strata, an absolute relative-bias of 3.038 using the single-set method was reduced to .112 using the multiple-set method. The variances of the variance estimates using the single-set method were much higher than the corresponding variances using the multiple-set method.

4. Conclusions

Our conclusion , therefore, is that the preferred procedure is to weight the individuals in each half-sample so that the half-sample estimates are representative of the population. We believe that an explanation for the similarity of the variance estimates produced by the two weighting methods found by Kish and Frankel (1968, 1970) and Simmons and Baird (1968) might be that:

(a) Calculations were not performed on domains of interest-particularly in which the n_{ijk} in each group were likely to be small, and

(b) The means of the strata are not likely to be very different for that particular set of data.

We believe that when variances are estimated for domains of interest - particularly when the strata means are not likely to be similar - it is advisable to reweight each half-sample. We note that, as was seen in formula (9), this does not involve the establishment of m+1 sets of weights (one for the total sample and one for each of the m half-samples). Instead, only two weights must be computed for each sample individual. The amount of extra work necessary in using this method, especially with the speed of modern computing facilities, is minimal.

By using the multiple-set weighting method, adequate variance estimates, in most of the situations considered in this sampling experiment, were obtained. Use of the single-set method led to an extremely high degree of bias and variability in certain of the situations considered. However, it should be recognized that these results only apply to the case of random sampling with replacement from L strata. For this reason, further research is necessary to determine the appropriateness of the single-set or multiple-set methods for more complex survey designs such as those used in the Health Examination Survey.

Table 1: Estimated values based on sampling experiments are presented for the absolute relative bias, $|\{\hat{E}[\hat{V}_B(\hat{\mu}_D)] - \tilde{V}(\hat{\mu}_D)\}/\tilde{V}(\hat{\mu}_D)|$, and variance, $\hat{V}(\hat{V}_B(\hat{\mu}_D))$ for domain 1, domain 2 and all sample individuals. Estimates are made using the single-set and multiple-set method of individual weighting in each of the thirteen situations described in the text.

ABSOLUTE RELATIVE BIAS

Situation	DOMAIN 1 Single -Set	DOMAIN 1 Multiple -Set	DOMAIN 2 Single -Set	DOMAIN 2 Multiple -Set	TOTAL SAMPLE Single -Set	TOTAL SAMPLE Multiple -Set
(i-1)	.949	.115	5.438	.209	.094	.124
(i-2)	1.355	.009	*	*	.116	.212
(i-3)	1.488	.136	*	*	.274	.397
(i-4)	1.069	.059	6.213	1.172	.041	.033
(i-5)	.474	.111	9.507	3.268	.063	.077
(i-6)	1.402	.113	4.192	.132	.092	.126
(i-7)	2.118	.110	3.619	.102	.079	.116
(i-8)	2.734	.047	2.988	.053	.087	.108
(i-9)	3.305	.080	14.772	.166	.099	.064
(i-10)	1.089	.069	5.624	.251	.025	.058
(i-11)	.117	.115	.125	.209	.149	.124
(i-12)	.068	.115	.303	.209	.147	.124
(ii)	60.157	.054	221.663	.045	3.038	.112

VARIANCE

Situation	DOMAIN 1 Single -Set	DOMAIN 1 Multiple -Set	DOMAIN 2 Single -Set	DOMAIN 2 Multiple -Set	TOTAL SAMPLE Single -Set	TOTAL SAMPLE Multiple -Set
(i-1)	.00026	.00007	.04763	.00199	.00005	.00004
(i-2)	.06322	.01098	*	*	.01098	.00791
(i-3)	.01032	.00243	*	*	.00246	.00257
(i-4)	.00112	.00032	.88485	.58082	.00022	.00024
(i-5)	.00012	.00006	2.49803	2.13784	.00005	.00005
(i-6)	.00053	.00009	.01190	.00076	.00005	.00004
(i-7)	.00118	.00013	.00472	.00041	.00005	.00004
(i-8)	.00204	.00023	.00218	.00021	.00005	.00005
(i-9)	.00032	.00001	.07618	.00028	.00001	†
(i-10)	.00025	.00006	.04748	†	.00004	.00004
(i-11)	.00007	.00007	.00144	†	.00004	.00004
(i-12)	.00008	.00007	.00190	†	.00004	.00004
(ii)	.00104	†	.43840	†	.00002	†

* some $n_{ijk} = 0$

† value < .00001

ENDNOTES

[1]The absolute relative bias of the estimated variance of $\hat{\mu}$ can be estimated from a sampling experiment as follows. For each repetition of the sampling experiment $\hat{\mu}$ is computed as if $\hat{V}_B(\hat{\mu})$, the estimated variance of $\hat{\mu}$ using the balanced half-sample technique. Then, let

(i) $E(\hat{\mu})$ = expected value of $\hat{\mu}$. This is estimated as $\tilde{E}(\hat{\mu}) = \frac{1}{r} \sum_{i=1}^{r} \hat{\mu}_i$, where r is the number of repetitions of the experiment and $\hat{\mu}_i$ is the estimate of μ for the i^{th} repetition.

(ii) $V(\hat{\mu})$ = variance of $\hat{\mu}$. This is estimated as $\tilde{V}(\hat{\mu}) = \frac{1}{r-1} \sum_{i=1}^{r} (\hat{\mu}_i - \tilde{E}(\hat{\mu}))^2$.

With r large, this value is regarded as the "target value" of the variance estimator $\hat{V}_B(\hat{\mu})$.

(iii) $E(\hat{V}_B(\hat{\mu}))$ = expected value of the half-sample estimation technique. This is estimated by

$$\tilde{E}[\hat{V}_B(\hat{\mu})] = \frac{1}{r} \sum_{i=1}^{r} \hat{V}_{B_i}(\hat{\mu}), \text{ where } \hat{V}_{B_i}(\hat{\mu})$$

is the half-sample estimate of the variance of $\hat{\mu}$ produced during the i^{th} repetition of the sampling experiment.

The absolute relative bias is defined as

$$\left| \frac{E[\hat{V}_B(\hat{\mu})] - V(\hat{\mu})}{V(\hat{\mu})} \right| .$$

This is estimated by

$$\left| \frac{\tilde{E}(\hat{V}_B(\hat{\mu}) - \tilde{V}(\hat{\mu})}{\tilde{V}(\hat{\mu})} \right| .$$

This relative measure places the bias into the context of the target value. It is the estimated bias as a proportion of the estimated target value.

[2]The variance of the balanced half-sample estimates of the variance of $\hat{\mu}$, $V[\hat{V}_B(\hat{\mu})]$ is estimated by

$$\tilde{V}[\hat{V}_B(\hat{\mu})] = \frac{1}{r-1} \sum_{i=1}^{r} \{\hat{V}_{B_i}(\hat{\mu}) - \tilde{E}[\hat{V}_B(\hat{\mu})]\}^2 .$$

ACKNOWLEDGMENTS

This research is based, in part, on the author's Ph.D. dissertation which was undertaken at UCLA. In this regard, computing assistance was obtained from the Health Sciences Computing Faculty (UCLA) under NIH Special Resources Grant RR-3. Other support was obtained at the University of North Carolina (Chapel Hill) through the U.S. Bureau of the Census under Joint Statistical Agreement JSA 76-93.

REFERENCES

Kish, Leslie and Frankel, Martin R. (1968). Balanced repeated replication for analytical statistics. Proceedings of the Social Statistics Section of the American Statistical Association: 2-10.

__________________ (1970). Balanced repeated replication for standard errors. Journal of the American Statistical Association, Vol. 65, No. 331: 1071-1094.

Simmons, Walt R. and Baird, James T., Jr. (1968). Pseudo-replication in the NCHS Health Examination Survey. Proceedings of the Social Statistics Section of the American Statistical Association: 19-30.

SOCIO-ECONOMIC DIFFERENTIALS IN MORTALITY IN MARYLAND 1959-61 and 1969-71.

Monroe Lerner and Richard N. Stutz

The Johns Hopkins University

The work necessary for the preparation of this paper was supported by grants (5 R01 HS 00110 and 5 T01 HS 00012) from the National Center for Health Services Research and Development, U.S. Department of Health, Education and Welfare.

I. Introduction

This report presents preliminary results of the third sub-study of a larger study of socio-economic differentials in mortality.1/ It presents mortality data for Maryland for 1959-61 and 1969-71 made available by the Maryland Center for Health Statistics.2/ Counties and State Economic Areas are the units of analysis.

The primary question to which both the larger study and this sub-study address themselves is, did socio-economic differentials in mortality, overall by age and cause of death, and in infant mortality change during the decade of the 1960s? As background: 1) The overall mortality rate declined significantly in Maryland as a whole from 1959-61 to 1969-71, continuing a trend of earlier years; 2) Some evidence exists to indicate that socio-economic differentials in mortality (as measured by methods used in the current investigation) were narrowing during the 1950s and probably earlier, and by 1960 these differentials were low, both in Maryland and in the U.S. as a whole; 3) Income in Maryland as a whole rose significantly between 1960 and 1970; and 4) During the decade social programs were in operation, in Maryland as elsewhere, to narrow existing socio-economic differentials in mortality as well as in other aspects of health status. The expectation of the investigators was that narrowing had probably continued in Maryland during the decade, with the differentials perhaps disappearing completely.

II. Methods

For the present study, Maryland's 24 counties (actually 23 plus Baltimore City as a separate unit) and its seven State Economic Areas were used as units of analysis. Median family income was used as an indicator of socio-economic status, and counties were grouped into twelve with highest and twelve with lowest median family incomes, ranked on the basis of the average for each county of its medians for 1959 and 1969. The upper-income twelve, with 88.6 percent of the state's population in 1969 and 90.4 in 1970, included the large population aggregates of Baltimore City (905,759 in 1970), Prince Georges (660,567), Baltimore County (621,077), Montgomery (522,809), and Anne Arundel (297,539).

State Economic Areas in Maryland, ranked in order of median family income, are: 1) Metropolitan Washington—Montgomery and Prince Georges; 2) South Central Maryland—Carroll and Howard; 3) Metropolitan Baltimore—Baltimore City, Anne Arundel, and Baltimore County; 4) North Central Maryland--Frederick, Harford, and Washington; 5) Southern Maryland—Calvert, Charles, and St. Marys; 6) Western Maryland—Allegany and Garrett; and 7) Eastern Shore—Caroline, Cecil, Dorchester, Kent, Queen Annes, Somerset, Talbot, Wicomico, and Worcester. When split by income (as above), the upper consist largely of counties in Metropolitan Washington, South Central Maryland, Metropolitan Baltimore, and North Central Maryland, while the lower are largely in Southern and Western Maryland and on the Eastern Shore.

Mortality rates were age adjusted by direct standardization (Shryock et al, 1971: 419-421), using the U.S. population of 1940 as standard. Deaths in 1959-61 were allocated to cause in accordance with the Seventh Revision of the International List of Diseases and Causes of Death, 1955, while 1969-71 deaths were allocated by the Eighth Revision, 1965, and adjusted for comparability to the Seventh.

III. Maryland's Population Characteristics

Maryland's population increased substantially between 1960 and 1970, from 3,100,689 to 3,922,399, or by 26.5 percent.3/ The majority of its growth was due to natural increase, but nearly one-half (12.4 percent) was due to net in-migration. Within the state itself, 20 of the 24 counties gained population during the decade; the exceptions were Baltimore City (net loss of 3.5 percent) and Allegany, Dorchester, and Somerset. (The latter three are classified in the present analysis as among the lower-income counties). However, the increases in population were far greater in the upper-income counties than in the lower, both percentage-wise and in absolute numbers, despite the large net loss in numbers in Baltimore City. Eight lower-income counties, including two Western Maryland and six Eastern Shore, experienced a net-outmigration of population, as did only Baltimore City and Cecil among the upper.

The state's already largely urban population in 1960 became more urban during the decade, rising from 72.7 to 76.6 percent urban. Large increases were registered in the percent urban among some upper-income counties, most notably Howard (because of the new city of Columbia), Harford, Anne Arundel, Prince Georges, Charles, and Cecil; in contrast, the percent urban declined for about one-half the lower-income counties. The upper-income counties included large population aggregations in the Baltimore and Washington Metropolitan Areas, e.g., Baltimore City (100 percent urban in 1970), Baltimore County (88.6), Prince Georges (92.3), and Montgomery (89.2), while four lower-income counties were wholly rural and contained no urban population whatever (Calvert, Queen Annes, Caroline, and Garrett).

Income rose in Maryland as a whole during the decade; thus median family income went from $6,309 in 1959 to $11,057 in 1969, an increase of 75.3 percent. By county, the percent increase was not quite as great in the upper-income counties as in the lower, so that the relative difference narrowed somewhat. Thus the weighted average for the upper-income counties rose from $6628 in 1959 to $11,702 in 1969 or 77 percent, in the lower from $4,341 to $7,983 or 84 percent. The excess of upper over lower dropped from 53 to 47 percent (even though in absolute numbers it increased from $2,287 to $3,719). The coefficient of variation of the income distribution for all counties dropped from 27.0 to 25.7 percent, i.e., the distribution converged slightly.

The state's educational attainment also increased: median school years completed among persons 25 and over rose from 10.4 to 12.1, percent with fewer than five years of schooling dropped from 7.7 to 4.5, and percent completing high school or beyond rose from 40.0 to 52.3. By county, the weighted average of median school years completed for the high-income counties in 1960 (10.4 years) exceeded the low (9.1) and although both groups of counties improved over the decade (to 11.7 and 10.9, respectively), this pattern persisted. However, both the absolute and relative excess narrowed; thus it was 13.9 percent in 1960 but only 7.5 in 1970, 1.3 years in 1960 but only 0.8 in 1970. The coefficient of variation declined from 13.5 to 9.2 percent; as with income, educational attainment converged to some degree.

Finally, the state's percent Negro rose from 16.7 to 17.8, while its percent "foreign stock" declined from 11.9 to 11.6. The proportion of its labor force engaged in manufacturing declined somewhat, while the proportion engaged in white-collar occupations increased.

IV. Mortality in Maryland as a Whole

As the top line of Table 1 shows, the age-adjusted death rate for Maryland as a whole decreased between 1959-61 and 1969-71, from 834.0 per 100,000 to 756.8, or by 9.3 percent. This continued a decline of many years; thus in 1940 Maryland's death rate was 1,218.4 per 100,000, by 1950 it had dropped to 952.4, and by 1960 it was 883.0.4/ (Grove and Hetzel, undated:664.)

By age (top tier of Table 2), mortality declined for all age-groups except 15-24 where there was actually a substantial increase (37.4 percent). The declines were sharpest at the younger ages—under 1 (32.3 percent) and 1-4 (30.0); however, they were substantial also (between 8.5 and 11.3 percent) at all other ages except 25-44, where there was only a small decrease (3.2 percent).

The top tier of Table 3 shows data for five major groups of leading causes of death: diseases of the heart, cerebrovascular diseases, malignant neoplasms, influenza and pneumonia, and accidents. When age-adjusted, they accounted for 73.1 percent of deaths due to all causes in 1959-61 and 70.6 percent in 1969-71. Mortality declined in the state as a whole for four of the five causes shown, especially (in relative terms) for influenza and pneumonia (22.4 percent); only for malignant neoplasms was there actually a slight increase (0.44 percent). In absolute terms, the most pronounced decline was easily for diseases of the heart, 51.3 deaths; this accounted for two-thirds of total decline.

V. Mortality in High and Low-Income Counties

Again in Table 1, mortality in both high and low-income counties decreased over the decade, but the decrease was larger for the high-income than for the low, i.e., 65.4 per 100,000 against 34.5, 8.1 percent against 4.1. In 1959-61 mortality in the low-income counties exceeded the high by 4.2 percent, but by 1969-71 rates for the two groups of counties diverged and the relative excess rose to 8.7 percent.5/ The coefficient of variation for the distribution as a whole increased (unlike its decrease for income and education) from 10.6 percent to 12.7 percent; thus the distribution became more dispersed, and a parallel increase in dispersion was evident for both mini-distributions (12 counties each) separately.

Although the income distribution by county converged over the decade while mortality diverged, their degree of inverse association changed little, with the coefficient of correlation at —.615 in the earlier period and —.611 in the later. However, because of the discrepant trends in dispersion of income and mortality, the slope of their regression line became steeper, changing from —.289 to —.320. Thus, an increase of $1,000 on the regression line for 1959 income was associated with a decrease in 1959-61 mortality of 28.9 deaths per 100,000 (age-adjusted rates), while a comparable increase in 1969 income was associated with a decrease of 32.0 deaths in 1969-71.

By age

As the bottom tier of Table 2 shows, during 1959-61 mortality in the low-income counties exceeded the high for all age-groups except 55-64 and 65-74, where the differences were small (1 and 5 percent, respectively), while in 1969-71 the low-income counties exceeded the high at all ages without exception. However the gap widened relatively only at under 1 and 45-54 and above; at 1-4 thru 25-44, the gap actually narrowed. The narrowing was particularly substantial at 15-24, where a sharp increase (46 percent) took place for the high-income counties, not matched by the low. The widening was most substantial at under 1, where a sharp decrease (34 percent) took place in high income again not matched by low, and at 65-74, where mortality in the high declined while it increased in the low.

By cause of death

The bottom tier of Table 3 shows that in 1959-61 the largest differential between the counties was for accidents, where low-income rates exceeded high by 64 percent, followed by cerebrovascular diseases (20 percent). For diseases of the heart, malignant neoplasms, and influenza and pneumonia mortality in the high-income counties exceeded the low.

By 1969-71, however, for each of the five causes alike, the high-income counties had gained relatively at the expense of the low, and mortality in the low-income counties now exceeded the high for all five causes. The relative gain was greatest for influenza and pneumonia, where the ratio of low to high increased by 31 percent; even for accidents, however, where it was already high, it rose to 69 percent.

However, the absolute gain for high-income counties over low was greatest for the two leading causes of death, diseases of the heart and malignant neoplasms. For diseases of the heart, mortality in the high-income counties dropped from 338.9 to 282.6, or by 56.3, while in the low the drop was only 8.9. Also, the direction of the differential was reversed, i.e., where in 1959-61 heart disease mortality was higher in the high-income counties by 27.7 deaths, in 1969-71 it was lower by 19.7. Considering the 24 counties as a whole, there was a slight negative correlation of mortality with income in the earlier period, —.176, but it increased substantially without changing direction by the later to —.350.

For malignant neoplasms, in 1959-61 mortality in the high-income counties was substantially above the low (22.2 deaths), but in 1969-71 it was actually less than in the low, by 1.1. This reversal took place because of the substantial increase in malignant neoplasms for the low-income counties not matched by an increase for the high. The change in direction of the excess mortality ratio was matched by a change in direction of the correlation coefficient; it went from +.200 to —.305.

VI. Mortality by State Economic Areas

Table 4 shows mortality by State Economic Areas ranked according to income. Much the same pattern is evident here, i.e., rates in the high-income Economic Areas were lower than those in the low in both periods. The most substantial improvements took place in North Central Maryland (Frederick, Harford, and Washington), South Central Maryland (Carroll and Howard), Western Maryland (Allegany and Garrett), and Metropolitan Washington (Montgomery and Prince Georges).

VII. Infant Mortality

The infant mortality rate for Maryland as a whole declined from 1959-61 to 1969-71, from 27.0 to 19.0 or by 29.6 percent (see Table 5). The decline was most evident for the high-income counties and the relative excess of low-income over high widened, from 13.1 to 26.6 percent. This trend parallels the widening in mortality at all ages combined and at 0-1. However, unlike the situation for overall mortality, the coefficient of variation for the entire distribution decreased from 20.0 to 18.7 percent, and a parallel decrease occurred for both mini-distributions (12 counties each) separately. Thus, although the differential between upper and lower-income counties widened substantially, the relative dispersion actually narrowed.

The negative correlation between income and infant mortality increased significantly, from —.483 to —.616. The slope of the regression line, however, decreased slightly, from —.164 to —.158. Thus an increase of $1,000 on the regression line for 1959 income was associated with a decrease in 1959-61 infant mortality of 1.64 deaths per 1,000 live births, while a comparable increase in 1969 income was associated with a decrease of 1.58 deaths in 1969-71.

Table 6 shows infant mortality rates by color. In both 1959-61 and 1969-71, and for whites and nonwhites alike, infant mortality rates in the low-income counties exceeded the high. The relative excess was greater for nonwhites than whites in both periods (12.9 percent for nonwhites and 0.4 for whites in 1959-61, and 7.48 and 11.7, respectively, in 1969-71). The decline for nonwhites was substantial in the high-income counties, from 42.7 to 25.4 or by 40.5 percent, but only minimal in the low, from 48.2 to 44.4 or by 7.9 percent. In 1969-71 the nonwhite rate exceeded white by 57 percent in the high-income counties and by 145 percent in the low.

VIII. Discussion

The rise in excess mortality among low-income counties over high reverses an earlier decline of at least a decade. From the data shown, it was concentrated at infancy, mid-life, and the older ages, and among deaths from heart disease and malignant neoplasms. At 15-24 the earlier narrowing continued.

To some unknown degree, the widening may reflect these factors:

1) By 1970, deaths due to communicable diseases with a significant inverse socio-economic gradient —tuberculosis and other infectious and parasitic diseases, influenza and pneumonia, and certain diseases of early infancy—had dropped to low levels (11.4 percent of all deaths in 1950, 10.1 in 1960, 6.6 in 1970). Their influence in the socio-economic differential for all causes was correspondingly lessened.

2) The benefits of the notable medical and surgical advances of the 1960s against heart disease and malignant neoplasms were likely to be more readily available in the metropolitan areas, even to the urban poor, than in the more isolated rural areas.

3) In general, migration has been from poorer to more affluent (and healthier) metropolitan areas; healthier migrants may have contributed measurably to lower mortality in the latter. In the present data, the correlation between net in-migration during the decade with percent decline in mortality was +.375.

The widening in mortality by no means indicates that the decade's social programs were unsuccessful. Perhaps without them it would have been greater; also, the programs may have most affected those aspects of health status unrelated to mortality.

Further analysis is indicated to determine whether the narrowing in mortality differentials at 15-24 was concentrated in either sex or race or in any cause of death, perhaps homicide and accidents. The increase in mortality at 15-24 was not restricted to Baltimore City, occurring among 10 upper and eight lower-income counties.

Finally, using counties as units of analysis imputes, perhaps unjustifiedly where populations are too heterogeneous, averages of aggregate characteristics to smaller units and even to each individual. Baltimore City has here been used as a single unit; in further analyses it will be disaggregated, where possible, into clusters of census tracts.

1/ The preliminary results of the first sub-study (using states as units of analysis) have been submitted for publication in Medical Care as M. Lerner and R.N. Stutz, "Narrowing the Gaps in Health Status, 1960 to 1970. I. Mortality", while preliminary results for the second, using census tracts in Baltimore City as units of analysis, has been published as: M. Lerner and R.N. Stutz, "Mortality Differentials Among Socio-Economic Strata in Baltimore, 1960 and 1973". Proceedings of the Social Statistics Section, American Statistical Association, 1975, pp. 517-522.

2/ Grateful acknowledgement for supplying these data is hereby extended to Dr. Jean Warthen, Director, Maryland Center for Health Statistics.

3/ Data on Maryland's population characteristics are derived from U.S. Bureau of the Census, County and City Data Book, 1967 and 1972, U.S. G.P.O., 1967 and 1973, respectively.

4/ The discrepancy between this 1960 rate (883.0) and the rate given earlier for 1959-61 (834.0) may result from the former being for a single year only, while the latter is for the average of three years; from the use of different age-groups in standardizing; or because these numbers are from different sources.

5/ Because of arbitrariness in using 12 counties each to represent upper and lower-income, the same comparison was also made for (separately) the top and bottom 11, 10, 9, 8, 7, and 6 counties. As with groups of 12, in each case the rates for higher and lower-income diverged between 1959-61 and 1969-71, with a resultant widening in excess, although its extent varied. In addition, the same comparisons were made using the rank ordering by income of the counties in 1960 alone as the basis for grouping counties. Again, separate comparisons were made of the top and bottom 12 through 6, and here also divergence appeared in all comparisons.

BIBLIOGRAPHY

Grove, Robert D. and Hetzel, Alice M.
1968 Vital Statistics Rates in the United States, 1949-1960. U.S.G.P.O.: Washington, D.C.

Shryock, Henry S., et al
1971 The Methods and Materials of Demography. U.S.G.P.O.: Washington, D.C.

TABLE 1.

Age-Adjusted Mortality Rates per 100,000 Population† by County
Twelve Highest and Twelve Lowest Median Family Income Counties††
Maryland, 1959-61 and 1969-71.

County	1959-61	1969-71	Percent Change
	(1)	(2)	(3)
Maryland	834.0	756.8	−9.26
High-Income Counties			
Montgomery	658.9	572.1	−13.17
Prince Georges	740.7	696.6	−5.95
Howard	800.5	671.9	−16.06
Baltimore	661.5	662.8	+ .20
Anne Arundel	771.8	735.9	−4.65
Harford	799.9	683.3	−14.58
Charles	943.5	849.0	−10.02
Carroll	663.2	611.2	−7.84
Frederick	818.9	727.1	−11.21
Baltimore City	984.6	959.7	−2.53
Cecil	764.7	779.8	+1.97
Washington	805.0+	703.3	−12.63
Weighted Average†††	810.0	744.6	−8.07++
Mean Average	784.4	721.1	−8.07++
Standard Deviation	98.1	99.8	—
Low-Income Counties			
Wicomico	790.0	804.1	+1.78
Calvert	798.2	793.1	− .64
Allegany	849.9	785.6	−7.57
St. Marys	800.0	811.3	+1.41
Talbot	864.6	781.3	−9.63
Queen Annes	867.9	737.2	−15.06
Kent	926.9	840.7	−9.30
Dorchester	845.1	931.8	+10.26
Caroline	894.0	826.5	−7.55
Worcester	833.3	883.7	+6.05
Garrett	795.2	658.1	−17.24
Somerset	980.7	891.2	−9.13
Weighted Average†††	843.9	809.4	−4.09++
Mean Average	853.8	812.1	−4.88++
Standard Deviation	55.8	69.6	—
Ratio: Low to High-Income Counties (Weighted Averages)	1.042	1.087	—

† Rates were adjusted by the direct standardization method (Shryock et.al., 1971:419-421), using the U.S Population of 1940 as the standard.

†† Counties are ranked in accordance with the mean of the median family incomes for 1959 and 1969.

††† Weighted in accordance with population with figures derived from the U.S. Census of Population for 1960 and 1970, respectively.

\+ Based on an average of 1959 and 1961 deaths.

++ Computed from the figures in Columns (1) and (2), rather than by averaging the percent change figures for individual counties.

Source of basic data:

Mortality: Unpublished data of the Maryland Center for Health Statistics.

Population Denominators: Various reports of the U.S. Bureau of the Census.

Median Family Income: U.S. Bureau of the Census, County and City Data Book, 1967 and 1972 editions, and Maryland Department of State Planning, "Maryland 1970 Social Indicators Series, Volume III: Income Characteristics, April, 1974.

TABLE 2.

Mortality Rates per 100,000 Population by Age
Twelve Highest and Twelve Lowest Median Family Income Counties††
Maryland, 1959-61 and 1969-71

Age Group	1959-61	1969-71	Percent Change
	(1)	(2)	(3)
Maryland			
All ages (adjusted)†	834.0	756.8	-9.26
By age-group:			
Under 1	2851.8	1930.4	-32.31
1-4	104.5	73.2	-29.95
5-14	44.2	39.2	-11.31
15-24	90.2	124.0	+37.47
25-44	240.4	232.8	-3.16
45-54	840.5	766.6	-8.79
55-64	1968.9	1786.9	-9.24
65-74	4277.7	3912.6	-8.53
75 and over	11100.0	9949.1	-10.37
High-Income Counties (Weighted Average)†††			
All ages (adjusted)†	810.0	744.6	-8.07
By age-group:			
Under 1	2837.4	1880.4	-33.73
1-4	102.6	73.8	-28.07
5-14	42.8	37.7	-11.92
15-24	83.2	121.5	+46.03
25-44	239.6	239.3	-.13
45-54	830.8	763.9	-8.05
55-64	1920.1	1739.0	-9.43
65-74	3982.9	3799.0	-4.62
75 and over	10973.9	9797.2	-10.72
Low-Income Counties (Weighted Average)†††			
All ages (adjusted)†	843.9	809.4	-4.09
By age group:			
Under 1	3199.2	2470.9	-22.77
1-4	131.4	75.9	-42.24
5-14	56.8	47.5	-16.37
15-24	138.9	150.1	+8.06
25-44	294.4	264.7	-10.09
45-54	846.5	852.5	+.71
55-64	1903.2	1823.6	-4.18
65-74	3798.9	4074.6	+7.26
75 and over	11266.5	10325.8	-8.35
Ratio: Low to High-Income Counties			
All ages (adjusted)†	1.042	1.087	+4.32
By age-group:			
Under 1	1.128	1.314	+16.49
1-4	1.281	1.028	-19.75
5-14	1.327	1.260	-5.05
15-24	1.669	1.235	-26.00
25-44	1.229	1.106	-10.01
45-54	1.019	1.116	+9.52
55-64	.991	1.049	+5.85
65-74	.954	1.073	+12.47
75 and over	1.027	1.054	+2.63

†Same as Table 1.
††Same as Table 1.
†††Same as Table 1.
Source of basic data: Same as Table 1.

TABLE 3.

Age Adjusted Mortality Rates per 100,000 Population† by Cause of Death
Twelve Highest and Twelve Lowest Median Family Income Counties††
Maryland, 1959-61 and 1969-71.

Cause of Death‡	1959-61	1969-71	Percent Change
	(1)	(2)	(3)
Maryland (Weighted Average)†††	834.0	756.8	-9.26
Diseases of the Heart	335.8	284.5	-15.28
Cerebrovascular Diseases	67.0	53.8	-19.70
Malignant Neoplasms	137.9	138.5	+.44
Influenza and Pneumonia	24.6	19.1	-22.36
Accidents	44.0	38.7	-12.05
Total, Five Causes	609.3	534.6	-12.26
Percent of all causes	73.06	70.64	—
High-Income Counties (Weighted Average)†††			
Diseases of the Heart	338.9	282.6	-16.61
Cerebrovascular Diseases	65.5	52.5	-19.85
Malignant Neoplasms	140.4	138.4	-1.42
Influenza and Pneumonia	24.9	18.8	-24.50
Accidents	41.0	36.3	-11.46
Total, Five Causes	610.7	528.6	-13.44
Percent of all causes	75.40	70.99	—
Low-Income Counties (Weighted Average)†††			
Diseases of the Heart	311.2	302.3	-2.86
Cerebrovascular Diseases	78.3	65.9	-15.84
Malignant Neoplasms	118.2	139.5	+18.02
Influenza and Pneumonia	22.5	22.3	-.89
Accidents	67.2	61.4	-8.63
Total, Five Causes	597.4	591.4	-1.00
Percent of all causes	70.79	73.07	—
Ratio: Low to High-Income Counties			
Diseases of the Heart	.918	1.070	+16.56
Cerebrovascular Diseases	1.195	1.255	+5.02
Malignant Neoplasms	.842	1.008	+19.71
Influenza and Pneumonia	.904	1.186	+31.19
Accidents	1.639	1.691	+3.17
Total, Five Causes	.978	1.119	+14.42

†Same as Table 1.
††Same as Table 1.
†††Same as Table 1.

‡Deaths in 1959-61 were allocated to cause in accordance with the Seventh Revision of the International Lists of Diseases and Causes of Death, 1955 while 1969-71 deaths were allocated in accordance with the Eighth Revision. The 1969-71 rates have been adjusted for comparability, by cause of death, to the Seventh Revision.

Causes of Death	International List Numbers 1959-61	International List Numbers 1969-71
Diseases of the Heart	400-402, 410-443	390-398, 402, 404, 410-429
Cerebrovascular Diseases	330-334	430-438
Malignant Neoplasms	140-205	140-209
Influenza and Pneumonia	480-493	470-474, 480-486
Accidents	E800-E962	E800-E949

Source of basic data: Same as Table 1.

TABLE 4.

Age-Adjusted Mortality Rates per 100,000 Population[†]
By State Economic Area[††]
Maryland, 1959-61 and 1969-71

State Economic Area	Median Family Income[‡]	1959-61 (Weighted Average)[†††]	1969-71 (Weighted Average)[†††]	Percent Change
	(1)	(2)	(3)	(4)
Metropolitan Washington	11,103	700.8	641.6	-8.45
South Central Maryland	8,589	719.0	639.9	-11.00
Metropolitan Baltimore	8,364	860.6	822.1	-4.47
North Central Maryland	7,492	807.5	702.4	-13.02
Southern Maryland	7,075	853.3	823.6	-3.48
Western Maryland	6,247	839.2	759.6	-9.49
Eastern Shore	6,218	840.2	824.7	-1.84

[†] Same as Table 1.

[††] State economic areas are relatively homogeneous subdivisions of states. They consist of groups of counties which have similar economic and social characteristics. The boundaries of these areas have been drawn in such a way that each state is subdivided into relatively few parts, with each part having certain significant characteristics which distinguish it from adjoining areas.

[†††] Same as Table 1.

[‡] State Economic Areas are ranked in accordance with the mean of the median family incomes for 1960 and 1970.

Source of basic data: Same as Table 1.

TABLE 5.

Infant Mortality Rates[†] by County
Twelve Highest and Twelve Lowest Median Family Income Counties[††]
Maryland, 1959-61 and 1969-71.

County	1959-61	1969-71	Percent Change
	(1)	(2)	(3)
Maryland	27.0	19.0	-29.63
High-Income Counties			
Montgomery	19.6	14.5	-26.02
Prince Georges	23.8	16.8	-29.41
Howard	32.5	19.1	-41.23
Baltimore	21.9	13.6	-37.90
Anne Arundel	24.6	16.6	-32.52
Harford	22.8	16.4	-28.07
Charles	39.4	22.7	-42.39
Carroll	26.2	16.6	-36.64
Frederick	23.4	16.2	-30.77
Baltimore City	33.3	26.0	-21.92
Cecil	27.2	16.9	-37.87
Washington	25.6	19.4	-24.22
Weighted Average[†††]	26.7	18.4	-31.09 [‡]
Mean Average	26.7	17.9	-32.96 [‡]
Standard Deviation	5.4	3.3	–
Low-Income Counties			
Wicomico	33.9	25.4	-25.07
Calvert	28.5	21.4	-24.91
Allegany	20.1	19.4	- 3.48
St. Marys	29.6	23.3	-21.28
Talbot	33.2	22.5	-32.23
Queen Annes	27.9	18.6	-33.33
Kent	39.7	20.5	-48.36
Dorchester	41.2	29.5	-28.40
Caroline	27.5	18.9	-31.27
Worcester	42.0	34.0	-19.05
Garrett	26.7	23.9	-10.49
Somerset	33.3	25.5	-23.42
Weighted Average[†††]	30.2	23.3	-22.85 [‡]
Mean Average	32.0	23.6	-26.25 [‡]
Standard Deviation	6.3	4.4	–
Ratio: Low to High-Income Counties (Weighted Averages)	1.131	1.266	–

[†] Deaths per 1,000 live births

[††] Same as Table 1.

[†††] Same as Table 1.

[‡] Computed from the figures in Columns (1) and (2), rather than by averaging the percent change figures for individual counties.

Source of basic data: Same as Table 1.

TABLE 6.

Infant Mortality Rates[†] by Color
Twelve Highest and Twelve Lowest Median Family Income Counties[††]
Maryland, 1959-61 and 1969-71.

Color	1959-61			1969-71		
	Low-Income Counties (Weighted Average)[†††]	High-Income Counties (Weighted Average)[†††]	Ratio: Low to High-Income Counties	Low-Income Counties (Weighted Average)[†††]	High-Income Counties (Weighted Average)[†††]	Ratio: Low to High-Income Counties
	(1)	(2)	(3)	(4)	(5)	(6)
White	22.5	22.4	1.004	18.1	16.2	1.117
Nonwhite	48.2	42.7	1.129	44.4	25.4	1.748

[†] Same as Table 5.
[††] Same as Table 1.
[†††] Same as Table 1.

Source of basic data: Same as Table 1.

SURVEY DESIGNS WHICH EMPLOY DOUBLE SAMPLING SCHEMES FOR ELIMINATING MEASUREMENT PROCESS BIAS

Judith T. Lessler, Research Triangle Institute

INTRODUCTION

The present paper describes some survey designs which employ double sampling schemes in order to reduce the measurement errors associated with estimates from a survey. Suppose one has available for use in a survey two measurement processes: a cheap-faulty measurement process and an expensive-accurate measurment process. If the net bias associated with the faulty measurement process is large, it may be advantageous to use a double sampling scheme for eliminating the measurement process bias.

The concepts employed are based upon the Census Bureau model for measurement errors [1,2] which may be briefly described as follows: There exists a population of N individuals. For each individual in the population, one wishes to measure their values on a set of p characteristics. For a particular measurement process, the measurement obtained for the i-th individual at the t-th trial of the survey is $\mathbf{Y}_{it}$. The subscript t indexes a series of repeatable trials of the measurement process (i.e. of the census or survey in question).

One can define the expected value over trials of the measurement for the i-th individual:

$$E_t \{\mathbf{Y}_{it} | U_i=1\} = \mathbf{Y}_i \quad , \tag{1}$$

where U_i is an indicator random variable denoting presence in the sample.

If we denote the "true" or actual values for the i-th individual as $\mathbf{X}_i$, then the expected measurement for the i-th individual may not be equal to the actual or "true" values for that individual. That is $\mathbf{X}_i \neq \mathbf{Y}_i$ and there may be a net bias in the measurement process. For example, if one wishes to estimate the population mean $\bar{\mathbf{X}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{X}_i$, there is a net bias in the measurement process if $\bar{\mathbf{X}} \neq \bar{\mathbf{Y}}$, where

$$\bar{\mathbf{Y}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{Y}_i \ .$$

We will term a measurement process which measures $\mathbf{Y}_{it}$ as a faulty measurement process. For a simple random sample of size n, the estimate of the population mean using the faulty measurement process is

$$\bar{\mathbf{y}}_t = \frac{1}{n} \sum_{i=1}^{N} U_i \mathbf{Y}_{it} \quad . \tag{2}$$

Now, $E(\bar{\mathbf{y}}_t) = \bar{\mathbf{Y}}$; and, assuming that there is no interaction between the sampling errors and measurement errors, i.e.,

$$E_t \{\mathbf{Y}_i | U_i = U_{i'} = 1\} = \mathbf{Y}_{ii'} = \mathbf{Y}_i \tag{3}$$

(see Koch [2]) then the mean square error matrix of $\bar{\mathbf{y}}_t$ is

$$\begin{aligned} MSE(\bar{\mathbf{y}}_t) = {} & \frac{1}{n} \; (\mathbf{SMV}) + (n-1)(\mathbf{CMV}) \\ & + \frac{1}{n}(\frac{N-n}{N}) \;\; \mathbf{BV} + \mathbf{BT} + \mathbf{TV} \\ & + \mathbf{B} \ . \end{aligned} \tag{4}$$

The first term is the measurement variance (or response variance), $\mathbf{MV}$ and consists of two terms: $\mathbf{SMV}$ the simple measurement variance and $\mathbf{CMV}$, the correlated measurement variance. This term arises because the measurements obtained for the i-th individual are not the same from trial to trial of the survey.

The second term is the sampling variance, $\mathbf{SV}$, and is due to the variability of the $\mathbf{Y}_i$ around $\bar{\mathbf{Y}}$. The sampling variance is composed of three components: $\mathbf{BV}$, the sampling variance of the individual bias terms, $\mathbf{B}_i = \mathbf{Y}_i - \mathbf{X}_i$ around the net bias $\bar{\mathbf{B}}$; $\mathbf{TV}$, the sampling variance of the true values; and $\mathbf{BT}$, the interaction between the individual bias terms and the true values.

The third term $\mathbf{B}$ is due to the net bias in the measurement process.

THE DOUBLE SAMPLING SCHEME IN GENERAL

The general double sampling scheme (DSS) is as follows: the survey is conducted in two phases. In phase 1, an initial sample is drawn and faulty measurements are obtained. In phase 2, a subsample fo the original sample is drawn and one of two schemes are employed:

1. Repeat measurement for each individual in the subsample are made using the faulty measurement process and the accurate values are obtained for each individual in the subsample. This scheme allows one to simultaneously estimate components of measurement error and eliminate the measurement bias. Assuming srs of size n and n_1 at the two phases of the survey, the unbiased estimate of $\bar{\mathbf{X}}$ is

$$\bar{\mathbf{z}}_t = \bar{\mathbf{y}}_{1t} - (\bar{\mathbf{y}}_{2t} - \bar{\mathbf{x}}_2) \quad . \tag{5}$$

If we assume that the measurements obtained at the two phases of the survey with the faulty measurement process are independent, the variance-covariance matrix of $\bar{\mathbf{z}}_t$ is given by

$$\begin{aligned} \dot{\mathbf{V}} = {} & \frac{1}{n} \{(\mathbf{SMV}) + (n-1)(\mathbf{CMV})\} \\ & + \frac{1}{n_1} \{(\mathbf{SMV}) + (n_1-1)(\mathbf{CMV})\} \\ & + \frac{n-n_1}{nn_1} \; \{\mathbf{BV}\} \\ & + \frac{1}{n} (\frac{N-n}{N}) \; \{\mathbf{TV}\} \ . \end{aligned} \tag{6}$$

2. One may measure only the accurate values in the subsample which allows one to eliminate the measurement process bias but not estimate the components or error. In this case, an unbiased estimator for $\bar{\underset{\sim}{X}}$ is

$$\bar{\underset{\sim}{w}}_t = \bar{\underset{\sim}{y}}_t - (\bar{\underset{\sim}{y}}_{st} - \bar{\underset{\sim}{x}}_s) \quad , \tag{7}$$

and

$$\begin{aligned} \underset{\sim}{V} = & \frac{n-n_1}{nn_1} \{(\underset{\sim\sim\sim}{SMV}) - (\underset{\sim\sim\sim}{CMV})\} \\ & \frac{n-n_1}{nn_1} \{(\underset{\sim\sim}{BV})\} + \frac{1}{n}(\frac{N-n}{N}) \{\underset{\sim\sim}{TV}\} . \end{aligned} \tag{8}$$

Greater detail concerning the DSS may be found in Lessler [3].

SPECIFIC SURVEY DESIGNS WHICH EMPLOY DSS's

The above DSS may be adapted to a variety of survey situations from the simple to the complex. The focus of the first scheme is to allow one to estimate the components of the mean square error of estimates using faulty measurement processes and to form estimates which are free of the net bias associated with the faulty measurement process. This is particularly important for the evaluation of ongoing surveys in which one wishes to estimate components of error in order that future adjustments can be made in the survey procedures that will accomplish a reduction of these errors. In addition, it would be useful for pilot studies of surveys in which alternative measurement processes including alternative questionnaires, types of interviewers, and other procedures are to be evaluated.

Two specific survey designs which employ the first form of the DSS are illustrated in the following:

A. A self-enumeration survey employing simple random sampling.

An original sample is drawn and the faulty measurements are obtained. In a second phase of the survey, the accurate values are obtained for members of the original sample as well as remeasurements with the faulty measurement process. For example, the faulty measurement process might be a mail survey in which individuals were queried as to certain demographic characteristics and their bank balance, income, health expenses, etc. A subsample is drawn and remeasurements obtained by the mail survey. In addition, record checks are done to obtain the accurate values.

The following model for the faulty measurements is assumed. Here, α indicates the phase of the survey:

$$Y_{i\alpha t} = X_i + B_{i\alpha t} \quad . \tag{9}$$

We assume the following:

1. $$E_t(Y_{i\alpha t}) = Y_i \quad , \alpha = 1, 2 \tag{10}$$

2. $$E_t(B_{i\alpha t}) = B_i \quad , \alpha = 1, 2 \tag{11}$$

3. The measurement process is equally variable from trial to trial so that

$$E_t\{(B_{i\alpha t} - B_i)^2\} = \gamma_i^2 \quad \text{for } \alpha = 1, 2. \tag{12}$$

4. The measurement process for each individual in the sample is statistically independent of that for any other individual and is statistically independent between phases for a particular individual, i.e.,

$$Cov(B_{i\alpha t}, B_{i'\alpha t}) = Cov(B_{i\alpha t}, B_{i\alpha' t}) = 0$$

for $i \neq i'$, $\alpha \neq \alpha'$.

The variance of $\bar{z}_t$ in terms of this specific model is

$$V(\bar{z}_t) = \frac{n+n_1}{nn_1}\{S_\gamma^2\} + \frac{n-n_1}{nn_1}\{S_B^2\} + \frac{1}{n}(\frac{N-n}{N}) \{S^2\} ; \tag{13}$$

where

$$S_\gamma^2 = \frac{1}{N} \sum_{i=1}^{N} \gamma_i^2 ;$$

$$S_B^2 = \frac{1}{N-1} \sum_{i=1}^{N} (B_i - \bar{B})^2 ;$$

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2 \quad .$$

Letting U_i and V_i be the indicator random variables indicating presence in the sample and subsample respectively, we have the following sample estimators for the above variance components.

1. The subsample variance of the true values s_x^2,

$$s_x^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{N} [V_i(X_i - \bar{x})]^2 \quad .$$

$$E(s_x^2) = S^2 \quad .$$

2. The subsample variance of the bias terms,

$$s_B^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{N} \{V_i[Y_{i2t} - X_i) - (\bar{y}_{2t} - \bar{x}_2)]\}^2 \; .$$

$$E(s_B^2) = S_\gamma^2 + S_B^2 \quad .$$

3. The between-phase within individual sum of squares s_w^2,

$$s_w^2 = \sum_{i=1}^{N} U_i V_i (Y_{i1t} - Y_{i2t})^2 \quad .$$

$$E\left(\frac{s_w^2}{2_{n1}}\right) = s_Y^2 \quad .$$

Thus we have the following set of estimators:

$$\hat{S}^2 = s_x^2 \quad ,$$

$$\hat{S}^2 = \frac{s_w^2}{2n_1} \quad ,$$

$$\hat{S}_B^2 = s_B^2 - \frac{s_w^2}{2n_1} \quad .$$

B. Survey Design Using Interviewers

In addition to the population of N individuals, let there be a fixed population of B interviewers, indexed by the subscript j, available for use in the survey. A simple random sample of size n is drawn for the first phase of the survey. Each individual in the sample is assigned at random to one of the interviewers. The interviewer structure is characterized by indicator random variables C_{ij} where

$$C_{ij} \quad \begin{array}{ll} 1 & \text{if the i-th individual is assigned to the j-th interviewer} \\ 0 & \text{otherwise.} \end{array}$$

For simplicity, we assume n = Br. The subsample and interviewer structure at the second phase is characterized by indicator random variables V_i and D_{ij} respectively. $Y_{ij\alpha t}$ is the measurement obtained for the i-th individual by the j-th interviewer at the α-th phase of the t-trial of the survey process. A specific model for the faulty measurements is:

$$Y_{ij\alpha t} = \bar{X} + H_i + \bar{B} + L_i + Q_j + (IQ)_{ij} + Z_{j\alpha t} + R_{ij\alpha t} \qquad (14)$$

where the effects in the model are defined using the following,

$$E_t(Y_{ij\alpha t}) = Y_{ij} \quad ,$$

$$Y_i = \frac{1}{N} \sum_{j=1}^{B} Y_{ij} \quad ,$$

$$\bar{Y}_{.j} = \frac{1}{N} \sum_{i=1}^{N} Y_{ij} \quad ,$$

$$\bar{Y}_{.j\alpha t} = \frac{1}{N} \sum_{i=1}^{N} Y_{ij\alpha t} \quad ,$$

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i = \frac{1}{NB} \sum_{i=1}^{N} \sum_{j=1}^{B} Y_{ij} \quad ,$$

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i \quad ,$$

$$B_i = Y_i - X_i \quad ,$$

$$\bar{B} = \frac{1}{N} \sum_{i=1}^{N} B_i \quad ,$$

which gives

$$H_i = X_i - \bar{X} \quad ,$$

$$L_i = (Y_i - X_i) - \bar{B} = B_i - \bar{B} \quad ,$$

$$I_i = H_i + L_i = Y_i - \bar{X} - \bar{B} = Y_i - \bar{Y} \quad ,$$

$$Q_j = \bar{Y}_{.j} - \bar{Y} \quad ,$$

$$(IQ)_{ij} = Y_{ij} - I_i - Q_j - \bar{Y} \quad ,$$

$$Z_{i\alpha t} = \frac{1}{N} \sum_{i-1}^{N} (Y_{ij\alpha t} - Y_{ij}) = \bar{Y}_{.j\alpha t} - Y_{.j} \quad ,$$

and

$$R_{ij\alpha t} = Y_{ij\alpha t} - Y_{ij} - Z_{j\alpha t} \quad .$$

Let

$$E_t\{Z_{j\alpha t}^2\} = \xi_j^2 \qquad (15)$$

and

$$E_t\{R_{ij\alpha t}^2\} = \eta_{ij}^2 \quad . \qquad (16)$$

If we assume that the measurement process associated with each interviewer is statistically independent of that associated with any other interviewer and is statistically independent from phase to phase for the same interviewer, but may be correlated within interviewers at a particular phase, then the variance of $\bar{z}_t$ in terms of this specific model is

$$V(\bar{z}_t) = \frac{2}{B}\{S_\xi^2\} + \frac{1}{nn_1}\left(\frac{n(N-r_1)+n_1(N-r)}{N-1}\right)\{S_\eta^2\} + \frac{n+n_1}{nn_1}\{S_{IQ}^2\} + \frac{n-n_1}{nn_1}\{S_B^2\} + \frac{1}{n}\left(\frac{N-n}{N}\right)\{S^2\}. \qquad (17)$$

Each variance component is defined as follows:

$$S_\xi^{\ 2} = \{\frac{1}{B} \sum_{j=1}^{B} \xi_j^{\ 2}\}$$

= Interviewer random effect variance

$$S_\eta^{\ 2} = \{\frac{1}{NB} \sum_{i=1}^{N} \sum_{j=1}^{B} \eta_{ij}^{\ 2}\}$$

= Interviewer individual interaction random effect variance

$$S_{IQ}^{\ 2} = \{\frac{1}{B(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{B} (IQ)_{ij}^{\ 2}\}$$

= Interviewer individual interaction fixed effect variance

$$S^2 = \{\frac{1}{N-1} \sum_{i=1}^{N} H_i^{\ 2}\}$$

= Sampling variance

$$S_B^{\ 2} = \{\frac{1}{N-1} \sum_{L=1}^{N} L_i^{\ 2}\}$$

= Sampling variance of the bias terms.

Given the above assumptions, the following set of estimators allows one to estimate each of these components.

1. $s_x^{\ 2} = \hat{S}^2$

2. Between-phase within interviewer within individual sum of squares, BPWII,

$$BPWII, = \sum_{i=1}^{N} \sum_{j=1}^{B} U_i V_i C_{ij} D_{ij} (Y_{ij1t} - Y_{ij2t})^2 .$$

$$E(BPWII) = 2r_1(S_\xi^{\ 2} + S_\eta^{\ 2}) \quad .$$

3. Between phase within interviewer between individual sum of squares, BPWIBI,

$$BPWIBI = \sum_{i \neq i'}^{N} \sum_{j=1}^{B} U_i U_{i'} V_i V_{i'} C_{ij} C_{i'j} D_{ij} D_{i'j} \times$$

$$[(Y_{ij1t} - Y_{i'j1t}) - Y_{ij2t} - Y_{i'2t})]^2 .$$

$$E(BPWIBI) = \frac{r_1(r_1-1)\ (r-1)\ 4N}{(n-1)\ (N-1)} \quad S_\eta^{\ 2} \quad .$$

4. Within phase within interviewer between individual sum of squares, WPWIBI,

$$WPWIBI = \sum_{i \neq i'}^{N} \sum_{j=1}^{B} U_i U_{i'} C_{ij} C_{i'j} [Y_{ij1t} - Y_{i'j1t}]^2 .$$

$$E(WPWIBI) = 2\ Br(r-1)\ \{S^2 + S_B^{\ 2} + S_{IQ}^{\ 2} + \frac{N}{N-1}\ S_\eta^{\ 2}\}.$$

5. Between phase between interviewer between individual sum of squares, BPBIBI,

$$BPBIBI = \sum_{i \neq i'}^{N} \sum_{j \neq j'}^{B} U_i U_{i'} V_i V_{i'} C_{ij} C_{i'j'} D_{ij} D_{i'j'} \times$$

$$[(Y_{ij1t} - Y_{i'j1t}) - (Y_{ij'2t} - Y_{i'j'2t})]^2 \quad .$$

$$E(BPBIBI) = \frac{4(r-1)r_1(r_1-1)}{n-1} BS_{IQ}^{\ 2} + \frac{(B-1)N}{N-1} \ S_\eta^{\ 2} \quad .$$

Thus we have the following set of estimators.

$$\hat{S}^2 = s_x^{\ 2} \quad ,$$

$$\hat{S}_\eta^{\ 2} = \frac{\dfrac{BPWIBI}{r_1(r_1-1)(r-1)4N}}{(n-1)(N-1)} \quad ,$$

$$\hat{S}_\xi^{\ 2} = \frac{BPWII}{2r_1} - \hat{S}_\eta^{\ 2} \quad ,$$

$$\hat{S}_{IQ}^{\ 2} = \frac{\dfrac{BPBIBI}{4(r-1)r_1(r_1-1)}}{(n-1)\ B} - (\frac{B-1}{B})\ (\frac{N}{N-1})\ \hat{S}_\eta^{\ 2},$$

and

$$\hat{S}_B^{\ 2} = \frac{WPWIBI}{2Br(r-1)} - \hat{S}^2 - \hat{S}_{IQ}^{\ 2} - \frac{N}{N-1} \ \hat{S}_\eta^{\ 2} .$$

The second form of the DSS is not directed at getting estimates of the components of error but, rather, at eliminating the net bias associated with the faulty measurement process. In addition, it should be noted that the correlated measurement variance makes a negative contribution to the variance of the estimate. This component, when present, is thought by many to make the largest contribution to the MSE of estimates using the faulty measurement process. As example of a survey which may employ the second form of the DSS is as follows:

C. A Multistage Cluster Sampling Design for the National Medical Care Expenditure Survey

RTI is in the process of designing a National Medical Care Expenditure Survey. The original specifications called for conducting household interviews in which the medical care expenditures of each member of the household are collected along with other data. Following this, the medical care provider and third party payors (TPP) were to be visited and the actual cost of the medical care was to be obtained. The obtaining of provider data and TPP data is an expensive procedure and cost savings could result if provider data are collected on a subsample basis and these data used to correct for the biases in the entire household interview data. The proposed plan for the sur

An original sample of households is drawn and interview values obtained for each individual in the household. A subsample of the households and individuals within households is drawn and the accurate values obtained for each individual in the subsample. The survey design is as follows:

Y_{ijkt} = The interview value obtained for the k-th individual, in the j-th household, of the i-th cluster, at the t-th trial of the interview process.

X_{ijk} = the accurate value obtained for the k-th individual, in the j-th household, of the i-th cluster.

M= number of clusters in population

$\bar{N}$ = N= average number of households in a cluster (or N_i if cluster sizes are not equal)

$\bar{L}$ = L= average number of individuals in a household (or L_{ij} if household sizes are not equal).

In this case, we are not interested in estimating the components of error and are only interested in eliminating the bias.

Assuming srs at all phases of sampling with an original sample of clusters of size m, an original household sample within each cluster of size n, and corresponding subsample sizes of m_1, and n_1 and within household subsamples of size ℓ_1, we have the following:

The estimator is

$$\bar{w}_t = \bar{y}_t - \bar{y}_{st} + \bar{x}_s \quad . \tag{18}$$

$$V(\bar{w}_t) = MV + BV + TV \; . \tag{19}$$

$$TV = \frac{1}{m}\left(\frac{M-m}{M}\right)\left\{\frac{1}{M-1}\sum_{i=1}^{M} Q_i^2\right\} + \frac{1}{mn}\left(\frac{N-n}{N}\right)\left\{\frac{1}{M(N-1)}\sum_{i=1}^{M}\sum_{j=1}^{N} H_{ij}^2\right\}$$

$$BV = \frac{m-m_1}{mm_1}\left\{\frac{1}{M-1}\sum_{i=1}^{M} K_i^2\right\} + \left[\frac{1}{m_1n_1}\left(\frac{N-n_1}{N}\right) - \frac{1}{mn}\left(\frac{N-n}{N}\right)\right]\left\{\frac{1}{M(N-1)}\sum_{i=1}^{M}\sum_{j=1}^{N} A_{ij}^2\right\} + \left[\frac{1}{m_1n_1\ell_1}\left(\frac{L-\ell_1}{L}\right)\right]\frac{1}{NM(L-1)}\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1}^{L} E_{ijk}^2\}$$

$$MV = \frac{m-m_1}{mm_1}\left\{\frac{1}{M}\sum_{i-1}^{M}\xi_i^2 - \frac{1}{M(M-1)}\sum_{i\neq i'}^{M}\xi_{ij'}\right. + \left[\frac{1}{m_1n_1}\left(\frac{N-n_1}{N}\right) - \frac{1}{mn}\left(\frac{N-n}{N}\right)\right]\frac{1}{M(N-1)}\sum_{i=1}^{M}\sum_{j=1}^{N}\eta_{ij}^2 + \frac{1}{m_1n_1\ell_1}\,\frac{L-\ell_1}{L}\,\frac{1}{MN(L-1)}\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1}^{L}\rho_{ijk}^2$$

The previous components are derived from the following model for the Y_{ijkt}:

$$Y_{ijkt} = \bar{X} + \bar{B} + Q_i + H_{ij} + L_{ijk} + K_i + A_{ij} + E_{ijk} + F_{it} + G_{ijt} + R_{ijkt} \quad .$$

Q_i = cluster effect = $\bar{X}_i$-X.

H_{ij} = household effect = $\bar{X}_{ij}-\bar{X}_i$.

L_{ijk} = individual effect = $X_{ijk}-\bar{X}_{ij}$.

K_i = bias effect for cluster i = $\bar{B}_i-\bar{B}$.

A_{ij} = bias effect for household = $\bar{B}_{ij}-\bar{B}_i$.

E_{ijk} = bias effect for individual = $B_{ijk}-\bar{B}_{ij}$.

$F_{it} = \bar{Y}_{it}-\bar{Y}_i$; $E_t\{F_{it}\}=0$; $E_t\{F_{it}^2\} = \xi_i^2$.

$G_{ijt}=\bar{Y}_{ijt}-\bar{Y}_{ij}-F_{it}$; $E_t\{G_{ijt}\}=0$; $E_t\{G_{ijt}^2\}=\eta_{ij}^2$.

$R_{ijkt}=Y_{ijkt}-Y_{ijk}-G_{ijt}-F_{it}$; $E_t(R_{ijkt})=0$;

$E_t(R_{ijkt}^2) = \rho_{ijk}^2$.

We will let the various components be denoted

$$S_Q^2 = \frac{1}{M-1}\sum_{i=1}^{M} Q_i^2,\quad S_H^2 = \frac{1}{M(N-1)}\sum_{i=1}^{M}\sum_{j=1}^{N} H_{ij}^2,\ \text{etc.}$$

In order to have a cost effective design for the survey, it is necessary to determine the optimum sizes of the sample and subsample. To do this one must have estimates for the following set of variance components:

(1) $S_\rho^2 + S_E^2$

(2) $S_\eta^2 + S_A^2$

(3) $S_\xi^2 - S_{\xi\xi'} + S_K^2$

(4) S_Q^2

(5) S_K^2 .

Suppose pilot study data is available in which interview values and accurate values are available for each individual in the pilot study. Then the above components can be estimated using the following sums of square and sums of products.

Summary of expected values of sums of squares and sums of products

	SSX	SSY	SXY
Between Clusters	$(m-1)\ s_Q^2 + (\frac{m-1}{n})(\frac{N-n}{N})\ s_H^2$	$(m-1)\ \{s_\xi^2 - s_{\xi\xi'} + s_Q^2 + s_K^2 + 2s_{QK}\} + (\frac{m-1}{n})(\frac{N-n}{N})\{s_\eta^2 + s_H^2 + s_A^2 + 2s_{HA}\}$	$(m-1)\ \{s_Q^2 + s_{QK}\} + \frac{m-1}{n}(\frac{N-n}{N})\ \{s_H^2 + s_{HA}\}$
Between Households	$m(n-1)\ s_H^2$	$m(n-1)\ \{s_\eta^2 + s_H^2 + s_A^2 + 2s_{HA}\}$	$m(n-1)\ \{s_H^2 + s_{HA}\}$
Between Individuals	$mn(L-1)\ s_L^2$	$mn(L-1)\ \{s_\rho^2 + s_L^2 + s_E^2 + 2s_{LE}\}$	$mn(L-1)\ \{s_L^2 + s_{LE}\}$

OPTIMUM ALLOCATION TO THE SAMPLE AND SUBSAMPLE

In each of the above cases, the variance of the sample statistics may be written in terms of the following simple expression

$$V = \sum_{k=1}^{K} \frac{V_K^2}{\ell_K} + V_o \tag{21}$$

where V_K^2 and ℓ_K are the variance and cost components associated with the k-th design level. In addition, simple linear cost functions may be used of the form

$$C = \sum_{k=1}^{K} C_K\, \ell_K + D_o \quad . \tag{22}$$

Thus, optimum allocation to sample and subsample for fixed cost may be obtained using the usual solution to the above linear forms. In addition, overall multipurpose allocations may be derived using a procedure proposed by Kish in 1974 [4].

REFERENCES

[1] Hansen, M.H., Hurwitz, W.N. and Bershad, M.A., "Measurement Errors in Censuses and Surveys," Bulletin of the Intervational Statistical Institute, 38, Part II (1962), 359-74.

[2] Koch, Gary G., " An Alternative Approach to Multivariate Response Error Models for Sample Survey Data with Applications to Estimators Involving Subclass Means," Journal of the American Statistical Association, 68 (December 1973), 906-913.

[3] Lessler, Judith T., A Double Sampling Scheme Model for Eliminating Measurement Process Bias and Estimating Measurement Errors in Surveys, Institute of Statistics Mimeo Series No. 949, University of North Carolina, August 1974.

[4] Kish, Leslie, "Optimal and Proximal Multipurpose Allocation" Proceedings of the Social Statistics Section, American Statistical Association, 1974, pp 111-118.

THE COHORT FACTOR IN THE GROWTH OF CRIME[1]

Irving Leveson, Hudson Institute

Introduction

Some explanations of the rise in crime suggest influences which are a characteristic of the period observed. A decrease in the availability of jobs or a decline in the likelihood of punishment might lead concurrently or with a short lag to an increase in crime in each age group. Other potential explanations, however, depend primarily on the characteristics which the individuals bring with them as they enter the ages at which crimes are committed in significant numbers. An increase in parental permissiveness, for example, might produce a new generation of children which is more likely to go on to commit crime at older ages.

Explanations of the latter type can be characterized by a cohort process in which crime increases as successive cohorts are replaced by others more likely to commit crime than their predecessors. Shifts in criminality from one cohort to the next result from the combined effects of the contributions of behavioral forces which it may not be possible to separately identify and measure.

The objective of this analysis is to determine the separate contributions to the rise of crime of cohort effects and period effects. The separation of the cohort from period effects has presented a significant problem in the interpretation of social data. While the basic concepts have long been understood, in the simplified empirical procedures which are in common use the cohort and period effects are intermixed. Examination of data on specific cohorts--looking at one age one year, the next age in the next year, etc.--as a means of trying to identify a constant group of persons under constant conditions also combines period and cohort effects since period forces can produce shifts and rotations in age profiles.[2] Significant improvements in methodology for estimating age, period and cohort effects in observations on individuals have occurred during the last several years.[3]

The present analysis separates age, period and cohort effects in aggregate data. Information over time by single years of age is examined, making use of the fact that a cohort effect implies that a rise at one age in one year will be associated with a rise at the next age in the next year. Shifts between successive cohorts, free of period effects, are estimated in order to determine their contribution to the growth of crime. The separation of cohort and period effects makes it possible to test and measure various notions about the way in which a cohort process works. At the same time it is possible to obtain clearer measures of the static age profiles--age patterns that exist in a constant population under constant conditions.

A View of the Crime Growth Process

The nature of a cohort process is illustrated in Figure 1. Panel A shows a rising trend in crime rates for 15 year olds. The trend line is shifted upward when a new generation appears which is more crime prone than its predecessors. The upward shift is a "cohort effect" and the underlying trend a "period effect." Panel B shows the effects of the same change on the crime rate of all persons age 15-24. As a result of the cohort effect, in one year the crime rate for 15 year olds rises, in the next year the rate for 16 year olds rises and so on. As time passes a larger proportion of 15-24 year olds is composed of persons with the greater tendency to commit crime. The average crime rate for all 15-24 year olds rises until all 15-24 year olds are from the new group. During this growth period a new trend is established which is the sum of both cohort and period effects. After the new level is reached for 24 year olds the slope of the crime trend will return to its original level but crime will grow from a higher rate.

Figure 1 deals only with a once and for all change. Cohort effects may also operate in the form of a trend factor which makes each succeeding group of 15 year olds more likely to commit crime than its predecessors. In Panel A of Figure 2 the crime rate of 15 year olds rises from year-to-year as the result of the continuous arrival of groups increasingly likely to commit crime. The observed increase in crime along the solid line is the sum of the effects of this cohort trend and the period trend of the dotted line. In a period during which such a cohort trend is being introduced the average crime rate of 15-24 year olds rises by increasing absolute amounts (Panel B). After the new trend has reached the 24 year olds the slope of the trend will continue at its high level as each age in the 15-24 year group is becoming more crime prone than its predecessors.

Social variables affecting crime which themselves have a trend would be expected to produce such a cohort trend effect. This can be expected to occur commonly and it is this form with which we will deal. It is apparent that in the presence of such patterns prediction based on observed trends may be inappropriate. If trends in social variables producing cohort effects are about to change or if the base period for prediction contains the effects of changes in trends which are not likely to continue, serious error may result.

Separation of Age, Period and Cohort Effects

Cohort and period effects are intermingled in observed patterns of crime rates by age as well as in trend data. Curve 1 in Figure 3 shows a static age-crime profile. If there were no underlying trend in crime rates within age groups there would still be a tendency for crime rates to rise and fall as persons entered more crime-prone years and moved out of them. Crime rates for a cohort which became 15 years of age in 1952, for example, might be greater in 1953 at age 16, rising from point A to point B. If crime rates rose in all age groups from one period to the next the age-crime profile would be higher each year than the year before. The crime rate of 16 year olds in 1953 would rise by the sum of the general rise BC and the movement along the age profile AB. Observed movements of crime rates under the influence of such a general trend would trace out a new and steeper curve (II) passing through points A and C.

The introduction of a cohort trend effect in addition to the general trend is shown in Curve III. The "class of 1953" begins with a higher crime rate than the "class of 1952." The difference is the sum of the greater general tendency for crime in 1953 than in 1952 (AD=BC) and the greater tendency for the new cohort to commit crime DE. The slope is parallel to Curve II because the new cohort is also subject to a trend. Observed relationships such as Curve II and Curve III differ from the static age profiles of Curve I by both period and cohort effects and the usual method of plotting values by age over time will not separate them.

We can statistically separate cohort effects from period effects by estimating the equation

$$I_{ij} = a_1 + b_{11} I_{i-1,j-1} + b_{12} P_j \qquad (1)$$

where I_{ij} is the arrest rate for the ith age in year j (for example 16 year olds in 1953), $I_{i-1,j-1}$ is the arrest rate for the preceding age in the preceding year (for example 15 year olds in 1952) and P_j is a measure of period forces affecting crime. The response coefficient b_{11} measures the importance of the cohort effect as the amount that the number of arrests will rise in one year when there is an increase of one arrest in the preceding age in the preceding year. The size of b_{11} will depend on the relative size of cohort and period effects.

Empirical Tests

Equation I was estimated for city arrest rates of all index crimes, index violent crimes, index property crimes and the seven categories of index crimes over the period 1952 to 1973. In the set of equations for each crime the arrest rate of persons age 25 and over was used to represent period effects. The arrest rate of persons one year younger in the preceding year measured the crime factors peculiar to a cohort. The analysis was performed for each age 16-24. The estimated equations for violent crime are presented in Table 1.

Typically, 95 percent of the variation in the age-specific arrest rates for violent crime was associated with the explanatory variables. Most of the estimates of response coefficients to cohort and period effects are statistically significant at very high levels. An increase of 100 violent crimes at one age is associated with a rise of 50-60 crimes in the next age in the following year. There is a tendency for cohort response rates to be lower after age 20. Property crime rates also show highly significant cohort responses (Table 2). They are somewhat smaller, however--typically .4-.5. A smaller response after age 20 is also evident.

Shifts in Age-Arrest Profiles

A measure of the contribution of cohort effects to the rise in crime can be obtained by examining shifts in age profiles. Gross age profiles in the data on arrest rates for 15-24 year olds from 1947-1973 are similar to those observed in other studies. The gross age profiles for violent and property crimes are shown in Figures 4 and 5. Arrest rates are plotted for each cohort. The year assigned to each curve is the year in which that cohort was age 15. The age-arrest profile of the "class of 70" is far higher than the "class of 52," reflecting the influence of both period and cohort effects. Because period effects produce a rise in arrest rates over time the shapes of the age profiles give a distorted picture of age patterns which would exist with other conditions unchanged. The patterns in the gross data show apparent tendencies for violent crime arrest rates to rise sharply with age in the teen years and then level off and for arrest rates for property crimes to decline with advancing age.

We can make use of the estimates of the strength of period effects to derive net age-arrest profiles which indicate the age patterns in arrest rates when period effects are removed. Net age-arrest rates (N_{ij}) were derived by standardizing the gross age-specific arrest rates to the 1973 period effects.

$$N_{ij} = I_{ij} + b_{12} \, (\hat{P}_j) \qquad (2)$$

The measure of period effects--the arrest rate for persons age 25 and over--was subtracted in each year from its 1973 value to obtain $\hat{P}$. This period difference was multiplied by the period response coefficient for each age group separately. The result was added to the gross arrest rate for that age in that year in order to obtain the arrest rate that would have been observed in that age in that year if the 1973 period effect had prevailed. The procedure was repeated for each crime using information specific to that crime.

Net age-arrest profiles with the period effects removed are shown for violent crimes in Figure 6. The differences between the net and gross patterns are striking. When period factors contributing to the rise in arrest rates are removed the violent crime age profiles are much closer together and no longer show a steep increase in the teenage years. Comparison of Figures 5 and 7 also indicate that for property crime the cohort profiles have become close together.

In order to estimate the extent of cohort shifts over time an equation was fitted to the net age-arrest profiles. The arrest rate is a function of age for each cohort and the level of the age-arrest relationship varies from year to year. ${}_nN_{ij}$, the net arrest rate for the i th age group in the j th year, is given by:

$${}_nN_{ij} = a_2 + b_{21} A + \Sigma \, (b_{22} D_1 + \ldots b_{2,n} D_{n-1}) \qquad (3)$$

where A is age and D_1 to D_{n-1} is a set of dummy variables which take on values of zero and 1 for each of the n-1 cohorts. Data were used for all cohorts which were included completely in the period 1952-1973.[4] The last group reaching age 24 in 1973 was 16 years old in 1965. Thus, the last dummy variable denotes the class of '65. The ratio of any year's cohort position to the position of the 1952 cohort is obtained by taking $e^{b_{2,n}}$.

Indexes of arrest rates attributable to cohort shifts are shown in Table 3. Values for the cohorts 1963-65 are significantly different from those of the class of '52. Between 1952 and 1965 cohort influences produced a rise in arrest rates for index property crimes of 18.1 percent, an average growth of 1.3 percent per year. Cohort shifts raised the violent crime arrest rate by even more, 28.6 percent, or 2.0 percent per year. The effect on homicides is particularly striking. Cohort shifts accounted for a rise of 40.3 percent in homicide arrest rates between 1952 and 1965, an annual growth rate of 2.6 percent. The indexes show a sharp jump after 1962. Cohort effects were about the same size for violent and property crimes between 1952 and 1960. After 1960, however, the violent crime effects were much greater.

These tests do not indicate whether cohort shifts had a greater impact after the early 1960s. Inspection of Figures 4 and 5 suggests, however, that a sharp jump also occurred in 1966 for both violent and property crimes and continued to 1968 for violent crimes. No upward shifts are indicated in the remainder of the decade.

Figure 6 shows a pattern of upward shifts in age profiles of violent crime arrest rates at a higher percentage rate for the youngest offenders. Tests of this pattern were conducted by estimating separate equations for each cohort and comparing successive age slope coefficients. The equations confirm the observation that the particularly rapid upward shift of the cohort profiles has contributed to the rapid growth of youth crime.[5]

[1]This paper is excerpted from Irving Leveson, The Growth of Crime, Hudson Institute, July, 1975, Chapter 5.

[2]The combination of a cohort and period effects has typically been used as a measure of cohort effects alone, both in aggregate analyses and in longitudinal studies of the behavior of individuals. See A. Joan Klebba, "Homicide Trends in the United States, 1900-1974," Public Health Reports, 90, No. 3 (May/June, 1975), pp. 195-204 and Marvin Wolfgang, Robert Figlio and Thorsten Sellin, Delinquency in a Birth Cohort, Chicago: University of Chicago Press, 1972. Klebba notes that the 15-19 year old population in 1972 had a higher homicide death rate than the cohort 5 years earlier and suggests that as a result 20-24 year olds can be expected to show an increase in homicide rates in the future.

[3]Important advances have been made using multiple classification analysis. See Karen Mason, et al., "Some Methodological Issues in Cohort Analysis of Archival Data," American Sociological Review, 38, No. 2 (April, 1973), pp. 242-258 and H. Winsborough, "Age, Period and Cohort Effects on Earnings by Race," Social Indicator Models, edited by Kenneth Land and Semour Spilerman, New York: Russell Sage Foundation, 1975, pp. 201-217.

[4]Equations with age in the linear form yielded nearly identical estimates of cohort shifts.

[5]Tests were conducted to verify that this pattern was not an accidental result of the way in which period effects were measured. Equations in Tables 1 and 2 were reestimated including alternative measures of youth unemployment as an additional period variable. The cohort response coefficients were affected insufficiently to account for this pattern. Furthermore, if it were the result of insufficient control for youth unemployment a rotation of age profiles would have been expected for property crimes as well.

FIGURE 1

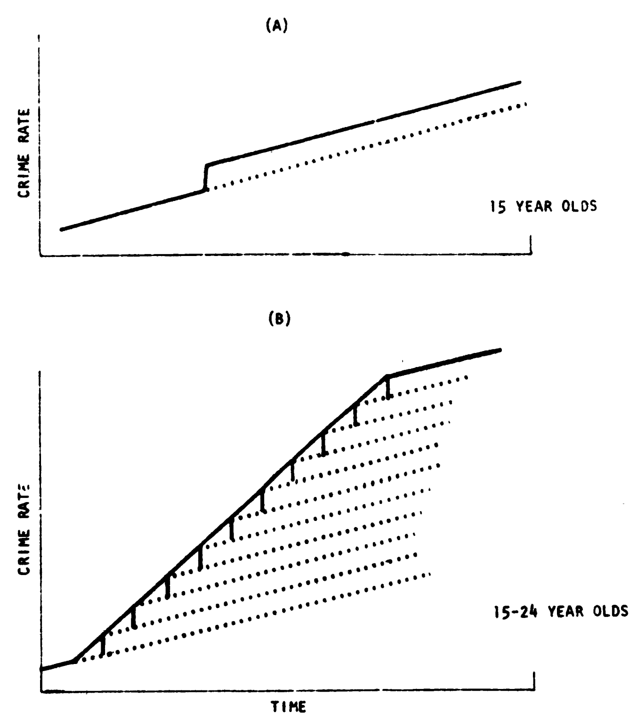


FIGURE 2

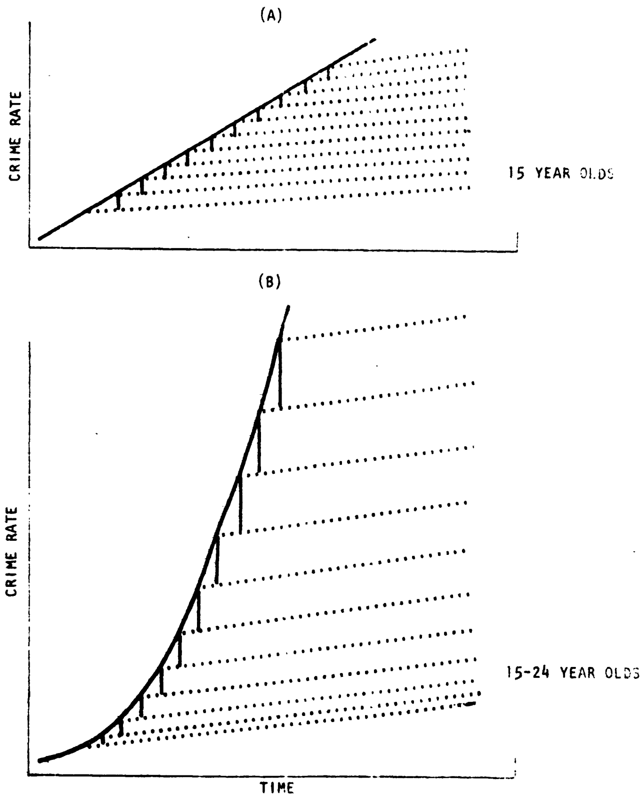


Table 1

RESPONSES OF VIOLENT CRIME ARREST RATES FOR PERSONS AGE 16-24 TO COHORT AND PERIOD EFFECTS

Age	Constant Term	Cohort Response	Period Response	Coefficient of Determination
16	-90.4	.903 (8.755)	1.885 (3.344)	.978
17	-171 6	.627 (5.107)	3.051 (3.887)	.969
18	-29.6	.621 (3.454)	1.927 (1.629)	.923
19	-20.3	.641 (5.454)	1.431 (2.13)	.945
20	-56.4	.736 (6.230)	1.296 (2.034)	.950
21	-85.0	.657 (5.027)	2.084 (3.139)	.949
22	-59.6	.482 (4.761)	2.156 (3.931)	.960
23	-67.1	.318 (1.545)	2.707 (2.932)	.878
24	-261.8	.348 (2.397)	4.448 (7.417)	.968

Note: t ratios are in parentheses.

FIGURE 3

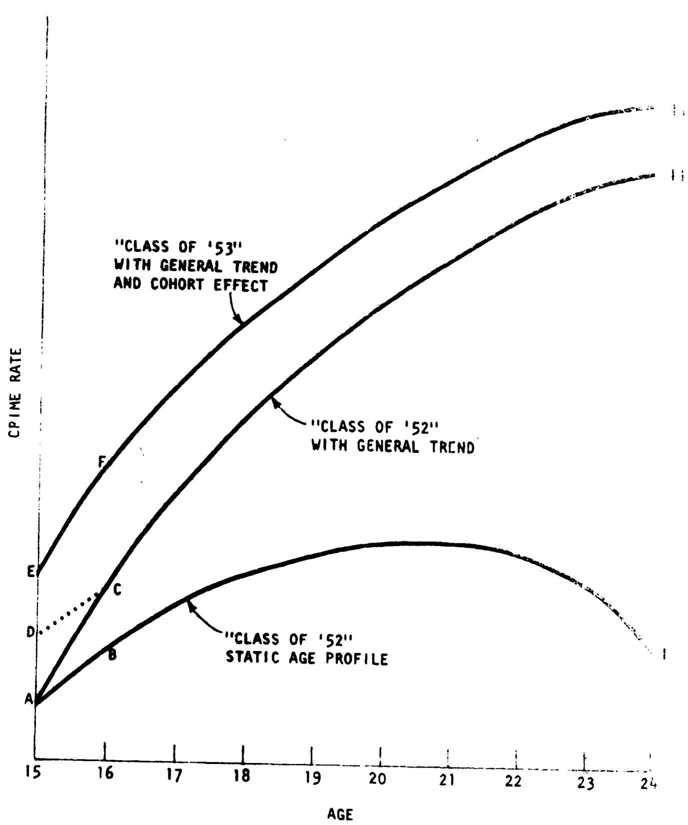


Table 2

RESPONSES OF PROPERTY CRIME ARREST RATES FOR PERSONS AGE 16-24 TO COHORT AND PERIOD EFFECTS

Age	Constant Term	Cohort Response	Period Response	Coefficient of Determination
16	650.6	.560 (4.518)	4.107 (2.086)	.879
17	-233.7	.294 (3.173)	8.566 (5.952)	.933
18	-344.1	.521 (3.701)	4.795 (2.598)	.941
19	-88.9	.544 (2.891)	3.771 (2.217)	.932
20	-234.6	.520 (3.527)	3.096 (2.324)	.947
21	-289.2	.461 (3.667)	3.783 (3.779)	.953
22	-126.2	.364 (3.571)	3.055 (4.001)	.966
23	-200.2	.128 (.916)	4.105 (5.270)	.957
24	-497.7	.289 (2.827)	4.948 (10.133)	.989

Note: t ratios are in parentheses.

FIGURE 4

GROSS AGE PROFILES OF CITY ARRESTS FOR PROPERTY CRIMES (PER 100,000 PERSONS)

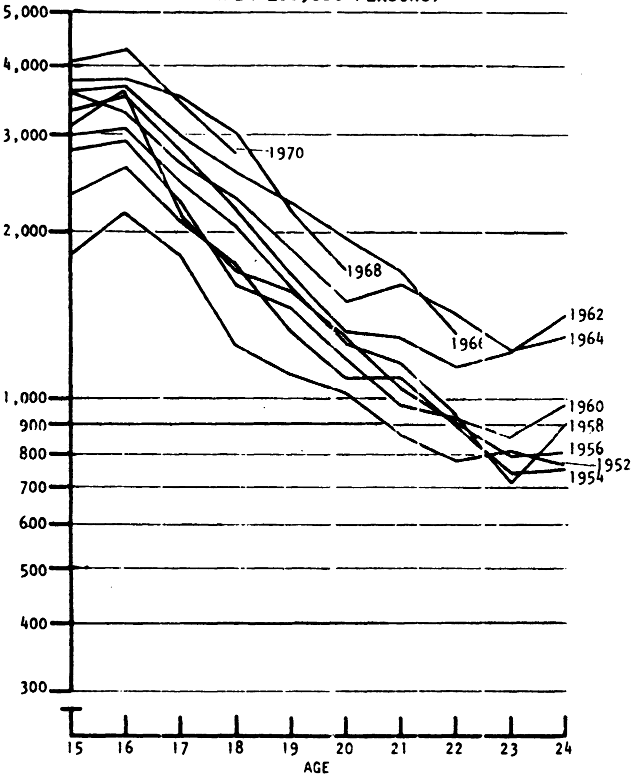


FIGURE 5

GROSS AGE PROFILES OF CITY ARRESTS FOR VIOLENT CRIMES (PER 100,000 PERSONS)

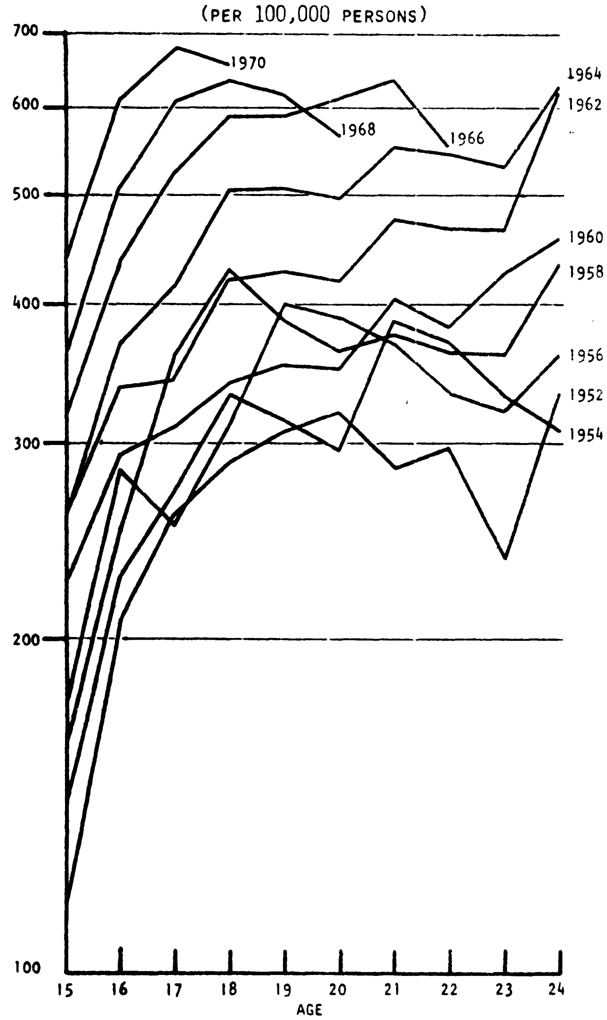


FIGURE 6

NET COHORT AGE PROFILES OF CITY ARRESTS FOR VIOLENT CRIMES
(PER 100,000 PERSONS)

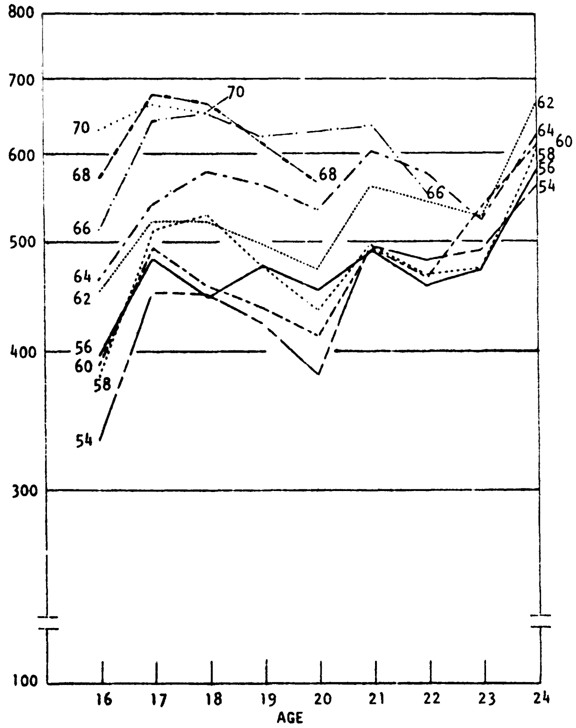


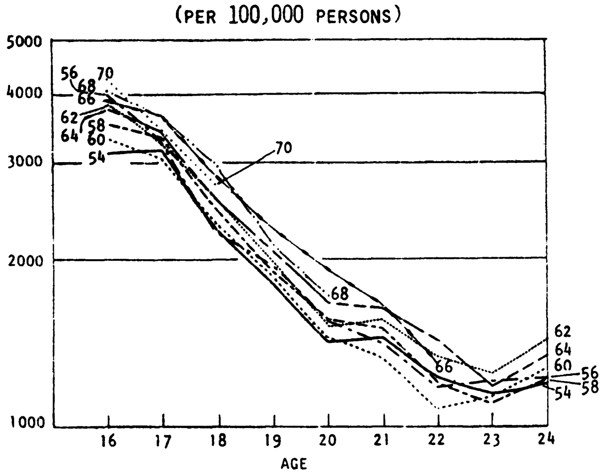
Table 3

INDEXES OF ARREST RATES
ATTRIBUTABLE TO COHORT SHIFTS
(1952 = 100)

Cohort	Violent	Property	Homicide
1952	100.0	100.0	100.0
1953	97.7	97.9	100.9
1954	99.6	102.2	102.1
1955	102.4	101.2	104.1
1956	107.5	110.8	116.0
1957	108.7	111.3	115.3
1958	107.5	109.0	116.3
1959	111.3	110.9	123.3
1960	108.8	108.2	118.5
1961	109.0	105.1	116.8
1962	111.9	106.2	122.8
1963	155.7	118.8	134.9
1964	125.9	120.1	136.0
1965	128.6	118.1	140.3

FIGURE 7

NET COHORT AGE PROFILES OF CITY ARRESTS FOR PROPERTY CRIMES
(PER 100,000 PERSONS)



OPTIMUM ALLOCATION IN STRATIFIED RANDOM MULTIPLICITY SAMPLING

Paul S. Levy
University of Illinois at the Medical Center
School of Public Health

This work considers estimation of dichotomous characteristics of rare populations by sample survey. In particular, we investigate optimum allocation when a network sampling rule (also called multiplicity rule) is used to link members of the rare population to enumeration units and when the enumeration units are chosen by stratified random sampling. For this situation, formulae are derived for optimum allocation under the restriction that an enumeration unit is linked to at most one member of the rare population, and the cost efficiency of network sampling rules is compared to that of conventional enumeration rules.

1. Introduction

This report is concerned with the estimation of the prevalence of dichotomous attributes among members of "rare" population groups where a "rare" population group is considered here to be one which consists of less than 3% of the total population. In the field of health statistics, we often are interested in making inferences about rare populations. For example, approximately 1% of the total U.S. population dies during a given year, and hence, decedents represent a rare population group. Another common problem is that of association between two disease conditions. For example, we may be interested in determining whether gout occurs more frequently among diabetics than among non-diabetics. Since recent surveys have estimated the prevalence of diabetes to be in the neighborhood of 3.0 per 100 persons [6], diabetics constitute a rare population group.

Often, characteristics of rare populations are examined not from random samples of the rare population groups but from "collections" of individuals from these groups selected for convenience. For example, association between two diseases is often examined from hospitalized patients and, as Berkson has pointed out [1], the same association between two conditions found among hospitalized patients may not exist in the general population. Thus, in order for valid inferences to be made concerning rare population groups, one is often necessarily confronted with the problem of seeking them out by sample survey.

The problem of locating members of rare population groups by sample survey for purposes of estimating their characteristics has received recent attention among survey statisticians. Sudman [7], suggested the use of stratification, Bayesian optimum allocation and sequential analysis to increase cost efficiency. Sirken, on the other hand, has developed the theory of network sampling (discussed below) as a tool for measurement of characteristics in rare populations [3], and has used it in a wide variety of applications [4], [5], [6].

The concepts, formulation and notation used here are similar to what was used by Sirken and Levy [5]. In particular, this work investigates again the effect which manipulation of enumeration rules (or counting rules) has on the precision of estimators derived from sample surveys. An enumeration (counting) rule is an algorithm for linking enumeration units with elements. For example, in household surveys for estimating deaths, an enumeration rule might specify that a death can be reported only in the household of the decedent. Another enumeration rule might specify that a death can be reported in the decendent's household or in the household of any sibling of the decedent. The former counting rule is an example of a conventional counting rule since it links every element (decedent) to only one enumeration unit (household), while the latter is a network counting rule since an element can be linked to more than one enumeration unit. The rationale behind network rules is that they can increase the "yield" of rare events in a sample survey which results in estimates having increased precision. In a series of papers, Sirken [3], [4], and Sirken and Levy [5], have developed expressions for unbiased estimates of characteristics and for the variances of these estimates when network counting rules are used.

In this report, methods are developed for optimum allocation in stratified random sampling when a network estimator is used to estimate the prevalence of a dichotomous attribute in a rare population group.

2. Method

Let us suppose that the rare population group consists of Y members or elements, X of which have some dichotomous attribute, A, and Y-X do not. Let us suppose further that the total population consists of L enumeration units grouped into H strata with L_h enumeration units appearing in the h^{th} stratum, $h = 1,\ldots, H$, and that within each stratum, X_h elements having attribute A and $Y_h - X_h$ elements not having attribute A are linked to enumeration units in the stratum by a conventional counting rule. For convenience, we assume that within each stratum (h), elements with the labels I_{hi}, $i = 1, \ldots, X_h$ have attribute A whereas elements with labels I_{hi}, $i = X_h + 1,\ldots, Y_h$ do not have attribute A.

Let us suppose that within each stratum, a simple random sample of ℓ_h enumeration units is drawn for purposes of estimating the prevalence, X/y of attribute A among members of the rare population. For each stratum, enumeration unit specific prevalence rates are given by:

$$R_{hx} = X_h / L_h$$

and

$$R_{hy} = Y_h / L_h$$

Likewise, for each stratum we define

$$\pi_h = L_h / L$$

$$\gamma_h = X_h / Y_h = R_{hx} / R_{hy}$$

and over all strata we define

$$R_y = Y/L$$

$$R_x = X/L$$

and

$$\gamma = X/Y = R_x / R_y$$

In this formulation, we will assume that within each stratum, the counting rule used specifies that each of the Y_h elements of the rare population is linked to at least one of the L_h enumeration units in the stratum but to no enumeration unit in another stratum. If the counting rule used is a conventional one, then each element would be linked to one and only one enumeration unit in the same stratum. On the other hand if the counting rule used is a network sampling rule, then an element may be linked to more than one enumeration unit. In addition, whether the counting rule used is a conventional or network rule, we place the important restriction that no enumeration unit is linked to more than one element of the rare population.

From the simple random sample of ℓ_h enumeration units in each stratum, we obtain the estimate γ' of γ given by:

$$\gamma' = \left(\sum_{h=1}^{H} L_h\, x_h/\ell_h\right) / \left(\sum_{h=1}^{H} L_h\, y_h/\ell_h\right) \tag{1}$$

where

$$y_h = \sum_{j=1}^{\ell_h} (\lambda'_{hi_j} + \lambda''_{hi_j})$$

$$x_h = \sum_{j=1}^{\ell_h} \lambda'_{hi_j}$$

$$\lambda'_{hi} = \sum_{\alpha=1}^{X_h} \delta_{h\alpha i} / W_{h\alpha},$$

$$\lambda''_{hi} = \sum_{\alpha=X_h+1}^{Y_h} \delta_{h\alpha i} / W_{h\alpha},$$

$\delta_{h\alpha i}=1$ if element $I_{h\alpha}$ is linked to enumeration unit i by the counting rule, $\delta_{h\alpha i} = 0$ otherwise, $i_1,\ldots,i_{\ell_h}$ represent the indices of enumeration units chosen in the sample,

and $W_{h\alpha} = \sum_{i=1}^{L_h} \delta_{h\alpha i}$ = the multiplicity of element $I_{h\alpha}$ (i.e., the number of enumeration units linked to element $I_{h\alpha}$ by the counting rule).

Since γ' is a ratio estimate based on stratified random sampling, its approximate variance is given by:

$$\sigma^2_{\gamma'} = \left(\sum_{h=1}^{H} \pi_h^2\, S^2_{hz'}\, (L_h-\ell_h)/(\ell_h\, L_h)\right)/R_y^2 \tag{2}$$

where

$$S^2_{hz'} = S^2_{hx} + \gamma^2 S^2_{hy} - 2\gamma\, S_{hxy}$$

and where S^2_{hx}, S_{hy}^2 and S_{hy} are within stratum variance and covariances with respect to the distribution of elements having attribute A and not having attribute A among enumeration units in the stratum.

For enumeration rules based on network sampling in which an enumeration unit is linked to at most one element, Sirken and Levy /̲ 5 /̲ have shown that:

$$S^2_{hx} = R_{hx}\,(E_{hx} - R_{hx})$$

$$S^2_{hy} = R_{hy}\,(E_{hy} - R_{hy}) \tag{3}$$

and

$$S_{hxy} = R_{hx}\,(E_{hx} - R_{hy})$$

where

$$E_{hx} = \left(\sum_{\alpha=1}^{X_h} 1 / W_{h\alpha}\right) / X_h$$

and

$$E_{hy} = \left(\sum_{\alpha=1}^{Y_h} 1 / W_{h\alpha}\right) \mid Y_h \quad .$$

Thus, $S^2_{hz'}$ is given by:

$$S^2_{hz'} = R_{hx} (E_{hx} - R_{hx}) + \gamma^2 R_{hy} (E_{hy} - R_{hy})$$
$$- 2\gamma R_{hx} (E_{hx} - R_{hy}) \quad (4)$$

3. Optimum Allocation

Let us suppose that the average cost per enumeration unit in stratum h is given by the equation:

$$c_h = c_1 + c_2 R_{hy} \bar{W}_{hy} \quad (5)$$

where c_1 is the cost component associated with screening a sample enumeration unit and determining whether it is linked to a member of the rare population, c_2 is the cost component assocaited with interviewing the member or members of the rare population linked to the sample enumeration unit, determining whether the individual has attribute A, and determining the total number of enumeration units linked to the individual (i.e., determining the multiplicity, $W_{h\alpha}$, of individual $I_{h\alpha}$), and

$\bar{W}_{hy} = \sum_{\alpha=1}^{Y_h} W_{h\alpha} / Y_h$ is the average multiplicity of elements in stratum h. Then, the total cost, c, for a survey of ℓ_h enumeration units in stratum h, h = 1, ..., H, is given by:

$$c = c_1 \ell_1 + c_2 \ell_2 + \dots + c_H \ell_H \quad (6)$$

Where c_h is given by equation (5) for h = 1, ..., H.

With cost function of the form given by equation (6), it is a well known result [2], that optimum allocation $\tilde{\ell}_h$, of sample to strata at fixed total cost, c, is given by:

$$\tilde{\ell}_h = c\pi_h \tilde{S}_h / \sqrt{c_h} \Big/ \sum_{h=1}^{H} \pi_h \tilde{S}_h \sqrt{c_h} \quad (7)$$

where

$$\tilde{S}_h = S_{hz'} / R_y \quad (8)$$

By substitution of (4), (5), and (8) into equation (7), we can obtain an explicit formula for $\tilde{\ell}_h$, and for a conventional rule, if we assume that the cost component, c_2, is the same as that for multiplicity rule, this explicit formula is given by:

$$\tilde{\ell}_h = \frac{c\pi_h \sqrt{R_{hy}} \sqrt{a_h} / \sqrt{c_1 + c_2 R_{hy}}}{\sum_h \pi_h \sqrt{R_{hy}} \sqrt{a_h} \sqrt{c_1 + c_2 R_{hy}}} \quad (9)$$

where

$$a_h = \gamma_h (1-\gamma_h R_{hy}) + \gamma (1- R_{hy}) (\gamma-2\gamma_h) .$$

If $\gamma_h = \gamma$ for all h, the explicit formula is given by:

$$\tilde{\ell}_h = \frac{c\pi_h \sqrt{R_{hy}} \sqrt{E_{hx} - \gamma E_{hy}} \Big/ \sqrt{c_1 + c_2 \bar{W}_{hy} R_{hy}}}{\sum_h \pi_h \sqrt{R_{hy}} \sqrt{E_{hx} - \gamma E_{hy}} \sqrt{c_1 + c_2 \bar{W}_{hy} R_{hy}}} \quad (10)$$

4. Comparison of Variances Under Different Counting Rules

If we make the simplifying assumptions that $E_{hx} = E_{hy} = E$, that $\bar{W}_{hy} = \bar{W}$ and that $\gamma_h = \gamma$ for h = 1, ..., H, it can be shown that the formula for the approximate variance of γ' (equation (3)) for the optimum allocation, $\tilde{\ell}_h$ at cost c reduces to the form given by:

$$\sigma^2_{\gamma'} = \frac{\gamma (1-\gamma) E}{cR_Y^2} \left[\sum_h \pi_h \sqrt{R_{hy}} \sqrt{c_1 + c_2 \bar{W} R_{hy}}\right]^2 \quad (11).$$

If we assume that the cost component, c_2, remains the same whether a conventional or network sampling rule is used, then the variance of γ' at total cost, c, for a conventional counting rule is given by equation (3) with E and $\bar{W}$ set equal to unity. Thus, the ratio of the variance γ' under network sampling to that under a conventional rule at the same total cost is given by:

$$\frac{E \left[\sum_{h=1}^{H} \pi_h (1 + c_2 \bar{W} R_{hy} / c_1)^{\frac{1}{2}}\right]^2}{\left[\sum_{h=1}^{H} \pi_h (1 + c_2 R_{hy} / c_1)^{\frac{1}{2}}\right]^2} \quad (12)$$

For most values of c_2, c_1, and $\bar{w}$ likely to encountered in practice, expression (12) will be approximately equal to E, and hence, for most applications, network sampling will result in estimates having lower variance at the same total cost than estimates based on a conventional counting rule.

5. Discussion

Although methodology for the use of network enumeration rules in stratified random sampling has been developed previously [4], this is the first work in which optimum allocation has been considered for network sampling. The scope of this work, however, is limited to the estimation of dichotomous attributes (proportions), and the methodology developed here depends on several restrictions which will be reiterated below.

Perhaps the most important restriction on the results developed here is to counting rules in which each enumeration unit is linked to at most one element. If this restriction were not made, algebraic expressions for the variances of estimates under network sampling rules become extremely complex and involved [3], [5]. Whether or not this restriction makes sense depends, of course, on the particular enumeration rule and the particular application. For example, in surveys for the prevalence of diseases which have a genetic component, one might expect a clustering of elements among particular enumeration units in which case, the expressions for optimum allocation developed here would not be valid. In other cases, the results developed here seem safe to use.

The other restriction on enumeration rules imposed here is that an element can be linked to enumeration units only in the same stratum. While this restriction is easy to impose, it could also decrease the yield of elements obtained in a sample and hence increase the sampling variance of the estimator.

For each stratum, the average cost per sample enumeration unit $c_1 + c_2 R_{hy} \hat{\bar{W}}_{hy}$ depends on three factors, namely c_1, the cost of screening an enumeration unit for purposes of determining whether any elements of the rare population unit are linked to it, c_2, the cost of obtaining the desired information concerning a sample element including the multiplicity of the element and $R_{hy} \bar{W}_{hy}$, the average number of elements per enumeration unit for a particular counting rule in a stratum. Although it was assumed here that the cost components, c_1 and c_2 are independent of the particular enumeration rule used, it should be recognized that this might not be a safe assumption to make when the process of determining the multiplicity of an element is complex or difficult and adds considerably to the cost.

Perhaps the most interesting result is that obtained for the ratio of the sampling variance of γ' under a network sampling rule to that under a conventional counting rule in the situation where γ_h is the same for all strata and the multiplicity parameters are the same over all strata (expression (12)). This ratio is equal to E, the inverse harmonic mean of the multiplicities multiplied by a factor which in most situations will be only slightly greater than unity. Thus, the parameter E, would give a rough indication of the extent to which the sampling variance can be reduced by use of a network counting rule over a conventional rule at equivalent field costs.

Finally, it should be mentioned that the important problem of measurement error in determining the multiplicities of elements, which was not treated here should be taken into consideration in choosing a particular counting rule, since errors in the multiplicities could introduce serious biases.

References

1. Berkson, J., "Limitation of the Application of Fourfold Table Analysis to Hospital Data," Biometrics Bulletin, 2, 47-53, 1946.

2. Hansen, Hurwitz, and Madow, "Sample Survey Method and Theory," Volume 1, John Wiley and Sons, New York, 1953.

3. Sirken, M.G., "Household Surveys with Multiplicity," Journal of the American Statistical Association, 65, 257-266, 1970.

4. Sirken, M.G., "Stratified Sample Surveys with Multiplicity," Journal of the American Statistical Association, 67, 224-227, 1972.

5. Sirken, M.G., and Levy, P.S., "Multiplicity Estimation of Proportions Based on Ratios of Random Variables," Journal of the American Statistical Association, 69, 68-73, 1974.

6. Sirken, M.G., Inderfurth, G.P., Burham, C.E., and Danchik, K.M., "Household Sample Survey of Diabetes: Design Effects of Counting Rules," American Statistical Association, Proceedings of the Social Section, 659-663, 1975.

7. Sudman, S., "On Sampling Very Rare Human Populations, Journal of the American Statistical Association, 67, 335-339, 1972.

PREDICTING INDIVIDUAL MIGRATION BEHAVIOR*

Wen Lang Li, Ohio State University

Migration prediction is an important and well-attended topic. The literature in migration studies is overwhelmed by this type of research. But, despite their diverse formulations, the basic of migration prediction can be summarized in the so-called "gravity model" (Shaw, 1975, Ch. 3). It was proposed by Zipf (1946) as follows:

$$M_{ij} = \frac{P_i\ P_j}{D_{ij}}\ K$$

where M_{ij} = net or gross migration between areas i and j
P_i = population in area i
D_{ij} = distance between areas i and j
K = constant term.

This type of prediction model is important because it attempts to estimate the volume of population flows among regions, which are indispensable for the projection of regions' populations. But, this type of prediction method is inadequate because it assumes that we have already known the regions' populations, P_i and P_j. In fact, the regions' population should be predicted, not as predictive variables. Demographers have come to realize that perhaps our understanding of migration behavior is not proper enough for predicting regional populations (Bogue, 1959).

There are some other models which attempt to "explain" migration behavior. The major difference of these models and gravity models is that their focus is on migration behavior, instead of the amount of migration. A migration rate is constructed based on whatever direction of movement: immigration, outmigration, or net migration. The units of analysis are mostly geographic: counties, states, or census tracts. Some socioeconomic characteristics of the areas are also collected. Then a multivariate analysis is performed, which generally is a linear regression, to derive the most acceptable prediction model of migration behavior.

Because of the units of analysis are geographic, the observations derived from the analysis have to be confined in the regional context. The results of migration research are hardly generalizable. Consequently, the majority of migrations is meaningful only in the historical sense. The traditional "push-pull" models have yielded little predictive results except to depict the regional socioeconomic correlates of migration. It would be truly an "ecological fallacy" if these observations are translated into migration predictions (Robinson, 1950).

Although it has been acclaimed as one of the meaningful explanations of migration behavior (Herrick, 1965), the push-pull model is at most a relabelling of the migration process. That people moves itself implies a departure from the area of distress to the areas of attraction. Numerous amounts of studies have endeavored to delineate the complexity of the "pull" and "push" forces, but the model is not able to take into account the behavioral components of the migrants. As Wolpert (1965:161) pointed out:

> Attempts at model building in migration research have largely focused on variables and surrogates such as distance and ecological characteristics of places exerting 'push' and 'pull' forces, to the exclusion of behavioral parameters of the migrants.

Migration as an Individual Behavior

Migration is basically an individual behavior. It is the individual who decides whether to move or to stay. Even in an area of natural disaster many people choose to stay. The importance of considering the decision mechanisms in the migration process is clear. Perhaps the precise mechanisms involved in the process varies from case to case. But, it is possible that certain regularities can be detected and generalized.

Cost-benefit model was proposed as an approach to study the migration decision process (Sjaastad, 1962). It attempts to calculate the monetary and employment payoffs as affecting migration behavior. Nevertheless, the classic utility concepts may not be useful in migration explanations. Many non-monetary forces are important and yet difficult to be directly measured. Obviously, a person's decision to migrate is dictated by a variety of socioeconomic constraints, for example, his stage of life cycle, his employment status, or his social contacts with other communities. Before a cost-benefit model can be fully developed, it is perhaps more proper that the study of migration behavior should begin with the theories of migration differentials.

Unfortunately, the theories of migration differentials are relatively underdeveloped. As aforementioned, model constructions in migration studies are predominantly in aggregate level. One of their distinctive features is to treat a population as homogeneous. The differential aspects of interregional population movement are not well attended in previous research (Li, 1970). The census data from which most generalizations of migration differentials are drawn do not have much behavioral measurements. The aggregate data make it impossible to extend further the theories of migration differentials. Except some scanty attempts (Beshers and Nishiura, 1961; Lee, 1966), many researchers seem to contend that a search for universal generalizations would be fruitless. Bogue (1959) concludes as follows:

> A little reflection convinces one that the search for universal migration differentials not only is doomed to failure but also fails to appreciate the reasons for migration selectivity. (Bogue, 1959: 504).

* This paper has been written with the partial support of the National Institute for Child Health and Human Development, HD-07094-01.

Such pessimistic contention may not only be due to the nature of migration data. The analytical methods employed to study migration differentials may also be a contributing factor. They are generally so primitive that an extensive study of the migration propensity is impeded. Multivariate analysis techniques have been used by previous researchers (Hamilton, 1959; Bogue, et. al., 1953; Tarver, 1961). In most cases the method is multiple linear regression which requires stringent assumptions and particular units of analysis. Yet migration is basically a binary variable: to move or not to move. Although the analytical techniques for this type of problems have been proposed in 1930's (Fisher, 1937), it is only in recent years that the method has received extensive attention (Cox, 1970).

The purpose of this paper is to introduce a multivariate analysis technique to migration studies. The technique is useful in analyzing migration behavior in the micro (individual) level. Its application will predict individual as well as aggregate propensity to migrate. The subsequent portion of this paper is divided into three parts: first, the structure of the model; next, the nature of the data and selected predictive variables; and finally, the research findings and conclusions.

The Model

The first step of the model is to use the linear discriminant function as proposed by R. A. Fisher (1936). We are given two groups of persons: migrants and nonmigrants. Every person is measured by k numbers of socioeconomic characteristics, or variables. It is assumed that, in the population from which the groups are drawn, the characteristics have a common multivariate normal distribution. The discriminant coefficients (which is analogous to regression coefficients) can be computed by:

$$b = V^{-1} d \qquad (2)$$

where V^{-1} is the inverse of the pooled covariance matrix and d the vector of differences between the pairs of means of the two groups. The constant term of the discriminant function is estimated by:

$$a = -\tfrac{1}{2}\sum_{i}^{k} b_i\,(\bar{x}_{i0} + \bar{x}_{i1}) - \ln(n_0/n_1) \qquad (3)$$

where 0 and 1 denote respectively migrant and non-migrant groups; $\bar{x}_i$ the ith variable's mean; and n the number of persons in each group. In other words, the constant term is a function of discriminant coefficients, the average values of the variables, and the proportion of migrants.

Based on the assumptions of homoscedasticity and multivariate normality, it is possible to show that a multiple logistic function can be derived from the linear discriminant function (Cornfield, 1967). And the predicted outcome corresponds to the probability of migration, such as:

$$p = 1/(1+ e^{-(a+\sum b_i x_i)}) \qquad (4)$$

where e is the base of natural logarithms. The multiple logistic function can be used for screening highly mobile individuals in the general population.

But the assumptions involved in this formulation are too stringent. Migrants' or nonmigrants' characteristics are rarely normally distributed, nor their covariance matrices are identical. To avoid such restrictions Walker and Duncan (1967) suggest a maximum likelihood approach. For a group of n individuals, the likelihood function of migration is:

$$L = \prod_{j}^{n} p_j^{\,y_j}\,(1-p_j)^{1-y_j} \qquad (5)$$

where y_j equals to one or zero depending on whether the jth individual is a migrant or not; and $\prod$ is the product sign. The function is to be maximized. Newton-Ralphson procedure is chosen. The coefficients from the linear discriminant function are used as initial values. Through successive approximation some refined estimates of the coefficients are obtained. Tests on the goodness of fit, such as maximum likelihood ratio, can determine whether the model outcomes are agreeable with the empirical data.

Data and Variables

The public use sample data of the 1970 census are used to test the model. From one-in-10,000 U.S. population, individual's information on socioeconomic activities in both 1970 and five years earlier in 1965 was collected. A question was asked about the place of residence in 1965. Together, these data make it possible to identify the migration and socioeconomic characteristics of every individual. Those whose states of residence differ between 1965 and 1970 are grouped as migrants; otherwise, nonmigrants. Of the total 20,196 persons in the sample, our analysis includes only the household's chief income recipients. Although they are not strictly the household heads, the chief income recipients are perhaps the major decision-makers of migration behavior. Other household members are excluded. Thus, the sample size for this study is reduced to 7,124 persons; roughly one person for one household in the U.S.

Three predictive variables of migration behavior are selected as predicated by the nature of the data. These variables were also shown to be highly related to migration in previous studies. The first variables is age, which is a clear determinant of migration behavior. Age is almost a perfect indicator for the stage of life cycle. It is also closely related to the duration of residence. Both concepts have been used as the predominant explanations for why people move (Rossi, 1955; Morrison, 1967). Thomas' classic report on migration differentials shows a definitively established generalization: young adults are more mobile than older persons (Thomas, 1938). As presented in Figure 1 our data clearly confirm this generalization. Migration behavior is indeed a decreasing function of the aging process.

Years of schooling are used as the second predictive variable of migration behavior. Many previous studies have observed that migration is positively related to education. The number of years one spends in schooling undoubtedly expand his information contacts with outside world beyond

Figure 1: Migration Rate by Age

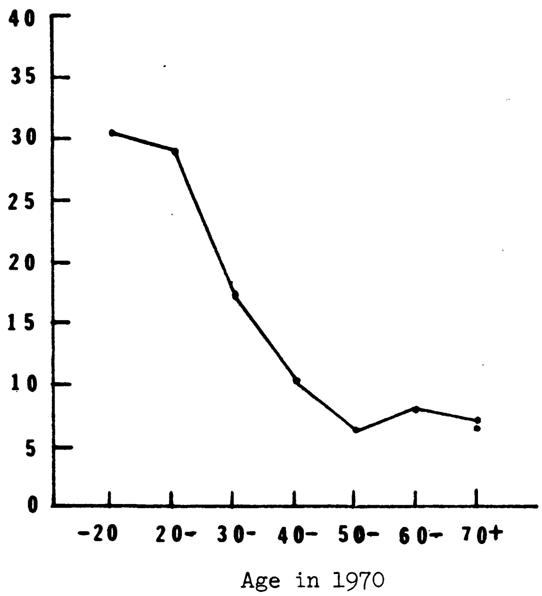
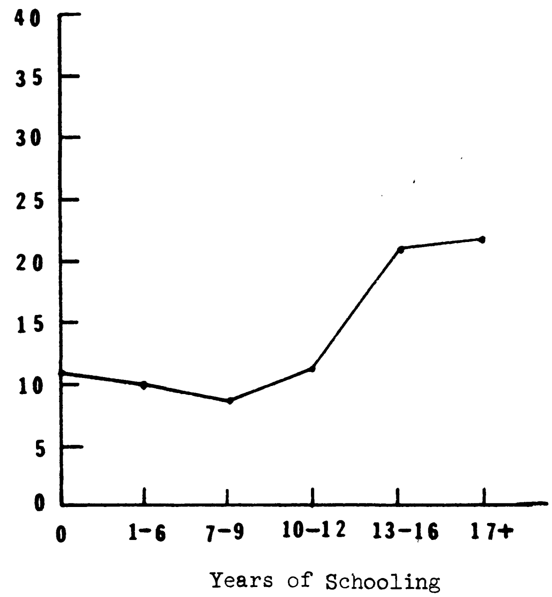


Figure 2: Migration Rate by Education



his locality. Wolpert (1965) has called our attention to the importance of "information fields" in regulating people's search behavior. In addition, Bogue et. al. (1957) conclude that the two factors that seem to contribute most to the mobility of the population are above average educational training and employment in white collar occupations. Again, our data clearly support this observation. Figure 2 shows that migration behavior is determined by the number of years of schooling.

The last predictive variable is employment status before migration occurs. The 1970 census provides the possibility of testing this relationship. A question was asked to every sample individual about his economic activity in 1965. As shown in Table 1, the results from the census data strongly suggest that migration and employment status are related as observed in many previous studies. Both Lansing and Mueller (1967) and Li (1976) have found that unemployed persons are much more likely to migrate than are employed persons, and that persons in schooling or in military services are perhaps the most mobile of all.

Empirical Results

The three predictive variables of migration behavior are denoted respectively by x_1, x_2, and x_3. Both age and educational attainment use

Table 1: Household Chiefs by Economic Activity and Migration Status: U. S., 1970

Economic Activity	Migrants	Nonmigrants	Percent Migrant
Employed	401	4,146	8.8
Unemployed	412	1,636	20.1
Services & Schooling	220	309	41.6
Total	1,033	7,124	35.2

straightforward measurements. Employment status is measured by: 0 for the employed, 1 for the unemployed, and 2 for being in colleges or military services. Premilinary results of the linear discriminant function, as described in Equation (2) and (3), are shown in Table 2. The constant term has a value of -1.691, which serves as a reference point for the comparison of discriminant scores among the sample individuals. The discriminant coefficients for the three predictive variables are -0.031, 0.046, and 1.176 respectively. The signs of the coefficients are consistent with what have been observed in the literature. Following the procedure described in Equation (5), a more refined version of the model is obtained. Compared to the results of the preliminary version, the values of the constant term and the discriminant coefficients decrease slightly The value of a is -1.436, whereas the coefficients for x_1, x_2, and x_3 respectively are -0.029, 0.038, and 0.803.

To assess the statistical significance of the discriminant coefficients, the following formulation is used to compute the standard errors of the estimates:

$$s = w \cdot m \qquad (6)$$

where w is the diagonal vector of V^{-1}, and m is $(1/n_0+1/n_1)$. The results of s vector are also shown in Table 2. Then t-scores are computed and used to determine the significance of the coefficients. It is noted that the three predictive variables are all statistically significant at 5 percent level. In other words, the results are quite consistent with the theories which were previously presented.

Table 2: Results of Model Implementation

Variable	b Coefficients	Standard error	t Score
Preliminary Version			
x_1	-0.031	0.002	-15.4
x_2	0.046	0.010	4.6
x_3	1.176	0.055	21.2
Constant a	-1.691		
Refined Version			
x_1	-0.029	0.002	-12.8
x_2	0.038	0.011	3.4
x_3	0.803	0.050	16.0
Constant a	-1.436		

The refined version of the model can be used to predict an individual's migration probability. Assume that a person is 25 years of age, college-graduated and employed, the model predicts that his migration probability would be 0.173, or 17.5 percent. If he happens to be unemployed, then the probability increases drastically to 31.9 percent. And, if he is in military services, his chance to move across state boundries is 48.9 percent.

Table 3 presents the simulated migration probabilities for a person who is 25 years of age in the U. S. The probabilities are calculated with various assumptions about his educational attainment and employment status. A general pattern emerges. It is noted that an increase of his educational level will result in an increase of roughly 2 percent points in migration probability. For example, among the employed the migration probability is 0.153 for a high-school graduate and 0.173 for college graduate. The difference is exactly 0.02. On the other hand, a change of the employment status, from being employed to unemployed, will almost double his migration probability; say, for an uneducated to change from 0.103 to 0.204. The results appear to indicate that employment status weights much more than educational attainment in determining migration behavior.

Table 3: Simulated Migration Probability for a Person at Age 25.

Educational Attainment	Employment status Employed	Unemployed	Schooling or Services
None	0.103	0.204	0.364
6th Grade	0.126	0.243	0.417
9th Grade	0.139	0.264	0.445
12th Grade	0.153	0.287	0.473
College Graduate	0.173	0.319	0.489

The predicted probability can be used as an instrument to screen the potential migrants. Note that the overall rate of migration is 14.5 percent as estimated from our sample. In other words, about 15 out of every 100 Americans are expected to migrate across state boundaries during 5-year period. Undoubtedly, any individuals whose predicted migration probability is higher than this figure would be judged as potentially active migrants. The average adult American is not a potential active migrant. Statistically speaking, he is about 46 years of age, received 11 years of schooling, and currently employed. His predicted probability of migration is only 8.7 percent, which is far less than 14.5 percent as expected. Nevertheless, if he happens to be unemployed, his migration probability would nearly be doubled. It increases to 17.5 percent. In this case, he would be a potentially active migrant.

An important use of the model is for regional population estimation and projection. As every individual's migration behavior can be predicted, so is the total population in an area. It is a simple case of summation. If a person's migration probability is 0.3, whereas another person is 0.7, then it is expected that one out of these two persons will be a migrant. Table 4 shows that through such aggregation the model can yield fairly reliable estimation of migrants for a region. As a case of illustration, the division of East South Central has 60 migrants in the sample; our model estimates it has 62. The migration rate is almost identical between the observed and the predicted.

Table 4: Observed and Predicted Migrants and Migration Rates by Division, 1970

Division	Number of Migrants Observed	Number of Migrants Predicted	Migration Rate (%) Observed	Migration Rate (%) Predicted
New England	71	65	16.4	15.1
Middle Atlantic	141	179	10.8	13.7
East North Central	151	188	10.9	13.7
West North Central	62	78	11.0	13.9
South Atlantic	199	159	18.6	14.9
East South Central	60	62	14.1	14.5
West South Central	94	96	14.4	14.8
Mountain	80	48	26.2	15.8
Pacific	175	157	17.7	16.0
Total	1,033	1,033	14.5	14.5

It is not only that a region's amount of migrants can be estimated, but also the demographic characteristics of the migrants can be ascertained. Given enough sample size, the model can yield an estimation of migrants by age, sex, and race for each region. This type of information is urgently needed in the projection of a region's population. As shown in Table 5, migrants are estimated by age for each division. Migration rates are then computed. The predicted migration rates are compared with the observed migration rates. In many cases the goodness of fit seems quite acceptable.

Some Cautious Remarks

A micro-predictive model as proposed in this paper has its merits as well as demerits. Although we have presented some pleasing results, they are by no means overwhelmingly satisfactory. The success of a prediction model depends on many factors. The most important one is perhaps the choice of "right" predictive variables. Obviously, the choice should be based on both theoretical and methodological considerations. Kendall (1966) has suggested a so-called "ratio-statistic method," which is unfortunately not usable because of the nature of our data. What this paper relies on is strictly theoretical considerations. It is therefore suggested that more behavioral variables need to be taken into account. And more sophisticate methods of variable-selection should be explored.

Though we have attempted to aggregate individual migration probabilities into a region's migration estimation, the major thrust of this paper is to present a micro model. No serious attempt has been made to relate the behavioral prediction with locational characteristics. As Lee (1966) correctly pointed out, a proper model of migration should simultaneously consider four dimensions: characteristics of origin, of destination, of intervening obstacles, and of migrants. This paper chooses to expand upon Lee's paradign through investigating first the migrants' characteristics as determinants of the migration decision. One may argue that the decision to move is a direct reflection of the locational characteristics. However, both the micro and the macro aspects of migration are indeed difficult to be separated. It is imperative that a behavioral model as presented in this paper should be integrated with a location model.

References

Beshers, James M. and Eleanor N. Nishiura
1961 "A Theory of Internal Migration Differentials," Social Forces 39: 214-218.

Bogue, Donald J.
1959 "Internal Migration," Pp. 486-505 in P. M. Hauser and O. D. Duncan (eds.) The Study of Population. Chicago, Ill. University of Chicago Press.

Bogue, Donald J., Henry S. Shryock, and Siegfried A. Hoermann
1957 "Subregional Migration in the United States, 1935-40," Vol. 1, Streams of Migration Between Subregions. Oxford, Ohio: Scripp Foundation, Miami University.

Cox, D. R.
1970 The Analysis of Binary Data. London: Methuen.

Cornfield, Jerome
1967 "Discriminant Functions," Review of the International Statistical Institute 35 (2): 142-153.

Fisher, R. A.
1936 "The Use of Multiple Measurements in Taxonomic Problems," Annals of Eugenics 7: 179-188.

Hamilton, C. Horace
1951 "Population Pressure and Other Factors Affecting Net Rural-Urban Migration," Social Forces 30 (December): 209-215.

Herrick, Bruce H.
1965 Urban Migration and Economic Development in Chile. Cambridge, Mass.: MIT Press.

Kendall, M. G.
1966 "Discrimination and Classification," Pp. 165-185 in P. R. Krishnaiah (ed.) Multivariate Analysis, N. Y.: Academic Press.

Lee, Everett S.
1966 "A Theory of Migration," Demography 3: 47-57.

Li, Wen Lang
1970 "A Differential Approach to the Study of Migration Streams," Rural Sociology 35 (December): 534-538.

1976 "A Note on Migration and Employment," Demography 13 (November): forthcoming.

Morrison, Peter A.
1967 "Duration of Residence and Prospective Migration: The Evaluation of a Stochastic Model," Demography 4: 553-561.

Robinson, W. S.
1950 "Ecological Correlations and the Behavior of Individuals," American Sociological Review 25: 352-357.

Rossi, Peter H.
1955 Why Families Move. Glencoe, Ill.: The Free Press.

Shaw, R. Paul
1975 Migration Theory and Fact. Philadelphia, Pa.: Regional Science Research Institute. Bibliography Series, No. 5.

Sjaastad, L. A.
1962 "The Costs and Aeturns of Human Migration," Journal of Political Economy 70 (Part II): 80-93.

Taeuber, Karl E.
1964 "White Migration and Socio-Economic Differences Between Whites and Suburbs," American Sociological Review 29(October): 718-729.

Tarver, James D.
1961 "Predicting Migration," Social Forces 39(March): 207-213.

Thomas, Dorothy S.
1938 Research Memorandum on Migration Differentials. New York: Social Science Research Council.

Walker, S. H. and D. B. Duncan
1967 "Estimation of the Probability of an Event as a Function of Several Independent Variables," Biometrika, 54(1): 167-179.

Zipf, G. K.
1947 "The P_1P_2/D Hypothesis: On the Inter-City Movement of Persons," American Sociological Review 11: 677-686.

THE ANALYSIS OF CYCLIC POINT PROCESSES WITH WALSH FUNCTIONS*

Walter S. Liggett, Jr., Tennessee Valley Authority

Transformation by Walsh functions is a technique useful for uncovering the interactions among the daily, weekly, and seasonal cycles often present in societal data. The structural and sampling properties needed for interpretation are simple, the former because Walsh functions are easy to visualize, the latter because the normal approximation applies whenever the total number of events is large. This paper demonstrates these properties including their extension multivariate point processes. This technique is illustrated by applying it to a series of robberies.

1. INTRODUCTION

For many series of events, the rate of occurrence is modulated by periodic phenomena of known frequency and also non-periodic phenomena. Robberies are one illustration. Although the frequencies are known, the effects on the rate of occurrence of the various phenomena are neither additive nor multiplicative but unknown. In other words, the rate function changes from period to period under influences of longer-period phenomena and non-periodic phenomena, influences that interact in an unknown way. For example, the weekday rate function for robberies differs from the weekend rate function, the winter from the summer. That the rate of occurrence is small also motivates the analysis to be discussed.

The data being considered can be regarded as a point process with interesting non-stationary characteristics. Lewis [8] illustrates point-process methods with admissions to a hospital emergency room. Some point process methods [3, 7, 8] are appropriate, but others [2, 4], meant for stationary point processes, are not. The data can also be regarded as a sparse contingency table with interesting interactions. For example, the table might have as dimensions an appropriate division of the day, day of the week, and week of the year. Thus, contingency table methods [1, 6] are appropriate.

Properly applied, transformation by Walsh functions shows the periodicities, the interactions, and how rapidly the rate function changes within periods. As discussed below, the Walsh transform must be matched to the frequencies present. Some Walsh coefficients indicate periodicities because they are averages over all periods of a function of the counts within a period. Other Walsh coefficients indicate interactions because they are differences between periods of a function of the counts within a period. Further, Walsh coefficients indicate behavior of the rate function within periods because they are based on successively finer divisions of the time scale.

* This research was supported by the Office of Naval Research under Contract N00014-74-C-0432 (NR 042-327) while the author was affiliated with the SIAM Institute for Mathematics and Society and the New York City-Rand Institute.

Thus, the Walsh transform gives at once coefficients that contraindicate smoothing by combining cells (using a courser division of the interval) and coefficients that contraindicate smoothing by ignoring interactions. In the analysis of sparse tables, comparison of these types of smoothing is important.

One further type of smoothing should be considered, combining events that are somewhat different. Extension of the Walsh-transform technique to this case is indicated.

Because Walsh functions have a range of ± 1, each Walsh coefficient is the difference between numbers of events in a two-way partition of the total interval. Consequently, the easiest interpretation of the Walsh coefficients is in terms of an additive model. When compared to multiplicative models, additive models have disadvantages [1]. However, because the coefficients are differences of counts, the normal approximation applies regardless of the cell size. Thus, the distribution theory for Walsh coefficients is simple.

The proper application of the Walsh transform to most data requires adding to the series periods that contain no events. For example, an eighth day must be added to the week and twelve weeks must be added to the year. These additions are needed because the Walsh transform only matches frequencies that are a power of two times some lowest frequency. Because of these additions, some of the Walsh coefficients must be adjusted so they can be compared to zero.

The Walsh transform is like analysis of variance with orthogonal contrasts. These techniques are identical if the ratios of the cycle lengths are a power of two, as required for proper application of the Walsh transform. Alternatively, orthogonal contrasts could have been used without adding zeros. This latter approach does not give uncorrelated coefficients as is usually the case in analysis of variance because the process is nonhomogeneous. Further, the results of the latter approach are not as easy to interpret because the resulting coefficients are not simple differences of counts.

The Walsh transform has attracted attention in signal processing because it is easily computed, requiring only $n \log_2 n$ additions for an n-point transform [9, 10, 11].

2. A SERIES OF ROBBERIES

The data that motivated this paper are descriptions of robberies (the taking of something from a person by force or threat of force) and purse snatchings that occurred in the Bronx during 1969 and 1970. These descriptions were recorded on a special form by the New York City Police. The recording was ordered and carefully supervised by the commander of the Bronx, who used the data for operational purposes. Beside time and place of occurrence, these descriptions include characteristics of the victims, perpetrators, and circumstances of the crime. Many series of events recorded by institutions have similar statistical properties.

For these data, the shapes of the daily, weekly, and seasonal patterns, as well as how they vary with event characteristics, are of interest. However, the intent of this paper is not to present a picture of robbery but only to illustrate the Walsh transform. The incidents used for illustration are single-victim, non-commercial incidents in which the victims were males between the ages of 18 and 52. The interval examined is the 52 week period from Sunday, July 27, 1969 to Saturday, July 25, 1970. This interval is chosen so that the first and fourth quarters are in daylight saving time and the second and third quarters are in standard time.

In this application of the Walsh transform, the day is divided into eight subintervals: 1:01 a.m. to 9:00 a.m., 9:01 a.m. to 1:00 p.m., 1:01 p.m. to 3:15 p.m., 3:16 p.m. to 5:20 p.m., 5:21 p.m. to 6:45 p.m., 6:46 p.m. to 8:30 p.m., 8:31 p.m. to 10:15 p.m., and 10:16 p.m. to 1:00 a.m. The unequal subintervals are justified because the shortest period expected in robbery data is one day. Thus, the time scale for days can be replaced by any monotone function of time. The eight subintervals are chosen to equalize the number of single-victim non-commercial incidents in each subinterval. Because the data contain ties due to rounding of event times, this cannot be done exactly. Since the equalization is for victims of all sexes and ages, equalization for a particular victim sex-age category such as adult males is not expected.

To meet the requirement that the ratios of the periods must be a power of two, a day with zero events is added to each week and three weeks with zero events are added to each quarter. The zero-event day follows Sunday, becoming the second day of the week. The three zero-event weeks follow the last week of the quarter. Including the divisions of the day, this produces 4096 (2^{12}) subintervals. The Walsh transform is applied to the counts in these subintervals.

Because of the added zero-event sub-intervals, some comparisons computed by the Walsh transform are between subintervals of unequal length. The Walsh coefficients that compare weeks within a quarter compare either eight weeks to five weeks or seven weeks to six weeks. Thus, they will not be zero when the weeks of the quarter are the same. Similarly, the Walsh coefficients that compare days within a week compare four days to three days. Thus, they will not be zero when the seven days of the week are all the same. For example, the Walsh coefficient that is the difference between the number of events on Sunday, Friday, and Saturday and the number on Monday, Tuesday, Wednesday, and Thursday does not account for the unequal number of days in the two groups.

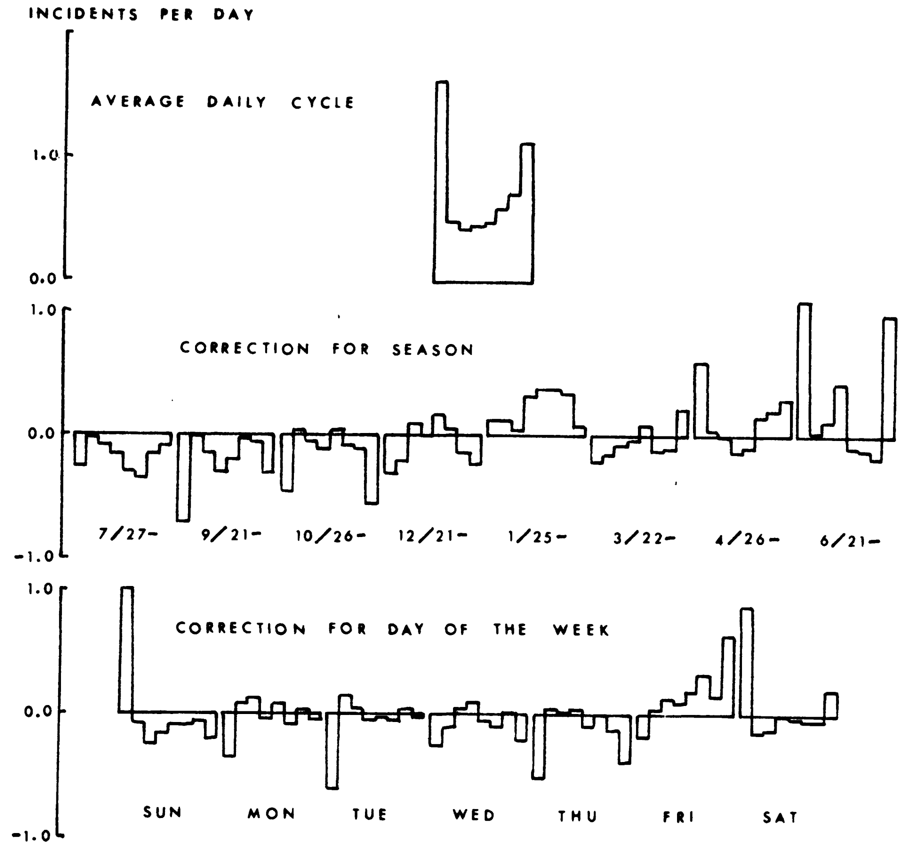
The proper adjustment is normalization by the lengths of the subintervals compared. Walsh coefficients can be obtained in four steps: by evaluating for each day a function of the occurrence times in that day, by evaluating for each week a function of the daily values obtained for that week, by evaluating for each quarter a function of the weekly values obtained for that quarter, and by evaluating for the year a function of the quarterly values obtained. Each of these functions is either a sum of the values for the next shortest period or a difference of two sums of the values. Adjustment for the added zero-event subintervals is needed when a difference of sums of daily or weekly values is involved. The proper adjustment is division of each sum by the number of actual days or weeks it contains. In the above example, the number of events occurring on Sunday, Friday, and Saturday should be divided by three and the number in the second group by four. This is equivalent to creating an adjusted coefficient by replacing the zero-event subintervals with average days or average weeks and then transforming. This second procedure is identical to the procedure actually followed.

The Walsh transform applied to the robbery data is the one described by Manz [9]. The 42 coefficients that have the largest adjusted values are given in Table 1. The coefficient numbers (which start with zero) are given in both decimal and binary form. Next, the Walsh coefficients obtained with the added zero-event subintervals are given. Finally, the adjusted

Table 1. The Largest Coefficients in the Robbery Data.

Coefficient Number Decimal	Binary	Walsh Coefficient	Adjusted Coefficient
0	0	2135	-
1	1	-447	-
4	100	157	178
11	1 011	173	152
128	10 000 000	-81	224
255	11 111 111	-455	-150
639	1 001 111 111	161	156
876	1 101 101 100	-163	-153
1024	10 000 000 000	729	-
1025	10 000 000 001	-169	-
1026	10 000 000 010	245	-
1152	10 010 000 000	133	237
1160	10 010 001 000	-199	-183
1280	10 100 000 000	249	145
1284	10 100 000 100	135	152
1440	10 110 100 000	-129	-140
1663	11 001 111 111	195	171
1919	11 101 111 111	127	151
1961	11 110 101 001	-139	-144
2014	11 111 011 110	143	145
2047	11 111 111 111	167	-
2048	100 000 000 000	529	-
2049	100 000 000 001	-225	-
2052	100 000 000 100	123	142
2176	100 010 000 000	193	269
2508	100 111 001 100	185	174
2626	101 001 000 010	167	168
2680	101 001 111 000	-89	-140
2687	101 001 111 111	187	141
2695	101 010 000 111	169	157
2696	101 010 001 000	153	155
2815	101 011 111 111	195	149
2852	101 100 100 100	157	145
2943	101 101 111 111	115	161
3071	101 111 111 111	319	-
3072	110 000 000 000	591	-
3073	110 000 000 001	-175	-
3074	110 000 000 010	175	-
3200	110 010 000 000	171	255
3563	110 111 101 011	-143	-145
3931	111 101 011 011	-139	-147
4095	111 111 111 111	217	-

Figure 1. The Weekly and Seasonal Variation in the Daily Pattern.



coefficients are given for those Walsh coefficients that compare subintervals of unequal length. Coefficient 0 is not a difference but the total number of incidents.

Each Walsh coefficient (except 0) is the difference between two counts that can be considered independent Poisson random variables. Thus, the variance of each Walsh coefficient is estimated by the total count, 2135. Since the two counts are large enough for the normal approximation to apply, three standard deviations (which equals 139) has its usual meaning as a standard of comparison. Estimates of the variance of the adjusted coefficients vary from coefficient to coefficient. However, for the purpose of choosing the significant coefficients, the variances of the adjusted coefficients are nearly the same as those of the Walsh coefficients.

Consider the meaning of the coefficients in Table 1. Beside coefficient 0, the largest is coefficient 1024. This coefficient is the number of events that occurred in the first two and the last two subintervals of the day minus the number in the middle four subintervals. It shows that more incidents occur between 8:31 p.m. and 1:00 p.m. the next day than in the other part of the day. Since the subintervals are chosen to equalize the rate of occurrence, it follows that the reverse is true for victims in other sex-age categories. The other coefficients in Table 1 that are sums over all days of differences between parts of the day are coefficients 2047, 2048, 3071, 3072, and 4095.

The coefficients in Table 1 that show the variation of the daily pattern from quarter to quarter are coefficients 1, 1025, 1026, 2049, 3073, and 3074. Coefficient 1 compares the first half of the total interval (which ends January 24, 1970) with the second half. It shows an upward trend. The next largest in this group is coefficient 1026. It is a comparison between the middle half of the year and the first and last quarters. The quantities compared are sums over the two parts of the year of a function of the daily pattern. This function is the first two and last two subintervals of the day minus the middle four subintervals. It shows that the predominance of attacks between 8:31 p.m. and 1:00 p.m. the next day is more pronounced during daylight saving time.

The coefficients in Table 1 that show the variation of the daily pattern within the week are 128, 255, 639, 1152, 1280, 1663, 1919, 2176, 2687, 2815, 2943, and 3200. These coefficients are affected by the added zero-event subintervals. Coefficient 128 is the number of events on Sundays, Fridays, and Saturdays minus the number on Mondays, Tuesdays, Wednesdays, and Thursdays. This difference has a value of -81. Since each day of the week has an average of 305 incidents, the adjusted coefficient is 224. It shows that on a per day basis weekends have more events than weekdays.

Retransforming some of the Walsh coefficients provides an informative display of the data. The first graph in Figure 1 shows that the first and last subintervals of the day predominate. As noted above, coefficient 1024 reflects this. The second and third graphs are additive corrections to the first. The second shows the correction to the daily pattern for eight divisions of the year. As shown by coefficients 1025 and 1026, the large percentage of incidents in the first and last periods of the day is more pronounced in the last quarter of the year. The upward trend is also evident. The third graph shows the corrections by day of week. As would have been noted if coefficient 1152 had been discussed, the predominance of the first and last periods of the day is more pronounced on weekends.

Figure 1 does not display all the coefficients in Table 1. Some, such as coefficient 11, may indicate the need for finer graduation of the season. Others, such as coefficients 1284, 2626, 2680, and 2695, may indicate that the weekly pattern varies with season. The others may have a societal explanation that has not yet been uncovered. On the other hand, they may be large only by chance.

3. APPLYING THE WALSH TRANSFORM

To apply the Walsh transform, simple methods for interpreting the coefficients are needed. This section provides such methods, both an algorithm for constructing any Walsh function and the distributional properties of the coefficients. Next, application of the Walsh transform to multivariate point processes is discussed. Finally, the Walsh transform is compared, in terms of simplicity of interpretation, to the log-linear analysis of contingency tables and to spectral analysis.

For visualizing Walsh functions, defining them by specifying their sign changes is convenient. This is possible because the range of Walsh functions is ± 1. Denote the kth value of the ith Walsh function by $w(i,k)$, $0 \le i, k \le 2^M-1$. The construction can be thought of as starting at $k = 0$ and proceeding to larger values of k. The value of $w(i,0)$ is +1. Let the binary representation of i be $i_{M-1}\, i_{M-2} \ldots i_0$ $(i = \sum i_p 2^p)$, and express k as $j2^{M-1-p}$ for some odd integer j. The ith Walsh function changes sign between k-1 and k if and only if $i_p = 1$. Thus, i_0 determines whether the sign changes between the first and last half of the domain; i_1 determines whether the sign changes between the first and second quarters and between the third and fourth quarters; etc.

This definition gives the Walsh functions in an order different from the usual one [10, 11]. The ith function has i sign changes. Thus, the Walsh functions are ordered by their sequency [11], that is, by the number of sign changes they have. A fast algorithm that applies the Walsh functions in sequency order is given by Manz [9]. Note that the matrix with elements $w(i,k)$ is a symmetric Hadamard matrix. If it is multiplied by $2^{-M/2}$, it is orthogonal.

Separate sets of binary digits specify divisions within and between the periods present. For the robbery data, the first three digits $i_{11}\, i_{10}\, i_9$ specify divisions within the day; the second three digits $i_8\, i_7\, i_6$ specify divisions between days within the week; the next four digits $i_5\, i_4\, i_3\, i_2$ specify divisions between weeks within the quarter; and the last two digits $i_1\, i_0$ specify divisions between quarters. The coefficients corresponding to the margins (those that could have been obtained by superimposing the data in successive periods of one of the cycles present) have digits with the following property. The digits in the set that specifies the divisions of the period, say $i_{p+q} \ldots i_p$, are followed by digits equal to the last in the set, $i_p = i_{p-1} = \ldots$.

A 2^M-point Walsh transform can be thought of as a $2^{M'}$-point transform of a $2^{M''}$-point transform, where $M' + M'' = M$. This property is the basis for the fast algorithm. It is also useful in the interpretation of the Walsh coefficients as comparisons. Let $n(k_d,k_w,k_q,k_y)$ be the number of events in subinterval k_d of the day, day k_w of the week, week k_q of the quarter, and quarter k_y of the year. Let $w^{(M)}(p,p')$ $(0 \le p,p' \le 2^M-1)$ be the values of the 2^M-point Walsh functions. The following four steps produce Walsh coefficient i, where $i = i_d 2^9 + i_w 2^6 + i_q 2^2 + i_y$:

$$f_d(i_d,k_w,k_q,k_y) = \textstyle\sum_k w^{(3)}(i_d,k)\; n(k,k_w,k_q,k_y)\ , \tag{3.1}$$

$$f_w(i_d,i_w,k_q,k_y) = \textstyle\sum_k w^{(3)}(i_w',k)\; f_d(i_d,k,k_q,k_y),$$

$$i_w = \begin{cases} i_w' & \text{if } i_d \text{ even} \\ 7-i_w' & \text{if } i_d \text{ odd,} \end{cases} \tag{3.2}$$

$$f_q(i_d,i_w,i_q,k_y) = \textstyle\sum_k w^{(4)}(i_q',k)\; f_w(i_d,i_w,k,k_y),$$

$$i_q = \begin{cases} i_q' & \text{if } i_w \text{ even} \\ 15-i_q' & \text{if } i_w \text{ odd,} \end{cases} \tag{3.3}$$

$$f(i_d,i_w,i_q,i_y) = \textstyle\sum_k w^{(2)}(i_y',k)\; f_q(i_d,i_w,i_q,k),$$

$$i_y = \begin{cases} i_y' & \text{if } i_q \text{ even} \\ 3-i_y' & \text{if } i_q \text{ odd.} \end{cases} \tag{3.4}$$

If each of these equations were further decomposed into 2-point transforms, the result would be Manz's algorithm [9].

The adjustment for the added zero-event subintervals can now be seen. Consider (3.2) first. No adjustment is needed if f_w is the sum of f_d (the case $i_w' = 0$). Otherwise, the proper comparison is given by

$$g_w(i_d,i_w,k_q,k_y) = f_w(i_d,i_w,k_q k_y) + (w^{(3)}(i_w',1)/7)\ \textstyle\sum_k f_d(i_d,k,k_q,k_y). \tag{3.5}$$

This is equivalent to normalizing by the number of days in the subintervals being compared.

Since the adjustment for the added weeks is similar, the equations for the adjusted coefficients equivalent to (3.1) - (3.4) are easily obtained. Let

$$a(i_w') = \begin{cases} [w^{(3)}(i_w',1)]/7 & \text{if } i_w' \neq 0 \\ 0 & \text{if } i_w' = 0 \end{cases} \quad (3.6)$$

$$b(i_q') = \begin{cases} [w^{(4)}(i_q',13) + w^{(4)}(i_q',14) \\ \quad + w^{(4)}(i_q',15)]/13 & \text{if } i_q' \neq 0 \\ 0 & \text{if } i_q' = 0. \end{cases} \quad (3.7)$$

Equation (3.1) remains the same. Equations (3.2) - (3.4) become

$$g_w(i_d,i_w,k_q,k_y) = \sum_k[w^{(3)}(i_w',k) + a(i_w')]\, f_d(i_d,k,k_q,k_y), \qquad (3.8)$$

$$g_q(i_d,i_w,i_q,k_y) = \sum_k[w^{(4)}(i_q',k) + b(i_q')]\, g_w(i_d,i_w,k,k_y), \qquad (3.9)$$

$$g\ (i_d,i_w,i_q,i_y) = \sum_k w^{(2)}(i_y',k) g_q(i_d,i_w,i_q,k), \qquad (3.10)$$

where the relation between (i_d,i_w,i_q,i_y) and (i_d,i_w',i_q',i_y') is given in (3.2) - (3.4). The adjusted coefficients can be computed from the Walsh coefficients.

In order to derive distributional properties for the Walsh coefficients, the series of events is modeled as a nonhomogeneous Poisson process. For the type of series being discussed, this assumption seems reasonable. The inhomogeneity is the most pronounced feature of such series, and the nonhomogeneous Poisson process is the simplest model with this feature. Note that adding zero-event subintervals does not make this model invalid.

The distribution theory for Walsh coefficients is simple as long as the number of coefficients considered simultaneously is small enough that the normal approximation applies. Approximate normality after transformation is a familiar property, occurring, for example, in ordinary time series analysis. Let

$$k = k_d 2^9 + k_w 2^6 + k_q 2^2 + k_y,$$

$$En(k_d,k_w,k_q,k_y) = \mu(k_d,k_w,k_q,k_y) = \mu_k,$$

$$f_i = f\ (i_d,i_w,i_q,i_y). \qquad (3.11)$$

The variance of the Walsh coefficients is

$$\text{var } f_i = \sum_k \mu_k \qquad (3.12)$$

Since $w(i',k)\ w(i'',k) = w(i,k)$, where i is the dyadic product of i' and i" (the binary digit i_p equals $i_p' + i_p''$ (mod 2)), the covariance of coefficients i' and i" is

$$\text{cov } (f_{i'},f_{i''}) = \sum_k w(i,k)\mu_k. \quad (3.13)$$

This can be estimated by the ith Walsh coefficient.

Consider the statistical problem of separating the adjusted coefficients that are not zero from those that are. Both because of the nature of the data and because of the adjustments for the added subintervals, conditional tests are appropriate. One approach is as follows: Tests for the coefficients that involve neither comparisons among days within a week ($i_w' = 0$) nor comparisons among weeks within a quarter ($i_q' = 0$) are conditioned only by the total number of events, $n(+,+,+,+)$. Tests for coefficients for which $i_w' = 0$ and $i_q' \neq 0$ and those for which $i_w' \neq 0$ and $i_q' = 0$ are conditioned on $n(k_d,+,+,k_y)$ $(= \sum_{j,k} n(k_d,j,k,k_y))$. Tests for the coefficients for which $i_w', \neq 0$ and $i_q' \neq 0$ are conditioned on $n(k_d,k_w,+,k_y)$. Even with the conditioning, the adjusted coefficients are approximately normal. Thus, given the variances, thresholds high enough to prevent many zero coefficients from being classified non-zero can be set using the Bonferroni inequality [6]. It can be shown that the variances of the adjusted coefficients in Table 1 are all nearly the same.

Since the Walsh transform fits an additive model to the data, extension to multivariate point processes by means of an additive model seems reasonable. However, this is not necessary. A particular Walsh coefficient can be compared among event types using a multiplicative model. Thus, the question of whether several rates of occurrence are proportional can be answered. If the number of event types is not too large, the distributional properties are simplified in the same way as in the univariate case.

For example, consider events of two types. Let the Walsh coefficients be f_i' and f_i'', respectively, and let $f_i = f_i' \mp f_i''$. The totals for each event type are f_0' and f_0''. The comparison appropriate to proportionality of the rates of occurrence is $f_i'/f_0' - f_i''/f_0''$. The statistic

$$c_i = \frac{f_0\, f_0'\, f_0''}{(f_0^2 - f_i^2)} \left[\frac{f_i'}{f_0'} - \frac{f_i''}{f_0''}\right]^2 \qquad (3.14)$$

is the chi-square statistic usually obtained for a fourfold table. This can be seen by letting $f_i' = n_{11} - n_{12}$ and $f_i'' = n_{21} - n_{22}$, where n_{11}, n_{12}, n_{21}, n_{22} are counts. The chi-square approximation for (3.14) does not require dividing the point process into subintervals with adequately large numbers of events. Thus, as in the univariate case, the Walsh transform allows large-sample approximations to be applied. The hypothesis that overall the two rates are proportional can be tested by comparing the largest c_i with a threshold provided by the Bonferroni inequality [6].

Many authors have concluded on the basis of structural properties that multiplicative models for contingency tables are superior to additive ones [1, 5]. In particular, the fact that the hierarchy of multiplicative models contains independence and conditional independence is an important advantage. For example, it can be argued that the model fit to the robbery data in Figure 1, which is an additive composition of the daily-weekly and daily-seasonal marginals, is not as helpful for interpretation as conditional independence, the analogous multiplicative model, would have been. Further, there are problems for which multiplicative models are unavoidable because testing a particular multiplicative model is suggested a priori.

For the problem considered in this paper, the arguments for multiplicative models do not seem important enough to outweigh the advantages of the Walsh transform. In this paper, the model is a device for finding irregularities that are either so interesting or so large that they should not be distorted by the smoothing. For this purpose, the fact that the hierarchy of Walsh-transform models provides variability in the sizes of the cells is important. For example, the choice of how many divisions are needed to represent the seasonal variation in the robbery data is included in the model building.

The advantage of the Walsh transform is the simplicity of its distributional properties. Limitations on the sparseness permissible with the Pearson chi-square are reviewed by Haberman [6]. Even less is known about the distributional properties of the likelihood chi-square when the expected number in each cell is small. Sparseness is an important property of the robbery data since the 2912 cells contain 1555 zeros, 863 ones, 321 twos, 109 threes, 35 fours, 18 fives, 7 sixes, 2 sevens, 1 eight, and 1 nine.

The guidance provided by the Walsh transform is needed in the analysis of the counting function of the process that results from superimposing several days. This analysis is important when the variation during the day is large as it is for the robbery data. This counting function, which can be treated as an empirical distribution function, might show sharp changes in the rate function that can be related to other phenomena. In such an analysis, the Walsh coefficients provide guidance on what days can be superimposed without risking a severely distorted result.

Spectral analysis of point processes also fits an additive model [7]. As mentioned above, this technique is needed for periodicities with unknown frequency. However, spectral analysis is not as suitable for assessing the interactions among periodicities as the Walsh transform. With spectral analysis, interactions must be detected in the spectrum by recognizing that the frequencies of some peaks are sums or differences of the frequencies of other peaks. Further, the form of the interaction is harder to visualize with spectral analysis.

These comparisons show that the Walsh transform lies between contingency table analysis and spectral analysis, modeling the interactions as in contingency table analysis and avoiding the problems of grouping the data as in spectral analysis.

REFERENCES

[1] Bishop, Y. M. M., S. E. Fienberg, and P. W. Holland, Discrete Multivariate Analysis, Cambridge, Mass.: The MIT Press, 1975.

[2] Brillinger, David R., "Statistical Inference for Stationary Point Processes," in M. L. Puri, ed., Stochastic Processes and Related Topics, New York: Academic Press, 1975, 55-99.

[3] Cox, D. R. and P. A. W. Lewis, The Statistical Analysis of Series of Events, London: Methuen, 1966.

[4] Cox, D. R. and P. A. W. Lewis, "Multivariate Point Processes," in L. M. LeCam, J. Neyman, and E. L. Scott, eds., Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics, Vol. III, Berkeley: University of California Press, 1972, 401-448.

[5] Darroch, J. N., "Multiplicative and Additive Interactions in Contingency Tables," Biometrika, 61 (August 1974), 207-214.

[6] Haberman, Shelby J., The Analysis of Frequency Data, Chicago: The University of Chicago Press, 1974.

[7] Lewis, P. A. W., "Remarks on the Theory, Computation, and Application of Spectral Analysis of Series of Events," Journal of Sound and Vibration, 12 (1970), 353-375.

[8] Lewis, P. A. W., "Recent Results in the Statistical Analysis of Univariate Point Processes," in P. A. W. Lewis, ed., Stochastic Point Processes, New York: Wiley-Interscience, 1972, 1-54.

[9] Manz, Joseph W., "A Sequency-Ordered Fast Walsh Transform," IEEE Transactions of Audio and Electracoustics, AU-20 (August 1972), 204-205.

[10] Pearl, Judea, "Application of Walsh Transform to Statistical Analysis," IEEE Transactions on Systems, Man, and Cybernetics, SMC-1 (April 1971), 111-119.

[11] Pichler, F., "Walsh Functions--Introduction to the Theory," in J. W. R. Griffiths, P. L. Stocklin, and C. van Schooneveld, eds., Signal Processing, New York: Academic Press, 1973, 23-41.

CONDITIONAL AGREEMENT SCORE WITH TWO RATERS

Y. S. Lin, The Pillsbury Company

I. INTRODUCTION

When measuring agreement among many raters with categorical data, researchers generally deal with the agreement score {coefficients such as Cohen's Kappa (1960), Guttman's λ (1941), Scott's π (1955), etc.} which have the following common properties:

1. No assumptions were made concerning a "correct" assignment of items to categories. Agreement scores were not measured relative to any "correct" or "standard" assignment of items, but only with respect to the rater's internal consistency.
2. All raters are weighted equally.

Since there are many situations where the first characteristic will not be appropriate, we consider a slightly different kind of agreement problem. In this paper we will discuss the situation where it is reasonable to assume the existence of a "correct" ("expert") or "standard" assignment of items to categories. In this situation we may wish to either (1) compare each non-standard rater's response to that of the standard, or (2) measure the conditional agreement given the response from the standard is in category i (i = 1,2,...,I). This paper deals with the latter approach, illustrated with a two-rater case in a categorical setting. Cohen's Kappa-type conditional agreement score is used. The asymptotic variance and covariance for the maximum likelihood estimate of the conditional agreement score are obtained under multinomial sampling assumption. Asymptotic confidence interval for each conditional agreement score and the difference between two conditional agreement scores can then be calculated. An example will be used for illustration.

II. DEFINITION OF THE CONDITIONAL AGREEMENT SCORE

Assuming that there are two raters categorizing N items among I categories with one of the two raters being considered as "standard", we would like to examine the agreement between the two raters for only those items which the standard (or correct) assignments is in category i (i = 1,2,...,I).

For the IxI contingency table, we propose the following conditional agreement measure, $K_{(R1=i)}$, for those items which the first rater (standard rater or row variable) assigned to the i^{th} category:

$$K_{(R1=i)} = \{(P_{ii}/P_{i.}) - P_{.i}\}/(1-P_{.i}). \quad (1)$$

Here, $P_{ii}/P_{i.}$ is the probability that each of the $(NP_{i.})$ items would be assigned to category i by rater 2 given that rater 1 already assigned them to category i. Since the expected chance agreement for i^{th} category by both raters is $P_{i.}P_{.i}$, the conditional expected chance agreement of rater 2 given that rater 1 is in category i is $P_{i.}$. Then normalize the difference $(P_{ii}/P_{i.}) - P_{.i}$, (the probability of rater 2 agreeing with rater 1 in category i beyond chance agreement given that rater 1 has been known to have assigned the item to category i), by $1 - P_{.i}$. The range of $K(R1=i)$ is between $-P_{.i}/(1-P_{.i})$ and +1. The upper bound of $K(R1=i)$ is achieved if and only if $P_{ii} = P_{i.}$, that is, all the items that rater 1 is known to have assigned to category i are also assigned to the same category by rater 2. When none of the items known to be assigned to category i by rater 1 is assigned to category i by rater 2, i.e., $P_{ii} = 0$, then $K(R1=i) = -P_{.i}/(1-P_{.i})$. The conditional agreement score is equal to zero if, and only if, $P_{ii} = P_{i.}P_{.i}$ (or $P_{ii}/P_{i.} = P_{.i}$), same as in the unconditional situation. The conditional agreement score as defined in (1) is identical to the one suggested by Coleman (1966) in an unpublished paper (see Light, 1971) and is discussed in Bishop, Fienberg, and Holland (1975). To obtain the conditional agreement score for rater 1 given rater 2 is in category i, we simply interchange $P_{i.}$ and $P_{.i}$ in equation (1).

III. ASYMPTOTIC DISTRIBUTION OF $\hat{K}(R1=i)$

Assuming the multinomial sampling model with $X_{i.}$'s (observed number of items assigned to category i by rater 1) fixed, and let $\hat{K}$ (R1=i) denote the maximum likelihood estimator of K(R1=i), then $\hat{K}(R1=i)$ is given by:

$$\hat{K}(R1=i) = (NX_{ii} - X_{i.}X_{.i})/\{X_{i.}(N - X_{.i})\}. \quad (2)$$

By applying the results given by Goodman and Kruskal (1972) we can easliy show that the asymptotic distribution of $(\sqrt{N}\{\hat{K}(R1=i) - K(R1=i)\})$ is normal with mean zero and variance:

$$\mathrm{Var}\{\hat{K}(R1=i)\}$$

$$=((P_{i.}-P_{ii})\{P_{ii}(1-P_{.i}-P_{i.}+P_{ii})$$

$$+(P_{i.}-P_{ii})(P_{i.}P_{.i}-P_{ii})\})/\{P_{i.}^3(1-P_{.i})^3\}. \quad (3)$$

Under the null hypothesis of independence the asymptotic variance given in (3) above is reduced to:

$$\mathrm{Var}_0\{\hat{K}(R1=i)\}=\{P_{.i}(1-P_{i.})\}/\{P_{i.}(1-P_{.i})\}. \quad (4)$$

In practice, we may wish to look at conditional agreement measures simultaneously. For example, we might want to find out whether one rater's agreement with the standard is the same given that responses from the standard are in categories i and ℓ ($i \neq \ell$). For this, it is necessary that we obtain the asymptotic covariance for $\hat{K}(R1=i)$ and $\hat{K}(R1=\ell)$. By using the general δ-method, it can be shown that the asymptotic covariance is given by:

$$\mathrm{Cov}\{\hat{K}(R1=i),\ \hat{K}(R1=\ell)\}$$

$$=\{1/(ND^2E^2)\}(ML(\phi_1-\lambda_1\lambda_2)+DL(\phi_2-\lambda_1\lambda_2)$$

$$+EM(\phi_3-\lambda_1\lambda_2)+DE\{\phi_4+ \quad (5)$$

$$(P_{ii}-\sum_j P_{ij}\Theta_{ij})(P_{\ell\ell}-\sum_j P_{\ell j}\Theta_{\ell j})\}),$$

where $D = P_{i.}(1-P_{.i})$, $E = P_{\ell.}(1-P_{.\ell})$,

$M = P_{ii}-P_{i.}P_{.i}$, $L = P_{\ell\ell}-P_{\ell.}P_{.\ell}$,

$\phi_1 = \sum_j P_{ij}(1-P_{j.}-P_{.i})(1-P_{j.}-P_{.\ell})+$
$\sum_j P_{\ell j}(1-P_{j.}-P_{.i})(1-P_{j.}-P_{.\ell})$,

$\lambda_1 = \sum_j P_{ij}(1-P_{j.}-P_{.i})$, $\Theta_{ij} = P_{j.}+P_{.i}$,

$\lambda_2 = \sum_j P_{\ell j}(1-P_{j.}-P_{.i})$, $\Theta_{\ell j} = P_{j.}+P_{.\ell}$,

$\phi_2 = \sum_j P_{ij}(1-P_{j.}-P_{.\ell})\Theta_{ij}+$
$\sum_j P_{\ell j}\Theta_{ij}(1-P_{j.}-P_{.\ell})$,

$\phi_3 = \sum_j P_{ij}\Theta_{\ell j}(1-P_{j.}-P_{.i})+$
$\sum_j P_{\ell j}\Theta_{\ell j}(1-P_{j.}-P_{.i})$,

$\phi_4 = \sum_j P_{ij}\Theta_{ij}\Theta_{\ell j}+\sum_j P_{\ell j}\Theta_{ij}\Theta_{\ell j}$.

From expressions (3) and (5), the asymptotic variance-covariance matrix can be constructed for $\hat{K}(R1=i)$, i = 1,2,...,I. As usual, the estimated large sample variances and covariances are obtained by substituting observed proportions for cell probabilities.

IV. EXAMPLE

The following data with two home economists judging 159 breakfast food items will be used for illustration.

Two judges were asked to rate the values of breakfast foods using a three-category scale (3=good, 2=medium, 1=poor value). The entire test lasts more than one month with at least 30-minute time intervals between tastings. The test results are summarized in the following table. Table 1 gives the actual counts by cell. (Note: The value concept of a product generally is the combination of three elements: usage occasion, quality of the product, and unit price).

Table 1 Cell Frequencies of 159 Breakfast Foods as Rated by Two Judges Regarding Their Value

	Judge 2			
Judge 1	Good	Medium	Poor	Total
Good	63	7	5	75
Medium	7	24	14	45
Poor	4	3	32	39
Total	74	34	51	159

In the example above, estimated Kappa coefficient $\hat{K}_2$ = .6077 and asymptotic standard error $S(\hat{K}_2)$ = .056, we concluded that the home economists agreed more often than they would by chance.

By using the conditional measures, K(R1=i), we can attempt to localize the agreement. Suppose that home economist 1 can be assumed as the standard against whose judgment we wish to compare the rating of home economist 2. Using the expression (2) we have:

$$\hat{K}(R1=1) = .701,$$

$$\hat{K}(R1=2) = .406,$$

$$\hat{K}(R1=3) = .736.$$

This shows that home economist 2 agrees most with home economist 1 when the latter assigns the breakfast foods to the good and poor categories. The 95% asymptotic Bonferroni simultaneous confidence limits for each K(R1=i) we get:

$$K(R1=1) : \{.518, .884\},$$

$$K(R1=2) : \{.213, .639\},$$

$$K(R1=3) : \{.519, .954\}.$$

The estimated asymptotic variance-covariance matrix for $\hat{K}(R1=i)$, i = 1,2,3, are given by:

.0058	.0006	.0009
	.0064	.0005
		.0076

Using the estimated variance-covariance, the 100(1-α)% confidence limit for the difference between two conditional agreement measures can be constructed, e.g., the 95% asymptotic Bonferroni simultaneous confidence intervals for differences between two conditional Kappas are given by:

$$K_{(R1=1)} - K_{(R1=2)} : \{.043, .547\},$$

$$K_{(R1=1)} - K_{(R1=3)} : \{-.295, .223\},$$

$$K_{(R1=2)} - K_{(R1=3)} : \{.057, .603\}.$$

REFERENCES

Cohen, J. "A Coefficient of Agreement for Nominal Scales," Education and Psychological Measurement. 20, (1960), pp. 37-46.

Guttman, L. "An Outline on the Statistical Theory of Prediction," in P. Horst (Ed.) The Prediction of Personal Adjustment. New York: Social Science Research Council, (1941).

Scott, W. A. "Reliability of Content Analysis: The Case of Nominal Scale Coding," Public Opinion Quarterly. 19, (1955), pp. 321-25.

Light, R. J. "Measures of Response Agreement for Qualitative Data: Some Generalizations and Alternations," Psychological Bulletin. 76, No. 5, (1971), pp. 365-77.

Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. Discrete Multivariate Analysis--Theory and Practices. MIT Press, (1975).

Goodman, L. A. and Kruskal, W. "Measures of Association for Cross Classifications, IV: Simplification of Asymptotic Variances," JASA. 67, (June 1972), pp. 415-21.

DEPENDENT VERSUS INDEPENDENT VERIFICATION

John S. Linebarger, Carl Jablin, William C. Davie
Bureau of the Census

Introduction

Human error is a part of any human process. To prevent a deterioration in quality, verification schemes have long been used as a means of checking on the truth or accuracy of the various elements that make up that process.

This paper discusses one such verification process, that being the verification of the coding of Industry and Occupation responses obtained in the Current Population Survey (CPS), a survey carried out monthly to primarily obtain estimates of labor force characteristics of our changing population. More precisely, the paper will compare and discuss two verification schemes, a dependent scheme and an independent scheme, which were carried out simultaneously for a period of ten months on the CPS Industry and Occupation coding operation. Based on the analysis of these two schemes, it was concluded that the independent is the better verification scheme to use on the CPS Industry and Occupation data. Consideration was given to timing, bias, record keeping, costs, etc., in attempting to arrive at the optimum scheme. Before dependent and independent verification can be discussed, each term should be defined:

Dependent Verification - A "verifier" reviews an Occupation or Industry code previously assigned by a producer -- with the verifier having complete knowledge of the producer-assigned code.

Independent Verification - At least one other person assigns an Occupation or Industry code without prior knowledge of what was done by the producer and the outcomes are compared.

Until recent years, most quality control programs used the dependent verification process for inspection. The verifier could be influenced by the work of the original producer; thus, many times the verifier failed to identify a substantial proportion of defective items. The corrector of rejected work could also be influenced by work initially done by the producer as well as changes made by the verifier. In most theoretical plans that were dependent in nature, the assumption was made that inspection was carried out with little or no error. A review of data from several coding operations has shown this assumption to be invalid.

In quality programs where it was suspected that dependent inspection might fail to uncover a substantial proportion of defectives, independent verification has been used. Two schemes that have been used are: (1) three independent codings with majority rule, and (2) two independent codings with adjudication of differences. With the majority rule device, three sets of independently produced items were compared and the code assigned by two out of three coders was considered the correct code. Studies indicate that when a particular item was assigned a code independently by three equally qualified coders, it was highly unlikely that codes in agreement were incorrect.1/

In the two-way system, an independently precoded item was matched to the production coder entry, and agreements were accepted as correct. Disagreements were sent to an adjudicator who decided whether the precoder was incorrect, the coder was incorrect, or both were incorrect. This latter plan is not discussed within the context of this paper.

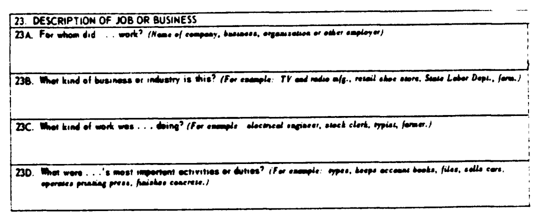
The attached chart shows some of the advantages and disadvantages that should be considered in selecting either a dependent or an independent verification scheme.

Background

The Census Bureau has conducted the CPS monthly since 1942. The Industry and Occupation (I&O) data is collected to reflect employment trends, unemployment rates, occupation mobility and other labor force data for the various industries and occupations throughout the country.

Labor force questions are asked with reference to the week containing the 12th day of the month. Hence, the survey is conducted during the week following the week containing this 12th day. Results of the survey for a particular month must be released to the public on the first Friday of the following month. This time-frame allows eight working days to assign approximately 75,000 I&O codes, with small amounts of the work early in the period and large amounts late in the period. This coding is performed at the Census Bureau's Jeffersonville, Indiana facility.

Below is a facsimile of the I&O questions on the CPS questionnaire from which the basic I&O information for an individual is obtained. This information, along with the type of ownership of the industry, listings of the large companies for the individual's geographic location, and the respondent's age, sex and education provide the basis for assigning the I&O codes.


23. DESCRIPTION OF JOB OR BUSINESS

23A. For whom did . . . work? (Name of company, business, organization or other employer)

23B. What kind of business or industry is this? (For example: TV and radio mfg., retail shoe store, State Labor Dept., farm.)

23C. What kind of work was . . . doing? (For example: electrical engineer, stock clerk, typist, farmer.)

23D. What were . . .'s most important activities or duties? (For example: types, keeps account books, files, sells cars, operates printing press, finishes concrete.)

1/ U. S. Bureau of the Census, United States Censuses of Population and Housing, 1960: Quality Control of Preparatory Operations, Microfilming, and Coding, Washington, D. C., 1965, p. 33.

Historically, there has been some type of industry coding at the Census Bureau since the 1850 Decennial Census. A method of occupation coding was introduced in the 1910 Decennial Census. The three-digit industry and the three-digit occupation codes currently used by the Bureau have been in use since 1960, with additional codes being added for the 1970 Census. Before the 1960 Decennial Census, dependent verification was used to control the quality of the I&O coding at the Bureau. Then, prior to the 1960 Decennial Census, a study was conducted to determine the number of coders required to successfully employ an independent system. This study, in which five different people coded the same items, showed that a three-way match provided the correct code with sufficient accuracy when the three persons were equally qualified.2/

Thus, for the first time in I&O coding, an independent plan with a three-way match was employed in the 1960 Census. Subsequently, independent plans were used for the 1970 Decennial Census, the 1970 Current Employment Survey, and the 1973-1974 Consumer Expenditure Survey. Estimated error rates on these surveys ranged from 2.7 percent to 8.9 percent. However, these error rates cannot be realistically compared to CPS or to each other as each coding requirement and group of coders differed.

CPS Verification Plans

From August 1942 to September 1975 the CPS I&O coding was verified on a dependent basis. In 1967 the coding operation was transferred from Washington to the Census Processing Office in Jeffersonville, Indiana. From this point until the independent plan was instituted, the qualification level was 1.0 percent on a code-pair basis with no verification of documents containing referral codes. A referral code occurs whenever specified in the coding manual or whenever a coder decides that he/she cannot give the item a specific code. Referrals are forwarded to a specialized group of coders. The observed error rates during this time were consistently less than 1.0 percent. Based on a review of data from the 1973-1974 Consumer Expenditure Survey, this error rate would have been estimated at 1.0 to 2.0 percent on an individual code basis.

A small pilot study was carried out to develop and test the necessary computer programs and to determine the costs of the clerical operations for an independent verification scheme. Although the test did not contain a representative sample of the qualified CPS I&O coders and did not contain any documents that were referred, the test did give some feeling for costs of an independent verification program and some idea of the size of the error rates for production coders. This study showed the error rate level to be around 3.0 percent, which was higher than the dependently derived error rate of 1.0 percent stated above.

2/ This study, conducted in 1959, was not published and is only in tabular form.

In late 1974, personnel from several divisions within the Bureau jointly designed an independent verification system for the industry and occupation coding operation. Because of the short timeframe allowed by the constraints of the CPS operation, the verification was designed as a post-survey operation. Therefore, the verification plan is necessarily a process control type plan rather than a lot acceptance sampling plan.

With this plan, feedback of errors cannot be given during a particular month of coding but rather it is given prior to the CPS coding for the next month. It was felt that this would not be a drawback because the majority of the coders had been coding industry and occupation for several years. Rapid feedback is of primary importance when a coder initially begins an operation.

Rather than concentrate on the feedback tool, it was felt that the entire training package should be reviewed and revised. A larger test deck of responses with their respective codes was devised for use in training new coders and those coders that require retraining. A review was made of the training instructions and coding instructions to determine what revisions might be made to eliminate ambiguity in these materials. Because many of the current coders had been on the job for several years, it was felt that they had probably attained their accuracy level and the greatest impact could be made on any newly hired coders.

The independent verification plan was designed as follows: a 10 percent sample of each coder's work is systematically selected throughout the coding operation. These CPS documents are kept separate from all other documents. Also included in the sample are 10 percent of the referral cases. As soon as enough sample documents have been accumulated, these documents are sent to microfilming operations as a unique work unit. Such work units are identified by a range of identification numbers set aside for the QC sample documents.

Following production microfilming, the QC sample work units are microfilmed again. Paper copies are generated from the microfilm with the coded areas masked so that the subsequent coding can be done independently. Each work unit is coded twice by the regular CPS coders with care taken that no coder codes the same document twice.

The two independent codings are done on FOSDIC-readable sheets that are microfilmed and read via FOSDIC equipment (the Census Bureau's film optical reading equipment). A computer match by document serial number and person number is made and the particular codes are matched. If there is a three-way agreement in code, the code is considered correct and no error is assigned. Similarly, if all three codes differ, the response is said to be vague, no code is said to be better than another and no error is assigned. However, when two of the codes agree and a third code disagrees, the two codes in agreement are said to be correct and an error is assigned to the person assigning the differing code (including referrals).

Coders are considered to be qualified as long as their error rates remain less than 7.5 percent, and their work is verified only via the independent system. When a coder becomes disqualified, in addition to the independent verification, dependent verification is performed in the non-sample work to assure that no poor quality work is released. As soon as the error rate is brought back within reasonable limits, the dependent verification is dropped.

Another change made in the new system is the definition of an item of error: each three-digit occupation code and each three-digit industry code is unique. In the past system, because of the relationship between some codes (e.g., a fireman for the railroad industry receives the occupation code 456, while a fireman in the mining industry receives an occupation code of 452), the two three-digit codes were linked and the unit item was defined to be the entire six-digit industry and occupation code. This was changed because it was felt that, in some cases, errors could be camouflaged by the use of the six-digit definition. It should be noted that this change should only serve to decrease the effective error rate.

The early results of the independent plan show the error rate to be considerably higher than shown by the dependent plan (6 percent versus less than 1 percent). While the error rate is 4.5 percent when the coder assigns a code, the coders are erroneously referring items approximately 15 percent of the time. While this error does not directly give an erroneous code, it does cause increased burden upon the referralists, and was therefore defined as an error. At the present time, approximately 13 percent of the I&O items are referred, of which one-seventh should not have been.

One feature of the independent plan is the utilization of the Bureau's FOSDIC and computer facilities for matching purposes. The FOSDIC equipment has a very low misread rate and the computer no-match rate is low (1.5 percent). Though there are no estimates for a manual matching operation, it is expected to be low also. Utilization of the computer also allows the compilation of numerous summary tables for use in decision-making. This tabular capability also allows the pinpointing of problem codes.

It is felt that the inclusion of the referral cases provides a more complete evaluation of a coder's error rate. No provision is made for estimating the quality of the codes assigned by the referral pool; however, early results show that when the production coder gives a referral as a minority code, the referral pool assigns the same code as the two independent coders 75 percent of the time.

Contrary to expectations, the classification of a minority referral as an error has not decreased the referral rate; in fact, the referral rate for documents has increased from the 20 percent level to around 25 percent. Thus, the desired decrease in referral rate has not occurred; this problem has not been fully dealt with and no reasons for the increase in referral rate come readily to mind.

Concurrent Verification

For a period of ten months, the industry and occupation coding operation for CPS was verified via both the dependent and the independent schemes. An initial 10 percent sample was selected for independent verification and dependent sample verification was performed on the remaining work. Records were maintained for each method of verification for the coders.

The average error rate for the dependent plan was 1.2 percent, while for the independent scheme it was 6.0 percent. These two error rates are significantly different at the 99 percent confidence level. A correlation coefficient was calculated to determine if there was a relationship between the two overall error rate estimates for a particular coder; it was found to be low (0.56). This can be partially ascribed to the difference in unit item definition for the two verification schemes and the exclusion of referral cases in the dependent scheme.

The error rate estimates for the independent plan, while somewhat higher than anticipated, were more in line with past experience. With the complexity and ambiguity within the I&O coding operation, it seems unrealistic that the error rate could have been below the 2 percent level. Therefore, it was decided that the independent verification plan gave a more credible estimate.

As the verification plans were utilized in a live, working operation rather than an evaluative atmosphere, no precise estimate of bias was derived. It was assumed that the dependent plan gave a downwardly biased estimate of the production coder error rate.

In regard to consistency, while the error rates for an individual coder fluctuated with the dependent verification scheme, "poor" coders consistently had high error rates and "good" coders consistently had low error rates in the independent scheme, however variable.

The costs for the independent verification plan were higher than for the dependent plan as the dependent plan was an integral part of the ongoing coding operation and had been for many years. However, it is felt that after the independent operation becomes smooth running, the costs for the independent operation, while probably still exceeding those of a dependent operation, will become more acceptable. Also, the institution of the use of the computer increased the cost.

Problems

The delay in the feedback of types of errors and error rates is an undesirable feature of the present independent verification plan for CPS. Due to the short processing time-frame there is no practical method of speeding up the timeliness of the feedback of results. The importance of timeliness of feedback must be weighed against the importance of accuracy and completeness of feedback.

A second problem arises in that, although coders are required to reach a qualification error level, the process control plan for qualified coders could permit a coder to perform sub-standard work for two months before an action is taken. However, due to the unreliability of the dependent plan, it would seem that little if any quality is lost due to independent verification.

During dependent verification the estimated coding error rates would have ranged between 1 and 2 percent on an individual code base. In independent verification this estimated error rate is approximately 5 percent on an individual code base. This is probably a result of the unreliability of dependent verification. Additionally, independent verification provides a measure of the percentage of codes incorrectly referred. There is no practical way to do this using a dependent plan.

Probably the most serious problem associated with independent verification is the increased cost. The hours expended to accomplish the independent verification increase the total hours expended by 30 percent. Due to the importance of the figures that are published, a reliable estimate of the outgoing error rate may justify this expense. In other surveys where the significance of the I&O statistics is secondary these costs may not be justified. One possible solution to this problem could be a reduction of the I&O coding sample.

Conclusions

While the independent verification plan as incorporated in the coding operation may be more expensive, more disruptive, and less timely than the old dependent plan, the new plan, as designed, gives a more credible estimate of the coders' error rates, gives a less biased estimate, and gives an estimate which is more consistent with those derived in other industry and occupation coding operations.

Further research is needed in verification plans for all coding operations, not just for CPS. Feedback is essential to any coding operation and the problems are intensified in a post-operation such as the independent QC scheme for CPS. Research is needed to derive new or better methods for giving information back to coders concerning errors found in the work. Notification of errors is a natural aid in the constant education process.

Another area where research is needed is in the area of the referrals. Specifically, research is desired in deriving a method of evaluating the referral pool, a specialized group. Since the referral cases are necessarily ambiguous and difficult to code, the setting of quality standards should also be studied.

In conclusion, past experience at the Census Bureau has shown that independent verification is a more reliable and realistic estimator of quality. Although operationally less efficient, more work is necessary to develop an independent verification scheme that is practical to carry out and less costly to maintain. The independent verification plan utilized for CPS industry and occupation coding is the Bureau's first step in such an effort.

Advantages/Disadvantages of an Independent Verification Plan Versus a Dependent Verification Plan

Item	Independent Plan	Dependent Plan
Bias	• Objective Verification • Usually less biased since no one person's work is dependent upon another.	• Subjective Verification • History of some collusion • Rectifier of rejected work units has history of missing many original errors.
Timeliness	• Usually time consuming • Usually disruptive to process	• Usually quicker • Not disruptive to process
Recordkeeping	• Usually more people upon whom records must be kept.	• Requires little work if handled clerically.
Costs	• Usually more costly	• Usually less costly
Overall Summary	• More expensive, more disruptive, and more time consuming; however, 1. Less variable est. of errors 2. Less biased; more credible	• Less expensive, less disruptive, and less time consuming; however, 1. More variable est. of errors 2. More biased; less credible

VARIATIONS IN ENVIRONMENTAL QUALITY INDICATORS IN LARGE METROPOLITAN AREAS

Ben-chieh Liu, Midwest Research Institute

I. Introduction

Dissatisfaction over land use, natural resources extraction, and pollution damage to our natural environment by industrialization and urbanization has been growing in this country. According to the estimate of the Council on Environmental Quality [12], a total of $200 billion will be spent on pollution control between now and 1980 in order to maintain present air and water quality standards. Since resources are finite and environmental protection or pollution control is costly, it is necessary to ascertain that the last unit of control bought imposes no additional costs greater than the additional benefits.

As Fisher and Peterson [4] have pointed out, not only is the policy of internalizing not generally relevant to the management of natural environments (and direct government intervention is required since there are no market mechanisms to rectify), but also most of our advice cannot be implemented without gathering a great deal of useful information, of which very little already exists. One of the most detrimental features of the social sciences to date has been the absence of any generally agreeable and acceptable consensus set of either social welfare functions or social conditions. In addition, a problem is not likely to be solved until it has been perceived and identified as a problem. Although there exist thousands of decisionmakers within the private sector who are able and willing, and devoted to the enhancement of our environmental quality of life, they are not certain about the direction that their activities should take, just as many public decisionmakers are not always sure about the social, economic, political, and environmental impacts of their actions. In order to promote the general welfare and to enrich the environmental quality, there is an urgent need for a mechanism which can distinguish better from worse as stressed by Anderson [1], Bauer [2], Cohn [3], Fox [5], Sheldon and Moore [13], Sheldon and Parke [14], et al. As it now stands, the United States has neither a comprehensive set of social statistics that reflect changes in our values and measure social progress or retrogression, nor an integrated set of environmental indicators which can describe the environmental conditions and evaluate all environmental protection policies among the standard statistical metropolitan areas (SMSA's).

The search for environmental indicators is an attempt to obtain consistent information that will be useful to evaluate the past, guide the action of the present, and plan for the future. The empirical measures of various levels of environmental quality of life presented in this paper are aimed at the identification of strengths and weaknesses.

II. An Environmental Quality Model

Although it is generally understood that the need for environmental quality indicators is urgent because they are essential to the assessment of many aspects of social progress and social accounting, and are useful for national goal setting, project planning, priority ranking, program manipulation, and performance evaluation, there is no consensus as to what environmental quality is all about, how the quality indicators should be defined, and for whom and in what manner they should be constructed. This failure to reach a consensus can be substantially attributed to the absence of a commonly accepted social welfare function or value system.

Methodological development of environmental indicators and interest in the environmental quality concept development grew remarkably in the 1960's. The National Wildlife Federation has constructed Environmental Quality Indexes since 1969. Furthermore, the Environmental Protection Agency has been generating a variety of air, water and solid waste, and other environmental pollution indicators in the U.S. Instruction and model specifications in measuring environmental quality and impacts were given in the interim guidelines for implementing the National Environmental Policy Act (NEPA) in April 1970, by the Council on Environmental Quality, which since 1970 has been submitting to the President an annual report, Environmental Quality.

Although the Council on Environmental Quality (CEQ) was authorized to promote the development of indexes and monitory systems to determine the effectiveness of programs for protecting and enhancing environmental quality to sustain and enrich human life, the reports issued by CEQ have not reported environmental indicators in any form comprehensive enough for detailed regional analyses. Consequently, reviewers such as Mills and Peterson [10,267] have stressed that the CEQ should accord high priority to development of an adequate statistical appendix.

The model employed in this paper is taken from the one developed by the author [6] and is termed the quality of life production model. Given that the quality of life means happiness or a state of satisfaction and that the quality of life indicators represented by a set of statistics on economic, political, environmental, health and education, and social conditions may offer clues to human attitudes and behavior and social performance over time, the quality of life that each individual (i) attempts to maximize may be expressed as an output function with two factor inputs as arguments--the physical (PH) and the psychological (PS)--a portion of which he owns and shares with other people in the community at any given point of time (t):

$$QOL_{it} = f(PH_{it}, PS_{it})$$

The physical input consists of quantifiable goods, services, material wealth, etc., while the psychological input includes all subjective, spiritual, sociological, and anthropological factors such as community belongingness, esteem, self-actualization, love, affection, etc. Although the production function expressing the relationship between output and input factors of quality of life is known to be enormously complex (there are as many such factors as there are people), an aggregate homogeneous production function may be assumed for a metropolitan area as a whole. Since the psychological inputs are not readily quantifiable and hence rarely reported, the quality of life output may be taken at a particular point in time as a function of those social (SO), economic (EC), political and welfare (PW), health and education (HE), and environmental (EN) inputs which are quantifiable or:

$$QOL_{it} = F(EC_{it}, PW_{it}, EN_{it}, HE_{it}, SO_{it} \quad PS_{it})$$

The model proposed here is similar to the conventional production models employed to study the behavior of firms. The two axes, instead of being labeled as capital and labor per unit of time, are, respectively, the ordinal measures of the psychological inputs and the cardinal measures of the physical inputs. The iso-quant curves are hereby replaced by the iso-quality of life curves, and the budget lines are substituted for by the individual's capability curves which, in this case, would likely be concave to the origin. The optimal level of quality of life is produced only by combining both the physical and psychological inputs in such a form as to locate the tangency point between the iso-quality and the capability constraints to exchange and to acquire, while the major concern for a society is how to improve an individual's capability by shifting the constraint curve outward to the right.

To measure objectively the output level of quality of life as subjectively perceived by individuals, we may start with the cardinal measures of the physical inputs by holding constant the psychological inputs. Given this, an environmental model ideally should take into account factors other than pollution, climate, and recreational facilities such as natural endowments and conservation, resource availability and accessibility, etc. However, the scarcity of comparable data for SMSA's prevents those representative variables from being selected and included. Thus, the environmental model in this paper encompasses only such variables affecting our urban quality of life as the air, visual, noise, solid waste and water pollution, climatological and recreational factors. All types of pollution are grouped under the individual and institutional environment because they are different by-products of various human activities.

The other quality measure in this model is the natural environment component which includes five climatological variables and two recreational variables: sunshine days, inversion frequency, thunderstorms, high and low temperatures, areas of parks and recreational areas, and miles of trails. Parks and recreational areas have come to play an ever-increasing, important role in our city life. As a result, this variable is used twice in the environment component, serving as a determinant of visual pollution and a factor of natural environment as well.

All variables, except the parks and recreational areas, miles of trails, and sunshine days, in this paper have adverse effects on our environmental quality, and are negative inputs to our daily life. Thus, 17 variables mentioned depict mostly our urban environmental "bads" rather than "goods." They are chosen for the following reasons: they make us alert to our environmental problems, compare the quality of our environment, and judge the efforts made to reduce and eliminate the pollutants. It should be noted, however, that evidence suggests that the direct effects of pollution on property, on human health, and on the quality of life are varied [7,9,15].

Among the individual concerns in our environmental quality, this model is thus identified with the determinants made up of individual and institutional environments (IIE) and natural environment (NE). While the former set of variables are entirely strategical and policy-oriented variables, the latter include most uncontrollable inputs of climate considerations, i.e.,

$$EN_{it} = f(IIE_{it}, NE_{it})$$

While some variables are represented by published official sources, some are denoted in the firsthand 1970 data collected and computed by the author [6]. The data for 1970 were collected for the 65 large SMSA's with populations between 500,000 and 1,000,000, and the standardized "Z" values were computed for all factors. On the basis of the percentile distribution of the "Z" values, SMSA's were divided into five groups and assigned points of 5, 4, 3, 2, or 1, respectively, for outstanding (A); excellent (B); good (C); adequate (D); and substandard (E). Factors within the same subcategory were then weighed equally to derive a subcategory score, and the subcategory scores were weighed equally to obtain a subcomponent score. Finally, the average of the subcomponent scores was taken to show the composite index for each SMSA, which was subsequently rated by the indexes in comparison with those of other SMSA's.

III. Empirical Results

The environmental quality of life indicators in this study concern both individual and institutional environment and natural environment. Air, visual, noise, water, and solid waste pollution are by-products of the postindustrialized society. Their existence and the attempts at eradication not only impose a heavy financial burden on our society, but they are also hazards to human health, animal fertility, crop production, etc. Thus, relative indicators for these five categories were constructed based on the absolute indicators obtained from various public and private sources. The individual and institutional environments among the metropolitan areas are evaluated jointly on 10 different factors.

The natural environment is evaluated from five climatological and two recreational factors. The factors included in this component are fewer than desirable and are far from being complete because of the lack of empirical statistics. Nevertheless, these factors provide basic information for a fairly accurate judgment on urban environment for all metropolitan areas. Table 1 presents all statistical results. The most important findings in this study and their implications are broadly delineated in the following:

1. This study of environmental quality in large SMSA's indicates that the Pacific region stands at the top of the listing. All but two SMSA's in the Pacific region are rated either "outstanding" or "excellent." In fact, California has four outstanding SMSA's, or about 40.0 percent of the total of 11 rated "A." They are Sacramento, San Bernadino-Riverside-Ontario, San Diego and San Jose. However, Los Angeles-Long Beach and Anaheim-Santa Ana-Garden Grove SMSA's fall only in the average category. The best of "A"-rated SMSA's is Sacramento which obtained an environmental quality index appreciably greater than the others; i.e., -0.20 or about two standard deviations above the mean of -1.03. Next to Sacramento in environmental quality is Seattle-Everett. Portland SMSA ranked ninth in the race for better environment quality. The other "A"-rated SMSA's are Miami, Honolulu, Phoenix, Allentown-Bethlehem-Easton, other "A"-rated SMSA's are Miami, Honolulu, Phoenix, Allentown-Bethlehem-Easton, and Springfield-Chicopee-Holyoke.

The geographic distribution of ratings shows also the existence of a concentrated pattern among the SMSA's that received unfavorable ratings. Many large SMSA's in the East North Central and Middle Atlantic regions were rated either substandard or only adequate. However, the lowest environmental rating among the 65 SMSA's was found in Pittsburgh. This resulted primarily from its extremely high level of total suspended particulates, visual and water pollution.

Chicago SMSA has the second lowest index, -1.82. This SMSA had a number of environmental problems, but its water and air pollution was among the worst. While Anaheim-Santa Ana-Garden Grove SMSA has water pollution indexes as low as 0.68, Chicago had an index of about 26 times as high as the best area in California. In addition, this SMSA also suffered from a lack of recreation areas and facilities.

2. Although there is a clear pattern of the distribution of the environmental quality indexes among regions, the overall variation in environmental quality among the large sized SMSA's does not appear to be very critical. Except for the SMSA's mentioned in the preceding section, most of the remaining SMSA's received quality indexes which are not significantly different from each other in aggregate values; the coefficient of variation is about 0.34 (0.345/1.034). This environmental inequality problem in the large SMSA's is relatively more serious than that in the medium or small sized SMSA's. Nevertheless, the environmental inequality problem among these 65 SMSA's is much less discernible than the social inequality problem as observed in the SMSA's. For a paper on social quality, see Liu [9].

3. Although it is normally expected that the levels of objectively measured environmental quality vary from region to region and from component to component, it is very interesting to note that only a few of the 65 SMSA's have consistently high or low ratings among all factors under consideration. For instance, even though San Bernadino-Riverside-Ontario ranked first in the natural environmental quality, this SMSA showed serious visual and noise pollution and solid waste problems. The most serious problem in Houston was the insufficient recreation areas and facilities--it ranked 53rd. On the contrary, the SMSA's rated substandard on the overall scale also showed comparatively favorable ratings in many environmental considerations. For instance, Cleveland compared very well in visual and noise pollution and in parks and recreational areas; Detroit ranked 14th in visual pollution and Indianapolis 17th in noise pollution and 33rd in parks and recreational areas; and Louisville even scored 3rd in solid waste generation.

4. Pollution and environmental damages have been increasingly attacked by opponents to economic growth and industrialization. Economists have aptly used pollution as an illustration of externalities. The trade-off between economic activities and environmental deterioration, or the degradative changes in our ecosystem, have been generally accepted. The author observed

that the trade-off phenomenon seems to be more significant in the large sized metropolitan areas than in the medium sized areas. Chicago ranks as an SMSA with substandard environmental quality, but outstanding economic health. Honolulu and Springfield-Chicopee-Holyoke were revealed to be opposite cases. The third typical case was found in Portland, where both economic and environmental quality was outstanding in 1970.

5. The Spearman rank order correlation coefficient (r) obtained between the individual and institutional (IIE) and the natural environmental component (NE) for the 65 SMSA's is 0.17, indicating that there is virtually no correlation between the two components employed for environmental quality evaluation among regions. In addition, this finding tends to be supportive of the basic prerequisite in the development of social indicators, that the selected variables should be as independent of each other as possible.

IV. Implications and Concluding Remarks

Empirically the model systematically evaluated the varying environmental elements among the U.S. urban areas, and constructed the first set of environmental quality indexes for the nation's large metropolitan areas.

While geographically this paper found a concentration pattern of environmental inequality in favor of the Pacific and against the East North Central and Middle Atlantic regions, this inequality problem among the 65 large sized SMSA's is not as serious as other quality of life components such as social, health and education. However, the trade-off hypothesis between economic growth or industrial development and environmental degradation has been observed in many cases among the large SMSA's, especially those in the manufacturing regions. The implication of those findings is that on the whole, people in the large SMSA's were still enjoying a relatively homogeneous environmental quality of life as of 1970. Policies probably would be better to focus on the preservation of this homogeneity in general and the improvement in the substandard North Central and Middle Atlantic regions in particular.

This paper has also found that in this country there is neither a perfect SMSA offering the best of environmental quality nor a worst area suffering substandard environmental illness in all 17 institutional and natural environment considerations. For policy decisionmakers, it indicates that there is always an area (or areas) requiring special attention and extra effort in order to balance the overall environmental quality of life within each SMSA. The environmental well-being is a notion for multidimensional concepts. Thus, at the present time it is not only theoretically controversial to consider a sole indicator for the overall environmental quality, it is also empirically difficult to single out an index for the environmental measurement due to the lack of consensus in weighting between the institutional and the natural components.

Since there are definite regional concentration patterns and inequalities in the environmental quality, a more thorough investigation of input factors in the adequate or substandard regions should reveal the cause-effect relationship, and the potential trade-offs between economic and environmental objectives. Consequently, better policy alternatives and feasible remedies may be recommended.

REFERENCES

1. Anderson, J. C., "Causal Models and Social Indicators: Toward the Development of Social Systems Models," American Sociological Review Vol. 38, No. 3 (1973).
2. Bauer, R. (ed.), Social Indicators, Cambridge: The MIT Press (1966).
3. Cohn, Wilbur, The Quality of Life and Social Indicators (New York: NBER, 1972).
4. Fisher, A. C. and F. M. Peterson, "The Environment in Economics: A Survey," Journal of Economic Literature, Vol. 14, No. 1, pp. 1-33, March 1976.
5. Fox, K. A., Social Indicators and Social Theory (New York: John Wiley and Sons, 1974).
6. Liu, B.C., Quality of Life Indicators in U.S. Metropolitan Areas, 1970 (Washington, D.C.: Government Printing Office, 1975a, and N. Y.: the Praeger Publisher, 1976).
7. Liu, B. C., "Functions of Air Pollution Damage on Human Health," Proceedings of Air Pollution Control Association (1975b).
8. Liu, B. C., "Variations in Social Quality of Life Indicators in Large Metropolitan Areas," Proceedings of the American Statistical Association, Social Statistics Section (1975c).
9. Liu, B. C. and S. H. Yu, "Mortality and Air Pollution: Revisited," Journal of Air Pollution Control Association (1976).
10. Mills, E. S. and F. M. Peterson, "Environmental Quality: The First Five Years," The American Economic Review, LXV, 3, pp. 259-268, June 1975.
11. Mills, E. S., "Economic Incentives in Air Pollution Control," in Economics of Air Pollution, Harold Wolzin (ed.) (New York: W. W. Norton and Company, Inc., 1966).
12. President's Council on Environmental Quality Environmental Quality 1972: Third Annual Report (Washington, D.C.) (1972).
13. Sheldon, E. and W. Moore, Indicators of Social Change: Concepts and Measurements (New York: Russel Sage Foundation, 1968).
14. Sheldon, E. and R. Parke, "Social Indicators," Science, Vol. 188, No. 4189, pp. 693-699 (1975).
15. Waddell, T. E., The Economic Damages of Air Pollution (Washington, D.C.: EPA Environmental Research Center, 1974).

Table 1

ENVIRONMENTAL QUALITY OF LIFE INDICATOR RATING AND RANKING FOR LARGE SMSA'S 1970

	Overall	Individual and Institutional Environment (IIE)	Natural Environment (NE)
Akron, Ohio	-.9667 C 23	-2.4333 C 23	0.5000 C 25
Albany-Schenectady-Troy, N.Y.	-1.2917 D 53	-2.6333 D 46	0.0500 D 30
Allentown-Bethlehem-Easton, Pa-N.J.	-.6167 A 9	-1.9333 A 2	0.7000 B 14
Anaheim-Santa Ana-Garden Grove, Ca.	-1.0500 C 33	-2.7000 D 50	0.6000 B 20
Atlanta, Ga.	-1.2833 D 52	-2.4667 C 35	-0.1000 E 58
Baltimore, Md.	-1.2667 D 50	-2.7333 D 52	0.2000 D 41
Birmingham, Ala.	-1.4250 E 59	-2.7000 D 51	-0.1500 E 59
Boston, Mass.	-1.2500 D 48	-3.0000 E 53	0.5000 C 24
Buffalo, N.Y.	-1.2000 D 45	-2.5000 C 37	0.1000 D 48
Chicago, Ill.	-1.8167 E 64	-3.3333 E 61	0.3000 E 64
Cincinnati, Ohio-Ky.-Ind.	-1.0333 C 30	-2.1667 B 8	0.1000 D 47
Cleveland, Ohio	-1.4250 E 60	-3.2000 E 62	0.3500 C 35
Columbus, Ohio	-1.0917 C 38	-2.4333 C 31	0.2500 D 38
Dallas, Texas	-.9083 B 21	-2.2667 B 16	0.4500 C 29
Dayton, Ohio	-1.3167 D 56	-3.1333 E 60	0.5000 C 23
Denver, Colo.	.9917 C 24	-2.6333 D 47	0.6500 B 15
Detroit, Mich.	-1.7250 E 63	-3.4000 E 64	0.0500 D 55
Fort Lauderdale-Hollywood, Fla.	-1.0833 C 36	-2.7667 D 53	0.6000 B 19
Fort Worth, Texas	.8583 B 18	-2.1667 B 9	0.4500 C 28
Gary-Hammond-East Chicago, Ind.	-1.1750 D 43	-2.3000 B 23	0.0500 D 54
Grand Rapids, Mich.	-1.0333 C 31	-2.2667 B 17	0.2000 D 40
Greensboro-Winston-Salem-High Point, N.C.	-1.3000 D 54	-2.4000 B 28	-0.2000 E 63
Hartford, Conn.	-1.1250 C 40	-2.5000 C 38	0.2500 C 37
Honolulu, Hawaii	-.4583 A 4	-2.0667 A 4	1.1500 A 10
Houston, Texas	-1.0000 C 26	-2.1000 A 5	0.1000 D 46
Indianapolis, Ind.	-1.5250 E 61	-3.2000 E 63	0.1500 D 43
Jacksonville, Fla.	-1.2500 D 49	-2.3000 B 24	-0.2000 D 62
Jersey City, N.J.	-1.0167 C 27	-2.4333 C 32	0.4000 C 33
Kansas City, Mo. - Ks.	-1.1250 C 39	-2.3000 B 25	0.0500 D 49
Los Angeles-Long Beach, Ca.	-1.0583 C 34	-2.9667 E 57	0.8500 D 13
Louisville, Ky.-Ind.	-1.4167 E 58	-2.9667 E 57	-0.2000 E 61
Memphis, Tenn.-Ark.	-1.2083 D 47	-2.3667 B 26	-0.0500 D 53
Miami, Fla.	-.4167 A 3	-2.4333 C 33	1.6000 A 4
Milwaukee, Wis.	-1.0417 C 32	-2.2333 B 13	0.1500 D 42
Minneapolis-St. Paul, Minn.	-.9000 B 20	-2.1000 A 6	0.3000 C 36
Nashville-Davidson, Tenn.	-1.0833 C 37	-2.2667 B 18	0.1000 D 45
New Orleans, La.	-1.2667 D 51	-2.5333 C 42	0.0000 D 51
New York, N.Y.	-1.3333 D 57	-3.0667 E 59	0.4000 C 32
Newark, N.J.	-1.2000 D 46	-2.8000 D 55	0.4000 C 31
Norfolk-Portsmouth, Va.	-.8667 B 19	-2.2333 B 14	0.5000 C 22
Oklahoma City, Okla.	-.8250 B 15	-2.2000 B 10	0.5500 C 21
Omaha, Nebraska-Iowa	-1.3083 D 55	-2.5667 C 44	-0.0500 D 52
Paterson-Clifton-Passaic, N.J.	-1.0000 C 25	-2.4000 B 29	0.4000 C 30
Philadelphia, Pa.-N.J.	-1.0250 C 28	-2.5000 C 39	0.4500 C 27
Phoenix, Airz.	-.5917 A 8	-2.6333 D 49	1.4500 A 5
Pittsburgh, Pa.	-1.8667 E 65	-3.5333 E 65	-0.2000 E 60
Portland, Oreg.-Wash.	-.6500 A 11	-2.5000 C 40	1.2000 A 9
Providence-Pawtucket-Warwick, R.I.-Mass.	-.7667 B 14	-2.4333 C 34	0.9000 B 12
Richmond, Va.	-1.1333 D 41	-2.3667 B 27	0.1000 D 44
Rochester, N.Y.	-.7000 B 13	-2.0000 A 3	0.6000 B 18
Sacramento, Ca.	-.2000 A 1	-2.2000 B 11	1.8000 A 2
St. Louis, Mo.-Ill.	-1.5833 E 62	-2.7667 D 54	-0.4000 E 65
Salt Lake City, Utah	-1.0250 C 29	-2.5000 C 41	0.4500 C 26
San Antonio, Texas	-.8333 B 17	-1.8667 A 1	0.2000 D 39
San Bernadino-Riverside-Ontario, Ca.	.4750 A 5	-2.8000 D 56	1.8500 A 1
San Diego, Ca.	-.5333 A 6	-2.2667 B 19	1.2000 A 8
San Francisco-Oakland, Ca.	-.7000 B 12	-2.6000 C 45	1.2000 A 7
San Jose, Ca.	-.5333 A 7	-2.2667 B 20	1.2000 A 6
Seattle-Everett, Wa.	-.2667 A 2	-2.1333 A 7	1.6000 A 3
Spring-Chicopee-Holyoke, Mass.-Conn.	-.6167 A 10	-2.2333 B 15	1.0000 A 11
Syracuse, N.Y.	-1.1500 D 42	-2.2000 B 12	-0.1000 E 57
Tampa-St. Petersburg, Fla.	-1.0583 C 35	-2.4667 C 36	0.3500 C 34
Toledo, Ohio-Mich.	-1.1833 D 44	-2.2667 B 21	-0.1000 E 56
Washington, D.C.-Md.-Va.	.8333 B 16	-2.2667 B 22	0.6000 B 17
Youngstown-Warren, Ohio	-.9667 C 22	-2.5333 C 43	0.6000 B 16
Mean (X) =	-1.0342	-2.5015	-0.4331
Standard Deviation (s) =	0.3452	0.3577	0.5292

A = Outstanding (≤ x t s) •
B = Excellent (X + .28s ≤ B < $\bar{x}$ + s)
C = Good (X - .28s < C < X + .28s)
D = Adequate ($\bar{X}$ - s < D ≤ $\bar{x}$ = .28s)
E = Substandard (≤ X - s)

A STUDY OF THE FEASIBILITY OF HOPKINS RANDOMIZED RESPONSE MODELS

P. T. Liu, Johns Hopkins University
C. N. Chen, Taipei Family Planning Promotion Center
L. P. Chow, Johns Hopkins University

1. Introduction

Social and demographic surveys generally use interviews as a major tool of measurement. With increasing sophistication of modern society, and increasing fear of people of invasion of privacy, more and more difficulty is being experienced in interviews. Such difficulty is particularly serious when questions asked in a survey are considered personal and sensitive. Ironically, many of the contemporary socio-demographic problems for which surveys are attempted are sensitive; e.g., criminal abortion, teenage pregnancy, pre-marital or extra-marital sexual relationship, deliquency, truancy, use of drugs and crime.

Epidemiological and health surveys may encounter similar difficulties because of social stigma attached to some specific health conditions: leprosy, tuberculosis, alcholism and psychiatric conditions, just to mention a few.

To overcome such difficulties and to enlist greater cooperation from the respondents, Warner in 1965 developed an innovative survey technique which he called the "randomized response technique" (RRT). (1) Considerable theoretical work and some field studies of limited scope have since been undertaken by various investigators in the United States (2) and elsewhere. (3)

The RRT, although promising as an innovative survey technique, does have an inherent weakness; its efficiency of estimate is substantially lower than that of a conventional survey of comparable scope asking direct questions.

The authors have undertaken a series of studies including field tests to improve the efficiency of the RRT, and one of the models proposed is the multiple trial per respondent.(4) Theoretically speaking, the number of trials per respondent can be increased indefinitely so that the variance of estimate is reduced to that of a similar survey asking direct questions. In practice, such an increase can be done only at the cost of compromising the confidentiality of the respondent's identity; her cooperation, therefore, might deteriorate. As a solution to the above dilemma, the multiple answer model has been developed which has been given a code name of Hopkins RRT Model II. (5)

Other works of the authors on the subject include the development of two discrete quantitative models, which have been coded Hopkins RRT III (6) and IV (7).

Field studies to test the feasibility of these models have since been undertaken, and the current paper presents the experience with use of the multiple answer model in the field.

2. The Study

The field study was conducted in Taichung, Taiwan, during August 1973. One local area in Taichung City, the South District (urban), and another in the adjacent county, Wu-feng township (semi-urbanized) were purposely selected for the study. From the former area, a total of 180 married women, age 20-44 were randomly chosen, and from the latter, the sample size was 173. Successful interviews were conducted for 150 women in each of the study areas.

In addition, 53 names of women who were known to have had an induced abortion during the past 12 months were provided by an Ob-Gyn doctor in Taichung City, who is a close associate of the authors. All of these women lived in the city. By matching with respect to age and parity, 53 "matches" were chosen from among a list of post-partum women in the city. The names of the abortees and the matches were then mixed and randomly assigned to the interviewers who were kept "blind" as to who were known abortion cases and who were the matches. A total of 45 successful interviews were conducted in each group.

The randomizing device used in the field trial contained 50 balls of identical size; 15 red and 35 white ($\underline{p} = 0.3$). It is designed in such a way that five balls will drop into the neck at each trial. Each respondent was asked to repeat the procedures three times; it was a three-trial of five answer model (Figure 1).

3. The Results

The results of the study, the proportion of the respondents who were estimated to have had an abortion (abortion rate), and the standard errors of estimates are shown in Table 1.

It will be noted that the estimated abortion rates obtained with the multiple answer RRT model in the South District and Wu-feng Towhsnip combined, varied from 22.5% in the first trial to 35.8% in the third. The standard errors of estimates also varied from 3.54% to 3.90%. By pooling the results of three trials, the estimated abortion rate was 27.8%, with the standard error being substantially reduced to 2.83%.

The pooled estimate of 27.8% is roughly comparable to the rate obtained by the Taoyuan RRT Study, and is significantly higher than the rates obtained by any of the island-wide KAP surveys conducted previously. At the 1973 island-wide KAP survey, a highest abortion rate of 19.5% was obtained; this, however, was significantly lower than the pooled estimate obtained with this RRT model ($p < 0.005$).

The abortion rate obtained from the 45 known abortion cases was 75.0%. Although its

95% confidence interval failed to comprise the expected value of 100%, it came close to it. An earlier study in Taoyuan revealed that women, not infrequently, failed to report the abortion experienced during the past three months correctly. There were errors, both in reporting the timing of abortion and in reporting of the abortion event itself.

It can be seen that the efficiency of this multiple answer model is significantly higher than the conventional RRT model used under the Taoyuan RRT Study in which the single trial RRT model was used.

Further analyses on the differential abortion rates by selected demographic variables, such as urban-rural, age and parity, agreed with expectation except that the correlation between abortion rate and education appeared to be negative; women of no formal education showing the highest abortion rate (Table 2). This rather unexpected result was caused partly by the difference in age composition and partly due to a smaller sample size in each category. A study of larger scale should be undertaken for more conclusive results.

Immediately after the RRT trials, a post-RRT interview was conducted on each respondent by asking nine questions about her impression of the RRT. The results indicated that the RRT is feasible and procedures can be understood by most of the respondents (Table 3). Also, with this device, most of them indicated willingness to respond truthfully to a question which is far more sensitive than a question on induced abortion (Table 4).

Surprisingly, although close to 90% of the respondents said they felt that there might be a gimmick in the RRT, only 20% of them indicated that more than half or most of their neighbors and friends would feel that there is a gimmick (Table 5). One possible explanation of this inconsistency of response is the ambiguity in the wording of the question in Chinese language; the question may be interpreted as asking if there is some "mechanism which is inducing or soliciting the respondent to answer truthfully" and some respondents might think that the RRT, indeed, is a device to induce people to answer honestly.

4. Discussion

The interest of social researchers in randomized response technique seems to have increased recently. Some epidemiologists have also started exploring the feasibility of its use in various health and epidemiological surveys.

The RRT, has also invited some skepticism from other researchers. Some of their reservations, however, are based not so much on their experience with, but rather on their perception of, the utility of the technique. A few stories have also been circulated about unsatisfactory experiences with the use of RRT. While theoretical development of the RRT is being continuously pursued, efforts for improvement of the feasibility of the technique is called for.

Adequacy of interviewers is particularly critical in conducting surveys with the RRT. If interviewers are ill-trained and ill-prepared, if they themselves do not understand and are unconvinced of the technique, they will not be able to secure good cooperation from the respondents. The surprisingly high "non-response rate" experienced by the Taoyuan study (3), we suspect, was due partly to this factor. In the current field study, there was little difficulty reported by the interviewers in securing the full cooperation of the respondents.

Unsatisfactory experience has also been reported from some investigators with regard to use of the two unrelated question RRT model. Although the latter model is generally considered more efficient than a corresponding related question model, formulation of an adequate innocuous question is harder than one might expect. Asking, for example, "Were you born in April?" type of question is unsatisfactory because some respondents might fear possible revelation of sensitive information. Similarly, a question such as, "Was your mother born in April?" may not work simply because not all the respondents would remember the birthdays of their mothers. Another criticism is that more information may be lost rather than gained by use of RRT because the information obtained cannot be correlated, for example, with the individual's characteristics. This is true, but this fact provides the basic argument for use of RRT in protection of privacy. This weakness of RRT, however, is not critical; it is always possible to analyze the data based on smaller groups of specific characteristics.

There are a number of ways by which RRT can be made more feasible. Careful selection of the sensitive problem for study with RRT, careful formulation of the sensitive and innocuous questions, better method of presenting and administering the RRT questions, and use of less sophisticated randomizing devices are some of such possibilities. RRT may also be used, not to "replace", but rather to "supplement" the direct questioning.

The RRT is developed on the basic assumption that when the respondent's anonymity is assured, the respondent will be more willing to answer a sensitive question truthfully. This assumption is logical but needs more empirical evidence for support. There may be considerable variation in the response pattern to RRT's among people of different cultural or educational backgrounds, calling for more field tests in various countries.

References

1. Warner, S. L., "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," Journal of the American Statistical Association, 60(309):63-69, March 1965.

2. Abernathy, J. R., Greenberg, B. G. and Horvitz, D. G., "Estimates of Induced Abortion in the Urban North Carolina," Demography, 7:19-29, 1970.
3. Chi, I. C., Chow, L. P. and Rider, R. V., "The Randomized Response Techniques as Used in the Taiwan Outcome of Pregnancy Study," Studies in Family Planning, 3, No. 11, November 1972.
4. Liu, P. T. and Chow, L. P., "The Efficiency of the Multiple Trial Randomized Response Technique," Biometrics, (in press).
5. Chow, L. P., Liu, P. T. and Mosley, W. H., "New Randomized Response Technique for Study of Contemporary Social Problems," (Paper presented at the annual meeting of the American Public Health Association, November 4-8, 1973.)
6. Liu, P. T., Chow, L. P. and Mosely, W. H., "Use of the Randomized Response Technique with a new Randomized Device," Journal of the American Statistical Association, 70(350):329-332, June 1975.
7. Liu, P. T. and Chow, L. P., "A New Discrete Quantitative Randomized Response Model," Journal of the American Statistical Association, 71(353):72-73, March 1976.

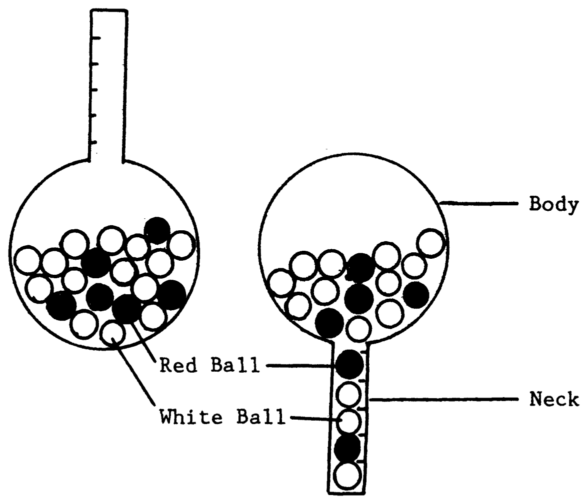


Figure 1. The Randomizing Device for the Multiple Answer RRT Model: Hopkins RRT II

TABLE 1

Proportion of Married Women of Childbearing Age in Taiwan Who Have Had Abortion: Comparison of the Results Obtained with RRT (Multiple Answer RRT Model and Single Trial of Conventional RRT Model) and Direct Questioning

Type of Survey	Proportion Having Had Abortion (%)	S. E. (%)	Sample Size	Year of Survey	Type of Survey
Randomized Response Technique:					
1. Multiple Answer Model: Taichung*					
(1) South District & Wu-feng Combined:					
(a) First trial	22.5	3.54	300	1973	Two townships
(b) Second trial	24.8	3.62	"	"	Married Women(MW)20-44
(c) Third trial	35.8	3.90	"	"	
(d) Pooled results: 3 trials	27.8	2.83	"	"	
(2) Known Abortion Cases: 3 trials	75.0	5.42	45	"	Women who are known to have had an abortion
(3) Matches of the Known Abortion Cases: 3 trials	36.0	5.16	45	"	Matches of the above known abortion cases
2. Single Trial Model: Taoyuan	28.2	4.69	692	1971	One County, MW 15-49
Direct Questioning:					
1. Island-wide KAP III	13.8	0.68	2558	1970	Taiwan, MW 15-44
2. Island-wide KAP IV	19.5	0.53	5588	1973	" " "
3. Taoyuan Abortion Study:					
(a) Repeated Interview:					
At Round 1	8.4	0.64	1861	1970	One County, MW 15-49
At Round 8	14.0	0.83	1737	1971	" " " "
(b) One-Shot KAP	12.7	1.00	1102	1971	" " " "

*The current study

TABLE 2

Proportion of Respondents Who Have Had Abortion Estimated by the Multiple Answer RRT Model and the Standard Error of Estimates by Selected Demographic Variables South District and Wu-feng Township, Taichung, Taiwan

Demographic Variable	Sample Size	Proportion Who Have Had An Abortion (Π)	S. E.
A. Urban-Rural:			
Urban*	150	0.360	0.0319
Semi-rural**	150	0.195	0.0279
B. Age Group:			
24 -	47	0.110	0.0428
25 - 34	137	0.255	0.0312
35 +	116	0.370	0.0364
C. Number of Live Births:			
0 - 1	44	0.045	0.0351
2 - 3	134	0.285	0.0323
4 +	122	0.350	0.0352
D. Education:			
No Formal	63	0.325	0.0484
Primary	186	0.270	0.0271
Jr. High +	51	0.240	0.0504
Total	300	0.278	0.0283

Note: Results of three trials combined
* South District of Taichung City
** Wu-feng township of Taichung County

TABLE 3

Difficulty or Ease in Understanding the RRT Procedures by Respondents Themselves and their Perception of Understanding of Neighbors/Friends by their Level of Education (In Percentage)

Understanding	Level of Education: No Formal	Primary	Jr. High +	Total
		Respondents Themselves		
Very easy	9.2	31.4	44.5	29.3
Easy	50.6	58.4	51.8	55.3
Difficult	25.3	6.6	3.6	10.1
Very Difficult	14.9	3.5	0.0	5.3
		Neighbors and Friends		
Will be understood by:				
Almost all	12.6	38.9	62.6	38.1
More than half	47.1	50.9	32.5	46.2
Less than half	28.7	5.8	1.2	9.8
Very few	9.2	4.0	2.4	4.7
Don't know	2.3	0.4	1.2	1.0
Total	100.0	100.0	100.0	100.0

Note: Total number of respondents was 396

TABLE 4

Feeling Safe to Answer a More Sensitive Question Such as Committing a Crime with the RRT

Response	Level of Education: No Formal	Primary	Jr. High +	Total
		Respondents Themselves		
Absolutely safe	47.1	52.7	62.7	53.5
Will respond after some hesitation	18.4	19.9	27.7	21.2
Will not take the risk	33.3	27.4	9.6	25.0
Don't Know	1.1	0.0	0.0	0.3
		Neighbors and Friends		
Most will	21.8	30.5	36.1	29.8
More than half	23.0	27.9	37.3	28.8
Less than half	19.5	13.7	10.8	14.4
Very few	33.3	23.0	10.8	22.7
None	1.1	4.4	2.4	3.3
Don't know	1.1	0.4	2.4	1.0
Total	100.0	100.0	100.0	100.0

TABLE 5

Do You Feel That There is a Gimmick in the RRT Method?

Response	No Formal	Primary	Jr. High +	Total
		Level of Education		
		Respondents Themselves		
Yes	87.3	88.1	90.4	88.4
No	12.7	11.9	9.6	11.6
		Neighbors and Friends		
Most	4.6	8.4	3.6	6.6
More than half	13.8	12.8	14.5	13.4
Less than half	23.0	18.6	13.3	18.4
Very few	34.5	45.6	54.2	44.9
None	18.4	12.8	13.3	14.1
Don't know	5.7	1.8	1.2	2.5
Total	100.0	100.0	100.0	100.0

GEOGRAPHICAL MOBILITY AS A SOCIAL INDICATOR: AN INTERNATIONAL COMPARISON

Larry H. Long
Celia G. Boertlein

U. S. Bureau of the Census

A great many people look upon social indicators simply as measures of the fundamental characteristics of a society, with an emphasis on trends in these measures. Such a broad concept raises the question of how to conceptualize and then measure these "fundamental characteristics." Ask any demographer to identify the basic demographic variables, and he or she will immediately answer that these are fertility, mortality, and migration. Of the three, fertility and mortality have been given extensive coverage in reports on social indicators that have been prepared by statistical offices in different countries.

In fact, the volumes on social indicators usually give numerous measures of fertility and mortality and include data from other countries in order to facilitate international comparisons of fertility and mortality. But no report on social indicators has ever included a measure of geographical mobility that can be compared with other countries. The reason for the lack of international comparisons of geographical mobility is simple: in countries that collect such statistics, migration is usually operationally defined as movement that crosses the boundary of some administrative area that seems convenient for the purposes of data collection.

The result is that we have statistics on movements between counties in the United States, between gemeinden in Germany, between parishes (församlingar) in Sweden, between boroughs and other local areas in England, and between local administrative areas in a number of other countries. The problem is that these areas vary greatly in size, shape, and significance from country to country and no one has devised a least common denominator that will enable one to compare moves between local areas in different countries. A great many statisticians, geographers, and other social scientists have worked on this problem without solving it. Indeed, the problem may never be solved.

An alternative that has been used in an earlier study [1] is simply to count all moves and not just those that involve crossing the arbitrary boundaries of local administrative areas. In a sense, the most basic measure of geographical mobility is the count of all moves from one residence to another during a specified interval of time. This measure represents the total amount of geographical mobility taking place in a country and can be almost universally applied.

The remainder of this paper evaluates recent developments in the production of data on residential mobility in different countries. The conclusion of the paper points out some important international differences in social structure that may be reflected by different levels of geographical mobility.

Residential Mobility

Data on residential mobility can rather easily be obtained with census or survey questions of the type: "Did you live at this address one year ago (or five years ago), on _____ (date)?" The answer is either "yes" or "no." In the United States data of this type have been available since 1948, and at least seven other countries now collect and publish statistics on residential mobility.

The resulting data (shown in Table 1 at the end of the paper) confirm what most people have long suspected but could not prove: people in the United States really do change residence more frequently than Europeans. But the interesting result from the new data is the demonstration that Canadians and Australians change residence as frequently as Americans do. A frequently cited statistic is that about 20 percent of the U. S. population moves in one year. When we exclude movement from abroad, this figure is reduced to 18.6 percent and compares with about 11 or 12 percent in Great Britain and Japan, 9.1 percent in Taiwan, and 4.3 percent in Ireland. Over a five-year period, over 45 percent of the U. S. population changes residence at least once, and this figure is about the same in Canada and Australia. About 36 percent of the populations of Japan and Great Britain move in five years.

Because of differences in coverage and in processing of data, great significance should not be attached to small differences. Some of these small differences in census practices can be corrected for by adjusting U.S. data (see [3]). Furthermore, some of the differences implied in Table 1 represent differences in age structure; Britain, for example, has a very old population relative to Japan's. The differences arising from the varying age compositions can also be taken into account [3]. Finally, small differences can also arise from the fact that the one-year data for the United States and Australia are derived from surveys, whereas the other data come from censuses or large samples of population taken in connection with censuses.

When all these considerations are taken into account the big picture remains: that industrialized nations vary much more in terms of residential mobility than in terms of almost any of the usual variables that are included in reports of social indicators. Certainly the differences among countries shown in Table 1 are many times greater than the fertility or mortality differentials which receive so much attention. In fact, in very few ways do industrialized nations differ to a greater degree than in terms of the geographical mobility of their populations.

One should note that the length of the interval over which mobility is measured makes an important difference, for the longer the interval, the smaller the differences among countries will appear. A long interval tends to mask the effect of repeat moves.

In spite of large differences, several interesting regularities appear in the geographical mobility patterns in the different countries. In each country the rate of moving is high among young children. The rate reaches a low point around age 15 or 16 and then rapidly rises to a peak around age 22. In the United States this peak represents a mobility rate of around 50 percent; that is, about 50 percent of Americans can be expected to change residence between their twenty-second and twenty-third birthdays. After this peak is reached, rates of moving steadily decline with age, except for a small increase in the rate around age 65.

An important point to take note of is the tendency for differences among countries to be least at the ages when mobility is highest. That is, there are almost universal forces which cause children to leave their parental home in their late teens to go to work, to go away to continue their education, or to get married. At any rate, differences among countries in terms of residential mobility are least between the late teens and the late twenties.

Number of Moves in a Lifetime

Rates of moving at each age allow us to calculate the probable number of moves a person can expect to make in a lifetime. The methodology is discussed in [3]. The rates shown in Table 1 imply that an average resident of the United States, Canada, or Australia probably makes 12 to 13 moves in a lifetime, compared with about 8 moves per person in Great Britain, 7 moves per person in Japan, 6 moves per person in Taiwan, and 3 or 4 moves in a lifetime for an average resident of Ireland.

From these data, one can see that in the high-mobility countries--the United States, Canada, and Australia--a person probably makes between three and four times as many moves in a lifetime as the average resident of Ireland. Once again, one can see that differences in levels of geographical mobility are far greater than most other demographic, social, and economic differences ordinarily noted in reports on social indicators.

The differences can be highlighted even more dramatically than the comparisons made above. Of the 12 or 13 moves an average American, Canadian, or Australian makes in a lifetime, about three will be made as a child moving with one's parents. This figure is nearly equal to the number of moves an average resident of Ireland will make in a lifetime.

The low rate of residential mobility in Ireland cannot be attributed to Ireland's being a rural country. Controlling for this factor, we still find very low rates of residential mobility in Ireland. We may note that although the rate of moving for Ireland as a whole is about 5.1 percent (including movers from abroad), the rate is highest--7.5 percent--in Dublin county, which is a reasonable approximation to the metropolitan area of the city of Dublin. In metropolitan areas of the United States, the annual rate of moving sometimes reached 30 percent around 1970.

Hence, the general differences among the countries in Table 1 indicate a degree of differences that roughly applies to both the metropolitan and nonmetropolitan parts of countries, even though the metropolitan areas have higher rates of residential mobility than the nonmetropolitan areas.

Another way of standardizing for the effect of differences in population composition is to examine rates of moving specific for occupation. These data are not yet available for 1970, but 1960 data [2] for the United States, Canada, Great Britain, and Japan indicate that the high-mobility countries have the highest rates at each occupational level. If the 1970 data were available, they would show that white-collar workers in each of the countries expect to move more frequently than other workers.

Long- and Short-Distance Moves

The differences in rates of residential mobility imply higher rates of movement over both short and long distances. That is, the higher rates of residential mobility in the United States, Canada, and Australia reflect a greater propensity to move short distances as well as a greater propensity to move long distances.

This conclusion is possible because the countries that collect statistics on residential mobility divide moves into those within and those between different administrative areas--e.g., counties and states in the United States. From this and other evidence we know that the volume of migration within a country varies inversely with the size of the areal units used to measure migration. And we can show that the areal units used in the United States (counties or states) are larger than the areal units used in other countries and also indicate a higher rate of migration in the United States.

This process can be extended to compare rates of migration in the United States with rates of inter-area movement in European countries. The process is simply to show that rates of inter-county or inter-state migration in the United States are higher than rates of migration in other countries in spite of the fact that the U.S. counties or states are larger than the areal units used in other countries to measure migration.

With this process, we cannot, however, be certain of the degree of difference among countries in the propensity to engage in short-distance as compared with long-distance movement. For the countries shown in Table 1, we can be reasonably certain that the high rates of residential mobility in the United States, Canada, and Australia reflect a higher propensity to move short distances as well as long distances. We make these observations partly to point out that geographical size of a country has relatively little to do with explaining the rate of residential mobility, for the high rates of residential mobility in the United States, Canada, and Australia reflect high rates of moving over short as well as long distances.

Trends in Geographical Mobility

Many persons seem to insist that a statistic can be considered a social indicator only if trend data are available. Data on residential mobility were first collected in 1948 in the United States and in 1960 and 1961 in Canada, Great Britain, and Japan. These data are shown in Table 2, along with the 1970 or 1971 figures. Because of differences in tabulation practices, we had to limit the 1960-70 comparisons to persons 15 years old and over, and for this reason the 1970 data in Table 2 are slightly different from the figures in Table 1.

The data clearly show stability in the rates of residential mobility in Canada and Great Britain. Residential mobility has increased in Japan, from a one-year rate of 9.5 percent in 1960 to 12.8 percent in 1970. This rise in residential mobility in Japan in the 1960's appears to be part of a gradual increase in internal migration that has been underway since the early 1950's. This increase in the volume of geographical mobility within Japan is associated with continued industrialization and movement of population from agricultural occupations.

The data in Table 2 may indicate the beginning of modest declines in the rate of residential mobility in the United States. Both the one-year and five-year comparisons show small decreases in the rate of residential mobility of the U.S. population.

The rate of migration within the United States increased greatly around the time of World War II (see [2]), but the annual data indicated no significant year-to-year changes in rates of moving over short- or long-distances between 1948 and the late 1960's, where there appeared some evidence that rates of moving might be declining. The apparent decline indicated in Table 2 is small, however, and difficult to interpret.

For the four countries that had data on residential mobility in 1960 and 1970 the differences have decreased. The convergence has been brought about as the lowest rate (in Japan) rose and the highest rate (in the United States) declined slightly.

Some Explanations

What do the United States, Canada, and Australia have in common that gives them a high rate of internal migration? There is no obvious answer to this question. If enough countries collected statistics on residential mobility, one could undertake multivariate analyses whereby the rate of moving could be correlated with a host of other national characteristics like those routinely published in the United Nations Demographic Yearbook. Because of the limited number of countries currently providing data on residential mobility, elaborate multivariate approaches are not possible, and we can only rely on ad hoc explanations of the observed differences.

Surely part of the explanation of high rates of migration in the United States, Canada, and Australia is the fact that each is a "nation of immigrants" and has attracted the geographically mobile segments of the populations of other countries. The effect of current immigration on rates of geographical mobility in the three countries is small, for even when we exclude movers from abroad, the United States, Canada, and Australia have high rates of geographical mobility.

High rates of internal migration in these countries may indicate a long-run dynamic built into populations which are almost entirely descended from long-distance migrants. Most persons in the United States, Canada, and Australia either crossed an ocean themselves or are descendents of persons who crossed an ocean. Past mobility may generate future mobility because of the exposure (direct or vicarious) with diverse places. Such knowledge about earlier migrations can make the possibility of mobility seem more readily apparent to potential movers. In a sense, Americans, Canadians, and Australians learn about mobility simply through knowledge of their ancestors.

We do know that persons who have moved once are likely to move again, and this effect may persist across generations. Hence, currently high rates of internal migration in the United States, Canada, and Australia may indicate the intergenerational transmission of a propensity toward migration.

Another characteristic shared by the United States, Canada, and Australia is a history dominated by a frontier that needed to be settled. Each country has taken measures like the Homestead Act in the United States to encourage settlement of the frontier. Each has also sought to exploit resources that were abundant but widely scattered. Throughout much of their history, the United States, Canada, and Australia have considered themselves underpopulated, and each has adopted policies to encourage persons to move to developing regions. These effects, too, may extend across generations.

A third characteristic of the United States, Canada, and Australia is that successive waves of settlement established numerous urban centers that were widely separated. Today each country has an urban structure that is not clearly dominated by one metropolis. Instead, numerous metropolitan areas in each country compete for industry and migrants, and this competition may keep the overall migration rate high. The decentralized urban pattern creates a diversity of regional markets, and many corporations in the United States have a practice of repeatedly moving their executives and managers from place to place in order to provide exposure to the many regional centers in which big corporations operate.

In these ways, sheer geographical size may contribute to a high rate of internal migration simply by offering more places to move to. But geographical size of a country has less explanatory power in accounting for rates of short-distance movement, and the United States, Canada, and Australia have high rates of short-distance movement as well as high rates of long-distance migration. Of course, it is possible that long-distance and short-distance moves are mutually reinforcing, so that a readiness to move long distances and a history of such movement is conducive to frequent short-distance movement. In other words, the short-distance mobility rate in the United States, Canada, and Australia may be high partly because these countries have high rates of long-distance migration, but these relationships cannot be statistically demonstrated.

We are thus left with the somewhat unsatisfactory conclusion that the United States, Canada, and Australia have high rates of short- and long-distance moving now because they have had high rates in the past. And they probably had high rates in the past because they were immigrant countries that attracted the geographically mobile element of European populations. These populations and their descendants have participated in successive waves of internal movements that established numerous regional population centers which have competed for migrants and thereby kept the level of mobility high.

Implications

If rates of geographical mobility are a social indicator, what are they indicating? A tendency in recent years has been to look upon geographical mobility as a force producing alienation and the breakdown of community structure. This theme appears in Alvin Toffler's Future Shock, which asserted that increased geographical mobility was part of the onrush of events which people were unable to adjust to. This theme is even more forcefully stated in Vance Packard's A Nation of Strangers, published in 1972 and long on the best-seller list. The same theme has appeared in subsequent books hypothesizing an increase in loneliness, including Ralph Keyes' We, the Lonely People (1972) and Suzanne Gordon's Lonely in America (1973).

There is little evidence to support such assertions simply because residential mobility rates are not increasing in the United States. On the contrary, they are probably beginning to decline. If the absence of an increase in rates of residential mobility were more widely known, perhaps there would be less of a tendency to link residential mobility with presumed increases in alienation or loneliness.

We should note, however, that high rates of residential mobility can affect the degree of participation and involvement in community activities. Some evidence to this effect was produced when in November 1974 the Census Bureau included for the first time a question on residential mobility and migration in its voting supplement to the Current Population Survey. The data [4] showed that persons who had recently moved were less likely to have registered and less likely to have voted than persons who had not moved.

The data from this survey showed that the higher-than-average rate of moving among persons in their early twenties explained part (but not all) of the lower-than-average voting rate among persons at this age group. If it is true that the proportion of eligible voters who actually vote is lower in the United States than in the democracies of Western Europe, then part of the explanation may be that at any given moment a larger proportion of the United States population consists of persons who are newcomers to a place. In these and other ways, residential mobility may influence age patterns of participation and differences among countries.

References

[1] Long, Larry H. "On Measuring Geographic Mobility," Journal of the American Statistical Association 65 (September, 1970), pp. 1195-1203.

[2] Long, Larry H. "Migration Differentials by Education and Occupation: Trends and Variations," Demography 10 (May, 1973), pp. 243-258.

[3] Long, Larry H. and Celia G. Boertlein, The Geographical Mobility of Americans: An International Comparison. U. S. Bureau of the Census, Current Population Reports, Series P-23, No. 64, U. S. Government Printing Office, 1976.

[4] U. S. Bureau of the Census. Current Population Reports, P-20, No. 293, "Voting and Registration in the Election of November 1974," U. S. Government Printing Office, 1976.

TABLE 1. Percent of Population Residentially Mobile in Seven Countries: Around 1970

Country	Percent moving in one year[1]		Percent moving in five years[2]	
	Including movers from abroad	Excluding movers from abroad	Including movers from abroad	Excluding movers from abroad
Australia	(NA)	15.7	51.4	48.4
Canada	(NA)	(NA)	46.6	44.3
Great Britain	11.8	11.1	37.2	35.9
Ireland	5.1	4.3	(NA)	(NA)
Japan	12.0	12.0	35.9	35.8
Taiwan	(NA)	9.1	(NA)	(NA)
United States	19.2	18.6	47.0	43.2

NA Not available.

[1]Persons one year old and over.

[2]Persons five years old and over.

TABLE 2. Percent of Population Residentially Mobile in Four Countries: Around 1960 and 1970

	1960	1970
ONE-YEAR INTERVAL		
Great Britain	11.9	11.6
Japan	9.5	12.8
United States	19.9	18.7
FIVE-YEAR INTERVAL		
Canada	46.0	47.0
Great Britain	36.2	36.6
United States	49.5	46.3

Note: Data refer to persons 15 years old and over and include movers from outside the countries.

A Multiple Logit Analysis of a Family Planning System

Robert W. Makuch and Daniel H. Freeman, Jr.
Yale University
Nancy S. Henley and Sagar C. Jain
U. of North Carolina

1. Introduction

This study comprises the second phase of an investigation by Henley et al. (1976) on data from a family planning program in the state of Haryana, India. Briefly, the study was designed to evaluate the extent to which various inputs, within the boundary of fixed constraints, affect program output. The output was the delivery of contraceptive devices; the input was both financial and technical; the constraints were socio-economic and demographic variables. Over 150 data variables were obtained for each of the 97 geographical blocks composing Haryana; however, many of these variables were eliminated on the basis of known unreliability, level of detail they represented, or large numbers of missing values. Thirty-seven independent variables and four measures of program output (IUD acceptors, tubectomies, vasectomies, and number of condoms distributed) remained. The first phase focused on summary statistics such as four-year averages across the period studied, 1968-1971. Five rural blocks and two urban blocks were dropped from the analysis due to 1 or more missing values; thus, 90 blocks remained for analysis. A maximum R-squared improvement technique was applied to ascertain the relative importance of the variables on the program output, and the objective was simply to obtain equations which had as many statistically significant variables as possible.

The second phase involved a multiple logit analysis which focused on the IUD acceptance rates to describe yearly patterns within the data. In particular, a linear model was constructed to examine both the relationship among the input variables in the context of fixed constraint variables, and the interaction between the constraint and year (trend) effect. The linear model was of the form

$$\underset{\sim}{K}\ln\underset{\sim}{\pi} \doteq \underset{\sim}{X}\underset{\sim}{\beta} \qquad (1)$$

where $\underset{\sim}{K}$, $\underset{\sim}{X}$ are known matrices subject to the condition that the matrix of first partials $\underset{\sim}{K}\ln\underset{\sim}{\pi}$ with respect to the elements of $\underset{\sim}{\pi}$ has full row rank and $\underset{\sim}{X}$ has full column rank, $\underset{\sim}{\pi}$ is a vector of multinomial probability parameters, $\underset{\sim}{\beta}$ is a vector of parameters to be estimated, $\ln\underset{\sim}{\pi}$ denotes the logarithm of the elements of the vector $\underset{\sim}{\pi}$, and '$\doteq$' means 'is approximated by'. The regression coefficients $\underset{\sim}{\beta}$ were estimated using a regression approach developed by Grizzle, et al. (1969). This methodology entails a weighted least squares approach to determine a BAN estimator $\hat{\underset{\sim}{\beta}}$ of $\underset{\sim}{\beta}$. Moreover, minimum modified chi-square statistics described by Neyman (1949) were employed to assess the adequacy of the model in fitting the data and to test hypotheses of interest.

The choice of independent variables (program inputs) was based on a preliminary optimal regression analysis (see Hocking, 1976), and the algorithm used for selecting input variables was described by Furnival and Wilson (1974). The procedure allowed the selection of the 'best' and several 'nearly best' subsets according to a specified criterion. Therfore, a comparison of the possible regression equations in terms of a summary statistic measuring the aspects of the adequacy of each equation was needed to aid in deciding which subsets to choose. The estimate C_p of the standardized total squared error Γ_p was chosen as the selection criterion. This estimate is defined as

$$C_p = \frac{RSS_p}{\hat{\sigma}^2} - (n-2p)$$

where p = number of variables in the reduced regression equation, n = number of data points, RSS_p is the residual sums of squares for the particular p-variate regression equation under consideration, and $\hat{\sigma}^2$ is the residual mean square from the full term p'-variate equation ($p' \geq p$). Employment of C_p statistics rather than the R-squared criterion was favored since C_p plots are more amenable to graphical analysis. Furthermore, this statistic served to confirm the observations of Henley et al. who utilized the R^2 statistic.

In summary, our analysis was two-fold. On the one hand, the optimal regression procedure was implemented so that the most important sets of input variables could be determined. The second part of the analysis embodied these results in constructing a linear model to represent a parsimonious, yet adequate, description of the data.

2. Overall (Optimal) Regression

The data consisted of responses gathered at four 1-year intervals for 85 urban and rural blocks. Based on the earlier investigation, 27 independent variables (defined in Table 1) were used for the overall regression. Observations on one block were assumed to be independent of those on every other block. Furthermore, a no-interaction linear model was postulated and a logit transformation of the IUD acceptance rate Y (see Table 1) served as the dependent variable. Residual plots were examined for each of the 27 independent variables and, as a result, log transformations of variables X4, X5, and X26 were performed. Finally, an inspection of the normal probability plots indicated that variables from this set of data could be regarded as samples from a normal distribution.

TABLE 1
Variables Considered for Use in Overall Regression

Variable	Variable Name
X1	Male illiteracy-percent-1971
X2	Female illiteracy-percent-1971
X3	Newspaper subscribers/1000 population
X4	Radios/1000 population (pop.)
X5	Agricultural composite
X6	% of total workers in agriculture-1971
X7	% vaids who motivate for family planning
X8	Dais/100,000 pop. (4 year average)
X9	Auxiliary nurse midwives (ANMs) - 1971
X10	Male physicians/100,000 pop.
X11	IUD referrals from 'other' sources/ total referrals (4 year average)
X12	Health expenditures/1000 pop. (4 year avg.)
X13	Extension educators/100,000 pop.
X14	Family planning field workers/100,000 pop. (4 year average)
X15	Auxiliary nurse midwives/100,000 pop. (4 year average)
X16	Hospital and PHC beds/100,000 pop. (4 year average)
X17	Family planning media events/100,000 pop. (4 year average)
X18	Family planning educational groups/100,000 pop. (4 year average)
X19	Km paved road - 1971
X20	Extension educators trained/in place-1971
X21	Female physicians/100,000 pop.
X22	Family planning expenditures/1000 pop. (4 year average)
X23	Dais trained/in place - 1971
X24	Sterilization referrals from sources 'other' than health and family planning sources/total sterilization referrals (4 year average)
X25	Field workers - 1971
X26	Population density - 1971
X27	% scheduled caste - 1971
Y	# IUD acceptors for 4 years/# of eligible couples in year 4

After this preliminary inspection of the data, simple correlation coefficients for all pairs of variables were computed to implement the all possible regressions procedure. It is noteworthy that over 96% of the coefficients were between -.3 and +.3, thereby raising little concern that problems due to multicollinearity would arise in this analysis. From the optimal regression procedure, the 'best' (in terms of lowest C_p value for all p) and several 'nearly best' subsets of independent variables for each size p, p=1,...,27, were determined. The elements of the most desirable subsets of each size p, p=1,...,11, are given in Figure 1. We note that most of the same variables reappear as the subset size increases. This seems to indicate that these variables were good regressors for the dependent variable, and this interpretation was useful when we constructed our model to analyze time trends.

FIGURE 1
Best and Nearly Best Subsets of Size p, p=1,...,11, and their elements

Set	Elements in Set (denoted by subscripts of variables in Table 1)
A	22
B	12 22
C	3 22
D	3 22 24
E	3 14 22
F	3 14 15 22
G	3 10 14 22
H	3 10 12 14 22
I	2 3 10 14 22
J	2 3 10 15 22
K	2 3 10 14 15 22
L	3 10 11 12 14 15 24
M	3 8 10 11 12 14 15
N	3 10 11 12 14 15 21 24
O	3 8 10 11 12 14 15 24
P	3 8 10 11 12 14 15 21 24
Q	2 3 10 11 12 14 15 21 24
R	2 3 8 10 11 12 14 15 21 24
S	2 3 10 11 12 14 15 21 24 27
T	2 3 4* 10 11 12 14 15 21 24
U	2 3 4* 6 10 11 12 14 15 21 24
V	2 3 4* 10 11 12 14 15 21 24 27
W	1 2 3 4* 10 11 12 14 15 21 24

* denotes a transformed variable

3. Modeling For Time Trends

Ten rural blocks were randomly chosen to illustrate a modeling procedure for time trends. Let n_{ij} be the number of IUD acceptors for year i (i=1,2,3,4) in block j (j=1,...,10), and N_j be the number of eligible couples in year 4 for block j. Then define $p_{ij} = n_{ij}/N_j$ to be the proportion of IUD acceptors for year i in block j, where

$$\underset{1\times 50}{\tilde{p}'} = (\tilde{p}_1', \ldots, \tilde{p}_{10}')$$

and

$$\tilde{p}_j' = (p_{1j} \; \cdots \; p_{4j} \;\; 1-\sum_{i=1}^{4} p_{ij}) \; .$$

Choose

$$\underset{40\times 50}{\tilde{K}} = \tilde{K}^* \otimes \tilde{I}_{10\times 10} \; , \qquad \text{where}$$

$$\tilde{K}^* = \begin{bmatrix} 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

and $\otimes$ denotes the Kronecker product, to form the vector

$$\underset{40\times1}{\ell} = K\ln p' = \begin{bmatrix} \ell_{1,1} \\ \ell_{2,1} \\ \ell_{3,1} \\ \ell_{4,1} \\ \ell_{1,2} \\ \vdots \\ \ell_{4,10} \end{bmatrix} .$$

Each $\ell_{i,j} = \ln \dfrac{p_{ij}}{1-\sum_{i=1}^{4} p_{ij}}$ corresponds to an element of the multiple logit response vector ℓ for year i in block j, and variation among these elements can be investigated by fitting linear regression models defined in (1). Briefly stated, the modeling procedure can be characterized by writing

$$E_A\{K\ln p\} = K\ln\pi \doteq X\beta$$

where 'E_A' means 'asymptotic expectation', so that BAN estimates $\hat{\beta}$ of β and minimum modified χ^2 statistics for testing hypotheses of interest by the method of weighted least squares can be obtained.

Patterns in the data suggested that 1) the proportion of IUD acceptors decreased as we moved from year 1 to year 4 and 2) these yearly trends were different in some blocks. Taking these patterns into account, the vector ℓ was fit to the right hand side of (1) by constructing the design matrix X_1. The first ten columns of X_1 spanned the space of all constraint variables, and the next thirty columns of ten identical matrices pertained to the constraint-by-year, or trend, space. Hence, our model initially lacked any input variables since the rank of X_1 was equal to the number of responses (i.e. the model was saturated).

$$\underset{40\times40}{X_1} = \left[\begin{array}{cccccccccccccccccccccccc}
1&0&0&0&0&0&0&0&0&0&0&0&0&&&&&&&&&&&\\
1&0&0&0&0&0&0&0&0&0&1&0&0&&&&&&&&&&&\\
1&0&0&0&0&0&0&0&0&0&1&1&0&&&&&&&&&&\underset{\sim}{0}&\\
1&0&0&0&0&0&0&0&0&0&1&1&1&&&&&&&&&&&\\
1&1&0&0&0&0&0&0&0&0&&&&0&0&0&&&&&&&&\\
1&1&0&0&0&0&0&0&0&0&&&&1&0&0&&&&&&&&\\
1&1&0&0&0&0&0&0&0&0&&&&1&1&0&&&&&&&&\\
1&1&0&0&0&0&0&0&0&0&&&&1&1&1&&&&&&&&\\
1&0&1&0&0&0&0&0&0&0&&&&&&&0&0&0&&&&&\\
1&0&1&0&0&0&0&0&0&0&&&&&&&1&0&0&&&&&\\
1&0&1&0&0&0&0&0&0&0&&&&&&&1&1&0&&&&&\\
1&0&1&0&0&0&0&0&0&0&&&&&&&1&1&1&&&&&\\
1&0&0&1&0&0&0&0&0&0&&&&&&&&&&\cdots&&&&\\
1&0&0&1&0&0&0&0&0&0&&&&&&&&&&\cdots&&&&\\
1&0&0&1&0&0&0&0&0&0&&&&&&&&&&\cdots&&&&\\
1&0&0&1&0&0&0&0&0&0&&&&&&&&&&\cdots&&&&\\
1&0&0&0&1&0&0&0&0&0&&&&&&&&&&&&&&\\
1&0&0&0&1&0&0&0&0&0&&&&&&&&&&&&&&\\
1&0&0&0&1&0&0&0&0&0&&&&&&&&&&&&&&\\
1&0&0&0&1&0&0&0&0&0&&&&&&&&&&&&&&\\
1&0&0&0&0&1&0&0&0&0&&&&&&&&&&&\cdots&&&\\
1&0&0&0&0&1&0&0&0&0&&&&&&&&&&&\cdots&&&\\
1&0&0&0&0&1&0&0&0&0&&&&&&&&&&&\cdots&&&\\
1&0&0&0&0&1&0&0&0&0&&&&&&&&&&&\cdots&&&\\
1&0&0&0&0&0&1&0&0&0&&&&&&&&&&&&&&\\
1&0&0&0&0&0&1&0&0&0&&&&&&&&&&&&&&\\
1&0&0&0&0&0&1&0&0&0&&&&&\underset{\sim}{0}&&&&&&&&&\\
1&0&0&0&0&0&1&0&0&0&&&&&&&&&&&&&&\\
1&0&0&0&0&0&0&1&0&0&&&&&&&&&&&&&&\\
1&0&0&0&0&0&0&1&0&0&&&&&&&&&&&&&&\\
1&0&0&0&0&0&0&1&0&0&&&&&&&&&&&&&&\\
1&0&0&0&0&0&0&1&0&0&&&&&&&&&&&&&&\\
1&0&0&0&0&0&0&0&1&0&&&&&&&&&&&&&&\\
1&0&0&0&0&0&0&0&1&0&&&&&&&&&&&&&&\\
1&0&0&0&0&0&0&0&1&0&&&&&&&&&&&&&&\\
1&0&0&0&0&0&0&0&1&0&&&&&&&&&&&&&&\\
1&0&0&0&0&0&0&0&0&1&&&&&&&&&&&&0&0&0\\
1&0&0&0&0&0&0&0&0&1&&&&&&&&&&&&1&0&0\\
1&0&0&0&0&0&0&0&0&1&&&&&&&&&&&&1&1&0\\
1&0&0&0&0&0&0&0&0&1&&&&&&&&&&&&1&1&1
\end{array}\right]$$

Since one aim of our analysis was to explain away as much variation in the data as possible using the program inputs, some of the ten rural blocks were grouped together to allow for the addition of these input variables to the model. The chi-square test statistics in Table 2 served as a guide in this grouping procedure.

TABLE 2
Test Statistics for Preliminary X_1 Model

Source of Variation		Degrees of Freedom	χ^2	p
Block	1 vs 2	1	36.74	0
	2 vs 3	1	25.88	0
	3 vs 4	1	9.75	0
	4 vs 5	1	.13	.72
	5 vs 6	1	11.68	0
	6 vs 7	1	40.00	0
	7 vs 8	1	.97	.32
	8 vs 9	1	14.07	0
	9 vs 10	1	20.27	0

One possible grouping is given in Table 3. Another possible grouping arrangement was identical to this grouping except that the last two blocks were not combined into one group. However, in order to utilize as many program inputs as possible, the combination of ten blocks into five groups as indicated in Table 3 was chosen for further analysis.

TABLE 3

Number of Rural Blocks in Each Group	Range of Overall Non-acceptance Rate Within Each Group
1	85%
1	90%
4	90.4%-91.4%
2	93.7%-94.6%
2	94.7%-95.6%

Based on the overall regression analysis, 9 input variables were considered. They were:

1. Family planning expenditures
2. Family planning field workers
3. Auxiliary nurse midwives
4. Female medical doctors
5. Sterility referrals
6. Health expenditures
7. Dais
8. Family planning media events
9. Extension educators

Since data on these variables were collected for each of the four years studied, 36 possible regressors were available. However, the overall regression suggested that family planning expenditures (X22), family planning field workers (X14), auxiliary nurse midwives (X15), and health expenditures (X12) were good candidates for year-by-variable interaction. Adopting this suggestion, we introduced parameters into the model for each of the four years for each of these four variables.

$$
\underset{\sim}{X}_2 = \underset{40\times 40}{} \left[\begin{array}{llllll}
1\,0\,0\,0\,0 & 0\,0\,0 & & & & \\
1\,0\,0\,0\,0 & 1\,0\,0 & & & & \\
1\,0\,0\,0\,0 & 1\,1\,0 & & & \underset{\sim}{0} & \\
1\,0\,0\,0\,0 & 1\,1\,1 & & & & \\
1\,1\,0\,0\,0 & & 0\,0\,0 & & & \\
1\,1\,0\,0\,0 & & 1\,0\,0 & & & \\
1\,1\,0\,0\,0 & & 1\,1\,0 & & & \\
1\,1\,0\,0\,0 & & 1\,1\,1 & & & \\
1\,0\,1\,0\,0 & & & 0\,0\,0 & & \\
1\,0\,1\,0\,0 & & & 1\,0\,0 & & \\
1\,0\,1\,0\,0 & & & 1\,1\,0 & & \\
1\,0\,1\,0\,0 & & & 1\,1\,1 & & \\
1\,0\,1\,0\,0 & & & 0\,0\,0 & & \\
1\,0\,1\,0\,0 & & & 1\,0\,0 & & \\
1\,0\,1\,0\,0 & & & 1\,1\,0 & & \\
1\,0\,1\,0\,0 & & & 1\,1\,1 & & \\
1\,0\,1\,0\,0 & & & 0\,0\,0 & & \\
1\,0\,1\,0\,0 & & & 1\,0\,0 & & \\
1\,0\,1\,0\,0 & & & 1\,1\,0 & & \\
1\,0\,1\,0\,0 & & & 1\,1\,1 & & \\
1\,0\,1\,0\,0 & & & 0\,0\,0 & & \\
1\,0\,1\,0\,0 & & & 1\,0\,0 & & \\
1\,0\,1\,0\,0 & & & 1\,1\,0 & & \\
1\,0\,1\,0\,0 & & & 1\,1\,1 & & \\
1\,0\,0\,1\,0 & & & & 0\,0\,0 & \\
1\,0\,0\,1\,0 & & & & 1\,0\,0 & \\
1\,0\,0\,1\,0 & & & & 1\,1\,0 & \\
1\,0\,0\,1\,0 & & & & 1\,1\,1 & \\
1\,0\,0\,1\,0 & & \underset{\sim}{0} & & 0\,0\,0 & \\
1\,0\,0\,1\,0 & & & & 1\,0\,0 & \\
1\,0\,0\,1\,0 & & & & 1\,1\,0 & \\
1\,0\,0\,1\,0 & & & & 1\,1\,1 & \\
1\,0\,0\,0\,1 & & & & & 0\,0\,0 \\
1\,0\,0\,0\,1 & & & & & 1\,0\,0 \\
1\,0\,0\,0\,1 & & & & & 1\,1\,0 \\
1\,0\,0\,0\,1 & & & & & 1\,1\,1 \\
1\,0\,0\,0\,1 & & & & & 0\,0\,0 \\
1\,0\,0\,0\,1 & & & & & 1\,0\,0 \\
1\,0\,0\,0\,1 & & & & & 1\,1\,0 \\
1\,0\,0\,0\,1 & & & & & 1\,1\,1
\end{array}\right.
$$

Input Variables Used in the Remaining Columns

1. X12 for years 1 thru 4
2. X14 for years 1 thru 4
3. X15 for years 1 thru 4
4. X22 for years 1 thru 4
5. X8 (4 year average)
6. X13 (4 year average)
7. X13 (4 year average)
8. X24 (4 year average)

To adequately describe the grouping of the blocks in Table 3, 5 columns to distinguish the 5 groups and five identical three-column matrices to span the constraint-by-year interaction space were also included. Finally, four of the next most important variables were represented by one column each in $\underset{\sim}{X}_2$ to again form a saturated model. This resulted in the design matrix $\underset{\sim}{X}_2$ as shown.

Based on significance tests, only non-significant constraint-by-year (trend) effects and group means were collapsed since it was of interest to determine the importance of the program inputs on the family planning system. As a result, the trend parameters of the last two groups and the constraint parameters of the first three groups were collapsed, as both χ^2 tests were not significant ($p>.25$). An intermediate model was formed after taking these results into account and including the ninth variable, extension educators (X13), as another parameter. Statistical tests for this intermediate model appear in Table 4.

TABLE 4

Test Statistics for Intermediate Model Analysis of Variation

Source	Degrees of Freedom	χ^2	p
Model	35	6020.58	0
Constraints	2	34.60	0
Trend	12	308.57	0
Across groups for year:			
2	4	110.93	0
3	4	106.68	0
4	4	36.33	0
Significance of trends within each group			
Group 1	3	18.17	0
2	3	59.63	0
3	3	34.74	0
4	3	30.28	0
Input variables	21	373.27	0
Family planning expenditures	4	56.83	0
Health expenditures	4	51.18	0
Family planning field workers	4	457.50	0
ANMs	4	47.15	0
Family planning media events	1	17.00	0
Dais	1	25.97	0
Sterility referrals	1	10.63	0
Extension educators	1	13.09	0
Female doctors	1	.01	.91
Error	4	1.88	.76
Total	39	6022.46	

Moreover, one degree of freedom tests on parameters for the number of family planning workers in year three, the amount of health expenditures in years two and three, and the amount of family planning expenditures in year four were non-significant (p>.18) . Another model was fit to the data excluding these parameters as well as the non-significant parameter for female doctors. Statistical significance of the parameters for this new model were examined and only one parameter (the number of ANMs in year three) was not statistically important. Our next model was then constructed by removing the parameter for ANMs in year three and constructing the design matrix $\underset{\sim}{X}_4$.

$$\underset{\substack{\sim \\ 40\times 30}}{X_4} = \left[\begin{array}{ccccccccccccccc}
1&0&0&0&0&0&&&&&&&&& \\
1&0&0&1&0&0&&&&&&&&& \\
1&0&0&1&1&0&&&&&&\underset{\sim}{\mathbf{0}}&&& \\
1&0&0&1&1&1&&&&&&&&& \\
1&0&0&&&&0&0&0&&&&&& \\
1&0&0&&&&1&0&0&&&&&& \\
1&0&0&&&&1&1&0&&&&&& \\
1&0&0&&&&1&1&1&&&&&& \\
1&0&0&&&&&&&0&0&0&&& \\
1&0&0&&&&&&&1&0&0&&& \\
1&0&0&&&&&&&1&1&0&&& \\
1&0&0&&&&&&&1&1&1&&& \\
1&0&0&&&&&&&0&0&0&&& \\
1&0&0&&&&&&&1&0&0&&& \\
1&0&0&&&&&&&1&1&0&&& \\
1&0&0&&&&&&&1&1&1&&& \\
1&0&0&&&&&&&0&0&0&&& \\
1&0&0&&&&&&&1&0&0&&& \\
1&0&0&&&&&&&1&1&0&&& \\
1&0&0&&&&&&&1&1&1&&& \\
1&0&0&&&&&&&0&0&0&&& \\
1&0&0&&&&&&&1&0&0&&& \\
1&0&0&&&&&&&1&1&0&&& \\
1&0&0&&&&&&&1&1&1&&& \\
1&1&0&&&&&&&&&&0&0&0 \\
1&1&0&&&&&&&&&&1&0&0 \\
1&1&0&&&&&&&&&&1&1&0 \\
1&1&0&&&&&&&&&&1&1&1 \\
1&1&0&&&&\underset{\sim}{\mathbf{0}}&&&&&&0&0&0 \\
1&1&0&&&&&&&&&&1&0&0 \\
1&1&0&&&&&&&&&&1&1&0 \\
1&1&0&&&&&&&&&&1&1&1 \\
1&0&1&&&&&&&&&&0&0&0 \\
1&0&1&&&&&&&&&&1&0&0 \\
1&0&1&&&&&&&&&&1&1&0 \\
1&0&1&&&&&&&&&&1&1&1 \\
1&0&1&&&&&&&&&&0&0&0 \\
1&0&1&&&&&&&&&&1&0&0 \\
1&0&1&&&&&&&&&&1&1&0 \\
1&0&1&&&&&&&&&&1&1&1
\end{array}\right]$$

Input Variables Used in the Columns

1. X12 for years 1 & 4
2. X13 (4 year average)
3. X14 for years 1,2,&4
4. X15 for years 1,2,&4
5. X17 (4 year average)
6. X22 for years 1,2,&3
7. X24 (4 year average)
8. X8 (4 year average)

This model fit quite well (χ^2 = 11.55 on 10 degrees of freedom) and so valid test statistics, shown in Table 5, were also obtained. All the parameters were highly significant (p<.05) and so no further removal of any of the parameters could take place without the fit of the model becoming unsatisfactory. As a result, this model was deemed the final model.

TABLE 5

Test Statistics for Final Model $\underset{\sim}{X}_4$

Analysis of Variation

Source	Degrees of Freedom	χ^2	p
Model	29	6010.91	0
Constraints	2	35.76	0
Trend	12	279.47	0
Across groups for year:			
2	4	134.73	0
3	4	140.44	0
4	4	45.69	0
Significance of trends within each group Group			
1	3	18.21	0
2	3	81.69	0
3	3	40.26	0
4	3	44.06	0
Input variables	15	366.25	0
Family planning expenditures	3	98.65	0
Health expenditures	2	59.91	0
Family planning field workers	3	164.65	0
ANMs	3	54.81	0
Family planning media events	1	8.83	0
Dais	1	31.39	0
Sterility referrals	1	51.18	0
Extension educators	1	15.40	0
Error	10	11.55	.32
Total	39	6022.46	

4. Conclusion

One thing that is immediately apparent from this table is that the input variables have a relatively large effect on the family planning system in comparison to the trend or constraint variables considered individually. Secondly, within the set of input variables, the magnitude of the χ^2 statistics reflects the importance of the number of family planning field workers, the amount of family planning expenditures and overall health expenditures, and the number of auxiliary nurse midwives. Of course, we must keep in mind that these results are preliminary as only ten rural blocks were examined. However, it is worthwhile to point out that these conclusions were similar to those in the investigation of Henley et al.. Furthermore, the methodology described here can be extended to analyze all the rural and urban blocks to see if similar results would be obtained.

REFERENCES

Furnival, G.M. and Wilson, R.W., Jr. (1974). Regression by leaps and bounds. Technometrics 13, 499-512.

Grizzle, J.E., Starmer, C.F. and Koch, G.G. (1969). Analysis of categorical data by linear models. Biometrics 25, 489-504.

Henley, N.S., Jain, S.C., and Wells, H.B. (1976). Relative importance of program input and environmental constraints to family planning programs in Haryana, India. Presented at the 1976 Annual Meeting of the Population Association of America (in Montreal).

Hocking, R.R. (1976). The analysis and selection of variables in linear regression. Biometrics 32, 1-49.

Neyman, J. (1949). Contributions to the theory of the χ^2 test. Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability. Berkeley and Los Angeles: University of California Press. 239-72.

Mallows, C.L. (1967). Choosing a subset regression. Bell Telephone Laboratories, unpublished report.

Mallows, C.L. (1973a). Some comments on C_p. Technometrics 15, 661-75.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans. Amer. Math. Soc. 54, 426-82.

THE 1974 POST-ENUMERATION SURVEY OF LIBERIA - A NEW APPROACH

Eli S. Marks and John C. Rumford, U. S. Bureau of the Census

Fifteen days after the beginning of the 1974 census of population and housing in Liberia, a post-enumeration survey was conducted. Though the short interval of time between the census and evaluation survey is unusual, what made the Liberian effort different was the simple and relatively inexpensive survey and implementation system used.

In the past, post-enumeration surveys have proven difficult, time consuming, and expensive to carry out. Because of this, many countries, particularly the developing nations, have avoided post-enumeration surveys in census evaluation planning. This is unfortunate because no census, regardless of how carefully planned and executed, is perfect. While perfection cannot be obtained, completeness of enumeration can and should be estimated.

The primary purpose of a post-enumeration survey (PES), is to provide estimates of census coverage. All censuses suffer from two types of coverage errors. The first, and most frequent, is undercoverage, and the second is overcounting. Undercoverage errors are caused by such things as missing persons in enumerated living quarters, missing living quarters entirely or, in some cases, missing whole localities. Overcounting is almost always of much smaller magnitude and results from enumerators overlapping at area boundaries, reporting of persons by more than one household and enumeration of persons who should not have been enumerated, such as (names entered for) nonexistent individuals or persons who died before or were born after the census date.

The traditional PES used to estimate these errors involves reenumerating a sample of census enumeration areas (EA's) or parts of EA's several months after the census, using a specially selected, well-trained group of enumerators. These enumerators are usually provided with preenumeration intelligence derived from the census, and each is instructed to reconstruct the population in the evaluation area as of the census date. At the close of the reenumeration, a case-by-case matching with the census questionnaires is carried out, and a full field follow-up "reconciliation" is made of all non-matching persons.

The principle underlying this methodology is that the PES will be much better than the census and, therefore, estimates derived from this survey are a "standard" to which the census can be compared and eventually adjusted.

However, experience in the United States and elsewhere has suggested that PES estimates are not necessarily better than census results (U. S. Bureau of the Census, 1960, Marks and Waksberg, 1966, and Marks, 1973). Moreover, in the Liberian context, evidence existed that the techniques associated with traditional PES systems were inappropriate. This evidence was primarily the experience gained from four years of conducting the Liberian Fertility Survey, a national, multi-round, household survey, that used case-by-case matching and other techniques common to traditional post-enumeration survey designs (Rumford, 1970). During this survey, it was repeatedly demonstrated that a single enumeration system, using experienced, well-trained enumerators and supervisors, failed to enumerate many persons. Moreover, it was found that, when enumerators were grouped and tested by age, experience, and education, the older, more experienced, and better educated enumerators missed about the same number of persons as did their younger and less experienced colleagues with average education (Rumford, 1972). Confronted with this evidence, it was decided to use a new PES approach in Liberia based on dual system estimation.

Dual System Estimation

There are four main methodological elements in this new approach as implemented in Liberia, which depart from the traditional PES method. The first is emphasis on independence between the census and PES; the second is the use of one-way matching to reduce "geographic out-of-scope" error; the third is providing for a very brief time interval between census and PES in order to minimize the problems of tracing migrants; and the fourth is the elimination of field verification of unmatched (Census or PES) enumerations. The third and fourth features were introduced in Liberia for reasons of cost and operational feasibility. The first two features are attempts to reduce the effects of two of the three major biases of dual system estimation.

As used in the current paper, "dual system estimation" involves (1) collecting data from a sample of the target population with two independent data collection systems (in census evaluation, the Census and the PES); (2) matching the reports of the two systems to determine which of the sample individuals were reported by both systems; (3) using the proportion matched of all cases reported in one system as an estimate of the completeness of reporting (the "coverage rate") in the <u>other</u> system. Thus, the proportion matched of all <u>PES</u> cases is used as the estimate of the completeness of the <u>Census</u> coverage.

Independence

It will be noted that the purpose of the PES is to provide an estimate of the completeness of the Census coverage. The PES estimate of Census coverage will not be biased by the erroneous omission of some sample cases or by the erroneous inclusions of some nonsample cases (e.g., due to the PES enumerator getting outside the boundaries of the sample segment), provided the PES coverage errors are <u>independent</u> of the Census coverage errors--i.e., provided the probability of a PES case being matched to the Census (found to have been enumerated in the Census) is the same for the cases erroneously omitted from or erroneously included in the PES as it is for the cases correctly included. Note that independence implies nothing about causality. Obviously, whether a person is or is not reported in the PES cannot causally affect his probability of being reported in a census taken prior to the PES. However, the classes of persons with low probabilities of being reported in the PES could also (apart from any direct causal connection) have lower (or higher) probabilities of being reported in the Census than the classes of persons with high probabilities of being reported in the PES. There is, in fact, evidence from recent post-enumeration surveys in Paraguay and in Korea that persons who change their places of residence between the Census and the PES tend to have lower probabilities of enumeration in both the Census (taken before their change of residence) and the PES than persons who do not move between the Census and the PES.

Correlation between PES and Census errors may be "direct" or "indirect". Indirect correlation results from the fact that the probability of certain individuals or classes of individuals being enumerated is high or is low for both Census and PES.[1] Direct correlation involves a causal relationship between Census and PES errors--i.e., the fact of a person being enumerated or not being enumerated in the Census actually changes the probability of his being enumerated in the PES.

Since indirect correlation is an inherent feature of the population and the reporting methods, its control involves careful selection of the basic PES procedures. The newer techniques of taking a PES, used in Korea and Paraguay, were developed primarily for the purpose of eliminating a major source of indirect correlation.

The preservation of direct statistical independence is largely a matter of avoiding collusion between the two enumeration systems in the field. One of the steps that can be taken to do this is to postpone the selection of the sample areas until after the Census enumeration is complete. In Liberia, the PES sample was selected immediately prior to national census day. However, the sample enumeration areas were not identified to the field officers until the census enumeration was completed 5 to 10 days later. When the field officers were notified, they were instructed to immediately impound the completed census questionnaire workbooks for the EA's scheduled to be reenumerated. These workbooks remained in custody throughout the PES enumeration.

In another effort to prevent direct correlation, the PES enumerators were recruited from County Inspectors and District Supervisors. These workers were selected because they were familiar with the general census enumeration procedures, but took no active part in the census enumeration at the EA level. The enumerators were briefed on the PES questionnaires but they were not given any additional training nor were they provided with any preenumeration intelligence. Moreover, they used listing sheets and EA maps that were duplicates of those provided to the census enumerators. A time limit of 72 hours was imposed for the enumeration (the same time target prescribed for the original census enumeration) and the PES, like the Census, was conducted on a de jure basis.

After completing the enumeration, the PES questionnaire workbooks were transferred to a census regional officer. However, no review was made at regional headquarters. Instead, the original census questionnaire workbooks were transmitted to national headquarters for review and matching by a completely separate and specialized group. In spite of these elaborate precautions, and as a tribute to the ingenuity of Man, two of the thirty-two sample EA's were compromised and had to be eliminated.

It should be stressed that although it is not sufficient for unbiased PES estimates, technical and administrative independence is essential. Without it, the Liberian PES system cannot be used.

Other Biases

In addition to correlation bias, dual system estimates are subject to matching bias and out-of-scope (or erroneous inclusion) bias. Matching bias is the result of "erroneous matches" and "erroneous nonmatches". Erroneous matches will increase the number of PES cases considered to be enumerated in the Census and will, consequently, result in an overestimate of the completeness of Census coverage and erroneous nonmatches lead to an underestimate. The overall matching bias depends on the "net matching bias", which is the difference between the number of erroneous matches and the number of erroneous nonmatches.

"Out-of-scope" error is the result of improper inclusion of cases in the PES or the Census. Erroneous inclusions in the Census are duplicate enumerations and enumerations of persons who should not have been enumerated--e.g., persons who died before or were born after the Census date, diplomatic personnel or other persons excluded from the Census by definition and enumeration of fictitious persons, either deliberately (i.e., enumerator "curbstoning", the completion of Census entries without interviewing every household) or accidentally (e.g., entry of a dog or other household pet under the mistaken impression that it was a child). Erroneous inclusions in the PES include the PES enumeration of persons who should not have been enumerated in the Census (as described above) and also the enumeration of nonsample persons. Enumeration of nonsample persons results from boundary difficulties (i.e., the PES enumerator enumerating households actually located outside the sample segment) and from improper handling of "migrants" (persons who move between the Census and the PES). On the latter, "migrants" can be sampled either on the basis of where they were on the Census date or of where they are at the time of the PES interview. Thus, an out-of-scope error occurs in a PES which samples on the basis of residence on the Census date, if the PES enumerator includes someone who moved into the sample segment after the Census date. For PES samples based on location at the time of the PES interview, an out-of-scope error occurs if the PES enumerator includes someone who moved out of the sample segment before the PES date.

Handling of Migrants

Persons who move into or out of a sample segment between the census date and the time of the PES enumeration represent a particularly difficult problem. Prior to 1970, all PES sampling was based on the person's location on the census date. However, this tended to produce a correlation bias because (1) migrants tend to have a larger census omission rate than nonmigrants and (2) persons who move away from an area prior to the PES are likely to be missed by the PES. Persons who move away tend to be omitted from the PES because (a) where a whole household moves, neighbors may be able to furnish only very incomplete information about the individual household members; and (b) if an individual moves out of a household, the remaining members of the household may be very vague about the date that he left.[2/] The reasons for poor census enumeration are not immediately obvious. Part of the difficulty may be with the fact that census enumeration is dragged out over a long time period and, therefore, the canvass for many of the migrants occurs after the migration. However, the higher omission rates for migrants occur also in de facto censuses in areas where 80% or more of the census enumeration is actually completed on the census date. It may be hypothesized that the causal mechanism may be the fact that many of the "migrants" (particularly in a de facto Census) have only tenuous connections with the household where they were staying or living on the census date and are, therefore, not mentioned when the enumerator asks for "all the people staying (or living) in this household".

There are two ways of reducing the correlation bias due to migrants. One of these has been introduced only recently into PES work--in the PES of the 1970 Census of Korea and the PES of the 1972 Census of Paraguay. This method involves sampling migrants on the basis of where they are at the time of the PES interview. In Paraguay, two samples of approximately equal size were used in the 1972 PES, one asking about people in the sample segment on the census date and the other asking about people in the sample segment at the time of the PES interview. The second sample gave nearly 4 times as many migrants (people who moved into sample segments) as the number of migrants (people who moved out of sample segments) given by the first sample, although the number of nonmigrants was about the same (only 7% difference) for the two samples.

While basing the sampling of migrants on their residence at the time of the PES will improve the reporting of migrants and thus reduce the correlation bias due to migrants, it considerably increases the difficulties (and the errors) in matching the migrants to the Census. That is, in order to search for a migrant in the Census files, it is necessary to have his address at the time of the Census, with the kind of precision and detail that permits an accurate determination of the enumeration area (EA) in which he should have been enumerated. While people will usually know their former addresses, they frequently cannot furnish the kind of address information needed to determine the EA.

The other method of reducing correlation bias due to the migrants is to reduce the number of migrants. This can be done by reducing the time lag between the Census and the PES. This is the method adopted for Liberia, where the time lag between the Liberian census date and the PES enumeration was set at fifteen days. This period was long enough to allow the census enumerators to canvass and clear their areas and yet, was short enough to limit opportunities for migrations, births, and deaths to occur. The PES asked whether anyone listed on the questionnaire was born after the national census day and eliminated such persons from the tabulations. No such accommodation was considered necessary for deaths, since the rarity of this event in a ten to fifteen day period would make errors from this source negligible. However, it should be pointed out that, if the period between the Census and the PES is longer, a specific question on deaths may be needed. It should also be pointed out that the short time period increases the danger of overlapping the two enumeration systems in the field if strict precautions are not taken.

Use of One-Way Matching

A second problem is control of the geographic out-of-scope errors, which occur at the boundaries of sample evaluation areas. These errors can be particularly troublesome in heavily populated urban centers where no easily identified natural or man-made boundaries exist. In these areas, it is relatively easy for a census or PES enumerator to erroneously extend his canvassing area beyond the limits of the designated sample area. Because the Liberian system is based on maintaining strict independence between the Census and the PES, it is not desirable to use information obtained during the actual Census enumeration to help locate the EA boundaries. To minimize the effect of boundary errors, the Liberian PES used "one-way matching", with a provision for searching for cases in adjoining EA's where the possibility of boundary errors existed.

One-way matching was used in Liberia primarily because of its relative simplicity and economy. The technique is most effectively employed in circumstances where one of the two data sources has records covering the entire population. With one-way matching, it is only necessary to determine the exact matching status of each report for one of the two record systems (the one with data for a sample only). As implemented in the Liberian Post Enumeration Survey, the PES records were compared to the census records and each person was categorized as "found" (or "matched") or "not found" in the census listings.

After the initial matching within the sample EA, the unmatched PES cases were searched for in adjacent EA's. For the PES cases that were matched in an adjacent EA, an attempt was made to determine whether the error was made by the Census or the PES. However, the determination of whether the error was made by the Census or the PES was not very successful. In most cases, an acceptable determination of the correct EA boundaries requires a further field visit and it was not felt that the expense of such a visit was warranted since the number of cases involved was small.

A major purpose of a field reconciliation is to determine whether any of the PES cases, matched or unmatched, should not have been included either because the person was not supposed to be enumerated in the Census or because the person was a nonsample case who should have been enumerated outside the sample EA boundaries. Actually, since the only purpose of the PES is to determine the estimate of the proportion of the population enumerated in the Census, the bias due to a few cases erroneously included in the PES sample is minor, provided the same treatment is given to unmatched as to matched cases. Of course, there will be a bias if the cases near an EA boundary do not have exactly the same Census omission rate as those in the interior of an EA. In general, cases near an EA boundary will have a somewhat higher census omission rate and, also, a somewhat higher census duplication rate. However, both the difference and the number of cases involved will be small (unless the census maps and procedures are totally inadequate) and the bias will be trivial. The important thing is to extend the search for PES cases near an EA boundary to the adjacent EA's, so that no PES case will be called unmatched because the census or PES enumerator (or both) made a mistake in locating the EA boundary. It would, of course, be desirable to search all boundary cases, matched or unmatched, in the adjacent EA's to check on duplicate census enumerations. However, such duplication is rare in practice and the bias is minor of not searching adjacent EA's for cases matched within the sample EA.

PES RESULTS

The PES estimates of census completeness by age and sex appear in Table 1. Overall, the 1974 Liberian Census appears to have achieved about 90% coverage. The figures for particular age-sex cells are subject to fairly high variances (and, possibly, to biases in the reporting of age). However, the overall pattern represents more than variance and shows interesting differences from the underenumeration pattern typical of more industrialized countries. Thus, there is little or no overall difference between males and females in completeness of coverage for the 1974 Liberian Census. On the other hand, the United States censuses show better coverage of women than men largely due to the poorer enumeration of men in the age range 15 to 45, which also seems to hold for Liberia. For Liberia, the higher coverage of females 14 to 45 is balanced by the lower coverage for females under 10 and 45+ which does not obtain for the U. S. In fact, U. S. coverage of females under 10 and also of females 65 and over tends to be better than the coverage of males in the corresponding age groups. In the U. S. 1970 Census, the coverage of males dropped for ages 20-24 and remained low up to age 45 but no such decline appears for U. S. females ages 20 to 44. In Liberia the coverage of both sexes seems to drop off at ages 15 to 19 but then improves for ages 20 to 24 and continues to improve for males up to age 60 and for females up to age 45.

Some Defects of the Liberian PES

A key feature of the Liberian PES system is completing the PES field work within a short interval after the Census date. This is also a key weakness of the system.

Having a very short interval between the Census and the PES eliminates many of the very serious problems associated with substantial migration between the Census date and the PES date. It makes it possible to keep the PES costs down to relatively low levels and also to utilize in the PES a good part of the field and administrative structure set up on a temporary basis for the Census. On the other hand, it imposes certain requirements which may be serious handicaps. One of these is the difficulty of maintaining independence and administrative control between the Census and PES. As noted above, special instructions were issued that a PES enumerator not be told the location of his assignment until all of the Census enumeration for the evaluation area was complete and the Census schedules for the EA were in the hands of the regional supervisor. In spite of this precaution, there was a coincidence between Census and PES reports for two of the thirty-two sample EA's which could not possibly have occurred without someone altering either the Census returns to accord with the PES results or vice versa.

The fact that this contamination affected only 6% of the Liberian PES sample reflects the great efforts which were made to preserve independence while adhering to an extremely tight time schedule. With adequate planning and supervision, five days is sufficient time for completing a Census in 80% or more of the EA's of almost any country. Unfortunately, there are delays and slip-ups that will affect a minority of EA's in almost any country. An enumerator may have misunderstood his map (with or without the "help" of ambiguous boundaries) and failed to enumerate a whole section of his EA; or an entire EA or group of EA's may not have been assigned for enumeration; and these errors might not be detected until preliminary count figures are announced and complaints from local areas start pouring (or dribbling) in.

A short time interval between Census and PES imposes other restraints. The sample areas must be selected well before the Census, possibly before the work of setting up the Census EA's has been completed. (Note: There are always areas where some of the initial EA's are too big and must be split up and other areas where combinations of some small EA's are desirable). Once selected, the sample EA's must be kept confidential until the Census work in those EA's has been completed.

Also, the short time interval usually means that the PES schedules and procedures cannot be tested and revised under actual census conditions. This is likely to be particularly harmful with respect to the matching procedures. Most countries have some experience with other Census and PES operations. If not, it is possible to draw on experiences of other countries with similar conditions (e.g., in collecting and coding data on occupation and industry) or to dispense with or simplify the operation. Most developing countries (and not a few statistically advanced countries) have no directly applicable experience with modern matching methods. The matching operation is indispensable to a PES and simplification can pose very serious dangers of matching errors which raise questions about the validity of the entire PES. There were, in fact, serious defects in the design of the Liberian PES matching. In trying to keep the PES procedure simple and the questionnaire short, some information very useful for matching was omitted, notably relationships and alternative names. Both of these items are fairly easy to obtain and add only trivially to the total interview time. "Relationship" will be useful for matching in almost any culture (particularly those where most of the names are common ones and it is necessary to distinguish the Ali bin Muhammed who is Muhammed bin Muhammed's younger brother from the one who is his son). The need for alternative given names is culturally limited; and the fact that they may not be needed in the United States does not mean that they will not be needed in Liberia.

One problem of the Liberian PES system is particularly acute for de facto censuses. Providing for a short time interval between PES and Census is not satisfactory for dealing with the "floating population"--i.e., persons with no usual residence any place. In de jure censuses, such persons are to be enumerated on a de facto basis --i.e., where they happen to be staying on the census date. Even with a PES taken two weeks after the census date (as in Liberia), the probability is small of finding persons with no usual place of residence where they were on the census date.

Conclusion

The Liberian system is no panacea. When the final 1974 census results become available, all the available methods of demographic analysis will be required to refine and improve the coverage estimates derived from dual system estimation. Comparisons with previous census and survey results will be necessary to fully evaluate the complete census coverage picture.

Moreover, as indicated above, the system itself has several deficiencies. Nevertheless, with some minor modifications, such as the provision of additional matching information or a small extension in the time period between Census and PES (e.g., 30 days instead of 15), the Liberian PES system is applicable to and effective for countries which do a de jure Census with a very short Census enumeration period and also to countries with a de facto census where the matter of temporary sojourners in a location is minor (usually countries which are primarily rural and nonindustrial).

Where a country does a de facto census and has a substantial number of temporary "sojourners" (or persons with no usual place of residence) or where a short time interval between PES and Census is not feasible, one should consider the use of a PES sample based on where the individual is at the time of the PES. This requires determining for migrants the location on the Census date and doing the matching search at that location. While this kind of matching can be difficult and expensive, methods for simplifying the procedures and improving the results are being explored and may provide an answer in the near future.

We can be certain that there will continue to be problems of evaluating Census coverage, no matter which methods are used, and whether they involve a PES or other dual system estimation or do not. However, the need for methods of more adequate census evaluation and correction is painfully evident for many developing countries where recent censuses have been so inconsistent with all the known demographic facts as to be completely unacceptable.

No one method of census evaluation and correction has a complete answer. However, the post-enumeration survey is a powerful technique which should not be ignored or avoided. It is hoped that the simplicity and economy of the dual estimation system used in Liberia will help remove the post-enumeration survey from the luxury item list in census evaluation and place this important tool within reach of all census planners. It is at least a starting point from which census planners can go on to more sophisticated and more powerful (but not necessarily, more expensive) tools.

Acknowledgements

The Liberian Post-Enumeration Survey was conducted by the staff of the Ministry of Planning and Economic Affairs, Republic of Liberia. The funds for survey implementation were provided by the Government of Liberia and the United Nations Fund for Population Activities. Technical support was given by the United Nations' Office of Technical Cooperation, the African Census Program of the United Nations Economic Commission for Africa and the United States Bureau of the Census.

The opinions expressed and the techniques demonstrated in this paper are those of the authors and are not necessarily endorsed by any of the cited organizations or individuals within these organizations.

Footnotes

1. There can, in theory, be indirect negative correlation, in which individuals with a high probability of being reported in the Census have a low probability of being picked up in a PES and vice versa, but no actual instances of this have been noted to date.
2. Precise dating of "events" such as births, deaths and migrations is a major problem of vital

statistics measurement. It was, for example, the major reason for the development of techniques for estimating fertility which use "children ever born" (regardless of date) rather than children born during the past year.

References

1. Ahmed N., and K. Krotki, 1963. Simultaneous Estimations of the Population Growth: The Pakistan Experiment. Pakistan Development Review, 3, 37-65.

2. Chandrasekhar, C., & W. E. Deming, 1949. On a Method of Estimating Birth & Death Rates & the Extent of Registration. JASA, 44, 101-115.

3. Malawi National Statistical Office, 1973. Malawi Population Change Survey: Feb. 1970 - Jan. 1972. Zomba, Malawi: Govt. Printer

4. Marks, E. S., 1973. The Role of Dual System Estimation in Census Evaluation. Unpublished paper presented at the first meeting of the International Assn. of Survey Statisticians, Vienna, Austria, Aug. 1973. Scheduled to appear as Chap. 10 in Development in the Dual System Estimation of Population Size & Growth, edited by K. Krotki, U. of Alberta Press, 1976.

5. Marks, E. S., W. Seltzer & K. Krotki, 1974. Population Growth Estimation: A Handbook of Vital Statistics Measurement. N. Y., The Population Council.

6. Marks, E. S., & J. Waksberg, 1966. Evaluation of Coverage in the 1960 Census of Population Through Case-by-Case Checking. Proceedings of the Social Statistics Section, 1966, 62-70, American Statistical Association, Washington, D. C.

7. Rumford, J. C., 1970. Use of the Chandrasekhar-Deming Technique in the Liberian Fertility Study. Public Health Reports, 85, 965-973.

8. Rumford, J. C., 1972. Factors Influencing the Casefinding of Migrations in the Liberian Fertility Survey, Demography, 9, 431-441.

9. Rumford, J. C., Y. Heperkan & N. Fincancioglu, 1968. The principles & Preliminary Results of the Turkish Demographic Survey. Public Health Reports, 83, 573-582.

10. Seltzer, W. & A. Adlakha, 1969. On the Effect of Errors in the Application of the Chandrasekhar-Deming Technique. Paper delivered at The Population Council Seminar on Problems of Population Growth Estimation, Jan. 1969. Reprinted in Laboratory for Population Statistics Reprint Series, No. 14, April 1974.

11. U. S. Bureau of the Census, 1960. The Post-Enumeration Survey.1950, Technical Paper No. 4, Washington, D. C.

Table 1: Estimated Completeness of the Census by Age and Sex; Liberia, 1974

Age Groups	Both Sexes %	Male %	Female %
All ages	89.0	89.2	88.8
Under 5 years	86.1	88.0	84.0
5 - 9 years	88.3	88.8	87.8
10 - 14 years	89.0	88.7	89.4
15 - 19 years	84.6	83.8	85.4
20 - 24 years	88.2	86.7	89.2
25 - 29 years	90.2	88.8	91.3
30 - 34 years	90.2	89.1	91.0
35 - 39 years	90.9	89.8	91.8
40 - 44 years	92.8	92.3	93.5
45 - 49 years	91.4	92.4	90.2
50 - 54 years	92.7	93.8	91.0
55 - 59 years	93.9	95.7	91.5
60 - 64 years	89.7	92.5	85.4
65 years & over	92.3	93.1	91.1

Standard errors are as follows:

Both sexes, all ages - 1.5%
Male, all ages - 2.2%
Female, all ages - 2.4%

OPTIMUM RECALL PERIODS FOR ESTIMATING ACCIDENTAL INJURIES IN THE NATIONAL HEALTH INTERVIEW SURVEY

James T. Massey and Joe Fred Gonzalez, Jr., National Center for Health Statistics

Introduction

In 1975 the Health Interview Survey (HIS) in collaboration with the Consumer Product Safety Commission added an accident supplement to the HIS questionnaire in order to collect more detailed information about accidental injuries.

The HIS estimates are based on information collected in a nationwide sample of about 120,000 persons in 40,000 interviewed households. The sampling plan of the survey follows a multi-stage probability design which permits a continuous sampling of the civilian, non-institutionalized population of the United States. A detailed description of the HIS sample design can be found in (1). For acute conditions (includes accidental injuries) HIS usually uses a two-week recall period. The magnitude of the variances using a two-week recall period, however, limits the amount of detail that can be published.

Because of the detailed information desired for accidental injuries, it was decided that a longer recall period should be used if at all possible. Although it has been shown in previous recall studies (2-4) that large memory biases result when long recall periods are used, a six-month recall period was adopted with the stipulation that an analysis would be performed after the data were collected to determine the optimum recall period(s) to use for the final analysis. An accident or injury is counted in the HIS if it was medically attended or activity restricting for as much as a day. This paper will discuss the methodology and results used to determine the optimum recall period(s) for accidental injury data in the HIS.

Methods and Procedures

Annual estimates of accidental injuries based on recall periods of 1, 2, 4, 8, 13, 16, 20, and 26 weeks were computed. For each different recall period, annual estimates were obtained by inflating the number of accidents and injuries reported in the sample by the reciprocal of the respondent's probability of selection into the national HIS sample and by the ratio of a year (52 weeks) to the length of the recall period. The basic HIS estimator that was used contains several other adjustments and is described in (5).

A mean square error (MSE) criterion was selected to compare the annual estimates of accidental injuries for 5 of the 8 recall periods. The mean square error of an estimate X is defined as

$$\text{MSE }(X) = \text{VAR }(X) + [\text{BIAS }(X)]^2.$$

The variance component of the MSE(X) was estimated for each recall period using the balanced half-sample replication technique developed by McCarthy (6). The bias for a given recall period is given by $E(X)-\mu$, where μ is the "true" annual number of accidents or injuries for the total population and $E(X)$ is the expected number of accidents or injuries that will be reported for a given recall period.

The estimator used by NCHS is nearly unbiased conceptually and the bias estimated above is a memory or under-reporting bias. In general, one would expect the recall bias to increase as the length of the recall period increases while the variance decreases for lengthening recall periods since the longer recall periods provide a larger sample of accidents. The optimum recall period is the reference period which yields the smallest mean square error. The optimum recall period may not be the same for all statistics, but hopefully a period can be chosen which is optimum or nearly optimum for all accident and injury statistics.

In most studies of mean square errors an independent assessment is made to determine the value of the parameter μ (i.e., record check studies). Since the only available data for our study were sample data, the value of μ had to be either approximated or assumed. The sample results were also used to estimate E(X). This was accomplished for each statistic by plotting the estimated annual number of accidents and injuries associated with each recall period and fitting a model to the data. The following three models were tested (models 1 and 3 were suggested by Simmons (3)).

Model 1. $y = ae^{-bx^2+\varepsilon}$

Model 2. $y = ae^{-bx+\varepsilon}$

Model 3. $y = a+bx+cx^2+\varepsilon$

Since independent estimates of the variance for each of the statistics were available, the first two models were compared using both weighted and unweighted data. Although the estimates for each of the recall periods for a given statistic are clearly not independent, the covariances were not approximated for the variance-covariance matrix in the evaluation of the models.

The goodness of fit (GOF) criterion used to evaluate the models is given by

$$\text{GOF Index} = \sum_{i=1}^{n} \frac{(Y_i-\hat{Y}_i)^2}{\hat{\sigma}_i^2}$$

where

n = number of recall periods
Y_i = observed value for i^{th} recall period
$\hat{Y}_i$ = predicted value for i^{th} recall period
$\hat{\sigma}_i^2$ = estimated sampling variance of Y_i.

The method of least squares was used to estimate the coefficients for each of the models and the number of observations n was either 5 or 6, since variances were not computed for all 8 recall periods. Even though the observations for a given accident statistic are not independent and the distribution of the GOF Index is unknown, the index still provides a means of comparing the relative fit of the models. Once a satisfactory model had been obtained the predicted points along the curve were used to represent the expected number of accidents or injuries for the different recall periods. If one assumes that there is very little memory bias associated with recall periods of two weeks or less, any one of three points along the fitted curve can be used to approximate μ: the y-intercept, the one-week value, or the two-week value. Each of the values has a different appeal and the one-week value on the curve was arbitrarily chosen to represent the parameter μ. All of the biases were then obtained by subtracting the points along the fitted curve (for any recall period) from the one-week value. The use of one value as the parameter is an over-simplification since each recall period covers a slightly different accident population for the 1975 collection year. It is highly unlikely, however, that the parameter changes to any significant degree from one recall period to another and the use of a single value should not affect any of the conclusions. Thus, by squaring the bias and adding the variance, an estimated MSE was obtained. The MSE's were then divided by μ^2 to obtain a relative MSE for interpretive purposes.

Results

Annual estimates were computed for 23 types of accidents for the 8 different recall periods and for 5 types of injuries by age and recall periods. Estimates were also obtained for a number of subpopulations along with variances for 5 or 6 recall periods. For most of the statistics estimated an unexpected result occurred; the estimate based on a one-week recall period was smaller than the estimate based on a two-week recall period. Further investigation revealed several possible reasons for this unexpected behavior. The first and foremost involves the definition of an accident. Accidents are counted in the HIS survey only if they receive medical attention and/or cause some activity restriction for at least one day. It was discovered that all of the accidents that occurred during the reference period, but received medical attention or reduced activity after the reference period (usually during interview week), were not counted in the HIS tally. This problem is most severe for the weekend just prior to the interview week, which begins officially at midnight on Sunday. Thus, while the one-week estimate is affected the most, the estimates for the other recall periods are also affected. There is also apparently some "telescoping" into the standard HIS two-week reference period which tends to offset the bias just discussed. Here, telescoping refers to the situation where accidents that occur more than two weeks ago are reported as occurring in the two-week reference period. For most types of accidents the estimate based on a one-week recall period was approximately 5 percent less than the estimate based on a two-week recall period. Further research by NCHS is planned in this area.

Figure 1 shows the annual estimate of total persons injured as a percent of the estimate based on a one-week recall period by type of accident and recall period. Two conclusions can be immediately drawn from the curves shown in Figure 1. The first is the continuous underestimation of total persons injured as the recall period becomes longer. The estimate of total persons injured using a six-month (26 weeks) recall period is only about 60 percent of the estimate based on a one-week recall period. The second conclusion that can be drawn from Figure 1 is the relationship between memory bias and the type of accident--the more severe the type of accident, the smaller the memory bias. For example, the estimate of persons injured in motor vehicle accidents is 25 percent less for the six-month recall period than for the one-week recall period, while the drop-off for persons injured by one-time lifting or exertion is nearly 50 percent. The relationship between the severity of injury and memory bias can be seen even clearer by examining Figure 2. Figure 2 presents the annual estimate of injuries as a percent of the estimate based on a one-week recall

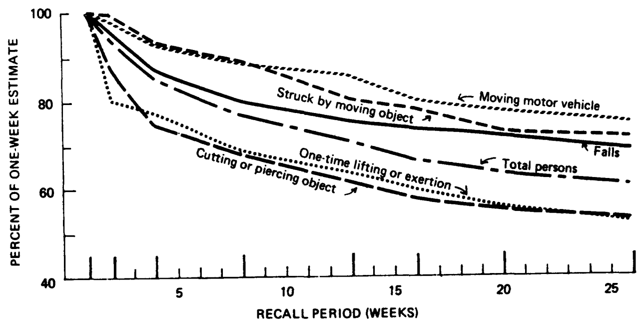


Figure 1. Annual Estimate of Total Persons Injured as a Percent of the Estimated Based on a One-Week Recall Period by Type of Accident and Recall Period.

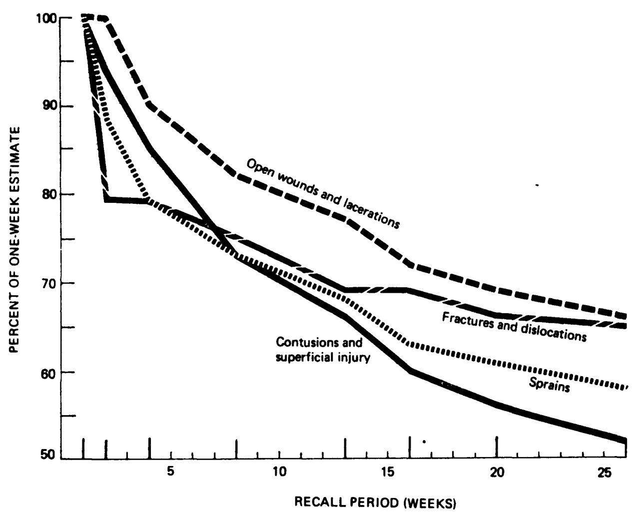


Figure 2. Annual Estimate of Injuries as a Percent of the Estimate Based on a One-Week Recall Period by Type of Injury and Recall Period.

period by type of injury and recall period. As the reference period extends to six months the estimate for open wounds and lacerations falls off by 25 percent, whereas the estimate for contusions and superficial injuries falls off by nearly 50 percent. Figure 2 also shows a significant drop in the estimates of injuries for the shorter recall periods. Figures 1 and 2 present a representative sample of the types of accidents and injuries studied.

Figure 3 is similar to Figure 2 except that curves are shown for age groups rather than types of injuries. The estimate of total number of injuries for persons under 6 using a six-month recall period was less than 50 percent of the estimate based on a one-week recall period. The age group with the next worst drop-off was the 6-16 year olds, while the age group with the smallest drop-off was the 17-24 year olds. A partial explanation seems to be the difference in the reporting of accidents and injuries by self-respondents versus the reporting by proxy respondents (in the HIS survey persons under 17 cannot respond for themselves). The 1974 HIS accident data seem to substantiate this finding by showing that the rate of injury for self-respondents is always greater than the rate of injury for proxy respondents, with the exception of the 65+ age group. One would also expect the fall-off curve for females to be less pronounced than the fall-off curve for males, since housewives often respond for their husbands. This, however, was not true for the accident data and no general conclusions can be drawn without further study.

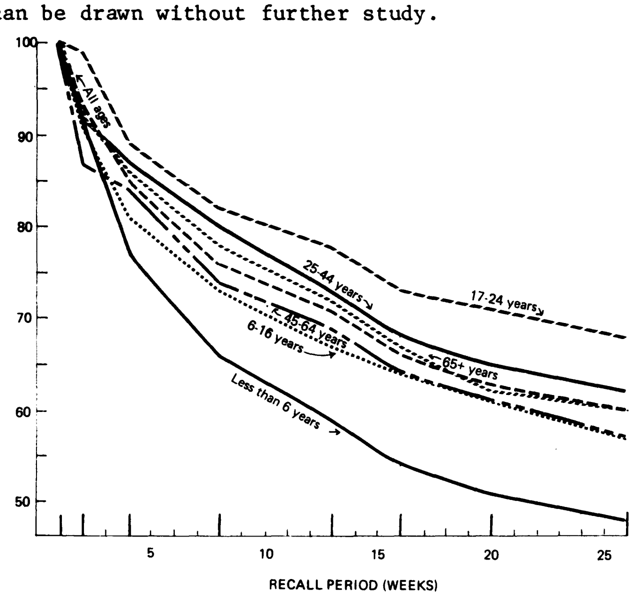


Figure 3. Annual Estimate of Injuries as a Percent of the Estimate Based on a One-Week Recall Period by Age and Recall Period.

The estimates shown in Figures 1-3 based on a one-week recall period are adjusted estimates (except for falls). As mentioned in the previous section, the estimate based on a one-week recall period was almost always less than the estimate based on a two-week recall period. The adjusted one-week estimates were obtained by approximating annual totals using the second week prior to the interview week rather than the first week prior to the interview week. The annual estimate based on a two-week recall period was doubled and then the estimate based on a one-week recall period was subtracted from it. The three models given in the previous section were tested for some 15 selected decay curves and the results are shown in Table 1. Although no one model was best for all 15 curves, the polynomial model $y=a+bx+cx^2$ provided a good fit for most of the curves and was chosen to estimate E(X) for each recall period and to generate the one-week value for μ.

Table 1. Comparison of Goodness of Fit Index for Accidental Injury Models

Type of Accident or Injury	Model			
	$Y=ae^{-bx^2}$ (Weighted)	$Y=ae^{-bx}$ (Weighted)	$Y=ae^{-bx}$ (Unweighted)	$Y=a+bx+cx^2$ (Unweighted)
Total Accidents, All ages	102.29	29.49	33.85	12.84
Ages 45-54	11.66	2.86	2.88	2.14
75+ Females	1.42	.35	.35	.47
Unemployed	.87	1.43	1.49	.95
Ages 25-34, $3000-$5000 Income	.90	.64	.79	.72
Moving Motor Vehicle	3.49	.86	.95	.79
Cutting or Piercing Instrument	18.89	6.58	7.35	2.51
Falls	8.11	2.89	3.81	1.69
Struck by Moving Object	4.14	.94	1.05	.48
Lifting or Exertion	7.42	1.24	1.48	1.51
Total Injuries	NC*	15.39	NC	15.63
Fractures and Dislocations	NC	1.19	NC	.90
Sprains	NC	3.79	NC	4.76
Lacerations	NC	3.86	NC	4.38
Contusions and Superficial Injuries	NC	5.86	NC	7.38

*Not computed

The results of the MSE analysis are shown in Tables 2 and 3 for selected types of accidents and injuries, respectively. The estimate of the number of persons injured or the number of injuries are shown for 5 recall periods along with the relative squared bias, the relative variance and the relative mean square error. The relative squared bias and relative variance components are additive and can be compared to one another to assess their contribution to the relative MSE. For the eleven types of accidents and injuries shown, the 2-week recall period was optimum (smallest relative MSE) in 7 of the 11 cases and the 4-week recall period was optimum for the remaining cases. Similar analysis for certain subdomains of the population was conducted and the results indicated that either the 2 or 4-week recall period was optimum in almost all cases. In general, the relative bias remained about the same for the subdomains while the relative variance increased. Thus, as the subdomain becomes smaller the more likely a longer recall period will become optimum. The squared bias for recall periods longer than 4 weeks was so large relative to the variance, however, that the optimum recall period was longer than 4 weeks in only one case.

Conclusions

The major conclusion drawn from this analysis is that the optimum recall period for collecting and analyzing accidental injury data is either a 2-week or 4-week period depending on the detail of the analysis. The memory bias for longer recall periods is quite large and totally unacceptable by NCHS standards. The study revealed several possible biases associated with the standard 2-week recall period used by HIS and additional research is planned. This study does provide additional evidence, however, that the 2-week reference

Table 2. Components of the Relative Mean Square Error of the Annual Number of Persons Injured by Type of Accident and Recall Period

Type of Accident	Recall Period				
	2 weeks	4 weeks	8 weeks	13 weeks	26 weeks
TOTAL ACCIDENTS					
Estimate (in thousands)	67,812	61,199	55,161	50,716	43,129
Rel-Bias^2 (%)	.086	.707	3.160	7.043	11.541
Rel-Var (%)	.086	.036	.019	.013	.006
Rel-MSE (%)	.172	.743	3.179	7.056	11.547
MOVING MOTOR VEHICLE					
Estimate (in thousands)	4,859	4,604	4,381	4,232	3,709
Rel-Bias^2 (%)	.023	.191	.897	2.153	4.894
Rel-Var (%)	.966	.392	.303	.156	.056
Rel-MSE (%)	.989	.583	1.200	2.309	4.950
CUTTING OR PIERCING INSTRUMENT					
Estimate (in thousands)	5,585	4,748	4,329	3,910	3,364
Rel-Bias^2 (%)	.119	.971	4.268	9.243	13.094
Rel-Var (%)	.980	.395	.180	.091	.049
Rel-MSE (%)	1.099	1.366	4.448	9.334	13.143
FALLS					
Estimate (in thousands)	4,919	4,434	4,059	3,822	3,468
Rel-Bias^2 (%)	.064	.520	2.289	4.971	7.135
Rel-Var (%)	1.339	.345	.176	.109	.051
Rel-MSE (%)	1.403	.865	2.465	5.080	7.186
STRUCK BY MOVING OBJECT					
Estimate (in thousands)	4,204	3,909	3,744	3,381	2,987
Rel-Bias^2 (%)	.042	.351	1.614	3.753	7.470
Rel-Var (%)	1.410	.646	.335	.175	.092
Rel-MSE (%)	1.452	.997	1.949	3.928	7.562
LIFTING OR EXERTION					
Estimate (in thousands)	5,444	5,245	4,638	4,327	3,534
Rel-Bias^2 (%)	.057	.480	2.236	5.300	11.438
Rel-Var (%)	.822	.413	.177	.115	.044
Rel-MSE (%)	.879	.893	2.413	5.415	11.482

Table 3. Components of the Relative Mean Square Error of the Annual Estimate of Injuries by Type of Injury and Recall Period

Type of Injury	Recall Period				
	2 Weeks	4 Weeks	8 Weeks	13 Weeks	26 Weeks
TOTAL INJURIES					
Estimate (in thousands)	71,844	65,418	59,219	54,544	46,481
Rel-Bias^2 (%)	.080	.667	3.091	7.286	15.359
Rel-Var (%)	.084	.034	.019	.013	.006
Rel-MSE (%)	.164	.701	3.110	7.299	15.365
FRACTURES AND DISLOCATIONS					
Estimate (in thousands)	7,130	7,133	6,750	6,264	5,884
Rel-Bias^2 (%)	.020	.167	.794	1.946	4.804
Rel-Var (%)	.703	.281	.136	.097	.043
Rel-MSE (%)	.723	.448	.930	2.043	4.847
SPRAINS					
Estimate (in thousands)	16,482	14,820	13,661	12,672	10,847
Rel-Bias^2 (%)	.071	.596	2.770	6.550	14.001
Rel-Var (%)	.363	.112	.058	.045	.021
Rel-MSE (%)	.434	.708	2.828	6.595	14.022
LACERATIONS					
Estimate (in thousands)	19,606	17,673	16,181	15,157	12,992
Rel-Bias^2 (%)	.071	.591	2.740	6.460	13.603
Rel-Var (%)	.338	.150	.070	.044	.022
Rel-MSE (%)	.409	.741	2.810	6.504	13.625
CONTUSIONS AND SUPERFICIAL INJURIES					
Estimate (in thousands)	13,659	12,393	10,657	9,634	7,635
Rel-Bias^2 (%)	.122	1.020	4.746	11.238	24.146
Rel-Var (%)	.368	.163	.066	.037	.020
Rel-MSE (%)	.490	1.183	4.812	11.275	24.166

period used to collect information on acute conditions in the HIS survey is close to optimum. A final check that was made to help determine the validity of the results obtained in this study was a comparison of the MSE's for motor vehicle accidents with the MSE's obtained in a previous NCHS study (4). The comparisons were quite favorable for all comparable recall periods.

References

1. U.S. National Health Survey: The statistical design of the health household interview survey. *Health Statistics*. PHS Pub. No. 584-A2. Public Health Service, Washington, D.C., July 1958.

2. Simmons, Walt R. and Bryant, E. Earl. "An Evaluation of Hospitalization Data from the Health Interview Survey," *American Journal of Public Health*, Vol. 52, No. 10, October 1962. (Data for 1958, 1959, and 1960)

3. Simmons, Walt R. "Analysis of Bias in Interview-Reported Hospital Experience," *Proceedings of the 36th Session of the International Statistical Institute*, Vol. 42, pp. 442-459, Sydney, Australia, 1967.

4. National Center for Health Statistics. "Optimum Recall Period for Reporting Persons Injured in Motor Vehicle Accidents," *Vital and Health Statistics*, PHS Pub. No. 1000-Series 2, No. 50. Public Health Service. Washington, D.C., April 1972.

5. National Center for Health Statistics. "Estimation and Sampling Variance in the Health Interview Survey," *Vital and Health Statistics*. PHS Pub. No. 1000-Series 2, No. 38. Public Health Service. Washington, D.C., June 1970.

6. National Center for Health Statistics. "Replication--An Approach to the Analysis of Data from Complex Surveys," *Vital and Health Statistics*. PHS Pub. No. 1000-Series 2, No. 14. Public Health Service. Washington, D.C., April 1966.

Acknowledgements

The authors would like to thank Miriam McDaniel, Chief, Sue Hsiung, and Ildy Shannon of the DHIS Programming Branch for their extra efforts in computing the estimates and variances shown in this report, and Mary Mills and Ann Conley for the typing of the manuscript. The authors also thank Guadalupe Gallegos of the Statistical Methods Staff who helped with the manual calculations and Naomi Forester of the Publications Branch, who provided the graphic presentations.

SEX DISCRIMINATION IN FACULTY SALARIES: A METHOD FOR DETECTION AND CORRECTION

George P. McCabe, Jr. and Virgil L. Anderson
Purdue University

1. INTRODUCTION

In addition to Federal law, common sense demands that persons having the same qualifications and performing the same work be paid the same salary. However, evidence of substantial sex discrimination in faculty salaries has been documented ([1], [2], [3], [6]). Using data from 78 universities, 168 four-year colleges and 57 two-year colleges, Darland, Dawkins, Lovasich, Scott, Sherman and Whipple [6] estimate that female faculty members are underpaid by an average of approximately $1500 annually.

Assuming that some sex discrimination does exist in a university, how can the administrators responsible for setting salaries best identify those discriminated against and remedy the situation? In some cases, detailed pay scales with precise formulas for salary determination may be feasible. However, factors such as differential marketability for various fields and quality of teaching, research and service should be taken into account in an institution attempting to attract and retain faculty members of the highest caliber. The methods presented in this paper provide a means for generating salary information which can be used by administrators to diagnose and correct instances of sex discrimination in salaries. It should be kept in mind that our purpose is not to develop a computer program for determining salaries but rather to provide information to decision makers so that they can more effectively deal with this difficult problem. We acknowledge the Office of Civil Rights of the Department of Health, Education, and Welfare, which through a letter to the Purdue University administration, provided the impetus and a certain degree of urgency in the development of these methods.

2. POPULATION

There are no random samples considered in this paper. The population of interest is assumed to be the faculty of a large university. This population would be considered a sample of faculties from similar universities if inferences are desired. The basic unit for job classification is the department-rank. Since our primary purpose is to study sex discrimination, only individuals in department-rank combinations having both male and female faculty members are considered. Thus, if the full professors in one department are all female while those in another are all male, these individuals are excluded from the study. For such individuals, it is impossible to distinguish between sex discrimination and salary differences due to differential market conditions for the departments in question. It should be noted that a substantial number of individuals may be excluded as a result of this criterion. Of course, it cannot be asserted that no discrimination is present in this excluded group. The proposed methodology, however, is inadequate for detecting such cases.

In addition to the above, deans, department heads, distinguished professors, lecturers, instructors and various types of visiting and temporary faculty are also excluded. Salaries of such individuals can usually be examined by comparison with the respective group means.

In summary, the methods to be presented in the next sections are appropriate for studying salaries of assistant, associate and full professors having a peer of the opposite sex in the same department-rank.

3. METHODOLOGY

For each of the three ranks, a regression equation for predicting monthly salary is computed. Variables used as predictors are: D-departmental designators, V-vita variables, A-articles, books and consulting and Q-questionnaire variables.

(D) Departmental designators are dummy variables indicating the department to which the individual is assigned. For joint appointments, the department which pays the largest percentage of the individual's salary is used.

(V) Vita variables include: a dummy variable indicating whether or not a doctorate is held, number of years in current rank, age, year highest degree completed, year hired, a dummy variable indicating an academic year or fiscal year employee, and a dummy variable indicating whether or not the individual is tenured (used for assistant professors only.) To account for some nonlinear effects of years in rank and age, quadratic terms for these variables are also included.

(A) "Articles" is the total number of articles published, truncated at 50. "Books" is the total number of books published, truncated at 5. "Consulting" is a dummy variable indicating whether or not the individual has engaged in outside consulting.

(Q) Questionnaire variables are obtained by compiling the results of questionnaires sent to departmental promotion committees (usually these consist of all full professors in the department.) Each faculty member studied is rated by all promotion committee members on teaching, research, service, marketability and overall. A seven point scale is used with a "no opportunity to observe" option available. Valid responses are averaged to obtain a score on each of these variables for each individual. Missing values are replaced by departmental averages except in the case of the marketability question. For a person about to retire, the lowest score is given. This procedure is consistent with cases where valid data was available for such individuals.

4. FORMAT AND USE OF RESULTS

The regression equations described in the previous section are used to compute residuals. Through a careful inspection of these residuals, administrators can pinpoint cases of possible discrimination. In addition, general patterns can be discerned and a measure of the earnings differential can be calculated.

For each department an ordered list of residuals is provided. By examining the relative positions of males and females in the list, along with the magnitude of the residuals, discrepancies can be detected. Of course, other factors not included in the regression equations contribute to the determination of salaries. The size of the residuals is an indication of the importance of these other factors. Clearly, salaries of all (female and male) individuals with large negative residuals should be carefully reviewed. Unusual patterns also indicate the need for further study of the salaries involved. In many cases, this procedure will uncover bad data or misclassified individuals. Such errors are then corrected and the equations recomputed.

In addition to the residual lists, residuals are categorized in a 2 x 5 table by sex and magnitude for each rank. Boundary points for the magnitude categories are approximately ±2% and ±5% of the average salary for each rank. From these tables of observed frequencies, "expected" frequencies can be computed in the usual manner. Comparisons of expected and observed frequencies give an indication of the extent of possible discrimination. The tables and lists of residuals with a brief explanation of the methodology, are transmitted to the University Equal Employment Opportunity Officer and the Provost. From these, deans and department heads receive lists of residuals for individuals in their schools and departments, respectively.

The residual tables provide an overall picture of the extent of possible sex discrimination. It is desirable, also, to summarize this information in a statistic which can be computed for each rank. Gastwirth [4], [5] has proposed the use of a measure, here denoted by G, based on the Wilcoxin test. Applied to the residuals, this statistic can be described as the probability that a randomly selected female residual is greater than or equal to a randomly selected male residual. A value of .5 corresponds to "equality" while smaller values are indicative of salary differentials. A standard error is easily calculated [5].

For institutions striving to improve the situation regarding sex discrimination, values of G can be calculated yearly. Progress toward a nondiscriminatory state can thus be monitored.

5. EXAMPLE

The procedures described above were applied to 1974-75 salaries of Purdue University faculty. On the basis of these results, it was decided to drop the articles, books and consulting variables. We believe that the information sought from these variables is more validly measured by the Q variables. Regression equation statistics for models using D, V and Q variables are presented in Table I.

Table I

REGRESSION EQUATION STATISTICS

YEAR	RANK	SAMPLE SIZE	SQUARED MULTIPLE CORRELATION	STANDARD DEVIATION ($)	COEFFICIENT OF VARIABILITY (%)
	Assistant	255	.630	101.	7.69
1974-75	Associate	131	.599	136.	8.07
	Full	75	.634	273.	12.28
	Assistant	245	.708	101.	7.10
1975-76	Associate	185	.635	141.	8.01
	Full	91	.776	220.	9.22

Most of the increase in correlations from 1974-75 to 1975-76 is probably due to correction of bad data.

Values of the statistic G ranged from .39 ±.07 to .55 ±.11 for the 1974-75 data. Since the results of our analysis were not available until after the 1975-76 salaries had been finalized, anticipated improvements should be observed in the 1976-77 data which is currently being analyzed.

The value ranges for the residual categories are given in Table II.

Table II

VALUE RANGES FOR RESIDUAL CATEGORIES

RANK	-2	-1	0	+1	+2
Assistant	<-75	-75, -30	-30, +30	+30, +75	>+75
Associate	<-100	-100, -40	-40, +40	+40, +100	>+100
Full	<-125	-125, -50	-50, +50	+50, +125	>+125

Since the detailed results of this analysis are private, the residual tables for the full model are not presented. To illustrate the use of such tables, however, the tables for regression equations using only D and V variables are given in Table III. Note that we do not recommend using residuals based on only these variables. The "expected" frequencies are in parenthesis.

Table III

RESIDUAL TABLES FOR DV MODEL
OBSERVED AND EXPECTED FREQUENCIES

RANK	SEX	-2	-1	0	+1	+2	Total
Full	Male	26	11	7	1	22	67
		(23)	(11)	(8)	(3)	(21)	
	Female	2	1	2	2	1	8
		(3)	(1)	(1)	(0)	(2)	
	Total	28	12	9	3	23	75
Associate	Male	23	14	9	19	25	110
		(24)	(17)	(28)	(19)	(22)	
	Female	6	6	4	4	1	21
		(5)	(3)	(5)	(4)	(4)	
	Total	29	20	33	23	26	131
Assistant	Male	24	39	64	32	27	186
		(32)	(39)	(60)	(25)	(30)	
	Female	20	15	18	2	14	69
		(12)	(15)	(22)	(9)	(11)	
	Total	44	54	82	34	41	255

If the results in Table III were based on the full equation, one would conclude that there is a problem with the assistant professors in the lower categories. In addition, the pattern for the associate professors deserves some attention.

Further analysis proceeds with inspection of residual lists for each department-rank combination. In Table IV, residual lists are presented. These particular lists do not correspond to any real data but have been constructed to illustrate the type of patterns that occur with real data.

Table IV

TYPICAL RESIDUAL LISTS

Department-Rank

1		2		3		4		5	
Sex	Residual	Sex	Residual	Sex	Residual	Sex	Residual	Sex	Residual
F	-105	M	-110	F	-101	M	-106	F	-100
M	-37	M	-48	M	-67	M	-54	M	-90
M	-30	M	-24	M	-19	M	-34	M	-73
M	-25	M	-4	F	32	M	20	F	-53
M	6	M	3	M	36	M	26	M	-49
M	95	F	10	M	48	M	27	M	-23
M	95	M	22	M	72	M	53	M	-7
		M	45			F	68	M	-3
		M	106					F	15
								M	45
								M	57
								M	80
								F	99
								F	102

Note that in all cases, the residuals add up to zero (there may be some small round-off errors.) Departments 1 and 3 should be asked to provide an explanation for the apparently underpaid females. In department 2, the pattern looks good while in department 4 some reverse discrimination may be present. The department 5 residuals, while not showing a particularly bad pattern, deserve some review.

We would like to point out that this methodology allows the administrators to look at all residuals, regardless of the sex of the individual.

6. CONCLUSIONS

The effectiveness of the procedures described in this paper ultimately depends on the use of the results by university administrators. Clearly, a sincere effort on the part of these individuals is required. With such an effort, we believe that responsible administrators can effectively use the data provided to diagnose and correct instances of sex discrimination in faculty salaries.

The methods presented in this paper, while devised for detecting sex discrimination, can also be effectively used to pinpoint possible cases of salary inequities for all faculty members.

REFERENCES

[1] Astin, H. S. and Bayer, A. E., "Sex discrimination in academe," Educational Record, 53: 101-118 (1972).

[2] Bayer, A. E. and Astin, H. S., "Sex differentials in the academic reward system," Science, 788: 796-802 (1975).

[3] Darland, M. G., Dawkins, S. M., Lovasich, J. L., Scott, E. L., Sherman, M. E. and Whipple, J. L., "Application of multivariate regression to studies of salary differences between men and women faculty," 1973 Social Statistics Section Proceedings of the American Statistical Association: 120-132 (1973).

[4] Gastwirth, J. L., "Measurement of earnings differentials between the sexes," 1973 Social Statistics Section Proceedings of the American Statistical Association: 133-137 (1973).

[5] Gastwirth, J. L., "Statistical measures of earnings differentials," The American Statistician 29: 32-35 (1975).

[6] Gordon, N. M., Morton, T. E. and Braden, I. C., "Faculty salaries: is there discrimination by sex, race and discipline," The American Economic Review 64: 419-427 (1974).

EFFECT OF USING A POVERTY DEFINITION BASED ON HOUSEHOLD INCOME

John M. McNeil, Douglas K. Sater and Arno I. Winard
Bureau of the Census

This paper describes the impact on the poverty count of moving from a poverty definition based on family or individual income to one based on household income.

The basic units in statistics on poverty are families and unrelated individuals. A family is comprised of all related persons who share the same residence; unrelated individuals are those persons who do not live with a relative. An unrelated individual may live alone, with a family, or with other unrelated individuals.

Under the current definition, the poverty status of a person who lives with one or more relatives is determined by the income of that person plus the income of those relatives with whom he or she lives. The poverty status of an unrelated individual who is 14 years of age or older is wholly determined by the income of that person (unrelated individuals under 14 years of age are excluded from the poverty universe). Under a definition based on household income, the poverty status of a person would be determined by the income of that person plus the income of any other persons with whom he or she lives. The change to a household-based concept would not affect the poverty status of persons who live alone or persons who live only with relatives, but it would affect the poverty count among unrelated individuals who do not live alone and among family members who live in a residence in which unrelated individuals are present.

There are good reasons for choosing families and unrelated individuals as basic economic units. Families exist as economic entities by virtue of traditional and legal bonds. Bonds between unrelated individuals who share a single residence are generally much weaker than familial bonds. This is so despite the fact that many unrelated individuals have strong ties to those with whom they live, e.g., unmarried couples who live together and regard themselves as a single economic unit, and families who consider the unrelated individual living with them as part of the family for economic purposes. Yet, although the bond may be weaker, households made up of unrelated individuals enjoy most benefits of economies of scale that characterize the economic situation of families. These benefits include the utilization of a single shelter, and the purchase and preparation of food in quantities.

Under the current definition, families and unrelated individuals over 14 years of age are classified as poor or non-poor by comparing their annual income with certain threshold measures. These threshold measures vary according to the size and composition of the family, the sex and age of the family head, and farm-nonfarm residence. For families, the poverty threshold increases as family size increases. An unrelated individual is a basic unit in himself, i.e., the size of the family is one. A household-based definition would simply treat all household members as family members.1/ As an example of the effect of a household-based definition, consider two households - household A and household B. Both are two person households containing one male and one female, but the persons in household A are related and the persons in household B are not. In 1974, the poverty threshold for the family in household A was $3,324. The poverty threshold for the male in household B was $2,658 and the poverty threshold for the female was $2,458. The combined poverty thresholds for the two persons in household B was more than 50 percent higher than the poverty threshold for the two person family in household A. Under the current definition, of course, the income of one person in household B does not affect the poverty status of the other person, i.e., one person in household B could have a very high income and the other person could still be classified as poor.

The poverty rate can be affected by changes in living arrangements and there have been significant changes in the pattern of living arrangements during the past few years. From 1970 to 1975, the number of households headed by a family member increased by only 8 percent while the number of single person households increased by nearly 30 percent and the number of households comprised of two or more unrelated individuals increased by nearly 50 percent. This latter arrangement, however, still represents a rather small proportion of all households. In 1975 there were 55½ million households headed by a family member, 15½ million single person households, and 1.6 million households comprised of two or more unrelated individuals.

The effect on poverty counts of a change to a household-based definition is shown in tables 1 and 2. Table 1 shows that, as of 1974, the number of persons in poverty would be reduced by about one and one-half million or by about 6½ percent if the family-based definition were replaced by a household-based definition. The change in definition would shift about one-half million family members and about one million unrelated individuals out of poverty.

The one-half million family members who would be shifted out of poverty represent less than 3 percent of all family members who are currently poor, but the one million decline in the number of poor unrelated individuals represents a shift of about 22 percent.

Table 2 shows that 887,000 unrelated individuals lived with families in March 1975. While 433,000 of these persons were counted as poor under the current definition, only 82,000 would be counted as poor under a household-based definition. The number of poor unrelated individuals who lived with one other unrelated individual would be reduced from 670,000 to 228,000 and the number of poor persons living in households comprised of

Table 1.—EFFECT OF USING A POVERTY DEFINITION BASED ON HOUSEHOLD INCOME ON THE ESTIMATED NUMBER OF PERSONS WHO WERE POOR IN 1974

(Numbers in thousands)

Characteristic	Current definition	Household-based definition	Difference
Total number of persons in poverty..........	24,260	22,678	1,584
Total number of family members in poverty...	19,440	18,919	521
Total number of unrelated individuals in poverty..................................	4,820	3,758	1,062

Source: Special tabulation from March 1975 Current Population Survey.

Note: The number of poor families with whom one or more unrelated individuals lived was estimated to be 213,000 under the current definition and 76,000 under a definition based on household income. The difference of 137,000 was multiplied by 3.8 (the average size of poor families) to obtain the estimate of 521,000 family members who would be shifted out of poverty by a change in definition.

exactly three or exactly four unrelated individuals would be reduced from 234,000 to 37,000. In group quarter households (those with five or more unrelated individuals) the number of poor persons would be reduced from 198,000 to 126,000. The decline of one million in the number of poor unrelated individuals represents a reduction of almost 70 percent in the number of poor unrelated individuals who do not live alone.

Persons under 14 years of age who do not live with relatives are excluded from the poverty universe. They do not have family income, of course, and income questions are asked only of persons 14 years of age or older. Table 2 shows that there were about 229,000 persons in this category in March 1975. The application of the household-based definition to this group would place 34,000 in poverty.

The data show that a household-based poverty definition would make only a modest impact on the total count of persons in poverty, but would have a significant and probably growing impact on the poverty rate among unrelated individuals. It is not possible to make the generalization that a household-based definition is more equitable than a family-based definition. The issue depends upon the economic relationship between or among the household members. A possible approach would be the addition of relevant questions to the March Current Population Survey which would allow tabulations of data for spending units. The spending unit concept is subject to some ambiguity, however, because persons may share some basic expenses and not others. In the absence of information on spending units, it would be useful to periodically publish poverty counts based on the household concept.

FOOTNOTE

1/ In the special tabulations prepared for this paper, the poverty status of the household members was determined by comparing the combined income of the household members with the applicable threshold. The thresholds varied by size of household, the sex and age or the household head, and farm-nonfarm residence. For example, the poverty threshold for a nonfarm household comprised of three unrelated individuals, at least one of whom was a male, was defined to be equal to the poverty threshold for a nonfarm family of three with a male head. If the combined income of the three unrelated persons in the household was less than the threshold, each of the persons was considered to be in poverty. If the combined income was at the threshold level or higher, none of the persons was considered to be in poverty.

Table 2.--POVERTY STATUS IN 1974 OF UNRELATED INDIVIDUALS BASED ON TWO DEFINITIONS; THE CURRENT DEFINITION AND ONE BASED ON HOUSEHOLD INCOME

(Numbers in thousands)

Unrelated individuals by age and composition of household	Total	Poor by current poverty definition	Poor by definition based on household income	Difference: current definition less household definition
Unrelated individuals 14 years of age or older	18,872	4,820	3,758	1,062
Living alone	13,939	3,286	3,286	—
Males	4,918	840	840	—
Females	9,021	2,445	2,445	—
Not living alone	4,933	1,534	472	1,062
Males	2,938	755	253	502
Females	1,994	779	220	559
Living with families	887	433	82	351
Aged 14 to 17	88	86	9	77
Aged 18 and over	799	347	73	274
Males	507	181	51	130
Females	291	166	22	144
Living with one other unrelated individual	2,832	670	228	442
Males living with a female	507	132	50	82
Females living with a male	507	211	50	161
Males living with a male	1,105	167	59	108
Females living with a female	713	160	68	92
Living with two or three other unrelated individuals	749	234	37	197
Males	529	146	24	122
Females	229	87	12	75
Living with four or more other unrelated individuals	465	198	126	72
Males	250	82	64	18
Females	215	116	62	54
Unrelated individuals under 14 years of age	229	(NA)	34	(NA)

Source: Special tabulation from March 1975 Current Population Survey.

AN ECONOMETRIC ANALYSIS OF NURSING HOME COSTS USING COMPLEX SAMPLE SURVEY DATA

Mark R. Meiners
National Center for Health Statistics

Introduction and Background

In 1973 the National Center for Health Statistics expanded its National Nursing Home Survey (NNHS) to collect data on the cost to facilities of providing long-term health care. The sample design for the cost data and the other facility characteristics was stratified random sampling. As with most large sample surveys of this sort, the principal goal of the sample design was to achieve a stated degree of precision on the estimation of various descriptive statistics (e.g., percentages, means, totals) for a minimum cost. There is, however, considerable interest in using these data in research which is orientated toward making analytical inferences either to suggest or confirm specific hypotheses.

One of the tools most widely used to accomplish analytical research is the multiple regression technique. This is particularly true of econometric research. Problems arise, however, in the direct application of the regression technique to data generated by a complex sample survey procedure because the randomness assumptions are violated. The 1973-74 NNHS, for example, was stratified into 26 certification (e.g., Both Medicare and Medicaid or Medicare only; Medicaid only; and Not Certified), bedsize groups.[1] Further stratification within each of these primary strata was accomplished by ordering by type of ownership, geographic region, State and county. The sample was then selected systematically after a random start within each primary stratum with the sampling fraction varying in an approximate inverse relation to the expected standard deviation of estimates for the stratum. Regression models which cross these primary stratum cannot, therefore, be assumed to have been generated from a simple random sample and additional estimation procedures must be considered.

Judging from what little can be found in the econometric literature to provide guidance on this subject, opinions vary as to what the problems are and, therefore, how to deal with them.[2-7] Central to the discussion is the use of weights in the estimation of econometric models. Weights are discussed in the context of two possible sources of bias which arise when stratified sample survey data are used in a regression analysis.

The first source of bias has to do with the estimate of the regression coefficient. In this context weights are used to compensate for the difference in sampling rates across strata as well as to adjust for nonresponse and post-stratification. The weights are generated as part of the sampling process and are used with the sample data to provide unbiased estimates of the characteristics of a finite population such as nursing homes as defined in the NNHS. In the literature there is little agreement on the need for this type of weighting when the objective is to estimate a regression model and most econometric texts do not discuss this issue. I suspect that the decision depends on whether the objective of the model is to estimate the parameters of some underlying universal law or alternatively to estimate the parameters of some finite population relationships. Further development of these ideas is beyond the scope of this paper. Both weighted and unweighted models, however, were estimated to provide some empirical evidence on the results of the alternative procedures.

The second source of bias has to do with the estimated variance of the regression coefficient. In econometric texts the most widely accepted need for weights with regression is the generalized least squares adjustments suggested to overcome the problems of heteroscedasticity and autocorrelation. These problems arise when the variance of the error terms are not all equal and/or the error terms are correlated. In these cases the observation in the model are transformed by weights which are designed to produce the desired properties of homogeneous variances and zero covariances in the error structure. Heteroscedasticity and autocorrelation are frequently encountered in dealing with cross-sectional data such as the NNHS where the sample has been chosen from groups stratified to reflect the different degrees of variability in the characteristics of interest. In practice, however, it is often quite difficult to obtain the information needed to properly transform the error matrix and at best approximate procedures are used. Fortunately, an alternative approach has been developed.

McCarthy, Kish, Frankel and others have suggested the technique of balanced repeated replications (BRR) as a viable approach to estimating the variance of regression coefficients when the data are from a complex sample survey design.[2, 7, 8-12] To implement this approach subsamples which replicate the sample design are formed by randomly selecting observations from the total sample, with each replicate subsample having approximately half the observations. The essence of the approach is quite straightforward. The regression model of interest is run on the total sample and on each half sample replicate. The estimated variance of a regression coefficient can then be calculated by the following formula:[13]

$$S^2_{\hat{B}} = \frac{1}{r} \sum_{i=1}^{r} (B_i - \hat{B})^2$$

$\hat{B}$ is the regression coefficient estimated from the total sample

B_i is the regression coefficient estimated using half sample replicate i r is the number of half-sample replicates utilized

A detailed exposition on features of the BRR approach is found in a NCHS (1966) publication written by McCarthy.[8] He has found that under quite general conditions analytical unbiased estimates of the standard errors of regression coefficients can be generated using this method.

Outline of the Paper

The purpose of this paper is to compare the empirical results of using BRR regression in analyzing the cost data from the 1973-74 NNHS to the simpler approach of ignoring all or part of the sample design and using ordinary regression. Two basic cost models are estimated using what amounts to four alternative approaches. Approach 1 takes into account all the survey design features by using the sample weights with BRR regression. This is considered the most complete approach and will serve as a point of reference. Approach 2 uses the sample weights but applies ordinary regression. Approach 3 uses BRR regression but does not use the sample weights. Approach 4 disregards the complex survey design and estimation procedures, and treats the data as though they were from a simple random sample, i.e., ordinary regression is applied with no sample weights.

The differences in results between approach 1 and the alternative approaches will be discussed with respect to the various estimation procedures and their potential policy implications. Similar results between approaches 1 and 2 would tend to show that the design effect is small and not a major concern in the estimation process. Similar results between approaches 1 and 3 would indicate that by applying the sample weights with ordinary regression both the type of bias mentioned in the introduction have been eliminated and it would not be necessary to use BRR regression. The comparison of approach 1 and 4 will show the potential differences which occur when the alternative assumptions behind the need for weights (to overcome the first type of bias mentioned in the introduction) are made.

Nursing Home Cost Models

Two major studies of cost functions for nursing homes have appeared in the literature and these serve as the basis for the two models estimated here. The first model is a version of the stock-flow model suggested by Skinner and Yett.[14] It is a multiplicative model estimated in log linear form, its dependent variable is total cost and it features two dimensional output. The second model is the hyperbolic version from the set of "classical" cost functions estimated by Ruckline and Levey.[15] It is an additive model estimated lineraly, its dependent variable is average cost (total cost per resident day) and it features the inverse of the number of beds in the facility as the size/capacity measure. The cost models are estimated for the subset of nursing homes certified by Medicare, i.e., those facilities in the 1973-74 NNHS which were certified by both Medicare and Medicaid or by Medicare only. The complete list of independent variables used in the cost models and a brief description of their purpose is given in table 1. It should be noted that because of BRR program limitations each model is limited to 12 independent variables. While this limitation points up an obvious area for future work it does not seriously hinder the model specifications. Each model includes the variable categories generally considered to be important in the estimation of cost functions for health related facilities, i.e., output or capacity, occupancy rate, scope of facility services, resident characteristics, quality indicator, factor prices, and standardizing measures for ownership and location.

Table 1. Variables used in cost model estimates and their description.

Variable	Description
ADM	ADM is the number of persons admitted to the facility in 1972. This variable is used as one of the two output measures in the total cost model. For a U-shaped average cost curve, the coefficient estimate must be positive.
LNRD	LNRD is the natural log of the total number of resident days of care provided in 1972. This variable is one of the two output measures in the total cost model. For a U-shaped average cost curve, the coefficient estimate must be greater than zero and less than 1.
BEDSINV	BEDSINV is the inverse of the number of beds in the facility in 1972. This variable is used as the size/capacity measure in the average cost model. A positive coefficient estimate indicates possible economies to scale.
OR	OR is the average occupancy rate for the facility in 1972. This variable is included in both the total and average cost model to account for capacity utilization. For both models a negative coefficient estimate is expected.
PROFIT	PROFIT is 1 if the facility is a proprietary facility and 0 if it is a voluntary or government facility. This variable is included in both the total and average cost model to account for the effect of proprietary vs. nonproprietary control.
NE NC S	NE, NC, S are 1 or 0 dummy variables used to represent the four geographic regions--North East, North Central, South and West (this last region is left out as is required when dummy vari-

ables are used). These variables are included in both the total and average cost model to account for regional variation in factor prices otherwise unaccounted for.

PBCERTMR PBCERTSN PBCERTIC — PBCERTMR, PBCERTSN, PBCERTIC are the proportion of beds in the facility certified as Medicare beds, Medicaid skilled nursing care beds, Medicaid intermediate care beds, respectively. These variables are included to account variation in the scope of services the facility provides. Because of BRR program limitation, only PBCERTMR and PBCERTIC were included in the total cost model.

KATZAB — KATZAB is the percent of residents in the facility who are totally independent in various activities of daily living or are dependent only in bathing as measured by the Katz patient assessment scale.[16,17] This variable is used in both the total and average cost model to account for resident mix characteristics. For both models a negative coefficient estimate is expected.

AHSLPN — AHSLPN is the average hourly salary of licensed practical nurses working in the facility. This variable is used in both the total and average cost model to account for variation due to wage levels. For both models a positive coefficient estimate is expected.

RNSHF — RNSHF is the number of shifts per day that the facility has a registered nurse on duty. This variable can take a value of 0 to 3. It is included in both the total and average cost model as a quality indicator. For both models a positive coefficient estimate is expected.

Analysis of Results

The results from estimating the total cost model and the average cost model using the alternative estimation approaches are presented in tables 2 and 3 respectively. Columns I and II of these tables comprise approach 1. Columns I and III comprise approach 2, columns IV and V comprise approach 3 and columns IV and VI comprise approach 4. These results are used in tables 2-A, 2-B, 3-A and 3-B to compare the effect of the alternative estimation approaches on the coefficient estimates, the standard error estimates, and the resulting inferential statistics.

Column I of tables 2-A and 3-A gives the ratio of weighted to unweighted coefficient estimates for the two models. While on average the ratio for each model indicates only about a 3 percent difference between the weighted and unweighted coefficient estimates, the variation of these ratios is considerable and in neither model is the weighted coefficient always higher or lower than the unweighted coefficient. For the total cost model the weighted coefficients ranged from 28 percent higher to 25 percent lower than the unweighted coefficients (table 2-A). For the average cost model the range was even wider with the weighted coefficients ranging from 74 percent higher to 50 percent lower than the unweighted coefficients (table 3-A). From a policy point of view it seems likely that such differences would be of concern. For example, using the total cost model with weights implies nearly a 2 percent greater increase in total costs with an increase of one in the number of shifts with a registerd nurse on duty than would be expected using the unweighted version. Also for the total cost model the weighted coefficients imply that being a proprietary facility results in about a 4 percent smaller decrease in total costs than would be indicated by the unweighted coefficient. Similar examples are available from the average cost model. For the weighted version a $1 increase in the average hourly salary of the LPNs increases average costs by $.93 less than for the unweighted version and being in the Northeast implies a $1.75 difference in impact between the weighted and unweighted versions.

The impact of the weights on the standard error estimates of the two models is also shown in tables 2-A and 3-A. Column II shows the weighted to unweighted standard error ratios when the BRR approach was used and column III shows these ratios when the simple random sample (SRS) approach was used. For the total cost model the weighted BRR standard errors average 22 percent higher than the unweighted BRR standard errors with the weighted estimates ranging from 105 percent larger to 11 percent lower than the unweighted estimates. The ratios of SRS standard errors for this model were considerably less variable. The weighted SRS standard errors ranged from 11 percent larger to 9 percent smaller than the unweighted SRS standard errors and were on average only 1 percent larger than the unweighted estimates. For these comparisons the results from the average cost model are similar to those for the total cost model in that the variability of the BRR standard error ratios is considerable larger than that for the SRS standard error ratios. The weighted BRR standard errors averaged 18 percent higher than the unweighted BRR standard errors with the weighted estimates ranging from 165 percent higher to 63 percent lower than the unweighted estimates. The weighted SRS standard errors averaged 5 percent lower than the unweighted SRS with the weighted estimates ranging from 52 percent higher to 44 percent lower than the unweighted estimates.

Other than the smaller differences between weighted and unweighted standard error estimates using the SRS approach rather than the BRR approach, there appears to be no systematic indication of whether weighted or unweighted estimates will be larger. For variables in both models examples can be found where the weighted standard error estimate exceeded the unweighted estimate when the BRR approach was used and was less than the unweighted estimate when the SRS approach was used. It is not possible, therefore, to infer that either weighted or unweighted standard errors would always be larger regardless of the estimation approach nor is it possible to infer that the BRR standard error ratio will be greater than one if the SRS standard error ratio is greater than one.

The design effects (the ratio of the BRR standard error to the SRS standard error) for the variables in the two models are presented in columns IV (weighted) and columns V (unweighted) of tables 2-A and 3-A. For both models the variability of the design effect is considerable. For the total cost model the weighted standard error design effect averaged 1.35 but ranged from 2.47 to .51. The unweighted design effect averaged 1.18 and range from 2.03 to .47. For the average cost model the weighted standard error design effect averaged 1.18 and ranged from 2.11 to .63 while for the unweighted version it averaged 1.05 and range from 1.75 to .61.

Tables 2-B and 3-B give the "t" statistics for the independent variables of the two cost models under the alternative estimation approaches. Column I is calculated by dividing the weighted coefficient by the weighted BRR standard error. Column II is calculated by dividing the weighted coefficient by the weighted SRS standard error. Column III is calculated by dividing the unweighted coefficient by the unweighted BRR standard error. Column IV is calculated by dividing the unweighted coefficient by the unweighted SRS standard error. The "t" statistics are marked to signify their level of significance. Those coefficients not marked can be considered significant only at levels below those normally acceptable. It should be noted that the degrees of freedom for these tests of significance depend on the estimation approach used. For the BRR approach the degrees of freedom are equal to the number of half samples (20) used in the analysis while for the SRS approach the degrees of freedom are the number of cases less the number of independent variables and the constant (603).

For the total cost model (table 2-B) the effect of the alternative estimation approaches on the significance tests is small. Each coefficient is significant at least at the .90 level regardless of the approach used and in many cases the level is much higher. Shifts do occur, however, in the level of confidence at which variables are considered significant when the alternative approaches to estimating standard errors used. The effect of these shifts becomes more noteworthy when the results for the average cost model (table 3-B) are considered. Taking into account both the weighted and unweighted versions of this model there are examples where the effect was such that when the SRS "t" statistic was significant at the 1 percent level the BRR "t" statistic was not significant and, alternatively, when the SRS "t" was not significant the BRR "t" was significant at the 5 percent level. These results indicate that to ignore the effect of the sample design on the standard error estimates can lead to erroneous inferences. The most striking example from a policy point of view is the coefficient (weighted or unweighted) for BEDSINV in the average cost model. Using SRS estimation procedures the positive sign of the coefficient and its "t" statistic indicate economies to scale, a result which, in this model, is shown to be questionable when the effect of the survey design is accounted for by using the BRR technique.

Summary and Conclusions

This paper compared the empirical results of using BRR regression in estimating two representative nursing home cost models to the simpler approach of ignoring all or part of the 1973-74 NNHS sample design and using ordinary regression. The effects of the alternative estimation procedures on the coefficient and standard error estimates and some potential policy implications were discussed.

The results showed that ignoring the sample design when using the 1973-74 NNHS with regression to analyze nursing home costs will effect both the coefficient and standard error estimates. None of the alternative short-cut methods gave results that consistently approximated the results gotten by using information about the sample design.

The effects of the sample weights on the coefficients and standard errors were analyzed by comparing ratios of weighted to unweighted estimates. For the coefficients and their standard errors, the ratios showed considerable variation with no apparent systematic way of predicting the magnitude of the ratio or whether it would be greater or less than one.

The design effect on the standard errors was analized by comparing the ratios of BRR standard errors to the SRS standard errors for both the weighted and unweighted versions of the two models. In all cases the individual design effects varied to such a large degree that the average design effect for a model could not be used to reliable adjust the SRS standard errors.

Finally, the effect of the sample design on inferences was considered by calculating the "t" statistics for the coefficients of the two models estimated under the alternative approaches. The results showed that the potential policy inferences were in some important instances effected by the estimation approached used. Therefore, if the goal is to estimate the cost function parameters for the finite population of nursing homes as defined in the 1973-74 NNHS the full sample design features must be considered in the estimation process.

Table 2. Summary of selected regression results for total cost model using estimation approaches 1 through 4, Medicare certified nursing homes, 1973-74 NNHS data.

Variable	I Wtd. Coef. Est.	II Wtd. BRR S.E. Est.	III Wtd. SRS S.E. Est.	IV Unwtd. Coef. Est.	V Unwtd. BRR S.E. Est.	VI Unwtd. SRS S.E. Est.
ADM	.00071	.00015	.00010	.00057	.00016	.00009
LNRD	.87923	.06109	.02693	.91416	.05247	.02580
OR	-.00736	.00121	.00091	-.00777	.00113	.00086
PROFIT	-.12937	.05571	.02980	-.17340	.03839	.02849
NE	.35548	.01727	.03404	.42311	.01640	.03488
NC	.19772	.04003	.03569	.20273	.04058	.03561
S	.08006	.03049	.03315	.07891	.02385	.03459
PBCERTMR	.09482	.03481	.02693	.08818	.03336	.02739
PBCERTIC	-.08252	.04020	.03674	-.07797	.02830	.03567
KATZAB	-.00335	.00163	.00066	-.00282	.00127	.00066
AHSLPN	.03770	.01005	.00920	.03726	.00490	.00916
RNSHF	.08007	.01245	.01457	.06241	.01402	.01603

Constant = 4.4107 R^2 = .80339   N = 616 (Weighted)

Constant = 4.1706 R^2 = .82339   N = 616 (Unweighted)

Table 3. Summary of selected regression results for average cost model using estimation approaches 1 through 4, Medicare certified nursing homes, 1973-74 NNHS data.

Variable	I Wtd. Coef. Est.	II Wtd. BRR S.E. Est.	III Wtd. SRS S.E. Est.	IV Unwtd. Coef. Est.	V Unwtd. BRR S.E. Est.	VI Unwtd. SRS S.E. Est.
BEDSINV	127.45	99.111	47.763	162.49	115.84	66.057
OR	- .2826	.0136	.0217	- .2895	.0302	.0213
PROFIT	-3.8678	1.5368	.8557	-4.1790	1.2582	.8427
NE	7.5954	.7556	.9708	9.3447	.6263	1.0258
NC	4.6790	.8715	1.0100	4.4155	.7598	1.0414
S	2.6089	.8755	.9394	2.5209	.6893	1.0141
PBCERTMR	2.3062	.7417	.8166	2.2627	.7574	.8201
PBCERTSN	1.8714	1.0347	.9867	1.3383	1.0250	1.0411
PBCERTIC	-1.9005	.9501	1.0915	-1.7406	.7742	1.1162
KATZAB	- .0576	.0393	.0187	- .0534	.0193	.0226
AHSLPN	.9149	.2168	.2605	1.8434	.5816	.4660
RNSHF	2.0497	.6297	.4050	1.1756	.2374	.2673

Constant = 34.561 R^2 = .35934   N = 616 (Weighted)

Constant = 35.163 R^2 = .36959   N = 616 (Unweighted)

Table 2-A. Ratio comparisons of weighted to unweighted coefficient and standard error estimates and design effects on standard errors (s.e. DEFF) for weighted and unweighted regressions, total cost model.

Variable	I Wtd./ Unwtd. Coef. Est.	II Wtd./ Unwtd. BRR S.E. Est.	III Wtd./ Unwtd. SRS S.E. Est.	IV Wtd. S.E. DEFF	V Unwtd. S.E. DEFF
ADM	1.25	.94	1.11	1.48	1.88
LNRD	.96	1.16	1.04	2.27	2.03
OR	.94	1.07	1.06	1.32	1.31
PROFIT	.75	1.45	1.05	1.87	1.34
NE	.84	1.05	.98	.51	.47
NC	.97	.99	1.00	1.12	1.14
S	1.01	1.27	.96	.92	.69
PBCERTMR	1.08	1.04	.98	1.23	1.22
PBCERTIC	1.06	1.42	1.03	1.09	.79
KATZAB	1.19	1.28	1.00	2.47	1.92
AHSLPN	1.01	2.05	1.00	1.09	.53
RNSHF	1.28	.89	.91	.85	.88
Average	1.03	1.22	1.01	1.35	1.18

Table 3-A. Ratio comparisons of weighted to unweighted coefficient and standard error estimates and design effects on standard errors (s.e. DEFF) for weighted and unweighted regressions, total cost model.

Variable	I Wtd./ Unwtd. Coef. Est.	II Wtd./ Unwtd. BRR S.E. Est.	III Wtd./ Unwtd. SRS S.E. Est.	IV Wtd. S.E. DEFF	V Unwtd. S.E. DEFF
BEDSINV	.78	.60	.72	2.08	1.75
OR	.97	.45	1.02	.63	1.42
PROFIT	.93	1.22	1.02	1.80	1.49
NE	.81	1.21	.95	.78	.61
NC	1.06	1.15	.97	.86	.73
S	1.03	1.27	.93	.93	.68
PBCERTMR	1.02	.98	1.00	.91	.92
PBCERTSN	1.39	1.01	.95	1.05	.99
PBCERTIC	1.09	1.23	.98	.87	.69
KATZAB	1.08	2.04	.83	2.11	1.17
AHSLPN	.50	.37	.56	.83	1.25
RNSHF	1.74	2.65	1.52	1.56	.89
Average	1.03	1.18	.95	1.18	1.05

Table 2-B. "t" statistics using weighted and unweighted BRR and SRS approaches total cost model.

Variable	I Wtd. BRR t Stat.	II Wtd. SRS t Stat.	III Unwtd. BRR t Stat.	IV Unwtd. SRS t Stat.
ADM	4.73***	7.10***	3.56***	6.33***
LNRD	14.39***	32.65***	17.42***	35.43***
OR	6.08***	8.09***	6.88***	9.03***
PROFIT	2.32**	4.34***	4.52***	6.09***
NE	20.58***	10.44***	25.80***	12.13***
NC	4.94***	5.54***	5.00***	5.69***
S	2.63**	2.42**	3.31***	2.28**
PBCERTMR	2.72**	3.52***	2.64**	3.22***
PBCERTIC	2.05*	2.25**	2.76**	2.19**
KATZAB	2.06*	5.06***	2.22**	4.27***
AHSLPN	3.75***	4.10***	7.60***	4.07***
RNSHF	6.43***	5.50***	4.45***	3.89***

*** significant at 1 percent level
** significant at 5 percent level
* significant at 10 percent level

Table 3-B. "t" statistics using weighted and unweighted BRR and SRS approaches, average cost model.

Variable	I Wtd. BRR t Stat.	II Wtd. SRS t Stat.	III Unwtd. BRR t Stat.	IV Unwtd. SRS t Stat.
BEDSINV	1.29	2.67***	1.40	2.46**
OR	20.78***	13.01***	9.58***	13.60***
PROFIT	2.52**	4.52***	3.32***	4.96***
NE	10.05***	7.82***	14.92***	9.11***
NC	5.37***	4.63***	5.81***	4.24***
S	2.98***	2.78***	3.66***	2.49**
PBCERTMR	3.11***	2.82***	2.99***	2.76***
PBCERTSN	1.81*	1.90*	1.31	1.29
PBCERTIC	2.00*	1.74*	2.25**	1.56
KATZAB	1.47	3.09***	2.77**	2.37**
AHSLPN	4.22***	3.51***	3.17***	3.96***
RNSHF	3.26***	5.06***	4.95***	4.40***

*** significant at 1 percent level
** significant at 5 percent level
* significant at 10 percent level

Notes and References

1. For a more detailed description of the survey design see pp. 16-18 of National Center for Health Statistics: Selected Operating and Financial Characteristics of Nursing Homes, United States: 1973-74 National Nursing Home Survey, Vital and Health Statistics, Series 13, No. 22, DHEW Pub. No. (HRA) 76-1773. Washington, U.S. Government Printing Office, Dec. 1975.

2. Cohen, M.S., Rea, A. and Lerman, R.I. A Micro Model of Labor Supply, BLS Staff Paper 4, Government Printing Office, 1970.

3. Cramer, J.S. Empirical Econometrics, Amsterdam: North-Holland, 1971, 141-143.

4. Hu, T.W. and Stromsdorfer, E.W. A Problem of Weighting Bias in Estimating the Weighted Regression Model, Proceedings of the Business and Economics Section of the American Statistical Association, 1970, 513-516.

5. Klein, L.R. and Morgan, J. Results of Alternative Statistical Treatments of Sample Survey Data. Journal of the American Statistical Association, 46, 1951, 442-460.

6. Klein, L.R. A Textbook of Econometrics. Evanston: Row, Peterson and Company, 1953, 406-415.

7. Roth, J. Application of Regression Analysis to Sample Survey Data: Unpublished memorandum, Social Security Administration, 1971.

8. National Center for Health Statistics: Replication: An Approach to the Analysis of Data From Complex Surveys. Vital and Health Statistics, Series 2, No. 14, Public Health Service Pub. No. 1000, Washington, D.C., U.S. Government Printing Office, April 1966.

9. National Center for Health Statistics: Pseudoreplication: Further Evaluation and Application of the Balances Half-Sample Technique, Vital and Health Statistics, Series 2, No. 31, Public Health Service Pub. No. 1000, Washington, D.C., U.S. Government Printing Office, January 1969.

10. Kish, Leslie, and Frankel, Martin. Balanced Repeated Replications for Standard Errors. Journal of the American Statistical Association, Sept. 1970, Vol. 65, No. 331, 1071-1094.

11. Frankel, Martin. Inferences From Sample Surveys: An Empirical Investigation, Ann Arbor, Michigan: Litho Crafters, Inc., 1971.

12. Simmons, Walt R. and Baird, James T. Jr. Pseudo-Replication in the NCHS Health Examination Survey. Proceedings of the Social Statistics Section of the American Statistical Association, 1968, 19-30.

13. To yield more information and improve the variance estimate the alternative variance formula $S_B^2 = \frac{1}{4r} \sum_{i=1}^{r} (B_i - B_i')^2$ was used in this paper. For this formulation the regression coefficient for each compliment half-sample (B_i') is also calculated.

14. Skinner, D.E. and Yett, D.E. Estimation of Cost Functions for Health Services: The Nursing Home Case. Paper presented at the 40th Annual Meeting of the Southern Economic Association, Atlanta, Georgia, November 1970.

15. Ruchlin, H.S. and Levey, S. Nursing Home Cost Analysis: A Case Study. Inquiry, IX, September, 1972, 3-15.

16. Katz, S., et. al. Studies of Illness in the Aged. The Journal of the American Medical Association, Vol. 185, No. 12, September, 1963, 914-919.

17. Katz, S., et. al. Progress in Development of the Index of ADL. The Gerontologist, Vol. 10, Spring, 1970, Part 1, 20-30.

CHOOSING A COLLEGE MAJOR AND LOOKING FOR A JOB

M. Melnyk, Kent State University

In recent years various studies of labor market for college graduates reveal sometimes substantial surplusses of graduates in various areas. According to a report by Carnegie Commission, "Nearly 30 percent of four-year male college graduates are now in blue-collar, sales, and clerical jobs, many of which do not make full use of their education."[1] A survey made by the Bureau of the Census for the Bureau of Labor Statistics indicates that 50.7 percent of college graduates got their first jobs in areas not directly related to their major field of study.[2]

In education, maladjustment between supply and demand is quite significant. The Government figures[3] indicate that in 1972 the number of holders of bachelor and master degree exceeded the additional demand for teachers in primary and secondary education by 26.2 percent. The above mentioned BLS survey[4] indicates that in October 1971 the percent of unemployed 1970 and 1971 bachelor's and advanced degree recepients ranged between 5.3 percent in business and commerce and 13.0 percent in humanities. On the other hand, projections in <u>Occupational Outlook Quarterly</u> and <u>Occupational Outlook Handbook</u>, both published by U.S.Department of Labor, as well as projections in the report by Carnegie Commission[5] indicate that in a number of areas the demand for college graduates is strong and on the increase.

There is, therefore, little doubt that quite a few high school graduates and college students chose wrong curricula. A pertinent observation on the subject has been made by Laure M. Sharp:[6] "But at times, the decision to major in a specific field is a fairly casual one, dictated by personal convenience rather than by a long-term occupational objective." In addition, a very large number of college graduates have a long waiting time between graduation and the first job[7] which is a good enough reason for suspicion that their job hunting methods are inadequate.

The problem of how to choose a major and how to find a job have been explored or touched upon by James A. Davis[8], Robert Calvert Jr.[9], Laure M. Sharp[10], Vera C. Perrella[11], and one or two others. Calvert's study is based on a survey of alumni by mail questionnaire. He got back about 11,000 usable questionnaires, but this was only 60 percent of questionnaires sent out. There is no indication in his study about the composition of those who did not respond. If those who did not respond are predominantly those who were not successful in finding satisfactory employment, then the results of the survey are of little value at least from this point of view.

In a similar situation Laure M. Sharp tried to reach a sample of 1,200 nonrespondents. She succeeded in obtaining only 697 usable responses. She seems to be satisfied with relatively few differences between respondents and nonrespondents, although it is not quite easy to see why. The sample of 697 is still only 58 percent of 1,200 people whom she tried to contact. In addition, in a sample of 697 there would be quite a few very small subsamples whose sampling variability would make a number of comparisons useless.

The present report provides some firm information in certain areas, but can be considered only a pilot study with respect to some other topics. The paper deals mainly with methods of choosing a major and with methods of choosing a job. The information comes from a survey of graduating seniors of Kent State University. The sample size was 500 of the total of about 2,000. The sampling error for sampling from dichotomous population is 3.8 percentage points for .95 confidence interval. The sampling errors for subsamples are larger, but in cases of too large a sampling error the results of the survey are simply not reported here.

The survey has been conducted by instructors and graduate assistants in the classrooms so that nonresponse problem has been eliminated. Some measure of proportionality of students in the sample with respect to the number of all graduating seniors in main areas of study has been achieved.

The results of the survey show that the choice of a major is in too many cases careless. This conclusion is supported by the fact that about 43 percent of students changed their majors. This figure is comparable with 38 percent obtained by Calvert.[12] Since, among others, difficulties in finding jobs in education were mentioned above, it may be noted that the percent of those who switched from education to something else is 25 percent for KSU and it is significantly higher that 15 percent in the Sharp[15] survey.

It would, of course, be interesting to see in detail which najors attract high school graduates and which become more popular as the students progress in their studies. Some information on this problem can be found in the book by Davis.[14] The results of the present survey indicate that substantial proportion of students changed their major within a major area. For example, 55 percent of students who changed their major and whose original major was a business major, made a change to another business major. Comparable percentage for education is about 52 percent and it is not significantly different from business. The percentage for humanities and arts is 36. It is 39 percent for social and behavioral sciences and 22 percent for mathematics and sciences. The survey also indicates that the greatest winner of those who change from one major to another is business. The greatest loser is sciences. Large sampling error does not permit more detailed comparisons.

Table 1 Decision to choose first major.

Decision to choose my first major was influenced by:	Rank numbers and percentages					Row totals
	1*	Per cent	2	Per cent	3	
My own considerations	423	82.3	11	6.2	1	435
Advice of parents	24	4.7	68	38.2	9	101
Advice of family & friends	17	3.3	40	22.5	8	65
My own friends	9	1.7	21	11.8	9	39
High school teachers	22	4.3	33	18.5	7	62
Other	19	3.7	5	2.8	3	27
Totals	514	100.0	178	100.0	37	---

*E.g.: Rank number 1 has been assigned to "My own considerations" by 423 students.

Table 2 Reasons for choosing first major.

I chose my first major for following reasons:	Rank numbers and percentages						Row totals
	1*	Per cent	2	Per cent	3	4	
Good paying job	66	13.1	50	23.2	10	5	131
Strong demand for graduates	75	14.9	47	21.8	14	3	139
Employment security	35	7.0	27	12.5	15	5	82
I love this area of study	235	46.7	23	10.6	12	-	274
Reasons other than employment	57	11.3	45	20.8	11	3	116
Want to go to graduate school	11	2.2	21	9.7	5	5	42
Other	24	4.8	3	1.4	1	3	31
Totals	503	100.0	216	100.0	68	24	---

*E.g.: Rank number 1 has been assigned to "Good paying job" by 66 students.

We note further that only 39 percent of high school graduates chose their major before entering college. Also the percent of changes of a major during junior year was about as high as during sohpomore and freshman year. It would certainly be interesting to find out why such a large number of changes persists into the junior year.

Let us now make a few observations on decisions concerning the choice of a major and decisions and reasons for change of a major. It appears from Table 1 that parents and high school teachers have little to do with the choice of a major by a high school graduate. The statement "My own considerations" in the choice of a major category was marked as the first choice 82 percent of times. Rank number two is given parents and high school teachers often. However, the proportion of students who checked off only one determinant of their decision to choose a major was 61 percent. Those who marked two determinants represent 33 percent; and only 6 percent checked off more than two determinants. Thus the percent of students who make decisions concerning the choice of a major on their own and who do not benefit from any advice on the matter must be considered excessively high. There is little doubt that this is the main reason for excessive changes of majors in subsequent years.

Why does a student choose a certain major? The reasons for choosing the first major can be inferred from Table 2. The reasons related to labor market were

Table 3 Decision to change major.

I changed my major mainly due to:	Rank numbers and percentages				Row totals
	1*	Per cent	2	Per cent	
My own decision	192	83.1	19	19.8	211
Advice of parents	16	6.9	15	15.6	31
Advice of family	5	2.2	10	10.4	15
My own friends	1	0.4	15	15.6	16
Professors or advisors	5	2.2	28	29.2	32
Other	12	5.2	9	9.4	21
Totals	231	100.0	96	100.0	---

*E.g.: Rank number 1 has been assigned to "My own decision" by 192 students.

Table 4 Reasons for change of major.

I changed major for following reasons:	Rank numbers and percentages				Row totals
	1*	Per cent	2	Per cent	
Better employment prospects	68	30.1	23	38.2	94
Low grades	37	16.4	9	13.2	46
Appeal of new major	87	38.5	16	23.6	103
Preparation for gradu te school	7	3.1	12	17.6	19
Other	27	11.9	5	7.4	32
Totals	226	100.0	68	100.0	---

*E.g.: Rank number 1 has been assigned to "Better employment prospects" by 68 students.

given rank 1 only 35 percent of times. However, rank 2 was given this category of reasons 45 percent of times. But even so it seems that a rather large proportion of high school graduates choose their major without too much preoccupation with the necessity of finding a job after four years of study. It may be mentioned again that 54 percent of students gave only one reason for their choice of a major. Two reasons were given by 31 percent of students and 10 percent of students gave three reasons.

The information concerning the decisions to change a major is provided in Table 3. Apparently, professors and advisors have little to do with students' decisions to change their majors. Their relative importance as secondary factor is somewhat greater, but again 73 percent of students checked off only one determinant of their decision to change a major.

The two major reasons for a change of a major are individual interest of the student and employment considerations. This can be seen in Table 4. How do these two reasons for a change of a major compare numerically with the reasons for a first choice of a major? The 35.0 percent of students who gave employment related reasons for the choice of the first major in Table 2 is statistically not significantly different from the 30.1 percent of students who marked "Better employment prospects" in Table 4. However, the 46.7 percent for "I love this area of study" in Table 2 is significantly greater than the 38.5 percent for "Appeal of new major" in Table 4. It appears also that the students do become more job oriented as they progress with their studies.

The above information suggests the following conclusions. Employment considerations figure strongly in the choice of the first major and in the change of a major and yet the students do not take advantage of experienced advisors either in the high school or in the college. A telephone interview of twenty high schools in Kent-Akron-Cleveland area indicates that the advisors are competent and use such publications as U.S.Department's of Labor Occupational Outlook Quarterly and Occupational Outlook Handbook. However, they seem to play a passive role. There is little doubt that much more could be done to make the students and their parents aware of career and employment prospects for various majors.

It is also obvious that college advisors should not restrict their activity to help students find a job. The percent of students who chose their major after entering college is in the vicinity of 60 percent. In addition, about 40 percent of students changed their major. Finally, the percent of students who are making these decisions entirely on their own, that is, without consultation with anybody, is unreasonably high by any standards. Therefore, college employment officers and advisors should play much more prominent role in providing these students with pertinent and useful information.

The information concerning job hunting methods is also quite revealing. About 44 percent of students did not look for

Table 5 Students who did not look for job during senior year.

I did not look for a job during senior year because I:	No	Per cent
intend to go to graduate or professionsl school	56	25.3
have a job waiting for me	35	15.8
was too busy with coursework	68	30.8
do not intend to work for awhile	46	20.8
Other	16	7.3
	221	100.0

Table 6 When did graduating seniors look for jobs?

	No	Per cent
Before September	15	5.4
Between September and November	39	14.0
Between December and February	80	28.8
After February	144	51.8
	278	100.0

Table 7 Job hunting methods. A comparison with BLS-BC survey.

Method	BLS-BC (1) Percent	KSU (2) Percent
College placement	17.6	37.2
Public employment service	1.5	2.4
Private employment agency	3.6	4.3
Newspapers	3.9	7.8
Direct application to employers	41.4	32.0
Friends or relatives	21.3	14.8
Other	10.7	1.5
	100.0	100.0

(1) The column lists percentages of recent college graduates who got their first jobs by one of the methods listed.
Source: Perella, Vera C., "Employment of Recent College Graduates," *op. cit.*, p.46.

(2) These are percentages of times a job hunting method was checked off by students who looked for jobs.

a job during their senior year. As may be seen in Table 5, of those who did not look for a job, 31 percent were too busy with their courses and 21 percent declared that they did not intend to work for awhile. Of those who looked for a job, a huge 52 percent (see Table 6) started looking for a job after February. The only explanation for this phenomenon of missed opportunities on such a large scale is that students are simply not aware of the fact that most intensive recruiting and interviewing occurs in late Fall and early Spring.

Of those who looked for a job, 26 percent did not get a single interview at the employer's headquarters. Only 29 percent got more than two interviews. These two percentages must be considered very large in spite of a difficult labor market due to economic recession. One explanation of these figures can be found in Table 6 where we note again that too many students started looking for a job too late. However, this is certainly not the only explanation. It is also important to have a closer look at job hunting methods.

The job hunting methods listed in the questionnaire can be seen in Table 8. The results of the survey indicate that a graduating senior used, on the average, 2.8 job hunting methods. But if the methods "Letters to employers," "Telephone calls to employers," and "Visited employers" are combined into one, the average number of methods used drops to 2.3. Of course, some areas lend themselves better to one method than to another (for example, graduates of college of education have to depend more on letter writing), but even so, an increase in the number and intensity of job hunting methods would undoubtedly land the students more interviews.

We note in Table 7 that the students' use of college placement service (37.2 percent) is much heavier than the proportion of first jobs (17.6 percent) secured with the help of this service by recent college graduates. Direct application to employers and reliance on friends and relatives are significantly lower for graduating seniors than for recent college graduates. Other differences cannot be declared significant due to small subsamples. However, the use of such means as employment services, newspapers, and "Other" is numerically very small.

Information in Table 8 suggests strongly that no major job hunting method should be disregarded. The proportions of interviews secured by such methods as employment services and newspaper ads are relatively high. These methods should be used much more often.

Teble 8 Job hunting methods that resulted in interviews at employers' place of business.

Method:	(1) No	(2) No	(3) Percent
College placement	191	92	48.2
Letters to employers	176	62	35.2
Telephone calls to employers	75	32	42.7
Visited employers	99	56	56.6
Answered ads	47	16	34.0
Family and friends	99	30	30.3
Professors	45	13	28.9
Other (state and private employment agencies, etc.)	58	20	34.5

(1) Number of times a method was checked off on the questionnaire.

(2) Number of times one or more interviews were attributed,to a job hunting method.

(3) Column (2) as percent of Column (1).

The importance of obtaining more interviews becomes more evident when it is realized that there is a relationship between the number of interviews and job offers. For example, of those graduating seniors, who got one interview, 32 percent got a job offer, but of those who got two interviews, 55 percent got a job offer. There is also a correlation between the number of interviews, and the number of job offers, although the size of the sample does not warrant precise comparisons.

Another piece of information should be of interest both to graduating seniors and to college advisors. As already mentioned, of those who looked for jobs, 26 percent - a very high proportion - did not get a single interview and only 45 percent got a job offer. Furthermore, the average salary of those who had 5 or more interviews and 3 or more job offers, was by $1,500 higher than the average salary of those with fewer interviews and job offers. It is, therefore, of basic importance for the graduating seniors to use more job hunting methods and to use them more efficiently in order to increase their chances to get more interviews and to get at least one job offer.

It may be noted that of those students who got a job, about 18 percent feel that their college studies are either irrelevant, or not quite pertinent to the kind of work they are going to do. This percentage is not too high, considering the difficult economic situation. But this observation can be qualified by answers to another question: only 56 percent of those who accepted a job offer, declared that they will like the job.

Another important purpose of this survey was to test the intensity of demand for graduates in different areas. This can be accomplished by relating the number of interviews, the number of job offers, and the salaries, to various majors. Certain kinds of graduates are in demand early before graduation, while others have to wait a year or more to find a job. Thus a survey of graduating seniors presents a unique opportunity to measure the intensity of demand for various majors and could become a major ingredient in forecasting such demand. The questionnaire used in this survey was designed with this purpose in mind, but very restricted budget did not permit to take a large enough sample in order to come up with valid conclusions.

Finally, some of the information on topics, discussed above, is being collected by many universities, but in most cases the information seems to be very spotty. A questionnaire, similar to the one used for this survey, should become a standard tool for the purpose of advising the students. And, as the above results indicate, considerable improvement in advising the students with respect to the choice of a major, a career, and methods of finding a job, is badly needed.

REFERENCES

The author would like to acknowledge financial assistance from KSU College of Business, without which this project would not be possible.

1. College Graduates and Jobs, A Report and Recommendations by the Carnegie Commission on Higher Education, New York: McGraw Hill, 1973, p.4.

2. Perrella, Vera C., "Employment of Recent College Graduates," Monthly Labor Review, Feb. 1973, p.44.

3. Frankel, Martin M. & Beamer, J. Fred, Projections of Educational Statistics to 1982-83, Washington, D.C.: U.S. Department of Health, Education, and Welfare, 1974.

4. Perrella, Vera C., "Employment of Recent College Graduates," op. cit.

5. College Graduates and Jobs, op. cit.

6. Sharp, Laure M., Education and Employment, Baltimore, Maryland: Johns Hopkins Press, 1970, p. 6.

7. See Perrella, Vera C., "Employment of Recent College Graduates," op. cit.

8. Davis, James A., Undergraduate Career Decisions, Chicago, Ill.: Aldine Publishing Co., 1965.

9. Calvert, Robert Jr., Career Patterns of Liberal Arts Graduates, Cranston, R.I.: Carroll Press, 1969.

10. Sharp, Laure M., Education and Employment, op. cit.

11. Perrella, Vera C., "Employment of Recent College Graduates," op. cit.

12. Calvert, Robert Jr., Career Patterns of Liberal Arts Graduates, op. cit., p. 28.

13. Sharp, Laure M., Education and Employment, op. cit., p. 7.

14. Davis, James A., Undergraduate Career Decisions, op. cit., pp. 4-77.

DISTRIBUTIONS OF RATIOS FOR SOME SELECTED BIVARIATE PROBABILITY FUNCTIONS

Paul W. Mielke, Jr., Colorado State University
John A. Flueck, Temple University

I. INTRODUCTION

Ratio estimators, R = Y/X of random variables (r.v.'s) X and Y, of population ratios have long been utilized in both theoretical and applied research. Examples from the theory of statistics include ratios of sample means in survey sampling, ratios of mean squares in experimental design, and likelihood ratios in testing. In the experimental areas, examples include crime indices in law enforcement, inheritance ratios in biology, effectiveness ratios in psychology, unemployment ratios in economics, stress ratios in engineering, precipitation ratios in meteorology, and survival and other ratios in demography.

However, despite this broad utilization, the distributional properties of the ratio R, generally have not been well developed. As Cochran (1963) has stated, "The distribution of the ratio estimate has proved annoyingly intractable because both X and Y vary from sample to sample. The known theoretical results fall short of what we would like to know for practical applications."

In this paper we will attempt to improve on this situation by reviewing some previously published literature on distributions of ratios and by presenting distributional and moment results for the ratio of two r.v.'s X and Y under selected bivariate structures (i.e., gamma, generalized gamma, beta of the first kind, log-normal, Weibull, beta-P, and beta-κ).

Greenwood and White (1910) initiated work on the distribution of R in their work on the opsonic index. They empirically generated sampling distributions of the opsonic index (ratio) under the assumption that X and Y are statistically independent. They noted that the resulting empirical distribution was skewed and had a mean >1.0. Pearson (1910), in a pioneering companion paper, presented Greenwood's initial theoretical effort along with the moments for R in the more general case where X and Y were correlated. However, he felt that "these approximate formulae would be practically unworkable if X and Y were correlated...." and hence concerned himself with an examination of the uncorrelated case. He concluded that (1) "if the distribution of both X and Y be symmetrical..., the distribution of indices must be skew" and (2) "the mean of the ratio of two numbers picked out of the same series is certainly greater than unity if the series be symmetrical, and will probably be always greater than unity even if it be not." Pearson went on to estimate the distribution of the opsonic index, for X and Y statistically independent, utilizing estimated moments to select and fit one of the Pearsonian type distributions. These estimated distributions were typically in close agreement with their empirical counterparts presented by Greenwood and White.

For the normal distribution, Merrill (1928) considered the ratio of two correlated normals using the method of moments and concluded "that for conditions ordinarily met in practice, the frequency distribution of the index, when both components follow closely the normal law, is sensibly normal." However, he also noted that "in cases where the correlation is high and the coefficients of variation are large, there may be a considerable deviation from normality." Craig (1929) presented a method of describing the distribution of R in terms of its semi-invariants (functions of the estimated moments of the joint distribution of X and Y with bounding restriction on X) and correctly indicated the persistent lack of normality of R and the condition needed for symmetry of its distribution. Geary (1930) followed with a "standardizing" of the ratio function (b+Y)/(a+X) [X and Y jointly distributed as $N(\mu_X,\mu_Y,\sigma_X,\sigma_Y,\rho_{X,Y})$] and showed it was approximately distributed N(0,1), provided a+X>0. He also presented the exact probability density function (pdf) of this ratio function for a=b=0, and (although not recognizing it as a member of the Cauchy family) commented on its nonnormality and infinite moments. Fieller (1932) presented and exact general solution to the distribution of the ratio of two jointly distributed normals with means μ_1 and μ_2, standard deviations σ_1 and σ_2, and correlation coefficient ρ. He also showed how Geary's and Merrill's results could be achieved under the assumption of a truncated normal distribution. Rao (1952) proved for a wide class of estimators (including R) that the standardized form was asymptotically distributed as N(0,1). Marsaglia (1965) presented the results of a study of the ratio of correlated normal r.v.'s utilizing computer drawn graphs of the pdf of R (producing symmetrical, nonsymmetrical unimodal, and bimodal pdf's). Press (1969) presented a similar study for correlated student's t r.v.'s. Rao and Garg (1969) presented the ratio of powers of the absolute value of two independent standardized normals and found the result to be generalized positive Cauchy distribution (a special case of the generalized beta of the second kind).

The ratio of two gamma r.v.'s was initiated by Kullback (1936) for the ratio of two independent gammas (i.e. a beta of the second kind). He also showed that Fisher's (1924) result for the distribution of the log of the ratio of two independent sample variances was a special case. Recently, Flueck, Holland, and Lee (1975) presented exact results for the density function of the ratio of two correlated gamma r.v.'s and illustrated some of the resulting properties of R with sixty-four computer drawn graphs.

The chi density function is closely related to the gamma, and Bose (1935) presented the density function for the ratio of two correlated chi r.v.'s. Krishnaiah, Hagis, and Steinberg (1963) presented the t^{th} moments of this ratio and extensive tables of its cumulative distribution function.

Rietz (1939) graphically presented pdf's for a special case of the ratio of two betas of the first kind (i.e., uniform r.v.'s given particular uncorrelated and correlated structure). Broadbent (1954) extended these results in presenting the ratio of both the uniform and triangular r.v.'s to an arbitrary independent positive r.v. Marsaglia (1965) presented results for the pdf of the ratio of sums of independent uniform r.v.'s and, even with relatively few terms, the ratio was quite closely approximated by the ratio of two independent normal r.v.'s.

The ratio of two independent log-normals, as well as the ratio of products, is given by Aitchison and Brown (1966). The resulting distributions are naturally log-normal.

Malik (1967) generalized Kullback's result for two independent gammas by deriving the density function for the ratio of two independent generalized gamma r.v.'s. Rao and Garg (1969) have indicated that the resulting pdf is a member of the family of generalized beta distributions of the second kind. Block and Rao (1973) subsequently generalized Malik's result by presenting the density and moments of the ratio of two independent distended gamma r.v.'s. The resulting density function was termed the distended beta.

In summary, the above brief history indicates that the distribution of the ratio of normal, student's t, gamma, chi, and a special case of the beta of the first kind r.v.'s has been presented for correlated X and Y.

II. SOME NEW RESULTS

The distribution of R for X and Y independent is conceptually attainable (e.g., Huntington, 1939). The situation when X and Y are correlated is more complex.

A. Gamma

Let X, Y and P be independent gamma r.v.'s with common scale parameter A and shape parameters $\alpha-\phi$, $\beta-\phi$ and ϕ, respectively $\{0\le\phi<\min(\alpha,\beta)\}$. As suggested by Cherian (1941) and David and Fix (1961), following Weldon, if U=X+P and V=Y+P, then the bivariate pdf of the r.v.'s U and V may be expressed as

$$f_{U,V}(u,v) = \frac{e^{-u-v}}{\Gamma(\alpha-\phi)\Gamma(\beta-\phi)\Gamma(\phi)} \cdot \int_0^{\min(u,v)} p^{\phi-1}(u-p)^{\alpha-\phi-1}(v-p)^{\beta-\phi-1}e^{p}dp, \quad (2.1)$$

where A is assumed equal to unity because it enters in the same form in both numerator and denominator of the ratio. It can be shown that equation (2.1) is equivalent to

$$f_{U,V}(u,v) = \begin{cases} \dfrac{u^{\alpha-1}v^{\beta-\phi-1}e^{-(u+v)}}{\Gamma(\alpha)\Gamma(\beta-\phi)} \cdot F_1^*(\phi,1+\phi-\beta,\alpha;\frac{u}{v},-u), & 0<u\le v, \\ \dfrac{u^{\alpha-\phi-1}v^{\beta-1}e^{-(u+v)}}{\Gamma(\alpha-\phi)\Gamma(\beta)} \cdot F_1^*(\phi,1+\phi-\alpha,\beta;\frac{v}{u},-v), & 0<v<u, \end{cases} \quad (2.2)$$

where

$$F_1^*(a,b,c;x,y)= \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \frac{(a)_{m+n}(b)_n}{(c)_{m+n}m!n!} x^m y^n, \quad |x|<1$$

is a "degenerate" two variable hypergeometric function (Gradshteyn and Ryzhik, 1965, p. 1067) and $(a)_b=\Gamma(a+b)/\Gamma(a)$. Thus, from the reproductive property of the gamma, U and V are gamma r.v.'s with marginal distribution shape parameters α and β, respectively, and dependence parameter ϕ. If $\phi=0$, then U and V are independent r.v.'s.

The nonnegative integral moments of U and V are given by

$$E(U^sV^t) = \sum_{j=0}^{s}\sum_{k=0}^{t} \binom{s}{j}\binom{t}{k}(\alpha-\phi)_{s-j}(\beta-\phi)_{t-k}(\phi)_{j+k}.$$

In particular, E(U)=Var(U)=α, E(V)=Var(V)=β, Cov(U,V)=ϕ, and $\rho(U,V)=\phi(\alpha\beta)^{-1/2}$. Let the ratio of the two gamma r.v.'s be defined as R=U/V. Then following the approach of Flueck and Holland (1974) and using a change of variable, the pdf of R is

$$f_R(r) = \begin{cases} \dfrac{r^{\alpha-1}(1+r)^{\phi-\alpha-\beta}}{B(\alpha,\beta-\phi)} F_1(\phi,\alpha+\beta-\phi,1+\phi-\beta,\alpha;\frac{r}{1+r},r), & 0<r\le 1, \\ \dfrac{r^{\alpha-\phi-1}(1+r)^{\phi-\alpha-\beta}}{B(\alpha-\phi,\beta)} \cdot F_1(\phi,\alpha+\beta-\phi,1+\phi-\alpha,\beta;\frac{1}{1+r},\frac{1}{r}), & 1<r, \end{cases} \quad (2.3)$$

where

$$F_1(a,b,c,d;x,y)= \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \frac{(a)_{m+n}(b)_m(c)_n}{(d)_{m+n}m!n!} x^my^n, \quad |x|<1,\ |y|<1$$

is a two variable hypergeometric function (Gradshteyn and Ryzhik, 1965, p. 1053) and

$$B(a,b)=\Gamma(a)\Gamma(b)/\Gamma(a+b).$$

A convenient computational form has been presented by Flueck, Holland and Lee (1975).

The integral moments of R are given by

$$E(R^s) = \begin{cases} \sum_{j=0}^{s} \binom{s}{j}\dfrac{(\alpha-\phi)_j(\phi)_{s-j}}{(\beta-j)_s}, & 0\le s, \\ \sum_{j=0}^{-s} \binom{-s}{j}\dfrac{(\beta-\phi)_j(\phi)_{-s-j}}{(\alpha-j)_{-s}}, & s<0. \end{cases}$$

In particular,

$$E(R) = \frac{\alpha\beta-\phi}{\beta(\beta-1)}, \quad 1<\beta, \quad \text{and}$$

$$Var(R)=\{\alpha\beta^2[(\alpha+\beta)\beta+\alpha-1]-2\beta[\alpha\beta(\beta+1)+(2\beta-1)(\beta-1)]\phi +[5\beta(\beta-1)+2]\phi^2\}/[(\beta+1)\beta^2(\beta-1)^2(\beta-2)], \quad 2<\beta.$$

If $(U_1,V_1),\ldots,(U_n,V_n)$ are n independent pairs from (2.2), then the analogous results for sums of gamma r.v.'s,

$$U^*=\sum_{i=1}^{n} U_i \text{ and } V^*=\sum_{i=1}^{n} V_i$$

can be shown to follow when the parameter set (α,β,ϕ) is replaced with $(n\alpha,n\beta,n\phi)$. As a consequence of these results and a general result by Rao (1952, p. 207), it follows that the asymptotic distribution of $R^*=U^*/V^*$ is normal with mean α/β and variance $\alpha(\alpha+\beta-2\phi)/n\beta$.

Using Stacy's (1962) generalization of the gamma, the previous results can be further generalized by constructing a bivariate pdf of generalized gamma r.v.'s given by

$$f_{W,Z}(w,z)=f_{U,V}(w^{\gamma},z^{\delta})\gamma\delta w^{\gamma-1}z^{\delta-1}$$

where $W=U^{1/\gamma}$, $Z=V^{1/\delta}$, $0 < \gamma$ and $0 < \delta$. If $\gamma=\delta$, then the pdf of the ratio r.v. $Q=W/Z$ is

$$f_Q(q)=f_R(q^{\gamma})\gamma q^{\gamma-1}.$$

B. Beta of the First Kind

Employing an approach similar to that of the previous section yields two distinctly different cases for a bivariate beta distribution of the first kind. The parameterizations used here follow Mielke (1975).

Case 1:

Let X, Y and P be independent beta r.v.'s with shape parameters $[p\gamma,(1-p)\gamma-\phi]$, $[q\gamma,(1-q)\gamma-\phi]$ and $(\gamma-\phi,\phi)$, respectively $\{0<p<1,\ 0<q<1,\ 0<\gamma,\ 0\le\phi<\min[(1-p)\gamma,(1-q)\gamma]\}$. If $U = XP$ and $V = YP$, then the bivariate pdf of U and V may be shown to be

$$f_{U,V}(u,v) = \begin{cases} \dfrac{u^{p\gamma-1}(1-u)^{(1-p)\gamma-\phi-1}v^{q\gamma-1}(1-v)^{(1-q)\gamma-1}}{B[p\gamma,(1-p)\gamma-\phi]B[q\gamma,(1-q)\gamma]} \\ \cdot F_1[\phi,1-(1-p)\gamma+\phi,\gamma-\phi-1,(1-q)\gamma;\frac{1-v}{1-u},1-v], \\ 0 < u \le v < 1, \\ \dfrac{u^{p\gamma-1}(1-u)^{(1-p)\gamma-1}v^{q\gamma-1}(1-v)^{(1-q)\gamma-\phi-1}}{B[p\gamma,(1-p)\gamma]B[q\gamma,(1-q)\gamma-\phi]} \\ \cdot F_1[\phi,1-(1-q)\gamma+\phi,\gamma-\phi-1,(1-p)\gamma;\frac{1-u}{1-v},1-u], \\ 0 < v < u < 1. \end{cases} \quad (2.4)$$

Here U and V are beta r.v.'s with marginal distribution shape parameters $[p\gamma,(1-p)\gamma]$ and $[q\gamma,(1-q)\gamma]$, respectively, and dependence parameter ϕ. If $\phi=0$, then U and V are independent r.v.'s.

The moments of U and V are given by

$$E(U^sV^t)= \frac{(p\gamma)_s(q\gamma)_t(\gamma-\phi+t)_s}{(\gamma)_{s+t}(\gamma-\phi)_s} = \frac{(p\gamma)_s(q\gamma)_t(\gamma-\phi+s)_t}{(\gamma)_{s+t}(\gamma-\phi)_t}.$$

In particular, $E(U)=p$, $\mathrm{Var}(U)=\frac{p(1-p)}{\gamma+1}$, $E(V)=q$, $\mathrm{Var}(V)=\frac{q(1-q)}{\gamma+1}$, $\mathrm{Cov}(U,V)= \frac{pq\phi}{(\gamma-\phi)(\gamma+1)}$, and $\rho(U,V)=\frac{\phi}{\gamma-\phi}\left[\frac{pq}{(1-p)(1-q)}\right]^{1/2}$. If $R=U/V$, the identity $U/V=X/Y$ conveniently yields the pdf of R given by

$$f_R(r) = \begin{cases} \dfrac{r^{p\gamma-1}B[(p+q)\gamma,(1-q)\gamma-\phi]}{B[p\gamma,(1-p)\gamma-\phi]B[q\gamma,(1-q)\gamma-\phi]} \\ \cdot F[1-(1-p)\gamma+\phi,(p+q)\gamma;(1+p)\gamma-\phi;r], \\ 0 < r \le 1, \\ \dfrac{r^{-q\gamma-1}B[(p+q)\gamma,(1-p)\gamma-\phi]}{B[p\gamma,(1-p)\gamma-\phi]B[q\gamma,(1-q)\gamma-\phi]} \\ \cdot F[1-(1-q)\gamma+\phi,(p+q)\gamma;(1+q)\gamma-\phi;\frac{1}{r}], \\ 1 < r. \end{cases} \quad (2.5)$$

where

$$F(a,b;c;x) = \sum_{n=0}^{\infty} \frac{(a)_n(b)_n}{(c)_n n!} x^n$$

is the Gaussian hypergeometric function of one variable (Gradshteyn and Ryzhik, 1965, p. 1039). The moments of R are given by

$E(R^s) = \dfrac{(p\gamma)_s(q\gamma)_{-s}}{(\gamma-\phi)_s(\gamma-\phi)_{-s}}$. In particular,

$E(R) = \dfrac{p\gamma(\gamma-\phi-1)}{(\gamma-\phi)(q\gamma-1)}$, $1<q\gamma$, and

$$\mathrm{Var}(R)=p\gamma(\gamma-\phi-1)[(\gamma-\phi)^2(\gamma-\phi+1)(q\gamma-2)(q\gamma-1)^2]^{-1} \cdot\{p\gamma(\gamma-\phi-1)[(1-q)\gamma-\phi]+[(1-p)\gamma-\phi](\gamma-\phi-2)(q\gamma-1)\},$$
$2 < q\gamma$.

Case 2:

Let X, Y and P be independent beta r.v.'s with paired shape parameters $[pq\gamma+\phi,(1-p)q\gamma-\phi]$, $[qp\gamma+\phi,(1-q)p\gamma-\phi]$ and $(pq\gamma,\phi)$, respectively $\{0<p<1,0<q<1,0<\gamma,0\le\phi<\min[(1-p)q\gamma,(1-q)p\gamma]\}$. If $U=XP$ and $V=YP$, then the bivariate pdf of U and V may be expressed as

$$f_{U,V}(u,v) = \begin{cases} \dfrac{u^{pq\gamma+\phi-1}(1-u)^{(1-p)q\gamma-\phi-1}v^{qp\gamma+\phi-1}(1-v)^{(1-q)p\gamma-1}}{B[pq\gamma+\phi,(1-p)q\gamma-\phi]B[qp\gamma+\phi,(1-q)p\gamma]} \\ \cdot F_1[\phi,1+\phi-(1-p)q\gamma,(p+q-pq)\gamma-1,(1-q)p\gamma;\frac{1-v}{1-u},1-v], \\ 0 < u \le v < 1, \\ \dfrac{u^{pq\gamma+\phi-1}(1-u)^{(1-p)q\gamma-1}v^{qp\gamma+\phi-1}(1-v)^{(1-q)p\gamma-\phi-1}}{B[pq\gamma+\phi,(1-p)q\gamma]B[qp\gamma+\phi,(1-q)p\gamma-\phi]} \\ \cdot F_1[\phi,1+\phi-(1-q)p\gamma,(p+q-pq)\gamma-1,(1-p)q\gamma;\frac{1-u}{1-v},1-u], \\ 0 < v < u < 1. \end{cases} \quad (2.6)$$

Now U and V are beta r.v.'s with marginal distribution shape parameters $[pq\gamma,(1-p)q\gamma]$ and $[qp\gamma,(1-q)p\gamma]$, respectively, and dependence parameter ϕ. Again, if $\phi = 0$, then U and V are independent r.v.'s.

The moments of U and V are given by

$$E(U^sV^t)= \frac{(pq\gamma)_{s+t}(pq\gamma+\phi)_s}{(q\gamma)_s(p\gamma)_t(pq\gamma+\phi+t)_s} = \frac{(pq\gamma)_{s+t}(pq\gamma+\phi)_t}{(q\gamma)_s(p\gamma)_t(pq\gamma+\phi+s)_t}.$$

In particular, $E(U) = p$, $\mathrm{Var}(U) = \frac{p(1-p)}{q\gamma+1}$, $E(V) = q$, $\mathrm{Var}(V) = \frac{q(1-q)}{p\gamma+1}$, $\mathrm{Cov}(U,V) = \frac{\phi}{\gamma(pq\gamma+\phi+1)}$, and $\rho(U,V) = \frac{\phi}{\gamma(pq\gamma+\phi+1)}\left[\frac{(p\gamma+1)(q\gamma+1)}{p(1-p)q(1-q)}\right]^{1/2}$.

If $R = U/V$, the identity $U/V = X/Y$ again yields the pdf of R given by

$$f_R(r) = \begin{cases} \dfrac{r^{pq\gamma+\phi-1}B[2pq\gamma+2\phi,(1-q)p\gamma-\phi]}{B[pq\gamma+\phi,(1-p)q\gamma-\phi]B[qp\gamma+\phi,(1-q)p\gamma-\phi]} \\ \cdot F[1+\phi-(1-p)q\gamma,2pq\gamma+2\phi;(1+q)p\gamma+\phi;r], \\ 0 < r \le 1, \\ \dfrac{r^{-pq\gamma-\phi-1}B[2pq\gamma+2\phi,(1-p)q\gamma-\phi]}{B[pq\gamma+\phi,(1-p)q\gamma-\phi]B[qp\gamma+\phi,(1-q)p\gamma-\phi]} \\ \cdot F[1+\phi-(1-q)p\gamma,2pq\gamma+2\phi;(1+p)q\gamma+\phi;\frac{1}{r}], \\ 1 < r. \end{cases} \quad (2.7)$$

The moments of R are given by

$E(R^s)=\dfrac{(pq\gamma+\phi)_s(pq\gamma+\phi)_{-s}}{(q\gamma)_s(p\gamma)_{-s}}$. In particular,

$E(R)=\frac{(pq\gamma+\phi)(p\gamma-1)}{q\gamma(pq\gamma+\phi-1)}$, $1<pq\gamma+\phi$, and

$Var(R)=(pq\gamma+\phi)(p\gamma-1)$

$\cdot\{q^2\gamma^2(q\gamma+1)(pq\gamma+\phi-2)(pq\gamma+\phi-1)^2\}^{-1}$

$\cdot\{(p\gamma-1)(pq\gamma+\phi-2)[(1-p)q\gamma-\phi]$

$+q\gamma(pq\gamma+\phi+1)[(1-q)p\gamma-\phi]\}$, $2<pq\gamma+\phi$.

C. Log-Normal

Let X and Y be dependent log-normal r.v.'s with marginal distribution shape parameters α and β, and scale parameters A and B, respectively. For brevity, let $U=X/A$ and $V=Y/B$. Then the bivariate pdf of U and V is

$$f_{U,V}(u,v) = \frac{1}{2\pi uv\alpha\beta(1-\phi^2)^{1/2}} \cdot e^{-\frac{1}{2(1-\phi^2)}\left[\left(\frac{\ell n u}{\alpha}\right)^2-2\phi\left(\frac{\ell n u}{\alpha}\right)\left(\frac{\ell n v}{\beta}\right)+\left(\frac{\ell n v}{\beta}\right)^2\right]},$$

$$0 < \min(u,v), \tag{2.8}$$

with dependence parameter $\phi\{0 < \min(\alpha,\beta), |\phi| < 1\}$. If $\phi=0$, then U and V are independent r.v.'s.

The moments of U and V are given by $E(U^sV^t)=e^{(s^2\alpha^2+t^2\beta^2+2\phi st\alpha\beta)/2}$. In particular, $E(U) = e^{\alpha^2/2}$, $Var(U)=e^{\alpha^2}(e^{\alpha^2}-1)$, $E(V)=e^{\beta^2/2}$, $Var(V)=e^{\beta^2}(e^{\beta^2}-1)$, $Cov(U,V)=e^{(\alpha^2+\beta^2)/2}(e^{\phi\alpha\beta}-1)$, and $\rho(U,V)=(e^{\phi\alpha\beta}-1)\{(e^{\alpha^2}-1)(e^{\beta^2}-1)\}^{-1/2}$. If $R=U/V$, then (2.8) yields the pdf of R given by

$$f_R(r)= \frac{1}{r\{2\pi(\alpha^2+\beta^2-2\phi\alpha\beta)\}^{1/2}} e^{-\frac{(\ell n r)^2}{2(\alpha^2+\beta^2-2\phi\alpha\beta)}},$$

$$0 < r. \tag{2.9}$$

The moments of R are given by $E(R^s)=e^{s^2(\alpha^2+\beta^2-2\phi\alpha\beta)/2}$. In particular, $E(R)=e^{(\alpha^2+\beta^2-2\phi\alpha\beta)/2}$, and $Var(R)=e^{\alpha^2+\beta^2-2\phi\alpha\beta}(e^{\alpha^2+\beta^2-2\phi\alpha\beta}-1)$.

If $(U_1,V_1),\ldots,(U_n,V_n)$ are n independent pairs from (2.8), then the analogous results for n^{th} roots of the products of log-normal r.v.'s

$$U^*=(\prod_{i=1}^n U_i)^{1/n} \text{ and } V^*=(\prod_{i=1}^n V_i)^{1/n}$$

follow when the parameter set (α,β,ϕ) is replaced with $(n^{-1/2}\alpha,n^{-1/2}\beta,\phi)$. A consequence of these results and a property of the log-normal distribution is that the asymptotic distribution of $R^*=U^*/V^*$ is normal with mean 1 and variance $(\alpha^2+\beta^2-2\phi\alpha\beta)/n$.

D. Weibull

The Weibull, beta-P, and beta-κ r.v.'s (the last two r.v.'s being restricted generalized beta distribution r.v.'s of the second kind, Mielke and Johnson, 1974), utilize a generalization of results presented by Farlie (1960).

Let X^* and Y^* be r.v.'s having closed form marginal cumulative distribution functions (CDF's). The scale parameters of X^* and Y^* will be eliminated by the simple transformations, $U=X^*/A$ and $V=Y^*/B$, in order to achieve brevity in exposition. Thus U and V are dependent r.v.'s having closed form marginal CDF's $x=F_U(u)$ and $y=F_V(v)$.

The joint CDF's of U and V can be expressed as $F_{U,V}(u,v)=xy\{1-g(x,y)\}$. Since $0\le F_{UV}(u,v)\le 1$, $g(x,y)\le 1\le g(x,y)+(xy)^{-1}$. The existence of $\frac{\partial g(x,y)}{\partial x}$, $\frac{\partial g(x,y)}{\partial y}$ and $\frac{\partial^2 g(x,y)}{\partial x\partial y}$ is assumed. Also, since $F_{U,V}(\infty,v)=F_V(v)$ and $F_{U,V}(u,\infty)=F_U(u)$, then $g(1,y)=g(x,1)=0$. The monotonicity of $F_{U,V}(u,v)$ implies $\frac{\partial}{\partial x}[xy\{1-g(x,y)\}]\ge 0$ and $\frac{\partial}{\partial y}[xy\{1-g(x,y)\}]\ge 0$, and $g(x,y)+x\frac{\partial g(x,y)}{\partial x} \le 1$ and $g(x,y)+y\frac{\partial g(x,y)}{\partial y} \le 1$.

The marginal CDF and pdf of a Weibull r.v. U are $F_U(u)=(1-e^{-u^\gamma})I_{(0,\infty)}(u)$ and $f_U(u)=\gamma u^{\gamma-1}e^{-u^\gamma}I_{(0,\infty)}(u)$, respectively, where $I_{(0,\infty)}(u)$ is an indicator function and $0<\gamma$. The marginal CDF and pdf of a second Weibull r.v., V, are identical except for replacement of the parameter γ by γ'. The $g(x,y)$ for this distribution is

$$g(x,y) = cx^m(1-x)^a y^n(1-y)^b \tag{2.10}$$

where m and n are nonnegative integers, $0<a$, $0<b$, and c satisfies the restrictions involving $g(x,y)$. Thus, a,b,c,m and n are dependence parameters. The bivariate CDF and pdf of Weibull r.v.'s U and V is

$$F_{U,V}(u,v) = xy-c\sum_{j=0}^{m+1}\sum_{k=0}^{n+1}\binom{m+1}{j}\binom{n+1}{k}(-1)^{j+k}(1-x)^{a+j}(1-y)^{b+k} \tag{2.11}$$

and

$$f_{U,V}(u,v) = \{1-c\sum_{j=0}^{m+1}\sum_{k=0}^{n+1}\binom{m+1}{j}\binom{n+1}{k}(a+j)(b+k) \cdot(-1)^{j+k}(1-x)^{a+j-1}(1-y)^{b+k-1}\}f_U(u)f_V(v), \tag{2.12}$$

respectively. The moments of r.v.'s U and V are given by $E(U^sV^t) = \Gamma[(s/\gamma)+1]\Gamma[(t/\gamma')+1]$

$$\cdot\{1-c\sum_{j=0}^{m+1}\sum_{k=0}^{n+1}\frac{\binom{m+1}{j}\binom{n+1}{k}(-1)^{j+k}}{(a+j)^{(s/\gamma)+1}(b+k)^{(t/\gamma')+1}}\}.$$

Let $R=U/V$. If U and V are bivariate Weibull r.v.'s with $\gamma=\gamma'$, then (2.12) yields the pdf of R given by

$$f_R(r)=\gamma r^{\gamma-1}[(r^\gamma+1)^{-2}-c\sum_{j=0}^{m+1}\sum_{k=0}^{n+1}\binom{m+1}{j}\binom{n+1}{k}(a+j)(b+k) \cdot(-1)^{j+k}\{(a+j)r^\gamma+(b+k)\}^{-2}],\ 0 < r. \tag{2.13}$$

The moments of R follow from the identity $E(R^s)=E(U^sV^{-s})$.

E. Beta-P

The marginal CDF and pdf of a beta-P r.v. U are $F_U(u)=\{1-(1+u^\theta)^{-\alpha}\}I_{(0,\infty)}(u)$ and $f_U(u)=\alpha\theta u^{\theta-1}(1+u^\theta)^{-(\alpha+1)}I_{(0,\infty)}(u)$, respectively, where $0<\alpha$ and $0<\theta$. The marginal CDF and pdf of a second beta-P r.v. V are identical except for replacement of the parameter set (α,θ) by (α',θ'). Using the $g(x,y)$ chosen for the Weibull (2.10), the bivariate CDF and pdf of beta-P r.v.'s U and V are again (2.11) and (2.12). The moments of r.v.'s U and V are given by

$$E(U^sV^t)=\alpha\alpha'[B(1+\frac{s}{\theta},\alpha-\frac{s}{\theta})B(1+\frac{t}{\theta'},\alpha'-\frac{t}{\theta'})$$
$$-c\sum_{j=0}^{m+1}\sum_{k=0}^{n+1}\binom{m+1}{j}\binom{n+1}{k}(a+j)(b+k)(-1)^{j+k}$$
$$\cdot B\{1+\frac{s}{\theta},\alpha(a+j)-\frac{s}{\theta}\}B\{1+\frac{t}{\theta'},\alpha'(b+k)-\frac{t}{\theta'}\}].$$

Again, let $R=U/V$. If U and V are bivariate beta-P r.v.'s with $\theta=\theta'$, then (2.12) again yields the pdf of R given by

$$f_R(r)=\alpha\alpha'\theta r^{\theta-1}[B(2,\alpha+\alpha')F(\alpha+1,2;\alpha+\alpha'+2;1-r^\theta)$$
$$-c\sum_{j=0}^{m+1}\sum_{k=0}^{n+1}\binom{m+1}{j}\binom{n+1}{k}(a+j)(b+k)(-1)^{j+k}B\{2,\alpha(a+j)$$
$$+\alpha'(b+k)\}F\{\alpha(a+j)+1,2;\alpha(a+j)+\alpha'(b+k)+2;1-r^\theta\}],$$
$$0 < r. \qquad (2.14)$$

Again the moments of R follow from the identity $E(R^s)=E(U^sV^{-s})$.

F. Beta-κ

The marginal CDF and pdf of a beta-κ r.v. U are $F_U(u)=\{1-(1+u^\theta)^{-1}\}^\alpha I_{(0,\infty)}(u)$ and $f_U(u)=\alpha\theta u^{\alpha\theta-1}(1+u^\theta)^{-(\alpha+1)}I_{(0,\infty)}(u)$, respectively, where $0<\alpha$ and $0<\theta$. The marginal CDF and pdf of a second beta-κ r.v., V, are identical except for replacement of the parameter set (α,θ) by (α',θ'). A convenient choice of $g(x,y)$ for this distribution is $g(x,y)=cx^a(1-x)^my^b(1-y)^n$ where m and n are positive integers, $0\le a$, $0\le b$, and c satisfies the restrictions involving $g(x,y)$. Again, a,b, c,m and n are dependence parameters. The bivariate CDF and pdf of beta-κ r.v.'s U and V are

$$F_{U,V}(u,v) = xy$$
$$-c\sum_{j=0}^{m}\sum_{k=0}^{n}\binom{m}{j}\binom{n}{k}(-1)^{j+k}x^{a+j+1}y^{b+k+1} \qquad (2.15)$$

and

$$f_{U,V}(u,v) = \{1-c\sum_{j=0}^{m}\sum_{k=0}^{n}\binom{m}{j}\binom{n}{k}$$
$$(a+j+1(b+k+1)(-1)^{j+k}x^{a+j}y^{b+k}\}f_U(u)f_V(v), \quad (2.16)$$

respectively. The moments of r.v.'s U and V are given by

$$E(U^sV^t) = \alpha\alpha'[(\alpha+\frac{s}{\theta},1-\frac{s}{\theta})B(\alpha'+\frac{t}{\theta'},1-\frac{t}{\theta'})$$
$$-c\sum_{j=0}^{m}\sum_{k=0}^{n}\binom{m}{j}\binom{n}{k}(a+j+1)(b+k+1)(-1)^{j+k}$$
$$\cdot B\{\alpha(a+j+1)+\frac{s}{\theta},1-\frac{s}{\theta}\}B\{\alpha'(b+k+1)+\frac{t}{\theta'},1-\frac{t}{\theta'}\}].$$

Let $R=U/V$. If U and V are bivariate beta-κ r.v.'s with $\theta=\theta'$, then (2.16) yields the pdf of R given by

$$f_R(r)=\alpha\alpha'\theta r^{\alpha\theta-1}[B(\alpha+\alpha',2)F(\alpha+1,\alpha+\alpha';\alpha+\alpha'+2;1-r^\theta)$$
$$-c\sum_{j=0}^{m}\sum_{k=0}^{n}\binom{m}{j}\binom{n}{k}(a+j+1)(b+k+1)(-1)^{j+k}r^{\alpha\theta(a+j)}$$
$$\cdot B\{\alpha(a+j+1)+\alpha'(b+k+1),2\}F\{\alpha(a+j+1)+1,\alpha(a+j+1)$$
$$+\alpha'(b+k+1);\alpha(a+j+1)+\alpha'(b+k+1)+2;1-r^\theta\}],$$
$$0 < r, \qquad (2.17)$$

and the moments of R follow from the identity $E(R^s) = E(U^sV^{-s})$.

III. CONCLUDING COMMENTS

This paper has reviewed a number of known results and presented some new results for the distributions of ratios of r.v.'s. The new results have (i) given extensions and simplifications of previous work involving the ratio of correlated gamma r.v.'s, (ii) extended Malik's work on the ratio of independent generalized gamma r.v.'s to the correlated case (iii) extended Reitz's and Marsaglia's results to the general case of the ratio of correlated beta r.v.'s of the first kind, (iv) extended Aitchison and Brown's work on the ratio of independent log-normal r.v.'s to the correlated case, and (v) presented additional results for ratios of dependent Weibull, beta-P, and beta-κ r.v.'s. In conjunction with this work, a number of new selected bivariate distributions are also presented.

In applications of specific ratio distributions [e.g., gamma (see Flueck and Holland, 1976) and log-normal (see Aitchison and Brown, 1966)], standard estimators will be appropriate. For the other ratio distributions, additional estimation procedures should be advantageous. Also, in some applications the use of bivariate distributions permitting different scale parameters (e.g., log-normal, Weibull, beta-P, and beta-κ) should allow closer modeling of the data.

REFERENCES

Aitchison, J. and Brown, J. A. C. (1966). The Lognormal Distribution. Cambridge: Cambridge University Press.

Block, H. W. and Rao, B. R. (1973). A beta warning-time distribution and a distended beta distribution. Sankhyā B 34, 79-84.

Bose, S. (1935). On the distribution of the ratio of variances of two samples drawn from a given normal bivariate correlated population. Sankhyā 2, 65-72.

Broadbent, S. R. (1954). The quotient of a rectangular or triangular and a general variate. Biometrika 41, 330-7.

Cherian, K. C. (1941). A bivariate correlated gamma-type distribution function. J. Indian Math. Soc. 5, 133-44.

Cochran, W. G. (1963). Sampling Techniques, 2nd edition. New York: Wiley.

Craig, C. C. (1929). The frequency function of y/x. Ann. of Math. 30, 471-86.

David, F. N. and Fix, E. (1961). Rank correlation and regression in a nonnormal surface. Proc. 4th Berkeley Symp. 1, 177-97.

Farlie, D. S. G. (1960). The performance of some correlation coefficients for a general bivariate distribution. Biometrika 47, 307-23.

Fieller, E. C. (1932). The distribution of the index in a normal bivariate population. Biometrika 24, 428-40.

Fisher, R. A. (1924). On a distribution yielding the error functions of several well known statistics. Proceedings of the International Mathematical Congress, Toronto, 2, 805-13.

Flueck, J. A. and Holland, B. (1974). The moments and distribution of the ratio of two gamma variables. 1973 Proceedings, Social Statistics Section, American Statistical Association, 252-5.

Flueck, J. A., Holland, B. (1976). Ratio estimators and some inherent problems in their utilization. J. Appl. Meteor., 15, 535-43.

Flueck, J. A., Holland, B., and Lee, R. (1975). Distribution of the ratio of correlated sums of gamma variables. 1974 Proceedings, Social Statistics Section, American Statistical Association, 285-91.

Geary, R. C. (1930). The frequency distribution of the quotient of two normal variates. J. R. Statist. Soc. 93, 442-6.

Gradshteyn, I. S. and Ryzhik, I. M. (1965). Table of Integrals, Series, and Products. New York: Academic Press.

Greenwood, M. and White, J. D. C. (1910). A biometric study of phagocytosis with special reference to the opsonic index. Biometrika 7, 505-30.

Huntington, E. V. (1939). Frequency distributions of products and quotients. Ann. Math. Statist. 10, 195-8.

Krishnaiah, P. R., Hagis, P. and Steinberg, L. (1963). A note on the bivariate chi distribution. SIAM Rev. 5, 140-4.

Kullback, S. (1936). The distribution laws of the differences and quotient of variables independently distributed in Pearson type III laws. Ann. Math. Statist. 7, 51-3.

Malik, H. J. (1967). Exact distribution of the quotient of independent generalized gamma variables. Canad. Math. Bull. 10, 463-5.

Marsaglia, G. (1965). Ratio of normal variables and ratios of sums of uniform variables. J. Amer. Statist. Assoc. 60, 193-204.

Merrill, A. S. (1928). Frequency distribution of an index when both the components follow the normal law. Biometrika 20A, 53-63.

Mielke, P. W. (1975). Conveient beta distribution likelihood techniques for describing and comparing meteorological data. J. Appl. Meteor. 14, 985-90.

Mielke, P. W. and Johnson, E. S. (1974). Some generalized beta distributions of the second kind having desirable application features in hydrology and meteorology. Water Resour. Res. 10, 223-6. Corrections 11, 606; 12, 827.

Pearson, K. (1910). On the constants of index distributions as deduced from the like constants for the components of the ratio, with special reference to the opsonic index. Biometrika 7, 531-41.

Press, S. J. (1969). The t-ratio distribution, J. Amer. Statist. Assoc. 64, 242-252.

Rao, C. R. (1952). Advanced Statistical Methods in Biometric Research. New York: Wiley.

Rao, B. R. and Garg, M. L. (1969). A note on the generalized (positive) Cauchy distribution. Canad. Math. Bull. 12, 865-8.

Rietz, H. L. (1936). On the frequency distribution of certain ratios. Ann. Math. Statist. 7, 145-53.

Stacy, E. W. (1962). A generalization of the gamma distribution. Ann. Math. Statist. 33, 1187-92.

A FEW EXERCISES WITH THE LIFE TABLE FUNCTIONS

S. Mitra, Emory University

Abstract

It has been observed that the average age of a stationary population is equal to the average number of years that remains to be lived by that population. This age or this number of years has been shown as approximately equal to the age at which the life expectancy is also equal to that age in number of years. Interestingly enough, this age can be obtained as the x coordinate of the point of intersection of the curves of the functions $x\ell_x$ and T_x. Also, the tangent to the curve of T_x, at this intersection, crosses the x axis at a point which corresponds to twice the average age of the stationary population. A biproduct of this investigation is an interesting boundary condition for the ratio of the expectation of life at birth to the average age. This ratio has been found to lie between $2\ell_{\bar{x}}$ and $2(\ell_{\bar{x}})^{\frac{1}{2}}$, where $\ell_{\bar{x}}$ is the proportion surviving from birth to the average age $\bar{x}$.

1. Definitions and Properties of Life Table Functions

The principal life table functions are conveniently defined as

(a) ℓ_x, the proportion of survivors from birth to age x, such that $\ell_0 = 1$;

(b) $\mu_x = -(1/\ell_x)\ (d\ell_x/dx)$, the force of mortality at age x;

(c) $T_x = \int_x^\alpha \ell_a da$, where α is the upper age limit at which $\ell_x = 0$;

(d) $\overset{0}{e}_x = T_x/\ell_x$ is the expectation of life at age x.

The following properties of the functions may be noted:

(a') $d\ell_x/dx$ is uniformly negative, that is to say, ℓ_x is a monotonically declining function. Further, the curve of ℓ_x has a point of inflection, i.e., $d^2\ell_x/dx^2$ equals zero for some $x = x'$. This results from the lower risk of death usually in the age interval 10-15, compared to other ages (Mitra, 1973).

(b') The force of mortality assumes its lowest value, that is, $d\mu_x/dx = 0$ for some $x = \tilde{x}$, also in the age interval 10-15 (Mitra, ibid.).

(c') $dT_x/dx = -\ell_x$.

(d') $\overset{0}{e}_x$ assumes its maximum value, that is, $d\overset{0}{e}_x/dx = 0$ for some $x = \hat{x}$, usually in the age interval 0-5.

From (d), $T_x = \ell_x \overset{0}{e}_x$, so that

$$dT_x/d_x = -\ell_x\ (-1/\ell x)(d\ell x/dx)\overset{0}{e}_x + \ell_x d\overset{0}{e}_x/dx$$

$$= -\ell_x(\overset{0}{e}_x\mu_x - d\overset{0}{e}_x/dx) \quad (1)$$

Equating (1) with (c') and solving

$$d\overset{0}{e}_x/dx = \overset{0}{e}_x\mu_x - 1 \quad (2)$$

The maximum life expectancy corresponds to age $\hat{x}$, where,

$$\overset{0}{e}_{\hat{x}} = 1/\mu_{\hat{x}} \quad (3)$$

It was shown earlier (Mitra, ibid.) that $\hat{x} < \tilde{x} < x'$.

2. Average Age of a Stationary Population

Interpreting the life table functions $\ell_x dx$ and T_x as the population sizes in the age intervals x to x+dx and ages x and above respectively in the corresponding stationary population, the average age can be defined as (Mitra, 1965)

$$m = \int_0^\alpha x\ell_x dx \Big/ \int_0^\alpha \ell_x dx = \int_0^\alpha x\ell_x dx/T_0 \quad (4)$$

Integrating by parts, the integral

$$\int_0^\alpha x\ell_x dx = -\int_0^\alpha x dT_x = xT_x]_0^\alpha + \int_0^\alpha T_x dx = \int_0^\alpha T_x dx \quad (5)$$

as $T_\alpha = 0$. Since $T_x = \ell_x\overset{0}{e}_x$, (5) can be rewritten as

$$\int_0^\alpha x\ell_x dx = \int_0^\alpha T_x dx = \int_0^\alpha \ell_x\overset{0}{e}_x dx \quad (6)$$

Accordingly, (4) can be expressed in two different ways, namely,

$$m = \int_0^\alpha x\ell_x dx/T_0 = \int_0^\alpha \ell_x\overset{0}{e}_x dx/T_0 \quad (7)$$

The last expression in (7) can be interpreted as the average number of years that remains to be lived in the stationary population. Thus, the total population T_0, that on an average, is m years old, can expect to live for an additional m number of years, again from the perspective of an arithmetic average.

3. Expectation of Life at Age m

Although, the average age of the population is m and so is the average life expectancy, it does not follow that the life expectancy at the exact age m is also m, although, intuitively, that seems to suggest itself. It can, however, be shown that, under certain simplifying assumptions, the equality of the two is, at least, a reasonable approximation. The assumptions are the same as those that are made to derive the expected value of a function of a variable, namely, that, the function has a Taylor's expansion and further, the higher moments or higher derivatives or both of the variable concerned, are relatively small. In other words, assuming Taylor's expansion of a function f(x) at x = m = E(x), and taking expected values,

$$E[f(x)] = f(m) + E \sum_{r=2}^{\infty} \frac{(x-m)^r}{r!} f^r(x)_{x=m} \tag{8}$$

since E(x-m) = 0. Further, when second and higher order moments or the second and higher order derivatives or both are small,

$$E[f(x)] = f(m) \tag{9}$$

It is well known that $\overset{0}{e}_x$ is a continuous and a smooth function of x, and hence the assumption of Taylor's expansion can be held. Also the derivative of $\overset{0}{e}_x$, decreases smoothly from a value of 0 at maximum life expectancy to a value, no smaller than -1 at the last age (see equation 2), because $\overset{0}{e}_x \mu_x$ is nonnegative in the entire age range. Accordingly, the second derivative is quite small (of the order of $1/\alpha$) and the higher order derivatives are even smaller. Thus, the second term of (8), namely, $\frac{V(x)}{2!} f''(x)_{x=m}$ is expected to be small for f(x) = $\overset{0}{e}_x$ and therefore, can be neglected for all practical purposes. The higher order terms will be even smaller and therefore,

$$E(\overset{0}{e}_x) = m \approx (\overset{0}{e}_x)_{x=m} = \overset{0}{e}_m \tag{10}$$

where ≈ stands for approximate equality. Alternatively, the average age can be approximately determined as the x or the y coordinate of the point of intersection of the line y = x and the curve y = $\overset{0}{e}_x$ (at which $\bar{x} = \overset{0}{e}_{\bar{x}}$). These parameters obtained from appropriate life tables (Keyfitz and Flieger, 1971) are presented in Table 1 for a few countries, covering a wide range of variation of patterns of mortality.

It may be noted that m is uniformly greater than $\bar{x}$ by about half a year for all mortality levels. Part of this bias must have resulted from the computational procedure based on the approximate equality

$$\int_a^{a+5} T_x dx \approx \frac{5}{2}(T_a + T_{a+5}) \tag{11}$$

in which, the right hand side uniformly over-estimates the integral because the curve of T_x is concave upwards. In any event, the empirical demonstration of the near equality of $\bar{x}$ and m, verifies the theoretical formulation arrived at in expression (10).

4. The Curves of T_x and $x\ell_x$

The definite integrals of the functions T_x and $x\ell_x$ over the entire age range were earlier shown to be equal (see expression 6), however, the curves that they generate are quite different and have some interesting features.

The derivative of T_x is $-\ell_x$ which is uniformly negative, assumes the value of -1 at age 0, slowly decreases thereafter and becomes equal to 0 at the highest age α. Simultaneously, the function T_x declines to zero from a maximum value of T_0.

The function $x\ell_x$, on the other hand, assumes its lowest value of zero at both extremes of the age interval. Its derivative

$$\frac{d}{dx}(x\ell_x) = \ell_x(1-x\mu_x) \tag{12}$$

indicates a maximum value of the function at some x = x'' where

$$x'' = 1/\mu_{x''} \tag{13}$$

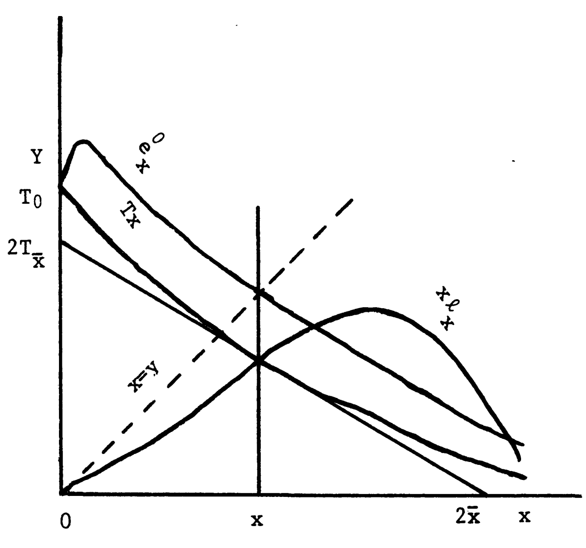


Figure 1. The Curves of $x\ell_x$ and T_x

at which the function $x\ell_x$ assumes its maximum value of $\ell_{x''}/\mu_{x''}$. Because of the nature of the functions ℓ_x and μ_x, the function $x\ell_x$ has only one maximum and thus the curve of the function is bell shaped.

The general nature of the curves of T_x and $x\ell_x$ may be studied from Figure 1.

a. The Point of Intersection

The point of intersection between $x\ell_x$ and T_x has an interesting feature. At this point,

$$T_x = x\ell_x \text{ or } x = \overset{0}{e}_x \tag{14}$$

and as was shown earlier, this happens at $x = \bar{x}$ where $\bar{x}$ is a close approximation of the average age m of the stationary population.

It can be shown that in most cases, $x < x''$ where the latter age corresponds to the maximum value of $x\ell_x$. This is so because, in the neighborhood of age $\bar{x}$, the life expectancy is usually a declining function of age so that, its derivative (see expression 2) $\overset{0}{e}_x\mu_x-1$ is negative at $x = \bar{x}$. Since $\bar{x} = \overset{0}{e}_x$,

$$\overset{0}{e}_{\bar{x}}\mu_{\bar{x}}-1 = \bar{x}\mu_{\bar{x}}-1 < 0 \tag{15}$$

so that (12) is positive at $x = \bar{x}$ and therefore, $\bar{x} < x''$.

b. The Tangent to T_x

The tangent to T_x meets the x axis at a point, the x coordinate of which is (because the derivative of T_x is $-\ell_x$),

$$x + T_x/\ell_x = x + \overset{0}{e}_x \tag{16}$$

Therefore, the tangent at $T_{\bar{x}}$ is $\bar{x} + \overset{0}{e}_{\bar{x}} = 2\bar{x}$ which must intersect the y axis at $2T_{\bar{x}}$. Since, the curve of $T_{\bar{x}}$ is concave upwards.

$$2T_{\bar{x}} < T_0 \tag{17}$$

(col. 6, Table 1), which means that more than half of the population is less than $\bar{x}$ years old. In other words, the median age is less than the average age, an expected result in view of the shape of T_x.

Incidentally, the area of the right triangle with sides T_0 and $2\bar{x}$ is also a close approximation of the area of the curve bounded by T_x as well as that by $x\ell_x$ over the age interval, $(0,\alpha)$. Also, the area of the right triangle with the tangent at (x, T_x) as the hypotenuse and, therefore, sides of lengths $x + \overset{0}{e}_x$ and $\ell_x(x+\overset{0}{e}_x)$ is given by

$$A(x) = \frac{(x\ell_x + T_x)(x + \overset{0}{e}_x)}{2} \tag{18}$$

At $x = 0$,

$$A(0) = \frac{1}{2}T_0^2 \tag{19}$$

Now,

$$\frac{dA(x)}{dx} = \frac{1}{2}\left[x\frac{d\ell_x}{dx}(x+\overset{0}{e}_x) + (x\ell_x+T_x)\overset{0}{e}_x\mu_x\right] \tag{20}$$

$$= \frac{1}{2}\left[-x\ell_x\mu_x(x+\overset{0}{e}_x) + (x\ell_x+T_x)\overset{0}{e}_x\mu_x\right] \tag{21}$$

$$= \frac{1}{2}\ell_x\mu_x(\overset{0}{e}{}_x^2-x^2) \tag{22}$$

Thus, the derivative is positive, as long as $\overset{0}{e}_x > \bar{x}$, vanishes at $x = \bar{x}$ and becomes negative thereafter. The maximum value of the area is therefore,

$$A(\bar{x}) = 2\bar{x}T_{\bar{x}} < \bar{x}T_0 \tag{23}$$

because of (17). Combining (19) and (23), the inequality

$$T_0^2 < 4\bar{x}T_{\bar{x}} = 4\bar{x}^2\ell_{\bar{x}} \tag{24}$$

is obtained, which can be rewritten as

$$\frac{T_0}{\bar{x}} < 2\sqrt{\ell_{\bar{x}}} \tag{25}$$

Noting from (17) that

$$2\ell_{\bar{x}}\bar{x} < T_0 \tag{26}$$

(25) can be combined with (26), to produce

$$2\ell_{\bar{x}} < \frac{T_0}{\bar{x}} < 2\sqrt{\ell_{\bar{x}}} \tag{27}$$

5. Average Age and Expectation of Life at Birth

Since $T_0 = \overset{0}{e}_0$ is the expectation of life at birth, (27) establishes an interesting inequality relationship between $\overset{0}{e}_0$ and the average age of the stationary population $\bar{x}$.

It may be mentioned in this context that usually, $\overset{0}{e}_0 > \bar{x}$ which, because of (27), requires that $\ell_{\bar{x}} > \frac{1}{2}$ (Table 1, cols 2, 4 and 5). However, it is also seen that the difference between $\overset{0}{e}_0$ and $\bar{x}$ decreases with $\overset{0}{e}_0$, and as could also be expected from the inequality relationship, $\ell_{\bar{x}}$ and $\overset{0}{e}_0$ are positively associated as well. It is therefore quite possible for expectations of life, lower than that shown in Table 1, to be less than $\bar{x}$. In fact, for the UN model life table (1956), level 0, the average age was found to be 23.5 years corresponding to the life expectancy of only 19.8 years and an

$\ell_{\bar{x}}$, as low as .37. Although, there are very few countries in the world today with a life expectancy that low, $\overset{0}{e}_0 > \bar{x}$ is at most, an empirical reality whereas, the actual mathematical relationship between the two can, at present, be expressed in the form of a boundary condition shown in (27). Further investigation is needed to find out the possibility or otherwise of improving this inequality relationship.

References

Keyfitz, N. & W. Flieger, 1971, Population: Facts and Methods of Demography, W.H. Freeman & Co.

Mitra, S., 1965, "Total Length of Life and Its Relation with the Parameters of Stable Population," Sankhyā, Vol. 27, Series B, pp. 189-192.

Mitra, S., 1973, "Optimum Values of Life Table Functions," Proceedings of the American Statistical Association, Social Statistics Section, pp. 356-358.

United Nations, 1956, Methods for Population Projections by Sex and Age, ST/SOA/ Series A, Population Studies, No. 25.

TABLE 1. Average Ages of Stationary Populations (m), Life Expectancies $\overset{0}{e}_0$, $\bar{x} = \overset{0}{e}_{\bar{x}}$, $\ell_{\bar{x}}$ and $T_{\bar{x}}$ for a Few Selected Countries (males only).

Country	$\overset{0}{e}_0 = T_0$	m	$\bar{x} = \overset{0}{e}_{\bar{x}}$	$\ell_{\bar{x}}$	$T_{\bar{x}}$
(1)	(2)	(3)	(4)	(5)	(6)
Cameron, 1964	34.2	28.0	27.6	.57	15.7
Indonesia, 1961 (incl. West Irian)	44.1	32.1	31.7	.66	20.9
S. Africa, 1961 (colored)	49.8	33.1	32.6	.74	24.1
Phillipines, 1960	55.4	35.7	35.2	.76	26.8
Chile, 1967	59.2	35.0	34.5	.84	29.0
Japan, 1966	68.5	36.6	36.0	.94	33.8
Denmark, 1967	70.7	37.4	36.8	.95	35.0

ASKING SINGLE WOMEN ABOUT THEIR CHILDREN: THE CENSUS BUREAU'S EXPERIENCE

Maurice J. Moore, U.S. Bureau of the Census

Census Surveys

The Bureau of the Census made its first inquiry into the childbearing experience of single (never-married) women in the decennial census of 1970 [1]. In both the 15-percent and the 5-percent samples all women 14 years old and over (with exceptions noted below) were subjects for the question "How many babies has she ever had, not counting stillbirths." In previous censuses where a similar question was asked, the question applied only to women who had ever been married.

The second venture, by way of a "pretest," was the June 1975 marital and fertility history supplement to the Current Population Survey (CPS) [2]. Single women 18 to 75 years old were asked about children already born, if the women were in that portion of the sample that was not to be interviewed again as part of the CPS. The experimental subsample included 832 women, about one-eighth of the entire sample of single women.

Finally, in June 1976, the question concerning the number of children born to date was extended to <u>all</u> single women 18 to 59 years old who fell into the Current Population Survey [3]. Detailed tabulations from the 1976 survey are expected to be published early in 1977.

In the 1970 census, about 60 percent of the population was enumerated by a mail-out/mail-back procedure. Both 100 percent forms and sample forms were distributed by mail in advance of census day [4]. The forms were to be filled out by a responsible member of the household and returned by mail to a data collection center. In these self-enumeration situations, if all sample items were fully reported, the number of children ever born would be reported for all single women 14 years old and over in the household.

Forty percent of the population was enumerated in 1970 by the conventional procedure of direct interviewing. Forms containing the 100 percent items were delivered by mail, with instructions to fill them out and wait for a census enumerator who would call to collect them. Enumerators also brought with them sample schedules to be filled out in selected households at the time of the visit [5]. In these direct interview situations, the enumerator's instructions stated that the question about children ever born was <u>not to be asked regarding a single woman unless</u> that woman had one or more children living in the household with her [6].

Response rates for single women in the 1970 census are shown in tables 1 and 2. About 70 percent of all single women had a report for number of children ever born, with somewhat lower rates of response occurring among women 30 years old or over and especially among those 45 years old or over. Table 1 indicates that the overall rate of response was about the same among Blacks and Whites: 71.8 percent for Blacks and 70.4 percent for Whites. Table 2 shows that, despite interviewer instructions to the contrary (which were stated in an unemphasized manner in the middle of a big instruction

Table 1.—Response Rates for Single Women to Questions on Children Ever Born: 1970 Census

Age of woman	All races		White		Black	
	Single women	Percent reporting	Single women	Percent reporting	Single women	Percent reporting
Total	15,604,425	70.6	13,328,860	70.4	2,051,428	71.8
15 to 17 years	5,553,582	70.4	4,753,084	70.7	729,471	68.7
18 and 19 years	2,804,666	74.0	2,410,507	74.1	354,934	74.1
20 and 21 years	1,763,105	74.7	1,504,522	74.5	228,016	76.4
22 to 24 years	1,267,771	73.7	1,053,724	73.1	188,379	77.3
25 to 29 years	827,906	71.5	646,466	70.5	161,636	76.0
30 to 34 years	435,897	68.3	338,370	67.3	88,101	73.0
35 to 44 years	672,255	67.1	546,154	66.1	114,850	71.8
45 years and over	2,279,243	62.9	2,076,033	62.9	186,041	63.9

Source: U.S. Bureau of the Census, <u>1970 Census of Population</u>, Subject Report PC(2)-3A, "Women by Number of Children Ever Born," table A-2.

Table 2.—Percent of Single Women Reporting on Number of Children Ever Born, by Type of Enumeration Area and Whether Own Children Are Present: 1970 Census

(Based on 5-percent sample)

Presence of children and residence	Mail out/mail back enumeration areas		Conventional enumeration areas	
	White	Black	White	Black
NO OWN CHILDREN PRESENT				
Number of women	8,613,333	1,174,607	4,563,274	591,464
Percent reporting	71.1	68.5	68.8	68.8
WITH OWN CHILDREN PRESENT				
Number of women	98,209	202,203	54,044	83,154
Percent reporting	83.4	91.7	79.0	91.9

Source: Unpublished records from the 1970 census

book), responses were obtained from a very large percentage of single women in the conventional enumeration areas. In fact, response rates were about the same both in mail-back and in conventional areas.

Table 3 presents rates of NONresponse (to emphasize differences) to the question on children ever born from the experimental portion of the June 1975 CPS and from the entire sample of single women in the June 1976 CPS. In the CPS, unlike the 1970 census, all data are collected in a direct interview situation, although the interview may take place by telephone and information about childbearing to date may be reported by the woman in question or by some knowledgeable person in the household. Rates of nonresponse were unexpectedly low at 9.9 percent overall in 1975 and 7.7 percent in 1976. These rates compare quite favorably with the corresponding nonresponse rates for ever-married women in the same surveys, 7.8 and 6.8 respectively. No consistent patterns of nonresponse for 1975 and 1976 clearly emerge when the data are classified by age, education, and region of residence. Overall, in both surveys, responses about previous childbearing were slightly harder to obtain for White single women than for Black single women, although the 1976 difference is marginal.

There is also available some evidence of a nonnumerical character regarding public tolerance for a question on past fertility of single women. During the 1975 survey, official observers of field interviews were asked to be alert to reactions to the material in the June supplement. While some of the interviewers and other Census Bureau field personnel seemed to be nervous about the prospect of asking single women questions about their past fertility, no observer reported any resistance from the persons interviewed. Moreover, to the best of the author's knowledge, neither the Census Bureau nor any other govenmental body has received any complaints from the public about the fact that single women were included in the survey concerning children born to date.

<u>Reliability of Fertility Data for Single Women</u>

At present there are very few ways to check on the overall reliability of census data on the fertility of single women. The Census Bureau has no experience prior to 1970 on which to base a comparison. A content reinterview study after the 1970 census did follow up on fertility reports for ever-married women, but no attempt was made to do the same for single women. Moreover, the only sure guarantee of a content reinterview is consistency, not accuracy. Other national fertility surveys have been limited to married women. Data on <u>illegitimacy</u> published in Vital Statistics are not strictly comparable to census data on <u>premarital fertility</u>. Hence, the census data permit only limited validation. Some attempts at this follow.

Table 4 shows mean numbers of children ever born to single women, as tabulated from the 1970 census. Rates for White women are substantially lower than those for Black women. Averages increase with age until a peak is reached for women in their thirties (who would have been in their highly fertile years during the "baby boom"); then the rates taper off with increasing age. The substantial difference by race is supported by information available from birth registrations [7]. The variation in fertility patterns by age parallels that among ever-married women.

Table 5 compares the results of reconstructing annual numbers of births from the 1970 census with births reported in Vital Statistics. Since the registration of births is estimated to be over 99 percent complete for the years shown in table 5 [8], the number of total births

Table 3. Nonresponse Rates for Single Women to the Item on Children Ever Born in the June 1975 and June 1976 Current Population Surveys

Subject	All races: Number of women (thousands)	All races: Percent not reporting[1]	White: Number of women (thousands)	White: Percent not reporting[1]	Black: Number of women (thousands)	Black: Percent not reporting[1]
June 1975[2]						
Total	1,348	9.9	1,115	10.5	214	7.5
Age of woman						
18 and 19 years	393	11.7	331	11.5	56	(B)
20 to 24 years	478	6.9	377	7.4	88	6.8
25 to 29 years	135	11.9	112	14.3	24	(B)
30 years and over	340	10.9	294	11.9	47	(B)
Years of school completed						
0 to 11 years	303	9.6	223	9.4	75	10.7
12 years	525	10.3	429	11.7	93	4.3
13 to 15 years	339	9.4	302	9.9	29	(B)
16 years or more	180	10.0	160	9.4	17	(B)
Region						
Northeast	400	12.5	338	13.0	62	(B)
North Central	355	9.6	316	10.1	39	(B)
South	368	10.1	264	11.0	99	8.1
West	224	5.4	196	6.1	14	(B)
June 1976						
Total	10,694	7.7	8,646	7.9	1,824	7.0
Age of woman						
18 and 19 years	3,286	8.6	2,742	8.2	494	11.1
20 to 24 years	4,186	7.0	3,385	7.5	702	5.0
25 to 29 years	1,351	6.3	1,030	6.5	280	5.3
30 to 59 years	1,871	8.8	1,489	9.3	348	6.9
Years of school completed						
0 to 11 years	2,015	7.7	1,345	9.3	635	4.2
12 years	4,409	8.1	3,555	8.1	781	8.6
13 to 15 years	2,783	7.3	2,403	7.2	306	8.7
16 years or more	1,486	7.3	1,344	7.5	102	7.7
Region						
Northeast	2,883	8.6	2,461	8.5	387	10.4
North Central	2,830	8.2	2,415	8.2	389	7.3
South	3,012	5.7	2,091	6.3	905	4.3
West	1,969	8.9	1,679	8.6	143	14.4

B Base less than 75,000.

1/ The term nonresponse rates shown in this table are based on single women in households where interviews were conducted. Interviews were not obtained from 4.7 percent of the occupied sample households in the June 1976 survey. Of the occupied sample households in which interviews were not obtained, slightly over half of the noninterviews (2.5 percent of all occupied households) were due to refusals to cooperate with any part of the Current Population Survey.

2/ Data limited to experimental portion of the sample.

Source: Forthcoming Bureau of the Census reports on the June 1975 and June 1976 Current Population Surveys.

recorded by Vital Statistics may be taken as the benchmark against which to judge other estimates.

Regarding total births to women of all races, including both legitimate and illegitimate births, the 1970 census consistently exceeds the Vital Statistics estimates for the years 1965 to 1969 by up to 3 percent. This suggests some small bias in the Census Bureau's reconstruction of the fertility histories of women. The number of births based on the reconstructed fertility histories of the 1970 census depends on the number of women counted in 1970. Since mortality will have made some small inroads on the number of women who were alive in each of the years from 1965 to 1969, one would expect census estimates to fall short of estimates based on vital registrations. Sampling variability, both in the census estimates and in those from Vital Statistics, could account for some of the relatively small differences in numbers of total births, but it is unlikely to account for the fact that the census numbers consistently differ from vital registration numbers in the direction opposite from that expected.

The number of first births estimated by the census for 1969 exceeds that from Vital Statistics by about 3 percent; otherwise census numbers fall short of Vital Statistics by up to 3 percent. Differences between the 1970 Census and Vital Statistics are somewhat more pronounced when racial differences are taken into account.

When the comparison between the 1970 Census and Vital Statistics concerns "illegitimate" first births, the basis for comparison is less clear. The only type of illegitimacy that can be inferred from census data is that which occurs prior to the first marriage of the child's mother, whereas Vital Statistics data include all reported instances of births out-of-wedlock, regardless of the mother's marital status at the time. In addition, census data include reports from women in all States of the union, but not all States report to the Office of Vital Statistics on the legitimacy status of births. In 1965, 16 States did not make such reports; by 1969, the number of nonreporting States had been reduced to 11. However, data reported by Vital Statistics do include estimates of the

Table 4.—Children Ever Born Per 1,000 Single Women: 1970 Census

Subject	All races	White	Black
15 to 17 years old:			
Women	5,553,582	4,753,084	729,471
Children per 1,000 women ...	23	10	105
18 and 19 years old:			
Women	2,804,666	2,410,507	354,934
Children per 1,000 women ...	67	32	302
20 and 21 years old:			
Women	1,763,105	1,504,522	228,016
Children per 1,000 women ...	117	53	531
22 to 24 years old:			
Women	1,267,771	1,053,724	188,379
Children per 1,000 women ...	217	110	813
25 to 29 years old:			
Women	827,906	646,466	161,636
Children per 1,000 women ...	428	208	1,306
30 to 34 years old:			
Women	435,897	338,370	88,101
Children per 1,000 women ...	595	263	1,871
35 to 39 years old:			
Women	337,144	267,297	63,510
Children per 1,000 women ...	593	271	1,939
40 to 44 years old:			
Women	335,111	278,857	51,340
Children per 1,000 women ...	460	236	1,659
45 to 49 years old:			
Women	334,549	289,137	40,708
Children per 1,000 women ...	352	203	1,379
50 years old and over:			
Women	1,944,694	1,786,896	145,333
Children per 1,000 women ...	224	157	1,008

Source: U.S. Bureau of the Census, 1970 Census of Population, Subject Report PC(2)-3A, table 65.

number of illegitimate births which occur in the nonreporting States [9]. If the Vital Statistics estimates can be accepted as a reasonably accurate estimate of illegitimate first births in the United States [10], then the 1970 census method of reconstructing fertility histories overestimates the number of illegitimate first births occurring in 1969 by about 13 percent and underestimates them by about the same amount for 1967 and 1968. The census overestimate for 1969 is all the more suspect, given the fact that census data do not include illegitimate first births occurring to women after the date of their first marriage. Thus, it appears that the 1970 census provided the basis for a reasonably accurate reconstruction of total births for each of the years from 1965 to 1969, but was less effective in terms of estimating the numbers of premarital births. To what extent deficiencies are due to the relatively greater sampling variability among the smaller numbers associated with illegitimacy, to deliberate misreporting of pertinent information on the census schedule, and to the Census Bureau's procedures for imputing values for missing information is something which, in practical terms, is undeterminable.

In another attempt to assess the validity of data collected by the Census Bureau, I have compared in table 6 the distributions and mean numbers of children ever born from the 1970 census with those from the June 1976 CPS. Among the percentages childless for the several age/race groupings, some differences are statistically significant; on the other hand, the percentages childless bear a marked similarity for the two different surveys. Even for average numbers of children ever born (shown in the last column), the similarities between the two surveys are quite strong, with the exception of the women 25 to 29 years old. Although there is no necessity for population paramaters of this sort to remain the same over a period of 6 years, such statistics generally do change slowly and the inner consistency of the census data provides some basis for confidence in the quality of the data.

Summary

In the course of this paper I have reported on the experience of the Census Bureau in an area of inquiry that is still in its infancy, at least as regards national level fertility

Table 5.—1970 Census and Vital Statistics Estimates of Annual Total Births, First Births, and Illegitimate First Births, 1965 to 1969.

Race and year	Total births (thousands)		First births (thousands)		Illegitimate first births (thousands)	
	1970 Census	Vital Statistics	1970 Census	Vital Statistics	1970 Census 1/	Vital Statistics
All Races						
1969	3,700	3,600	1,392	1,353	253	223
1968	3,540	3,502	1,280	1,311	191	213
1967	3,526	3,521	1,223	1,228	166	189
1966	3,606	3,606	1,192	1,224	147	NA
1965	3,848	3,760	1,154	1,160	141	NA
White						
1969	3,066	2,994	1,147	1,144	123	115
1968	2,943	2,912	1,064	1,104	90	113
1967	2,941	2,923	1,025	1,037	79	99
1966	3,017	2,993	1,012	1,043	77	NA
1965	3,194	3,124	978	988	73	NA
Black						
1969	567	543	217	199	120	104 2/
1968	535	531	191	187	94	100 2/
1967	526	544	177	174	83	90 2/
1966	528	558	160	165	66	NA
1965	590	581	159	155	64	NA

1/ Restricted to first births calculated as occurring before first marriage.

2/ Data refer to races "other than White."

NA Not available

Source: U.S. Bureau of the Census, 1970 Census of Population, Subject Report PC(2)-3B, "Childspacing and Current Fertility," tables 66 and 68; National Center for Health Statistics, Vital Statistics of the United States, Vol. I, for the stated years.

surveys. I have tried to support the position that inquiring into the previous childbearing experience of single women will not result in unacceptably high rates of nonresponse, at least when the questioning is limited to numbers and birthdates of children. Nevertheless, there are some good reasons for suspecting that the quality of Census data on premarital fertility is not quite as good as that on fertility after first marriage [11]. As more experience is gained in conducting fertility surveys among single women, we should come into a position to make better evaluations of the quality of the data.

Table 6.—Children Ever Born to Single Women 18 to 29 Years Old as Reported in the 1970 Census and the June 1976 Current Population Survey

Age and race	1970 Census					June 1976 Current Population Survey				
	Percent by number of children ever born				Children ever born per 1,000 women	Percent by number of children ever born				Children ever born per 1,000 women
	Total	0	1	2 or more		Total	0	1	2 or more	
18 To 24 Years Old										
All races	100.0	92.6	7.4		115	100.0	91.0	6.3	2.7	126
White	100.0	96.6	2.4	1.0	58	100.0	96.2	3.0	0.8	48
Black	100.0	68.8	19.5	11.7	494	100.0	64.7	23.1	12.2	518
ALL RACES										
18 and 19 yrs. old...	100.0	95.0	5.0		67	100.0	94.2	4.8	1.0	69
20 and 21 yrs. old...	100.0	91.9	5.7	2.4	117	100.0	91.0	6.0	3.0	129
22 to 24 yrs. old....	100.0	88.1	7.1	4.8	217	100.0	85.6	9.3	5.2	220
25 to 29 yrs. old....	100.0	81.1	8.7	10.2	428	100.0	80.9	9.8	9.3	358

Source: U.S. Bureau of the Census, 1970 Census of Population, Subject Report PC(2)-3A, table 65; forthcoming Bureau of the Census report on the June 1976 Current Population Survey.

FOOTNOTES

1. See U.S. Bureau of the Census, 1970 Census of Population, Subject Reports PC(2)-3A, "Women by number of Children Ever Born," and PC(2)-3B, "Childspacing and Current Fertility."

2. U.S. Bureau of the Census, Current Population Reports, Series P-20, "Fertility of American Women: June 1975" (forthcoming).

3. U.S. Bureau of the Census, Current Population Reports, Series P-20, "Prospects for American Fertility: June 1976" (forthcoming).

4. U.S. Bureau of the Census, 1970 Census of Population and Housing, PHC(R)-1, "Procedural History," pp. 1-6.

5. Ibid., pp. 1-7.

6. U.S. Bureau of the Census, "Questionnaire Reference Book, Census '70," p. 41.

7. National Center for Health Statistics, Vital Statistics of the United States 1971, Vol. 1-Natality, table 1-30.

8. Ibid., table 1-19. For a discussion of the completeness of birth registrations, see ibid., p. 4-13.

9. Ibid., pp. 4-10 and 4-11. For a discussion of the problems associated with the NCHS estimates of illegitimate births for nonreporting States, see Beth Berkov and June Sklar, "Metholological Options in Measuring Illegitimacy and the Difference They Make," Social Biology, Vol. 22, No. 4 (1975), pp. 356-71.

10. For comments regarding the quality of illegitimacy statistics derived from birth registrations, ibid,; National Center for Health Statistics, Vital and Health Statistics, Series 21, No. 15, "Trends in Illegitimacy: United States-1940-1965," p. 1-2; and Phillips Cutright, "Illegitimacy in the United States: 1920-1968," in Commission on Population Growth and the American Future, Demographic and Social Aspects of Population Growth, Charles F. Westoff and Robert Parke, Jr., eds., Vol. 1 of Commission Research Reports (Washington: U.S. Government Printing Office, 1972), pp. 429-38.

11. For a discussion of the quality of census fertility data collected from ever-married women in the June 1965 Current Population Survey, see Monroe G. Sirken and George Sabagh, "Evaluation of Birth Statistics Derived Retrospectively From Fertility Histories Reported in a National Population Survey: United States, 1945-64," Demography, Vol. 5, No. 1 (1968), pp. 485-503.

ON THE CHOICE OF A RANDOMIZATION TECHNIQUE WITH THE RANDOMIZED RESPONSE MODEL

Mark Moriarty, University of Iowa
Frederick Wiseman, Northeastern University

The randomized response model has been suggested as an appropriate data collection method when there is a belief that respondents may be unwilling to identify themselves as a member of some stigmatizing group, A. The basic model was first proposed by Warner [5] in 1965 and since that time a number of variations of the model have been proposed.

In Warner's original version, a respondent was presented with two statements:

S_1: I am in A

S_2: I am not in A

and asked to respond to either S_1 or S_2 depending upon a randomization whose outcome was known only to the interviewee. The two statements, S_1 and S_2, would be responded to with known probabilities P and 1 - P, respectively. An estimate of Π_A, the unknown proportion in the population in A is estimated by:

$$\hat{\Pi}_A = \frac{\lambda + P - 1}{2P - 1}$$

with an estimated variance of

$$v(\hat{\Pi}_A) = \frac{\lambda(1 - \lambda)}{nP^2}$$

where λ is the percentage of "yes" responses in a random sample of size n.

Simmons [4] proposed an alternative model which has been labelled the "Unrelated Question Method." Here, the respondent, after completing a randomization responds to either

S_1: I am in A

S_2: I am in Y

where membership in Y is assumed to be totally innocent and non-stigmatizing. Also, it is generally assumed that Π_y, the proportion in the population in Y, is known.

This paper will only consider the Simmons' model, but the principles to be outlined should hold for all randomization schemes. In the Simmons' model

$$\hat{\Pi}_A = \frac{\lambda - \Pi_Y + P\Pi_Y}{P}$$

and

$$v(\hat{\Pi}_A) = \frac{\lambda(1 - \lambda)}{nP^2} .$$

In the design of the model, the researcher has the liberty to choose the values of P and Π_Y.

Lanke [2] has concluded, based on the findings of Greenberg et al. [1], Moors [3], and his own research, that the following selection guidelines as to the value of P and Π_Y should be used: "choose P as large as you dare and choose Π_Y as small as you dare." However, Lanke states that while it is true that P should be chosen as large as possible, in practical terms the following must hold:

$$\Pr\ (A|\text{Yes response}) \leq \theta$$

else the respondent may either give false answers or fail to cooperate.

In the present case in which Π_Y is known:

$$\Pr(A|\text{Yes response}) = \frac{P\Pi_A}{P\Pi_A + (1 - P)\Pi_Y}$$

and based on Lanke's reasoning, if this value is low, the likelihood of cooperation and truthful reporting is likely to be high. However, for fixed Π_A and Π_Y, the researcher desires that the quantity be large because as P increases, Pr (A|Yes response) increases. That is, if we let P increase by X then it can be shown that:

$$\frac{P\Pi_A}{P\Pi_A + (1 - P)\Pi_Y} < \frac{(P + X)\Pi_A}{(P + X)\Pi_A + (1 - P - X)\Pi_Y}$$

for all relevant X. So the resulting conflict is that the researcher wants P to be large because of improved precision, but the larger the value of P, the less likely truthful reporting will result.

A Possible Solution

A possible solution to the conflict previously posed results when one considers that there are two relevant P values that must be considered. These are:

P = true probability that the respondent will be required to answer the sensitive question

P^* = respondent's perceived probability that he will be required to answer the sensitive question.

Thus, for the respondent, the relevant quantity and decision rule becomes:

if $\Pr^*(A|\text{Yes response}) = \dfrac{P^*\Pi_A}{P^*\Pi_A + (1 - P^*)\Pi_Y} \leq \theta$, cooperate and answer truthfully;

if $\Pr^*(A|\text{Yes response}) > \theta$, either fail to participate or participate and answer untruthfully if a "yes" response is required as a result of the randomization. Now if a randomizing device can be found for which:

$$P^* < P$$

a possible resolution can be found. That is, a good randomizing device will have as one of its properties that the true probability that the respondent will be required to answer the sensitive question is considerably larger than the perceived probability that the sensitive question will have to be responded to.

Empirical Results

In order to search for a randomizing device with the above characteristic, four devices were tested on a sample of n = 100 undergraduate business students at two major universities who had no prior instruction in probability. These devices were:

(1) roll two dice and add the totals
(2) pick one poker chip from a box containing one white, one red and ten blue chips
(3) spin a spinner
(4) pick three cards with replacement from a full deck.

For the respective devices, respondents were asked to assess the likelihood that:

(1) the sum of the dice will be between four and ten
(2) a blue chip will be picked
(3) the dial will land in the shaded area
(4) the cards selected will not all be the same color.

The results obtained can be found in Table 1 in the form of perceived frequency distributions. For each randomization procedure, there was a substantial amount of variation.

From an examination of Table 1, it appears that the first device--rolling two dice and adding the totals--has all the desirable properties of a good randomization device:

(1) it is easy to understand as only knowledge of simple addition is required;
(2) it produces good randomization;
(3) its P value is large, .83, but most respondents had $P^* \leq .80$ with median value of .70. Further, for only 9 percent of the sample was the perceived probability larger than the actual probability.

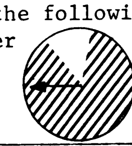
The second device, picking a poker chip from a box, had relatively little variability in the perceived probabilities with slightly more than two-thirds of the respondents assessing the true probability as being between .81 and .90.

Confusion existed for the third randomization procedure where the probability of landing in the shaded area was .875. A relatively large percentage, 29 percent, indicated that they thought the probability was actually less than .50. This phenomena is rather difficult to explain. However, one possible explanation is that some respondents misread the instructions and estimated, instead, the probability that the dial would land in the non-shaded area. This hypothesis, if valid, would cause one to eliminate such a device as a possible randomization procedure.

The fourth procedure, picking three cards from a full deck, produced a large amount of variability. The major drawback that it had was that approximately 33 percent of the respondents had $P^* > P$.

Table 1

Randomizing Devices Tested

Device 1	Device 2	Device 3	Device 4
Roll 2 dice and add the totals	Pick 1 chip from a box containing 1 white chip, 10 blue chips, and 1 red chip	Spin the following spinner 	Pick 3 cards with replacement from a deck of 52
Question: What is the probability that the total is between 4 and 10?	Question: What is the probability that you will pick a blue chip?	Question: What is the probability that the dial will land in the shaded area?	Question: What is the probability that the 3 cards will not be all red or all black?

Perceived probability	f_1	f_2	f_3	f_4
No idea	10	3	7	18
.01- .50	10	6	29	32
.51- .60	10	0	3	5
.61- .70	24	0	6	5
.71- .80	22	10	8	13
.81- .90	17	68	39	21
.91-1.00	7	13	8	6
TOTAL	100	100	100	100
Theoretical Probability	.833	.833	.875	.750

Example:

Suppose in a situation we hypothesize that $\Pi_A \doteq \frac{1}{10}$ and $\Pi_Y = \frac{1}{6}$ and we wish to estimate the proportion of students who have used LSD within the last year. Using the data from the previous study we would ask each respondent to "Flip two dice and add the totals," and then proceed according to the following instructions:

- If the sum is either 4, 5, 6, 7, 8, 9, 10 respond to "I have used LSD in the last year."
- If the sum is either 2, 3, 11, 12 respond to "My birthday is in either January or April."

For this randomization

$$\Pr(A|\text{Yes response}) = \frac{\frac{5}{6}\left(\frac{1}{10}\right)}{\left(\frac{5}{6}\right)\left(\frac{1}{10}\right) + \frac{1}{6}\left(\frac{1}{6}\right)} = .75.$$

However, the typical respondent does not use the true probability of answering the sensitive question, 5/6, but rather instead uses a value of .7 (see Table 1) and computes

$$\Pr^*(A|\text{Yes response}) = \frac{\frac{7}{10}\left(\frac{1}{10}\right)}{\frac{7}{10}\left(\frac{1}{10}\right) + \frac{3}{10}\left(\frac{1}{6}\right)} = .58.$$

As a consequence, based on Lanke's reasoning, it is likely that now greater cooperation and more truthful reporting will result.

Conclusion

It is obvious that randomization procedures exist for which $P^* < P$. This paper has shown the importance of using a device in which the perceived probability of answering the sensitive question is less than the actual probability. Based upon a limited sample of n = 100 students, the results suggest that one promising randomization device is one in which two dice are thrown and the relevant quantity is the sum of the two outcomes. Further investigation and extensions are clearly warranted in this area.

REFERENCES

1. Greenberg, Bernard G., et al., "The Unrelated Question Randomized Response Model: Theoretical Framework," Journal of the American Statistical Association, 64, (June, 1969), 520-39.

2. Lanke, Jan, "On the Choice of the Unrelated Question in Simmons' Version of Randomized Response," Journal of the American Statistical Association, 70, (March, 1975), 80-3.

3. Moors, J. J. A., "Optimization of the Unrelated Question Randomized Response Model," Journal of the American Statistical Association, 66, (September, 1971), 627-9.

4. Simmons, Walt R., "Response to Randomized Inquiries--A Technique for Reducing Bias," presented to the American Society for Quality Control, Washington, D.C., 1970.

5. Warner, Stanley L., "Randomized Response: A Survey Technique for Elimination Evasive Answer Bias," Journal of the American Statistical Association, 60, (March, 1965), 63-9.

SEXUAL EXPERIENCE, CONTRACEPTIVE USAGE, AND SOURCE OF CONTRACEPTION FOR NEVER-MARRIED WOMEN: ALBANY (NY) HEALTH REGION

Leo Morris, Judy Conn, George Stroh, and Melita Gesche
Center for Disease Control and New York State Health Department

INTRODUCTION

Twenty-three percent of never-married women aged 15-19 and 71% aged 20-44 in the Albany New York Health Region (AHR) were estimated, as of 1974, to have had some coital experience. This estimate is based on data obtained through interviews of an area-probability sample of the 15-to-44-year-old female population living in the 18-county AHR. The survey area was divided into 2 segments--the Albany-Schenectady-Troy Standard Metropolitan Statistical Area (SMSA)--which includes 4 counties and slightly over 50% of the population in the AHR; the remaining 14 counties of the health region were classified as non-SMSA. This division is essentially an urban-rural split of the health region. The survey included 2,291 respondents, of which interviews were completed for 2,059 (90%), both married and unmarried. In the SMSA, 87.4% of women eligible for interview were interviewed, and in the non-SMSA 94.0%. Refusals accounted for only 4.3% of the incomplete interviews in the SMSA and 2.8% of the incomplete interviews in the non-SMSA. The survey was conducted by the Bureau of Family Planning of the New York State Health Department with consultation from the Family Planning Evaluation Division of the Center for Disease Control, Atlanta, Georgia, in response to a need for information on correlates of planned and unplanned fertility, family planning practices, and the sexual experience and contraceptive practice of unmarried women, an important program target group. This paper deals with the 497 never-married women included in the survey.

The interest in data for never-married women, especially teenagers, is due to the increasing rate of out-of-wedlock births in recent years and a concurrent decline in marital fertility (1). In 1974, 13% of all births in the United States were out of wedlock; 36% of births to women aged 19 or younger were out of wedlock (2). In the mid-1960s, for women 15-19 years of age, 42% of legitimate births were estimated to be premaritally conceived (3). A study done in Massachusetts in the late 1960s shows this figure to be as high as 60% for that state (4). In 1973, 68% of all legal abortions were performed on unmarried women (5). Since 1972, with the total fertility rate below replacement reproduction, it has become apparent that an increasing proportion of unplanned fertility in the United States may be attributable to out-of-wedlock births (6).

The National Fertility Studies, which were conducted in 1955, 1960, 1965, and 1970, only surveyed women who had been or were then married (ever-married). The only comparable data for never-married women were obtained in 1971 through a national probability sample of the 15-to-19 year-old female population in the United States (7,8). For never-married women over 19, no survey data are available on sexual experience and contraceptive practice with the exception of studies involving special groups, such as college women.

METHODOLOGY

The basic survey design was a modification of the stratified, 2-stage, cluster sampling technique described by Serfling and Sherman and standard statistical formulas, as described by Kish, which take into consideration the design effect of a cluster sample, were utilized in considering the survey design and sample selection (9,10). Briefly, the Serfling-Sherman design calls for the ranking of each census tract or minor civil division in a survey area according to a set of socioeconomic criteria, then grouping the tracts into socioeconomic strata. Criteria for the stratification had been developed by the Division of Epidemiology and Preventive Health Services of the New York State Health Department (11). The New York stratification scheme is not self-weighting, so there was a deliberate use of different sampling rates for each stratum. The lower socioeconomic strata were oversampled as an independent sample was selected for each stratum. Thus, throughout this paper, results presented for any individual socioeconomic stratum are independent sample results, but all findings presented over socioeconomic strata, for the SMSA or non-SMSA or the total Albany Health Region, are based on a weighting scheme that takes into account the deliberate use of different sampling rates in each stratum.

RESULTS

As seen in Table 1, our investigation indicates that in the health region studied 23% of never-married women 15-19 years of age and 71% of never-married women 20-44 years of age have experienced coitus. The proportion of never-married women who had experienced intercourse did not differ between the SMSA and non-SMSA areas. For SMSA teenagers, there appears to be an inverse relationship between sexual experience and socioeconomic area. In the non-SMSA, no such relationship is seen; for never-married women over 19 no such relationship is evident for either area. Among 15-to-19 year-olds in the lower socioeconomic area of the SMSA, 46% of blacks were sexually experienced as compared to 29% of whites. However, there was no difference in rates of sexual experience between black and white never-married women over 19.

For the AHR, the likelihood that a young never-married woman will have experienced coitus rises from 16% for the 15-17 year-old age group to 49% for the 18-19 year-old age group. The sexual experience rate is somewhat higher for non-SMSA women than for SMSA women; for non-SMSA women a 3 to 1 differential between lower and upper socioeconomic strata was reported (35% versus 12%).

Table 2 compares the AHR results for 15-19 year-olds with the results of the 1971 U.S. survey. Sexual experience in the AHR in 1974 was slightly less than that seen in the United States in 1971. However, in the lower socioeconomic area of the

SMSA, sexual experience was higher in the AHR overall and for white teenagers, but lower for black teenagers.

Another variable used in the analysis of the National Survey as well as in the AHR survey was religion. Analysis of religious differences and sexual experience patterns is handicapped by small numbers in some categories, but data are available for upper and lower socioeconomic strata in both the SMSA and non-SMSA. As seen in Table 3, the differential in the rates of sexual experience between Protestant and Catholic teenagers is of approximately the same magnitude as that found in the United States in 1971, but is at a slightly lower level for both groups. Catholics tend to be below average and Protestants somewhat above in the proportion sexually experienced.

Of the sexually experienced never-married women 15-44 years of age, 50% were currently practicing contraception, 20% had at some time, and 30% had never done so. The group most likely to be using contraception and least likely to have never practiced contraception is the 20-24 year-old age group. This is essentially true for all socioeconomic areas in both the SMSA and non-SMSA. A greater proportion of non-SMSA women were currently practicing contraception than SMSA women, and this difference would be even greater if we consider that 73% of non-SMSA women who are not currently practicing contraception are not currently sexually active compared with only 55% of SMSA women.

Seventy-five percent of never-married women who are currently or who have ever practiced contraception had used oral contraceptives as the most recent method. For women 20-44 years-old the most recently used method of contraception was oral contraceptives for 86%. For teenagers, the most recently used method of contraception was oral contraceptives for 63% and condoms for 29%. Condoms were used to a greater extent by SMSA women than non-SMSA; whereas other methods such as foam, diaphragm, rhythm, douche, jelly, and withdrawal, were relied on to a greater extent by non-SMSA women. In the upper socioeconomic strata of the SMSA, teenage never-married women relied on oral contraceptives and condoms equally; but in the lower strata much greater reliance was placed on oral contraceptives. Twice as many women in the AHR reported that they had never practiced contraception, as in the 1971 U.S. survey (Table 4). Some of this difference appears to be related to the fact that withdrawal was very seldom reported as a method used in the AHR, compared with over 10% of the women so reporting in the U.S. survey. Further communication with those responsible for the U.S. survey revealed that they had used the term "withdrawal or pulling out" as a method of contraception; whereas, in the AHR we had used the term "withdrawal." Otherwise, the distribution of methods most recently used by SMSA women is consistent with that in the central cities of the United States. However, for non-SMSA women in the AHR, a much greater proportion of teenagers were relying on oral contraceptives than teenagers in the United States. Part of this difference may be explained by when the AHR survey was done, which was 3 years after the U.S. survey, a period in which family planning programs were extending into rural areas.

For all medical methods, 60% of women who had ever practiced contraception utilized private physicians as their source of contraceptives, and 31% attended planned parenthood clinics. Drugstores were the principal source of contraception for non-medical methods. Whereas 6.6% of never-married SMSA women who had ever used contraceptives reported that public clinics were their source of contraceptives, this overall proportion may be misleading without a control for socioeconomic status. Only women in the lower socioeconomic strata indicated public clinics as their source. In the lower socioeconomic stratum, 38% of women received contraceptives from public clinics, 30% from planned parenthood clinics, and 27% from private physicians.

In the AHR 22% of first births were either out-of-wedlock or premaritally conceived. Within the SMSA, the proportion of first pregnancies that were a result of post-marital conception decreases from 80% in the upper socioeconomic stratum to 66% in the upper middle and lower middle socioeconomic strata, to only 45% in the lower socioeconomic stratum. Fully 20% of ever-married women in the lower socioeconomic strata are estimated to have had a premaritally conceived legitimate birth, and an additional 24% have had an out-of-wedlock birth prior to marriage. These data are currently being analyzed by marriage cohort and age at first marriage.

If we focus our attention on out-of-wedlock births only, the planning status for all out-of-wedlock births in the period 1969-1974 is shown in Table 5. Planned births are those that were both desired and did not occur before they were planned, mistimed are those that occurred before planned but are still desired, and unwanted are "number failures." This terminology conforms to that of the published analyses of the 1965 and 1970 National Fertility Studies (12,13). Nineteen percent of the out-of-wedlock births in the time period covered in the AHR survey were reported as planned; in 1971 in the United States, 20% of premarital first pregnancies in women who are unmarried at the time of pregnancy outcome were reported as planned(14). There is essentially no difference by ethnic group in the reporting of planned out-of-wedlock births, but some differences in planning status are shown when comparing the lower socioeconomic stratum to all other socioeconomic stratas.

CONCLUSION

The data reported here give an overview of sexual experience, contraceptive usage, and source of contraception for never-married women in the AHR. With this data, it should be possible to estimate the need for effective contraceptive services for unmarried women, whatever their economic status, who are sexually experienced but do not have access to or are not currently using effective means of fertility control and who do not desire to be pregnant (18). The data on unplanned pregnancies and out-of-wedlock births should indicate areas of need for family planning services and the data on contraceptive usage and source of contraception provide information that is useful in assessing the provision of these services in both the public and private sector. Of special interest is the fact that 27% of SMSA women in the lower socioeconomic strata who had ever practiced contraception utilized private physicians as their source of contraception, and 66% of never-married women in other socioeconomic strata utilized private physicians. In general, the sexual experience of never-married 15 to 19 year-olds in the AHR, when compared by area of residence, religion, and other variables, is slightly lower than the experience reported in 1971 for never-married 15 to 19 year-olds in the United States. Although some recent studies have shown--and others hypothesized--that since 1971 sexual experience has increased for teenagers in the United States, the AHR appears to be an area where the rate of sexual experience in 1974 was only equal to or somewhat below the experience of similar women in the United States 3 years previously (15-17).

REFERENCES

1. Hinman AR, Stroh G, Gesche MC, Whitaker KF: Medical consequences of teenage sexuality. NY State J of Med 75(9):1439, August 1975

2. NCHS: Advance report-final natality statistics, 1974. Monthly Vital Statistics Report 24(11) Supplement 2, February 13, 1976

3. NCHS: Interval between first marriage and legitimate first birth, United States, 1964-66. Monthly Vital Statistics Report 18(12) Supplement, 1970

4. Whelan EM: Illegitimate and premaritally conceived first births in Massachusetts, 1966-1968. Social Biology 19(9), 1972

5. Center for Disease Control: Abortion Surveillance 1973. 1975

6. Anderson JE, Morris L: Planned and unplanned fertility in the Albany (NY) Health Region, 1969-1974. Family Planning Perspectives (forthcoming)

7. Kantner JF, Zelnik M: Sexual experience of young unmarried women in the United States. Family Planning Perspectives 4(4):1, October 1972

8. Kantner JF, Zelnik M: Contraception and pregnancy experience of young unmarried women in the United States. Family Planning Perspectives 5(1):11, Winter 1973

9. Serfling RE, Sherman IL: Attribute Sampling Methods for Local Health Departments. Public Health Service Publication No. 1230, U.S. Government Printing Office, Washington, DC, March 1965

10. Kish L: Survey Sampling. John Wiley & Sons, New York, Second Printing, April 1967

11. Teske R: Socioeconomic Stratification for Upstate New York. Unpublished Report, Division of Epidemiology and Preventive Health Services, New York State Department of Health, Albany, NY

12. Ryder NB, Westoff CF: Wanted and Unwanted Fertility in the United States, 1965 and 1970, in Westoff CF and Parke R, eds., Demographic and Social Aspects of Population Growth, U.S. Commission on Population Growth and the American Future, Vol I, USGPO, Washington, DC, pp 471-488

13. Westoff CF: The decline of unplanned births in the United States. Science 191:38-41, 1976

14. Zelnik M, Kantner JF: The Resolution of Teenage First Pregnancies. Family Planning Perspectives 6(2):74, Spring 1974

15. Udry JR, Bauman KE, Morris N: Changes in Premarital Coitus Experience of Recent Decade of Birth Cohorts of Urban American Women. J Marriage Family 37(4):783, November 1975

16. Vener AM, Stewart CS: Adolescent Sexual Behavior in Middle America Revisited, 1970-1973. J Marriage Family 36:728, 1974

17. Dryfoos JG: Women Who Need and Receive Family Planning Services at Mid-Decade. Family Planning Perspectives 7(4):172, July/Aug 1975

18. Morris L: Estimating the need for family planning services among unwed teenagers. Family Planning Perspectives 6(2):91, Spring 1974

TABLE 1

Percent of Never-Married Women 15-44 Years of Age Who Have Ever Had Intercourse, by Socioeconomic Area and SMSA Residence Albany (NY) Health Region: 1974

Socioeconomic Area	Age Group 15-19	20-44	(20-24)	15-44
A. SMSA				
Upper	19.1	64.1	(62.1)	30.5
Upper-Middle	19.0	78.1	(80.8)	38.9
Lower-Middle	27.5	59.0	(50.0)	38.9
Lower*	38.2	78.4	(88.5)	54.3
TOTAL SMSA (Weighted)	22.1	71.2	(72.0)	37.7
B. Non-SMSA				
Upper/Upper-Middle	12.2	73.9		34.4
Lower-Middle	37.9	70.8		46.7
Lower	20.0	55.6		29.4
TOTAL Non-SMSA (Weighted)	24.2	70.7		39.3
C. Albany Health Region				
TOTAL (Weighted)	23.0	71.0		38.4

*Lower Socioeconomic Area by Race:

	15-19	20-44	15-44
Black	45.8	77.3	60.9
White	29.2	78.6	47.4

TABLE 2

Percent of Never-Married Women 15-19 Years of Age Who Have Ever Had Intercourse, by SMSA Residence Albany (NY) Health Region, 1974 and United States, 1971

Residence	Albany Health Region (1974)	United States (1971)*
SMSA of <1,000,000	22.1	--
Central City	--	32.4
Remainder	--	23.0
Non-SMSA	24.2	26.2
SMSA-Lower Socioeconomic Area**		
Black	45.8	60.1
White	29.2	20.7
TOTAL	38.2	33.0

*Kantner JF and Zelnik M: Sexual Experience of Young Unmarried Women in the United States. Family Planning Perspectives 4(4):1, October 1972

**Comparison data for the United States is taken from the category, SMSA ≤150% of Poverty

TABLE 3

Percent of Never-Married Women 15-19 and 15-44 Years of Age Who Have Ever Had Intercourse, by Socioeconomic Area, SMSA Residence, and Religion Albany (NY) Health Region: 1974

Socioeconomic Area	15-19		15-44	
	Catholic	Protestant	Catholic	Protestant
A. SMSA				
Upper/Upper-Middle	17.7	16.7	31.7	35.9
Lower-Middle/Lower	20.3	46.6	34.2	63.6
TOTAL SMSA (Weighted)	18.3	23.1	32.2	41.8
B. Non-SMSA				
Upper/Upper-Middle	**	**	24.1	42.4
Lower-Middle/Lower	31.5	41.6	38.1	48.2
TOTAL Non-SMSA (Weighted)	17.0	33.3	31.6	45.5
C. Albany Health Region				
TOTAL (Weighted)	17.7	27.6	31.9	43.4
D. United States (1971)*	21.1	29.7		

*Kantner JF and Zelnik M: Sexual Experience of Young Unmarried Women in the United States. Family Planning Perspectives (4)4:1, October 1972
**Unweighted number of cases <20

TABLE 4

Percent of Sexually Experienced Never-Married Women 15-19 Years of Age According to Method Most Recently Used, by SMSA Residence: Albany (NY) Health Region, 1974 and United States, 1971

Method Most Recently Used	Albany Health Region (1974)		United States (1971*)		
			SMSA		
	SMSA	Non-SMSA	Central City	Remainder	Non-SMSA
Pill	34.0	45.0	30.3	15.4	17.7
IUD	0.5	0.0	1.8	1.1	0.7
Condom	27.4	5.8	22.4	31.2	29.0
Other**	4.7	14.6	28.4	37.6	36.7
Never Used	33.5	34.5	17.1	14.7	15.9
TOTAL	100.0	100.0	100.0	100.0	100.0

*Kantner JF and Zelnik M: Contraception and Pregnancy-Experience of Young Unmarried Women in the United States, Family Planning Perspectives 5(1): 11, Winter 1973
**Includes foam, diaphragm, rhythm, douche, jelly, and withdrawal

TABLE 5

Percent Distribution of Out-of-Wedlock Births, 1969-1974, by Planning Status for Ethnic Group and Socioeconomic Area Categories--Albany (NY) Health Region

Planning Status	TOTAL	Ethnic Group		Socioeconomic Area	
		White	Black	Lower	All Other
TOTAL	100.0	100.0	100.0	100.0	100.0
Planned	19.0	18.1	19.3	24.8	15.8
Mistimed	73.9	75.6	69.1	67.9	77.4
Unwanted	5.0	4.7	7.7	5.5	4.8
Unknown	2.0	1.6	3.9	1.8	2.0
Number of Births (Unweighted)	(74)	(46)	(24)	(48)	(26)
Planned Pregnancies: United States (1971)*	19.8	15.6	23.6		

*Proportion of premarital first pregnancies for women who were unmarried at time of pregnancy outcome (14)

EFFECTS OF BOUNDING ON TELESCOPING IN THE NATIONAL CRIME SURVEY

Linda R. Murphy and Charles D. Cowan, U.S. Bureau of the Census

Introduction

In a general population sample survey calling for respondent recall of events experienced prior to the interview, the type of memory error known as telescoping is of major concern. Telescoping is the tendency of the respondent to report events as occurring either earlier or later than they actually occurred. An event being reported as occurring earlier than it actually occurred is backward telescoping, whereas forward telescoping is reporting an event as occurring later. Further, both backward and forward telescoping can be either internal to the survey's reference period, or external. Internal telescoping occurs when the respondent correctly places an event within the reference period, but misinforms on the precise day, week, or month of occurrence. External telescoping occurs when the respondent erroneously places an event into the reference period. Telescoping is an important technical issue in a panel survey involving recall for two reasons. First, depending upon the magnitude, nature and direction, uncontrolled telescoping can result in serious response biases in survey estimates for a given time period; and second, various procedural efforts to control telescoping have a major impact on survey design and cost.

The National Crime Survey (NCS) is one such retrospective survey, involving individual respondents in a national, rotating sample of 72,000 housing units. The NCS rotation scheme is such that an incoming panel or rotation group of approximately 12,000 new sample units is introduced over each 6-month period, replacing about the same number of units which expire from the sample or rotate out during the 6-month period. Thus, at any given time, the NCS sample consists of a number of panels, or rotation groups, that are being interviewed for the first through the seventh time. (Due to the nature of the original introduction of the NCS sample, some small number of households may be interviewed 8 or 9 times through 1976 only, after which time the NCS rotation scheme will be fully operative.) Respondents are asked to report all incidents of specific types of criminal victimization that they have experienced during the 6 months preceding the interview month. The types of crime covered by NCS are completed and attempted assault, robbery, burglary, larceny, and auto theft. A self-respondent technique, as opposed to household respondent, is used to collect data for persons 12 and over. (Papers presented previously at ASA meetings [1, 9] explain the NCS design in greater detail.)

Developmental work during the 3-year period prior to the start of NCS in July 1972, provided some information about the nature, direction and magnitude of telescoping in victim surveys. Results of 3 reverse record check studies of known victims (4, 7, 8) indicated the presence of both forward and backward telescoping, the net effect being forward. These results also indicated the presence of both internal and external telescoping. The next stage of developmental work involved a victim survey supplement to the Quarterly Household Survey, a national probability sample of 18,000 households. Results of this effort (6) indicated that telescoping, both internal and external, was of such a magnitude as to justify incorporating special features into the NCS design and procedures that would minimize the effects of telescoping.

One NCS design feature, intended to minimize the effect of internal telescoping, both forward and backward, is the use of a rolling reference period, achieved by interviewing one-sixth of the sample each month. Data on incidents of victimization are then tabulated according to reported month of occurrence by 3-month calendar periods called Data Quarters. Data for one Data Quarter are collected over 8 months of interviewing. Because households are interviewed every 6 months, one-third of the sample is interviewed twice for the same Data Quarter. For example, a respondent interviewed in February, 1974 would be interviewed again in August, 1974; this respondent would contribute data for January, 1974 from the first interview, and for February and March 1974 from the second interview. In an effort to control forward external telescoping, an interviewing procedure called bounding was instigated for the purpose of minimizing the shifting of reports of crimes into the NCS reference period. Bounding is thus a procedure utilized to prevent the reporting of the same incidents in consecutive reference periods by eliminating reports of incidents that were also reported during the previous interview. The initial interviews at addresses in incoming rotation groups are used to bound subsequent interviews; they are not used to produce the estimates of victimizations. This is a very costly feature of the NCS design, since the data from incoming rotation groups are not used in tabulating results for publication.

The primary focus of this paper is to examine the effectiveness of the procedural and design features of NCS related to bounding, in controlling forward external telescoping, using data, for the first time, from NCS itself. This will be done by comparing estimates of victimizations based on bounded data from returning rotation groups with estimates based on unbounded data from incoming rotation groups. This estimate of the bounding effect using NCS data can provide a basis for reevaluating the cost-effectiveness of this aspect of the survey design, though in this paper we provide only a bivariate description of the data.

A second issue addressed in this paper is variation in forward external telescoping, by means of examining defferential effects of bounding by demographic and socioeconomic characteristics of respondents, as well as by characteristics of the incidents of victimization. We believe this is a critical issue to investigate in victimization surveys. If there are no significant differences in telescoping for either certain classes of respondents or for certain classes of events, then relationships and patterns would be unaffected by the inclusion of unbounded data in producing survey estimates, though levels of victimization would be affected. Thus it may be less crucial to maintain the bounded aspects of the NCS. On the other hand, if there are significant differences in telescoping by certain respondents or for certain incidents, then relationships and patterns would be distorted by including unbounded data in the survey estimates.

Two broad questions suggest themselves. First, do some groups of respondents telescope events more than others? Second, are some types of incidents telescoped more than other types? There are two feasible, but opposing hypotheses related to differential telescoping by type of incident. One hypothesis is that the more important, more serious, or more salient events are telescoped forward to a greater degree than the less important, etc., perhaps because the less important are more likely to be completely forgotten; the second hypothesis is that the less important, less serious, or less salient events are telescoped forward to a greater degree because the month of occurrence is less accurately recalled and therefore subject to greater recall bias.

A final aspect of the bounded NCS design discussed in this paper is the extent of actual bounding of interviews among households, and within households among persons, in repeat rotation groups.

Comparison of Victimization Rates

We have two estimates, total personal victimizations 1/ and total property victimizations, 2/ that are of primary interest to us. For each, we are comparing the total bounded victimization rate obtained from the returning rotation groups with the unbounded victimization rate estimated from the incoming rotation groups. The rates for each sample are produced using identical processing, weighting, and tabulation procedures, with appropriate adjustments to account for the fact that the incoming rotation group is approximately one-sixth the size of the bounded sample.

The first two tables in our report show the bounded and unbounded rates for total personal and total property victimizations for Data Quarters I-74 through I-75, plus a z-test of the difference between the rates. All rates reported in the tables are victimizations per thousand people or households. Tables 1 through 7 indicate for each of the Data Quarters under analysis (column 1) the weighted population sizes being represented by the bounded and unbounded samples (columns 2 and 3), the victimization rates estimated for the population from the bounded and unbounded samples (columns 4 and 5), the difference between the unbounded and bounded rates divided by the bounded rate and expressed as a percent (column 6), the standard errors associated with the two rates (columns 7 and 8), the standard error of the absolute difference between the rates (column 9), and the z-statistic testing whether the difference between the bounded rates is significantly greater than zero. The test is calculated as the ratio of the absolute difference between the unbounded and bounded victimization rates to the square root of the sum of the squared standard errors associated with each rate (3). The standard errors used were published in the 1973 Advance Report, Criminal Victimization in the United States by the Law Enforcement Assistance Administration (LEAA). Because of the large sample sizes, the z-statistic approximates the normal distribution, and can be used in conjunction with a table of normal areas and ordinates to determine the level of significance of the test. The test being performed is a one-tailed z-test, because the procedure of bounding as applied in NCS would only eliminate reporting of victimizations in two consecutive quarters. There is never a chance that victimizations would be added to the reports because of bounding. The null hypothesis, formally stated, is "there is no difference between bounded and unbounded victimization rates." The alternative hypothesis, formally stated, is that "unbounded victimization rates are greater than bounded rates." A z-value of greater than 1.64 means that we can be sure 95 out of a hundred times that the estimated differences are greater than zero, and thus are not due to sampling variation; similarly, a value greater than 1.28 means that 90 out of a hundred times, the results will not be due to sampling variation (except in the comparison of victimizations reported to police, Table 8, which is a two-tailed test, with z-values of 1.96 and 1.64 respectively).

The z-values clearly show that there are statistically significant differences in the bounded and unbounded personal crime victimization rates for each quarter (Table 1), demonstrating that bounding does eliminate a significant number of duplicate victimization reports. The same is true for bounded and unbounded property crime victimization rates (Table 2). The unbounded personal victimization rates average 43.8 percent higher than the bounded rates, ranging from 36.4 percent to 58.6 percent. The unbounded property victimization rates average 39.9 percent higher than the bounded rates and range from 35.0 percent to 44.1 percent.

Now the question is: How does telescoping affect subgroup estimates and estimates by type of crime? The following analysis again makes use of testing the difference between bounded and unbounded rates, and represents a preliminary look at the data. A more detailed analysis of the effects of bounding on telescoping for subestimates is planned as more data are collected. The first comparisons we make are by type of crime. Rates of assaultive violence without theft are 44 percent higher on average in the unbounded sample than in the bounded sample, and rates for personal theft without assault are 51 percent higher on average in the unbounded sample (Tables 3A and 3B). But comparing the rates quarter by quarter, there is no clear-cut pattern showing that unbounded rates are uniformly higher for the one type of crime over the other. The same result can be found comparing burglaries with larcenies (Table 4A and 4B). The unbounded rates for burglaries and larcenies are on average 40 percent higher than the bounded rates. But in some quarters the relative difference for burglaries is significantly ($\alpha < .10$) larger than for larcenies, and in other quarters it is smaller. Telescoping doesn't seem to consistently affect rates for one major type of crimes more than another.

Looking at subgroups of burglary, however, it becomes apparent that telescoping is much more prevalent for attempted entries than for actual entries. The unbounded sample rates for burglary: actual entry are on average only 32 percent higher than the bounded rates, whereas the unbounded sample rates for burglary: attempted entry are 66 percent higher than the bounded rates on average (Tables 5A and 5B). The same is true when comparing completed and attempted larcenies. Again the relative difference is higher for attempted than completed crimes. The average relative difference for the five quarters for attempted larcenies was 50 percent, while for completed larcenies it was only 40 percent (Tables 6A and 6B). So it is apparent that telescoping does have a differential effect on the rates of various subcategories of crimes.

There also seem to be some differences in telescoping by demographic characteristics of households for property crimes. The relative difference between bounded and unbounded rates for property crimes reported by one-person households is rather low, only 20 percent higher for the unbounded sample on average. This relative difference increases as the number of persons in the household increases, rising to a 51 percent greater reporting rate in the unbounded sample for households having six or more persons (Tables 7A-7D).

Another factor which may indicate saliency or importance, and thus influence telescoping, is whether or not the victimization was reported to the police. In 4 of the 5 Quarters examined, a significantly larger proportion of the property victimizations were not reported to the police in the unbounded sample (Table 8).

Two additional factors, total loss suffered in property victimizations and for personal victimizations, whether or not the offender was a stranger to the victim, were included in our analysis as possible indicators of saliency or importance. However, we found no consistent pattern associated with either of these variables over the 5 Data Quarters examined.

The figures presented in these tables are simply a faithful reporting of the degree to which telescoping occurs. At present, it is safe to conclude that telescoping would have a significant effect on victimization rates if the interviews were not bounded. Beyond that, we can point out that some crimes are telescoped to a greater degree than others, either according to the type of crime or the circumstances, or because of the demographic characteristics of the household. We do not have valid empirical information about <u>why</u> these factors affect telescoping.

<u>Qualifications to Comparisons</u>

Three qualifications should be noted with regard to the preceding analysis comparing bounded with unbounded data from NCS. The first qualification is that since respondents are interviewed every six months, the Data Quarters are not independent from one another, as there is some overlap of respondents from one Data Quarter to the next. Secondly, all data from returning rotation groups are considered and treated as bounded for purposes of the preceding analysis in this paper. However, since NCS uses a probability sample of addresses rather than designated households or persons, not all of the interviews conducted in returning rotation groups are subject to the actual interviewing procedure of bounding. For any interviews in a household to actually be bounded, the identical household must have been interviewed the previous enumeration period. Therefore, interviews in replacement households, and households that were noninterview or not in sample the previous period, are actually unbounded. However, data from these unbounded interviews are included with data from the bounded interviews because they are in returning rotation groups, and their exclusion may bias the sample.

The unbounded households in returning rotation groups comprise a sizeable portion of the interviewed sample (Table 9), averaging 13.3 percent over the five Collection Quarters, I-74 through I-75. Of these unbounded households, an average of 9.6 percent are replacement households, and 3.7 percent were previously noninterview or not in sample. These unbounded households contribute disproportionally more victimizations than do the actually bounded households. Though bounded households make up about 86 percent of the interviews, they contribute only 76 percent of the victimizations, while unbounded households, which comprise only 13 percent of the interviews, contribute 24 percent of the victimizations. This translates into a reporting rate of about 79 percent more victimizations from unbounded households than expected from their proportion of the sample.

Even more striking in terms of contributing victimizations, is the difference between types of unbounded households. Households that were previously noninterview or not in sample, while making up 4 percent of the interviews, contribute almost 6 percent of the victimizations. But replacement households, which primarily represent movers and make up about 10 percent of the interviews, contribute an average of nearly 18 percent of the victimizations, or 92 percent more than their expected proportion. Recalling the overall difference of about a 40 percent higher victimization rate for unbounded, incoming rotation groups than for bounded, returning rotation groups, these figures appear to indicate that something more than merely the lack of bounding may be related to the disproportionate reporting of victimizations among replacement households. It is conceivable that they actually experience victimization more frequently than non-movers for reasons associated with their mobility--perhaps they move to get away from crime. This question appears to warrant further investigation.

Admittedly the set of data used in the preceding discussion of unbounded data within returning rotation groups is somewhat lacking in refinement, being based on unweighted counts. However, the stability of the patterns is apparent and provides evidence that the effect of bounding is understated in comparisons of data between incoming and returning rotation groups, since returning groups include a substantial amount of unbounded data.

The third, further qualification is that even within actually bounded households, some interviews with individual household members are unbounded, either because the person is new to the household since the prior enumeration period, or because the person was previously noninterview. A special computer match of interviewed persons in Collection Quarters I-74 through I-75 with files for previous enumeration periods was performed for the purpose of determining correspondence and bounding of individuals within bounded households. Results of that operation indicate an average of about 95 percent bounded individual interviews (Table 10). Again, this pattern is quite stable over time; and again a difference in reporting victimizations between bounded and unbounded interviews is evident. An average of 7.9 percent of the bounded persons, and 10.5 percent of the unbounded persons reported one or more victimizations. These data also appear to provide evidence that the bounding effect is understated in comparisons of incoming with returning rotation group data.

Conclusion

The data presented in this paper strongly support the conclusion that NCS bounding procedures and design effectively reduce the memory bias of forward external telescoping. Our results, comparing bounded with unbounded sample data, are consistent with results from similar comparisons in the area of consumer expenditures (5). In that study, however, Neter and Waksberg point out that telescoping effects are compounded with conditioning effects in comparisons between unbounded data based on first interviews and bounded data based on second or later interviews. Evidence from the expenditure study (5) and also from a study of NCS panel bias (9) suggests that conditioning probably accounts for a much smaller portion of the observed differences in NCS than does telescoping.

Further, we can conclude that some variation in telescoping is associated with characteristics of victimization events. Our analysis indicated that telescoping was present for all major types of crimes, but in no discernible pattern. However, it did indicate a greater degree of telescoping for the subcategories of attempted larceny and attempted burglary than for the completed crimes. It also indicated a larger proportion of victimizations not reported to police in the unbounded sample than in the bounded. These results, considered alone, could be interpreted as evidence that the less serious, less important, or less salient events are more subject to the recall bias of forward telescoping than the more serious, etc. However, the finding of no pattern of association with total loss or victim-offender relationship, does not support this interpretation. Therefore, we can only conclude that some characteristics of events appear to be related to differential forward external telescoping. Finally, our evidence also indicates that some variation in telescoping is associated with household characteristics, but hardly any telescoping can be explained by respondent characteristics. This lack of establishing a strong relationship between respondent characteristics and the memory bias of telescoping, is consistent with findings by Gottfredson and Hindelang on total memory bias based on NCS Cities data (2). Most of the differences found in our analysis of demographic variables, including age, sex, race, education, tenure, and income, were tenuous at best.

We plan to investigate further the differential effects of bounding on telescoping by characteristics of victimization events, households, and respondents, and to test what biases would arise in the sample if unbounded households were excluded from tabulations based on returning rotation groups.

Footnotes

[1] Personal crimes encompass completed and attempted assault, including rape, and robbery.

[2] Property crimes encompass completed and attempted burglary, larceny, and auto theft.

References

(1) Argana, Marie G., Thompson, Marvin M., and Gerson, Earle J. "The Measurement of Crime Through Victimization Surveys: The Census Bureau Experience," Paper presented at the annual meetings of the American Statistical Association, December 27-30, 1973, New York.

(2) Gottfredson, Michael R. and Hindelang, Michael J., "A Consideration of Telescoping and Memory Decay Biases in Victimization Surveys," 1976 (unpublished).

(3) Hoel, Paul G., Introduction to Mathematical Statistics, John Wiley and Sons, Inc., New York, New York, 1962.

(4) Law Enforcement Assistance Administration, The San Jose Methods Test of Known Crime Victims, Statistics Technical Report No.1, LEAA, U.S. Department of Justice, June 1972.

(5) Neter, John and Waksberg, Joseph, U.S. Bureau of the Census, "Response Errors, in the Collection of Expenditures Data by Household Interviews: An Experimental Study," Technical Paper No. 11.

(6) Turner, Anthony G., Methodological Issues in the Development of the National Crime Survey Panel: Partial Findings," LEAA, December 1970 (unpublished).

(7) U.S. Bureau of the Census, "Victim Recall Pretest (Washington, D.C.)," Demographic Surveys Division, June 10, 1970 (unpublished).

(8) U.S. Bureau of the Census, "Household Survey of Victims of Crime Second Pretest (Baltimore, Maryland)," Demographic Surveys Division, November 30, 1970 (unpublished).

(9) Woltman, Henry and Bushery, John, "A Panel Bias Study in the National Crime Survey," Paper presented at the annual meetings of the American Statistical Association, August 25-28, 1975, Atlanta, Georgia.

TABLE 1: Total Personal Victimization Rates for Bounded and Unbounded Samples (Rates per 1000 persons)

Data	Population		Victimization Rate			Standard Errors			z of
Quarter	Bounded	Unbounded	Bounded	Unbounded	% Difference	Bounded	Unbounded	Difference	Difference
I/74	163799000	27299833	7.89	11.30	43.219	.268	0.834	0.876	3.892
II/74	164244000	27374000	8.90	12.31	38.315	.285	0.871	0.916	3.721
III/74	164861000	27476833	9.38	14.88	58.635	.292	0.958	1.002	5.491
IV/74	165344000	27557333	9.74	13.29	36.448	.297	0.904	0.952	3.731
I/75	165874000	27645666	8.55	12.17	42.339	.275	0.864	0.906	3.994

TABLE 2: Total Property Victimization Rates for Bounded and Unbounded Samples (Rates per 1000 households)

Data	Households		Victimization Rate			Standard Errors			z of
Quarter	Bounded	Unbounded	Bounded	Unbounded	% Difference	Bounded	Unbounded	Difference	Difference
I/74	71118300	11853050	102.77	138.75	35.010	1.247	3.407	3.628	9.917
II/74	71489200	11914866	104.09	149.97	44.077	1.249	3.502	3.718	12.339
III/74	72163700	12027283	114.99	156.65	36.229	1.292	3.548	3.776	11.033
IV/74	72565900	12094316	119.80	168.96	41.035	1.308	3.647	3.874	12.689
I/75	72686500	12114416	102.75	147.16	43.221	1.225	3.453	3.664	12.120

TABLE 3: Comparison of Bounded and Unbounded Personal Victimization Rates for Various Types of Crimes (Rates per 1000 persons)

3A - Type of Crime: Assaultive Violence--Without Theft

Data	Population		Victimization Rate			Standard Errors			z of
Quarter	Bounded	Unbounded	Bounded	Unbounded	% Difference	Bounded	Unbounded	Difference	Difference
I/74	163799000	27299833	5.36	8.12	51.493	.215	.689	.721	3.827
II/74	164244000	27374000	6.67	9.09	36.282	.243	.739	.778	3.112
III/74	164861000	27476833	6.67	10.75	61.169	.241	.812	.847	4.817
IV/74	165344000	27557333	6.76	8.83	30.621	.242	.724	.764	2.711
I/75	165874000	27645666	5.93	8.36	40.978	.222	.699	.734	3.311

3B - Type of Crime: Personal Theft--Without Assault

Data	Population		Victimization Rate			Standard Errors			z of
Quarter	Bounded	Unbounded	Bounded	Unbounded	% Difference	Bounded	Unbounded	Difference	Difference
I/74	163799000	27299833	1.91	2.48	29.843	.107	.391	.405	1.407
II/74	164244000	27374000	1.63	2.62	60.736	.106	.397	.411	2.410
III/74	164861000	27476833	2.04	3.44	68.627	.104	.431	.444	3.155
IV/74	165344000	27557333	2.33	3.52	51.073	.102	.434	.446	2.667
I/75	165874000	27645666	2.01	2.88	43.284	.100	.407	.420	2.074

TABLE 4: Comparison of Bounded and Unbounded Property Victimization Rates for Various Types of Crimes (Rates per 1000 households)

4A - Type of Crime: Burglary

Data	Households		Victimization Rate			Standard Errors			z of
Quarter	Bounded	Unbounded	Bounded	Unbounded	% Difference	Bounded	Unbounded	Difference	Difference
I/74	71118300	11853050	19.23	27.38	42.382	.567	1.598	1.696	4.807
II/74	71489200	11914866	22.60	33.34	47.522	.612	1.760	1.864	5.763
III/74	72163700	12027283	26.85	36.62	36.387	.664	1.837	1.954	5.001
IV/74	72565900	12094316	23.89	31.40	31.436	.625	1.698	1.809	4.152
I/75	72686500	12114416	20.65	29.13	41.065	.581	1.634	1.734	4.890

4B - Type of Crime: Larceny

Data	Households		Victimization Rate			Standard Errors			z of
Quarter	Bounded	Unbounded	Bounded	Unbounded	% Difference	Bounded	Unbounded	Difference	Difference
I/74	71118300	11853050	79.15	105.34	33.089	1.110	3.080	3.274	7.999
II/74	71489200	11914866	77.13	111.93	45.119	1.094	3.141	3.326	10.464
III/74	72163700	12027283	83.34	113.44	36.117	1.124	3.144	3.339	9.014
IV/74	72565900	12094316	90.78	129.74	42.917	1.163	3.295	3.494	11.151
I/75	72686500	12114416	77.60	112.30	44.716	1.085	3.124	3.307	10.494

TABLE 5: Comparison of Bounded and Unbounded Property Victimization Rates for Various Types of Crimes (Rates per 1000 households)

5A - Type of Crime: Burglary—Entry

Data	Households		Victimization Rate			Standard Errors			z of
Quarter	Bounded	Unbounded	Bounded	Unbounded	% Difference	Bounded	Unbounded	Difference	Difference
I/74	71118300	11853050	15.08	19.89	31.897	.503	1.362	1.451	3.314
II/74	71489200	11914866	17.85	25.28	41.625	.545	1.532	1.626	4.569
III/74	72163700	12027283	20.94	27.71	32.330	.587	1.598	1.703	3.976
IV/74	72565900	12094316	18.77	24.25	29.196	.554	1.492	1.592	3.443
I/75	72686500	12114416	16.09	20.43	26.973	.512	1.369	1.462	2.969

5B - Type of Crime: Burglary—Attempted Entry

Data	Households		Victimization Rate			Standard Errors			z of
Quarter	Bounded	Unbounded	Bounded	Unbounded	% Difference	Bounded	Unbounded	Difference	Difference
I/74	71118300	11853050	4.15	7.49	80.482	.282	.856	.901	3.708
II/74	71489200	11914866	4.75	8.06	69.684	.301	.880	.930	3.559
III/74	72163700	12027283	5.92	8.91	50.507	.327	.915	.971	3.078
IV/74	72565900	12094316	5.12	7.15	39.648	.307	.833	.888	2.287
I/75	72686500	12114416	4.56	8.70	90.789	.289	.903	.949	4.365

TABLE 6: Comparison of Bounded and Unbounded Property Victimization Rates for Various Types of Crimes (Rates per 1000 households)

6A - Type of Crime: Larceny--Completed

Data	Households		Victimization Rate			Standard Errors			z of
Quarter	Bounded	Unbounded	Bounded	Unbounded	% Difference	Bounded	Unbounded	Difference	Difference
I/74	71118300	11853050	73.98	97.72	32.090	1.078	2.990	3.179	7.469
II/74	71489200	11914866	72.57	105.22	44.991	1.065	3.072	3.252	10.040
III/74	72163700	12027283	77.50	104.87	35.316	1.089	3.058	3.246	8.432
IV/74	72565900	12094316	85.42	121.49	42.227	1.132	3.216	3.409	10.580
I/75	72686500	12114416	72.48	104.34	43.957	1.053	3.043	3.221	9.893

6B - Type of Crime: Larceny--Attempted

Data	Households		Victimization Rate			Standard Errors			z of
Quarter	Bounded	Unbounded	Bounded	Unbounded	% Difference	Bounded	Unbounded	Difference	Difference
I/74	71118300	11853050	5.17	7.62	47.389	.316	.862	.918	2.670
II/74	71489200	11914866	4.56	6.71	47.149	.295	.817	.869	2.475
III/74	72163700	12027283	5.84	8.57	46.747	.325	.900	.957	2.853
IV/74	72565900	12094316	5.36	8.25	53.918	.313	.884	.938	3.082
I/75	72686500	12114416	5.12	7.96	55.469	.307	.870	.923	3.078

TABLE 7: Comparison of Bounded and Unbounded Property Victimization Rates by Number of Persons in Household (Rates per 1000 households)

7A - Number of Persons in Household: 1 Person

Data	Households		Victimization Rate			Standard Errors			z of
Quarter	Bounded	Unbounded	Bounded	Unbounded	% Difference	Bounded	Unbounded	Difference	Difference
I/74	14402200	2400366	58.87	64.93	10.294	2.119	6.003	6.366	.952
II/74	14537300	2422883	62.20	76.44	22.894	2.166	6.397	6.754	2.108
III/74	14818900	2469816	66.58	82.49	23.896	2.215	6.515	6.881	2.312
IV/74	14924200	2487366	69.15	83.62	20.926	2.248	6.516	6.893	2.099
I/75	14939600	2489933	58.37	72.01	23.368	2.061	6.099	6.438	2.119

7B - Number of Persons in Household: 2 to 3 Persons

Data	Households		Victimization Rate			Standard Errors			z of
Quarter	Bounded	Unbounded	Bounded	Unbounded	% Difference	Bounded	Unbounded	Difference	Difference
I/74	34497900	5749650	89.09	121.93	36.862	1.621	4.783	5.050	6.503
II/74	34711200	5785200	91.72	131.01	42.837	1.637	4.916	5.182	7.582
III/74	35137300	5856216	103.25	133.00	28.814	1.720	4.927	5.218	5.701
IV/74	35417700	5902950	104.40	134.25	28.592	1.723	4.933	5.225	5.713
I/75	35436800	5906133	87.89	125.66	42.974	1.588	4.799	5.055	7.472

7C - Number of Persons in Household: 4 to 5 Persons

Data	Households		Victimization Rate			Standard Errors			z of
Quarter	Bounded	Unbounded	Bounded	Unbounded	% Difference	Bounded	Unbounded	Difference	Difference
I/74	16814600	2802433	140.65	186.12	32.328	2.845	8.067	8.555	5.315
II/74	16890500	2815083	138.87	200.83	44.617	2.825	8.289	8.757	7.075
III/74	16939100	2823183	156.23	217.86	39.448	2.945	8.557	9.049	6.811
IV/74	16974400	2829066	163.47	268.20	64.067	2.992	9.271	9.742	10.750
I/75	17052000	2842000	141.86	219.25	54.554	2.830	8.504	8.963	8.635

7D - Number of Persons in Household: 6 or More Persons

Data	Households		Victimization Rate			Standard Errors			z of
Quarter	Bounded	Unbounded	Bounded	Unbounded	% Difference	Bounded	Unbounded	Difference	Difference
I/74	5386200	897700	189.88	282.68	48.873	5.890	16.658	17.669	5.252
II/74	5336300	889383	188.39	303.44	61.070	5.886	17.084	18.070	6.367
III/74	5262600	877100	197.13	320.15	62.406	6.029	17.512	18.520	6.642
IV/74	5242700	873783	226.72	324.76	43.243	6.422	17.631	18.764	5.225
I/75	5245700	874283	202.24	281.56	39.221	6.103	16.928	17.995	4.408

TABLE 8: Comparison of Total Property Victimizations Not reported to Police for Bounded and Unbounded Samples

Data	Number of Victimizations		Percent Victimizations Not Reported to Police			Standard Errors			z of
Quarter	Bounded	Unbounded	Bounded	Unbounded	Difference	Bounded	Unbounded	Difference	Difference
I/74	7309020	9870360	70.672	72.575	1.903	.6000	.4778	.7671	2.481
II/74	7441220	10767050	66.954	68.769	1.815	.6128	.4779	.7771	2.335
III/74	8298200	11316020	65.504	65.132	-0.372	.5824	.4857	.7583	-0.491
IV/74	8693250	12271730	68.884	71.546	2.662	.5481	.4394	.7024	3.790
I/75	7468410	10728060	67.343	71.339	3.996	.6097	.4661	.7674	5.207

THE EVALUATION OF DIFFERENT COUNTING RULES AND WEIGHTING PROCEDURES FOR SURVEYS WITH MULTIPLICITY

Gad Nathan, Hebrew University, Jerusalem

1. Introduction and Notation

In a sample survey with multiplicity, (Sirken [3]), designed to estimate the total number, N, of events occurring in a population of L reporting units, each event is linked to one or more reporting units, by a well defined counting rule. The counting rule is defined by the values of the indicator variable, $\delta_{\alpha,i}$, which takes the value one if the α-th event is linked to the i-th reporting unit ($\alpha=1,\dots,N$; $i=1,\dots,L$), where:

$$(1.1) \qquad s_\alpha = \sum_{i=1}^{L} \delta_{\alpha,i} \geq 1 \quad (\alpha=1,\dots,N),$$

i.e. each event is linked to at least one reporting unit. The value of s_α is termed the multiplicity of the α-th event.

A probability sample, S, of reporting units is defined by the indicator variable $d_i(S)$, which equals 1 if the i-th reporting unit is in the sample, S.

Initially, it will be assumed that, if $d_i(S) = 1$, the vector $\underline{\delta}_\alpha = (\delta_{\alpha,1},\dots,\delta_{\alpha,L})$ is known for all $\alpha=1,\dots,N$, such that $\delta_{\alpha,i} = 1$, i.e., each reporting unit in the sample reports, for each event linked to it, on all other reporting units linked to the same event and there are no response errors.

If $E[d_i(S)] > 0$ ($i=1,\dots,L$), then a weighted linear multiplicity estimator of N is defined as:

$$(1.2) \qquad \hat{N}(S) = \sum_{i=1}^{L} \frac{d_i(S)}{E[d_i(S)]} \sum_{\alpha=1}^{N} w_{\alpha,i}(S)\delta_{\alpha,i} ,$$

where $w_{\alpha,i}(S)$ are any real numbers. The expectation of the estimator is:

$$(1.3) \qquad E[\hat{N}(S)] = \sum_{i=1}^{L}\sum_{\alpha=1}^{N} E[w_{\alpha,i}(S)|d_i(S) = 1]\delta_{\alpha,i} = \sum_{\alpha=1}^{N}\sum_{i=1}^{L} w_{\alpha,i}\delta_{\alpha,i} ,$$

where $w_{\alpha,i} = E[w_{\alpha,i}(S)|d_i(S) = 1]$.

Thus, in order that $\hat{N}(S)$ be an unbiased estimator of N for any matrix, $\Delta = ||\delta_{\alpha,i}||$ ($\alpha=1,\dots,N;i=1,\dots,L$), for which (1.1) holds, the following relationship must hold;

$$(1.4) \qquad \sum_{i=1}^{L} w_{\alpha,i}\delta_{\alpha,i} = 1 \quad (\alpha=1,\dots,N).$$

Sirken [3] has proposed using the weights:

$$(1.5) \qquad w_{\alpha,i}(S) = w_{\alpha,i} = 1/s_\alpha ,$$

for all S, and for all (α,i) such that $\delta_{\alpha,i} = 1$ (i.e. weighting by the reciprocal of the multiplicity). For these weights, referred to in [4] as unit weights, (1.4) obviously holds, so that N(S) is an unbiased estimator. In the following, it will be shown that, in a certain sense, this weighting is optimal for simple random sampling under the above assumption, in the absence of response errors. However alternative weighting procedures must be considered if there are response errors. The components of mean square error, taking response errors into account, are evaluated approximately as a function of certain parameters. This allows the comparison of alternative counting rules and weighting methods.

2. Optimal Weighting in the Absence of Response Errors

In the following, simple random sampling without replacement, of size ℓ, will be assumed, i.e.:

$$(2.1) \qquad E[d_i(S)] = \ell/L; \qquad E[d_i(S)d_j(S)] = [\ell(\ell-1)]/[L(L-1)], \ (i\neq j).$$

Let:

$$(2.2) \qquad X_i(S) = \sum_{\alpha=1}^{N} w_{\alpha,i}(S)\delta_{\alpha,i}$$

and

$$(2.3) \qquad X_i = E[X_i(S)|d_i(S)=1] = \sum_{\alpha=1}^{N} w_{\alpha,i}\delta_{\alpha,i} .$$

Then it can easily be shown that the variance of the estimate (1.3) is:

$$(2.4) \qquad Var[\hat{N}(S)] = (L/\ell)^2 E\{\sum_{i=1}^{L} d_i(S)[X_i(S)-X_i]\}^2 + \frac{L(L-\ell)}{\ell(L-1)}\sum_{i=1}^{L}(X_i-N/L)^2$$

Each of the two terms is non-negative and, for given values of X_i, the variance is minimal if, for all S such that $d_i(S) = 1$,

$$(2.5) \qquad X_i(S)-X_i = \sum_{\alpha=1}^{N} [w_{\alpha,i}(S)-w_{\alpha,i}]\delta_{\alpha,i} = 0.$$

But, for (2.5) to hold it is sufficient if, for all S, such that $d_i(S) = 1$:

$$w_{\alpha,i}(S) = w_{\alpha,i} \quad (\alpha=1,\dots,N;i=1,\dots,L). \tag{2.6}$$

Since, when (2.6) holds, the variance (2.4) depends only on the values of X_i, which in turn depend only on the values of $w_{\alpha,i}$, weighting which is independent of the sample is optimal, in the above sense.

For sample-independent weighting (i.e. if (2.6) holds) the variance (2.4) is minimized if $\sum_{i=1}^{L} X_i^2$ is minimal subject to (1.4). For given Δ, weights $w_{\alpha,i}$ can easily be found to attain this. However Δ must be assumed unknown, except for the row vectors $\underline{\delta}_\alpha=(\delta_{\alpha,1},\dots,\delta_{\alpha,L})$, such that $\sum_{i=1}^{L} d_i(S)\delta_{\alpha,i}\geq 1$ (i.e. for events linked to at least one reporting unit in the sample). But sample-independent weighting requires that for given α_o, the weights, $w_{\alpha_o,i}$, must be a function only of $\underline{\delta}_{\alpha_o}$. Thus the following minimax approach is implied:

Let $D(\underline{\delta}_{\alpha_o})$ be the set of matrices, Δ, for which the α_o-th row is $\underline{\delta}_{\alpha_o}$. For a given matrix Δ, let $W(\Delta)$ be the set of all vector functions:

$$\underline{w}: \{0,1\}^L \to R^L ,$$

for which (1.4) holds, i.e.:

$$\sum_{i=1}^{L} w_i(\underline{\delta}_\alpha)\delta_{\alpha,i} = 1. \tag{2.7}$$

Let:

$$f(\underline{w},\Delta) = \sum_{i=1}^{L} [\sum_{\alpha=1}^{N} w_i(\underline{\delta}_\alpha)\delta_{\alpha,i}]^2 . \tag{2.8}$$

Then, for a given vector $\underline{\delta}_{\alpha_o}$, the optimal weight vector, $\underline{w}^*(\underline{\delta}_{\alpha_o})$, is defined as that for which:

$$\min_{\underline{w}\in W(\Delta)} \max_{\Delta\in D(\underline{\delta}_{\alpha_o})} f(\underline{w},\Delta), \tag{2.9}$$

is attained.

It is easy to see that, for a given $\underline{\delta}_{\alpha_o}$,:

$$\max_{\Delta\in D(\underline{\delta}_{\alpha_o})} f(\underline{w},\Delta) = (N-1)^2 + \sum_{i=1}^{L} w_i^2(\underline{\delta}_{\alpha_o}) + 2(N-1) \max_{1\leq i\leq L} w_i(\underline{\delta}_{\alpha_o}), \tag{2.10}$$

for any $\underline{w}\in W(\Delta)$ such that $\Delta\in D(\underline{\delta}_{\alpha_o})$. The maximum is attained by Δ^* such that $\underline{\delta}^*_{\alpha_o} = \underline{\delta}_{\alpha_o}$ and, if $\max_{1\leq i\leq L} w_i(\underline{\delta}_{\alpha_o}) = w_{i_o}(\delta_{\alpha_o})$,

$$\delta^*_{\alpha,i} = \begin{cases} 1 ; & i=i_o \\ 0 ; & i\neq i_o , \end{cases} \tag{2.11}$$

$$(\alpha\neq\alpha_o;\ \alpha=1,\dots,N).$$

The minimum of (3.4), for $\underline{w}\in W(\Delta^*)$, is:

$$\min_{\underline{w}\in W(\Delta^*)} f(\underline{w},\Delta^*) = (N-1)^2 + \frac{1}{s_{\alpha_o}} + \frac{2(N-1)}{s_{\alpha_o}}. \tag{2.12}$$

This minimum is attained by $\underline{w}^*$ for which $w_i^*(\underline{\delta}_{\alpha_o}) = \frac{1}{s_{\alpha_o}}$, for all $i=1,\dots,L$, such that $\delta_{\alpha_{oi}} = 1$.

Thus the unit weighting (1.5) is optimal in the above minimax sense.

3. Weighting in the Presence of Response Errors

If, as usually happens in practice, events, for which $\delta_{\alpha,i} = 1$ and for which $d_i(S) = 1$, are sometimes under-reported, or if there is over-reporting of events, the optimality of the weights (1.5) is not necessarily attained. For the case where the reciprocals of the multiplicities are used as weights, an approximation to the total mean square error, including response and sampling error components, of the estimate (1.3) is given in Nathan [1], under certain simplifying assumptions. In the following a similar development for a general weighting scheme will be given.

The weighting can be defined by means of a variable, $Z_{\alpha,i}$, measurable for all i, such that $\delta_{\alpha,i}=1$ for any event, α, for which $\sum_{i=1}^{L} d_i(S)\delta_{\alpha,i}\geq 1$ (i.e. the event is linked to at least one reporting unit in the sample).

Let $Z_\alpha = \sum_{i=1}^{L} Z_{\alpha,i}\delta_{\alpha,i}$. Then for the weights:

$$(3.1) \qquad w_{\alpha,i} = \frac{Z_{\alpha,i}}{Z_\alpha},$$

the relationship (1.4) holds, so that, in the absence of response errors, $\hat{N}(S)$, would be unbiased. Thus if the variable $Z_{\alpha,i}$ is the number of persons linked to the event in the household, element weighting, as defined by Sirken and Royston [4], is obtained.

Response errors may occur both due to under-reporting or over-reporting of events and to errors in reporting the values of $Z_{\alpha,i}$ and of Z_α. For the present it is assumed that there is no over-reporting of events. This assumption is made to simplify the expressions and will be later relaxed.

In addition to the assumptions and notations of [1], it will be assumed that the weighting variables, $Z_{\alpha,i}$, are observed without error, if $d_i(S)=1$, i.e. the reporting unit itself reports correctly on its own value. However, response errors may occur in the reporting of Z_α. Let $Z_\alpha(i,t)$ be the value reported for Z_α by the i-th reporting unit at trial t. Unbiasedness will be assumed, so that:

$$(3.2) \qquad E_t[Z_\alpha(i,t)] = Z_\alpha \quad (i=1,\dots,L;\ \alpha=1,\dots,N)$$

The relative response variance of $Z_\alpha(i,t)$ is assumed to be independent of α and of i, so that:

$$(3.3) \qquad Var_t[Z_\alpha(i,t)]/Z_\alpha^2 = V^2$$
$$(i=1,\dots,L\ \alpha=1,\dots,N).$$

The sample estimate of N, at trial t, for sample S, is:

$$(3.4) \qquad \hat{N}(S,t) = \frac{L}{\ell}\sum_{i=1}^{L} d_i(S)\sum_{\alpha=1}^{N} w_{\alpha,i}(t)\delta_{\alpha,i},$$

where:

$$(3.5) \qquad w_{\alpha,i}(t) = \frac{Z_{\alpha,i}}{Z_\alpha(i,t)}.$$

If we denote the average of the weights, $w_{\alpha,i}$, for reports of relationship r, by:

$$(3.6) \qquad A_r = \frac{1}{N}\sum_{\alpha=1}^{N}\sum_{i=1}^{L} w_{\alpha,i}\delta_{\alpha,i,r}, \quad (r\epsilon C)$$

so that $\sum_{r\epsilon C} A_r = 1$, then, similarly to the development in (1) the bias of the estimate can be approximated by:

$$(3.7) \qquad B = N_o - N \doteq -N\sum_{r\epsilon C}\{1-(1+V^2)P_r\}A_r.$$

Neglecting the correlated response variance, the (simple) response variance, RV, can be approximated by:

$$(3.8) \qquad RV \doteq \frac{L}{\ell}N\sum_{r\epsilon C}[V^2 P_r B_r + (1+V^2)P_r(1-P_r)A_r]$$

where:

$$(3.9) \qquad B_r = \frac{1}{N}\sum_{\alpha=1}^{N}\sum_{i=1}^{L} w_{\alpha,i}^2\delta_{\alpha,i,r}$$

is the average of the squared weights, $w_{\alpha,i}^2$, for reports of relationship r. Similarly the sampling variance can be approximated by:

$$(3.10) \qquad SV = \frac{L-\ell}{\ell}[N(1+V^2)^2\{\sum_{r\epsilon C} P_r^2 B_r + \sum_{r,r'\epsilon C}\sum_{\alpha,\alpha'=1}^{N}\sum_{i=1}^{L} w_{\alpha,i}w_{\alpha',i}\delta_{\alpha,i,r}\delta_{\alpha',i,r}\} - N_o^2/L].$$

The above expressions for the components of the mean square error, (3.7), (3.8) and (3.10), are exactly the same, for the case of no over-reporting, as the expressions (3.2), (3.8) and (3.9) in [1], with $w_{\alpha,i}$ replacing $1/s_\alpha$. The above expressions are, in fact, a generalization of the case of unit weighting (i.e. $w_{\alpha,i}=1/s_\alpha$). They can thus be easily extended to the case where M additional "non-events" are liable to be reported in exactly the same way as in [1], with $1/s_\alpha$ replaced by $w_{\alpha,i}$.

4. Evaluation of the Components

The values of the parameters required to evaluate the components can be estimated from an evaluation survey, as specified in [1], or as proposed by Sirken and Royston [4]. However, the values of A_r and of B_r which are, of prime importance in the components of error and contribute importantly to the differences between counting rules and weighting methods, can, in some cases be obtained from approximations to the distributions of components of the weights $w_{\alpha,i}$ by simple single-parameter distributions.

For example, the empirical results of a multiplicity study on births in Israel, described fully in Nathan, Schmelz and Kenvin [2], show that certain household multiplicities were distributed approximately as Poisson distributions.

Thus, for the counting rule for which births are reported by the woman giving birth, by her sisters and by her mother (in that order of precedence, i.e. mothers only report in the absence of sisters in the household), the distributions of the number of households of sisters for networks with and without mothers (not residing with daughters) are given in Table 1. The fit to the theoretical Poisson distributions is good and the difference between the empirical and the theoretical distributions is not significant. The values of the components of relative error resulting from the use of this approximation, for various combinations of the parameters are given in Table 2, for unit weighting. A similar approximation for element weighting gives the results of Table 3.

References

[1] Nathan, Gad, "An Empirical Study of Response and Sampling Error for Multiplicity Estimates with Different Counting Rules," Journal of the American Statistical Association, Vol. 71 (1976), No. 357 (to appear).

[2] Nathan, Gad, Usiel O. Schmelz, and Jay Kenvin, "Multiplicity Study of Marriages and Births in Israel," National Center for Health Statistics, Series 2, No. 70, Washington, D.C., 1976 (in press).

[3] Sirken, Monroe, "Household Surveys with Multiplicity, Journal of the American Statistical Association," Vol. 65 (1970), pp. 257-266.

[4] Sirken, Monroe and Patricia N. Royston, "Design Effects in Retrospective Mortality Surveys." Proceedings Social Statistics Section, American Statistical Association, 1976.

Table 1: Empirical and theoretical distributions of household multiplicities (reports on births by women giving birth*)

Number of households of sisters $s_{\alpha,3}$	Separate household of mother $s_{\alpha,2} = 0$ (1) Empirical	Theoretical	$s_{\alpha,2} = 1$ (2) Empirical	Theoretical
Total	164	164	164	164
0	41	37.04	52	47.85
1	54	55.11	55	58.95
2	38	41.00	34	36.30
3	18	20.33	16	14.90
4+	13	10.52	7	6.00
Mean	1.488		1.232	
χ^2 (goodness of fit)	1.517		1.017	

*Source: Multiplicity study of marriages and births in Israel.

(1) Mother's mother resides in household with daughter or deceased.

(2) Mother's mother resides in separate household (without daughters).

Table 2: Components of mean square error for various values of basic model parameters - unit weighting.

Parameter/Component	Multiplicity rule alternatives							Conventional rule
Values of basic model parameters								
<u>Mean number of sisters' households</u>								
per network <u>without</u> household of mother - λ_o	1.50	1.80	1.50	1.80	1.50	1.50	1.50	---
per network <u>with</u> household of mother - λ_1	1.25	1.25	1.50	1.50	1.25	1.25	1.25	---
<u>Proportion of networks without household of mother - R</u>	.50	.50	.50	.50	.60	.50	.50	---
<u>Under-reporting probabilities</u> - $1-P_r$								
Event household (r=1)	.02	.02	.02	.02	.02	.02	.02	.02
Household of mother (r=2)	.09	.09	.09	.09	.09	.09	.09	---
Households of sisters (r=3)	.19	.19	.19	.19	.19	.15	.19	---
<u>Over-reporting rates</u> - $Q_r^{(1)}$								
Event household (r=1)	.02	.02	.02	.02	.02	.02	.02	.02
Household of mother (r=2)	.18	.18	.18	.18	.18	.18	.18	---
Households of sisters (r=3)	.20	.20	.20	.20	.20	.20	.15	---
Components of relative standard error (percentages)								
Total root mean square error - $\sqrt{MSE}/N$	4.12	4.02	4.06	3.96	4.10	5.00	3.93	5.37
Bias - B/N	1.14	1.06	1.00	0.92	0.83	3.07	0.49	0.04
Response standard error - $\sqrt{RV}/N$	2.39	2.44	2.41	2.46	2.37	2.34	2.29	0.96
Sampling standard error - $\sqrt{SV}/N$	3.16	3.02	3.10	2.96	3.24	3.19	3.15	5.29

Table 3: Components of mean square error for various values of basic model parameters - element weighting.

Parameter/ Component								
	Values of basic model parameters							
Mean number of reporting persons:								
in event household and in households of sisters (excluding woman giving birth) in networks without mother	3.40	4.00	3.40	4.00	3.40	3.40	3.40	3.40
in households of sisters in networks with mother	1.25	1.25	1.50	1.50	1.25	1.25	1.25	1.25
Proportion of networks without household of mother	.50	.50	.50	.50	.60	.50	.50	.50
Proportion of sisters in separate households	.90	.90	.90	.90	.90	.95	.90	.90
Under-reporting probabilities								
Event household	.02	.02	.02	.02	.02	.02	.02	.02
Household of mother	.09	.09	.09	.09	.09	.09	.09	.09
Households of sisters	.19	.19	.19	.19	.19	.19	.15	.19
Over-reporting rates								
Event household	.02	.02	.02	.02	.02	.02	.02	.02
Household of mother	.18	.18	.18	.18	.18	.18	.18	.18
Households of sisters	.20	.20	.20	.20	.20	.20	.20	.15
	Components of relative standard error (percentages)							
Total root mean square error - $\sqrt{MSE}/N$	4.17	4.09	4.05	3.65	4.02	4.12	5.63	3.63
Bias - B/N	1.91	1.86	1.72	0.72	1.56	1.86	4.26	0.07
Response standard error - $\sqrt{RV}/N$	2.56	2.59	2.58	2.58	2.57	2.56	2.50	2.45
Sampling standard error - $\sqrt{SV}/N$	2.68	2.57	2.61	2.48	2.67	2.62	2.71	2.68

A STATISTICAL MICRO-DATA SOURCE ON AFDC RECIPIENTS

Howard Oberheu and Mitsuo Ono
Social and Rehabilitation Service

Introduction

The National Center for Social Statistics (NCSS) of the Social and Rehabilitation Service (SRS) and its predecessor organizations have conducted detailed studies on the aid to families with dependent children (AFDC) program on an intermittent basis since 1948 covering demographic, social, and program characteristics and financial circumstances of recipients. AFDC, the nation's major Federal-State cash assistance welfare program, provides public assistance to needy families with children who are deprived of parental support and care, including those with fathers who are unemployed. (Reference 1) Since its inception in 1936, the AFDC family caseload has increased from 162 thousand in December 1937 to 3.6 million families in March 1976.

This paper provides key information on the only National source of AFDC "micro" data and observations on the 1975 AFDC study. Study results have been used as basic information on AFDC recipients, benchmarks to evaluate household survey data, and as policy oriented research material (References 2, 3, and 4). AFDC studies are conducted primarily for program administration, e.g., to analyze how long AFDC assistance is received and reasons for eligibility. The 1973 and 1975 studies have provided basic data on child support enforcement and the provision of services such as family planning. State agencies use data from these studies to make inter-State comparison. Also, many research agencies have used the micro-data computer tape files for modeling purposes.

Historical Perspective

AFDC studies were conducted with one or more study months in 1948 (June), 1953 (November), 1956 (January, February, March, or April), 1958 (October, November, or December), 1961 (December), 1967 (November or December), 1969 (May),1971 (January), and 1973 (January). The 1975 study has a May study month. Chart I shows the timing of these studies relative to the AFDC recipient caseload. Although studies have become known as a biennial series because of their frequency since 1967, there is no mandate that the studies be conducted at any specified time. A study is conducted based on need and availability of resources, not necessarily following a period of caseload growth.

Over time, the methodology and content of the studies have changed and expanded as the program changed and data needs increased. These changes, of course, do impart on the comparability of data. A list of items included in the various studies since 1967 is available on request. Most of the data elements have been included in more than one study while many items have been included in all studies, e.g., basis for eligibility, race, number in assistance group, and amount of grant. The primary focus of these studies was the AFDC family.

With respect to program changes, the recent separation of services and payments necessitates the use of two separate survey schedules for each sample case. New items such as those on child support have been recently incorporated. In addition, it has been possible to obtain information on all fathers associated with a case rather than identifying the status of only one father; resulting from more flexibility allowed by computer processing.

Another important difference, especially to the researchers, is that information prior to the 1967 study was reported by State agencies in summary reports, in effect providing only macro-data on recipient characteristics. For studies conducted in 1967 and thereafter, States were required to submit a survey schedule for each family in the study sample. This has resulted not only in a report series but also made available public use micro-data files. Of course, all case-specific identifying information such as names and case number have been removed from these public-use files. NCSS has provided copies of these public-use tapes to government agencies and to the National Technical Information Service.

Table I provides the sample size utilized in these studies. The 1967 study is the most recent that provided published estimates for all State AFDC caseloads. In 1973, States were given the option to supplement the minimum sample required to yield publishable State data. Only for the largest States were the sample sizes sufficient for data to be published for all years.

For the 1969 and 1971 studies, a sample size of 350 or more cases was considered sufficient to separately publish State data. With a confidence level of 95 percent, the sample would measure a difference greater than one percent (i.e., 49.5 & 50.5 percent). For State data published separately for the 1973 study, the sample was designed to measure that a characteristic of 0.5 percent was significantly different from zero (at 95% confidence level). For more details, see reference 2.

The 1975 AFDC Study

The 1975 AFDC Study is currently being completed. Due to State reporting delays and additional processing time needed, a completed computer file should be available at the end of September 1976.

As with other NCSS reports, this study picks up available administrative records in the State/local welfare agencies as the information source. Hence, information to be collected in the study is limited by the information contained in the case record. This constraint has become more acute with the introduction of the simplified method of eligibility determination which has limited available information only to those data items required for such a determination or retained as an historical caseload management record.

The sample frame for the 1975 AFDC study was similar to that utilized in prior studies. A minimum of 0.5 percent was required (allowing for six States to be published separately) with the option of supplementing the required sample to allow for reliable State estimates (exercised by 24 States).

As in 1973, the sample size was such that a characteristic of 0.5 percent was significantly different from zero (at 95% confidence level). Each State that chose the optional size sample computed the exact sample size for the study month using the formula:

$$n = \frac{N}{1+00116879N}$$

where n= number of families in the sample, and
N= number of families in the caseload in in the study month

This sample size is designed to measure a characteristic for each State using the same precision as the national sample.

Two innovations were introduced into the sample frame for the 1975 study. The first provided longitudinal information on AFDC families. Sample cases of the January 1973 study that received a payment in May 1975 were incorporated in the 1975 study sample. These cases which received a payment in January 1973 and May 1975 represent the universe of cases that received AFDC at two points in time spanning a 29 month period. These same cases which have been receiving AFDC over a period of several years are included in the regular 1975 AFDC Study sample along with cases that started receiving assistance after January 1973. Thus, the 1975 sample consists of "old" cases, which comprise the longitudinal segment, and "new" cases. Because the sample is drawn by the States, no allowance is made for interstate transfers of cases, i.e., an approved AFDC case discontinued because of a move to another State would not be included in the longitudinal segment.

The second innovation was to include AFDC-Quality Control (QC) sample cases for May 1975 in order to check reported data. After completion of the study, comparisons will be made between QC review data and study data to evaluate the accuracy of the reporting. Items which can be compared are as follows:

- Time of most recent opening
- Most recent action is approved application or redetermination
- Number of persons in assistance group (total,children, adults) and in household (total, others not in group)
- Deprivation factor
- WIN program registrants and participants
- Current employment status of caretaker relative and spouse included in assistance group
- Presence of nonassistance income and/or resources, by type
- Amount of AFDC payment

Although the primary effort for developing and conducting the study is with NCSS, considerable guidance is provided by State public welfare agencies because of the Federal-State relationship. In addition to State comments, study schedules were pretested in the Arlington, Virginia, Department of Human Resources. Also recommendations on content were provided by Federal agencies, e.g., OMB.

Preliminary Observations from the 1975 AFDC Study

Some observations are gleaned from preliminary edit reviews on the collection of data for several items relating to time on assistance and number of persons in the household. Additional data on prior openings for AFDC have likely resulted in more accurate reporting on date of most recent opening. For the AFDC study, caseworkers were instructed that payment lapses of three months or less were not to be counted as discontinuances and subsequent reopenings. For a number of cases, the additional data indicated the instruction was not properly followed but it was possible for corrections to be made.

In regard to number of persons in the household, prior studies reported numbers of persons in the AFDC assistance group and of other persons in the household. In addition, the 1975 study also obtained a separate listing of all adults and children in the household. Here the results are more obscure in regard to improved accuracy of the detailed listing versus the summary numbers in those few instances where the data were inconsistent. In some instances, some individuals were not listed while to a much lesser extent some individuals had not been included in the household summary but were listed independently. A concern of any such study is possibly that those individuals a caseworker was reluctant to list as a household member may have been excluded from both the household summary and the detailed listing. Such reluctance may carry over from the man-in-the house era or as a result of the present emphasis on an individuals privacy. In instances where inconsistencies occurred, State agencies were requested to provide the corrected data.

Some areas of concern regarding these studies are as follows. The biennial frequency of the study causes the data to be outdated for some purposes after a period of time prior to availability of results from a subsequent study. Conducting the studies during one month precludes analyzing seasonal variation of the data. Since the study universe includes only AFDC families currently receiving assistance, data on cases discontinued and those eligible and not receiving assistance are limited. The use of administrative records prohibits questions not directly related to the welfare process such as information on non-recipient household members or opinion type response of recipients. Hence, need exists to institute supplementary household surveys covering target populations in order to obtain required data.

Selected General Findings

As an example of information obtained from the studies, the following is provided:

- In line with the general population, AFDC families are more likely than not to live in metropolitan areas.
- The proportion of families that are black has been increasing.
- The size of the AFDC family has been decreasing and typically includes only a mother and two children at the present.
- Mothers that are not employed have a greater than even chance of having children under age six; fathers are most likely to be incapacitated for employment when not working.
- The proportion of families with children whose father was not married to the mother has increased while that for families with orphans has decreased.

Future Efforts

Closely associated with the characteristics studies is a major effort in SRS to develop a statistical data base of recipients of the various social welfare programs. Briefly, such a system, known as the Recipient Statistical Reporting System, will identify individuals and families in one or more programs, e.g., AFDC, food distribution, social services, Medicaid, etc., resulting in a micro-data file containing demographic and program characteristics. Such data can be used in the analysis of separate programs or of the receipt of multiple benefits.

The next characteristics study will be conducted in the spring of 1977. We appreciate the continued support of researchers and other interested data users in the development of these studies.

References:

1. Studies in Public Welfare, Paper No. 20, Handbook of Public Income Transfer Programs: 1975 Joint Committee Print, December 31, 1974.

2. Findings of the 1973 AFDC Study:

    Part I Demographic and Program Characteristics (June 1974)

    Part II-A Financial Circumstances (September 1974)

    Part II-B Financial Circumstances (September 1975)

    Part III Services to Families (October 1974)

    Part IV Discontinuances for AFDC Money Payments During 1973 (January 1975)

    Department of Health, Education, and Welfare, National Center for Social Statistics

3. Income Transfers As A Public Good: "An Application to AFDC" by Larry Orr, The American Economic Review, June 1976

4. Child Support, Data and Material, Prepared for Senate Committee on Finance, U.S. Government Printing Office, Washington,D.C. November 10, 1975

CHART I

FREQUENCY OF AFDC STUDIES RELATIVE TO THE NUMBER OF PUBLIC ASSISTANCE RECIPIENTS OF AFDC MONEY PAYMENTS, JUNE AND DECEMBER OF EACH YEAR, 1936 TO DATE

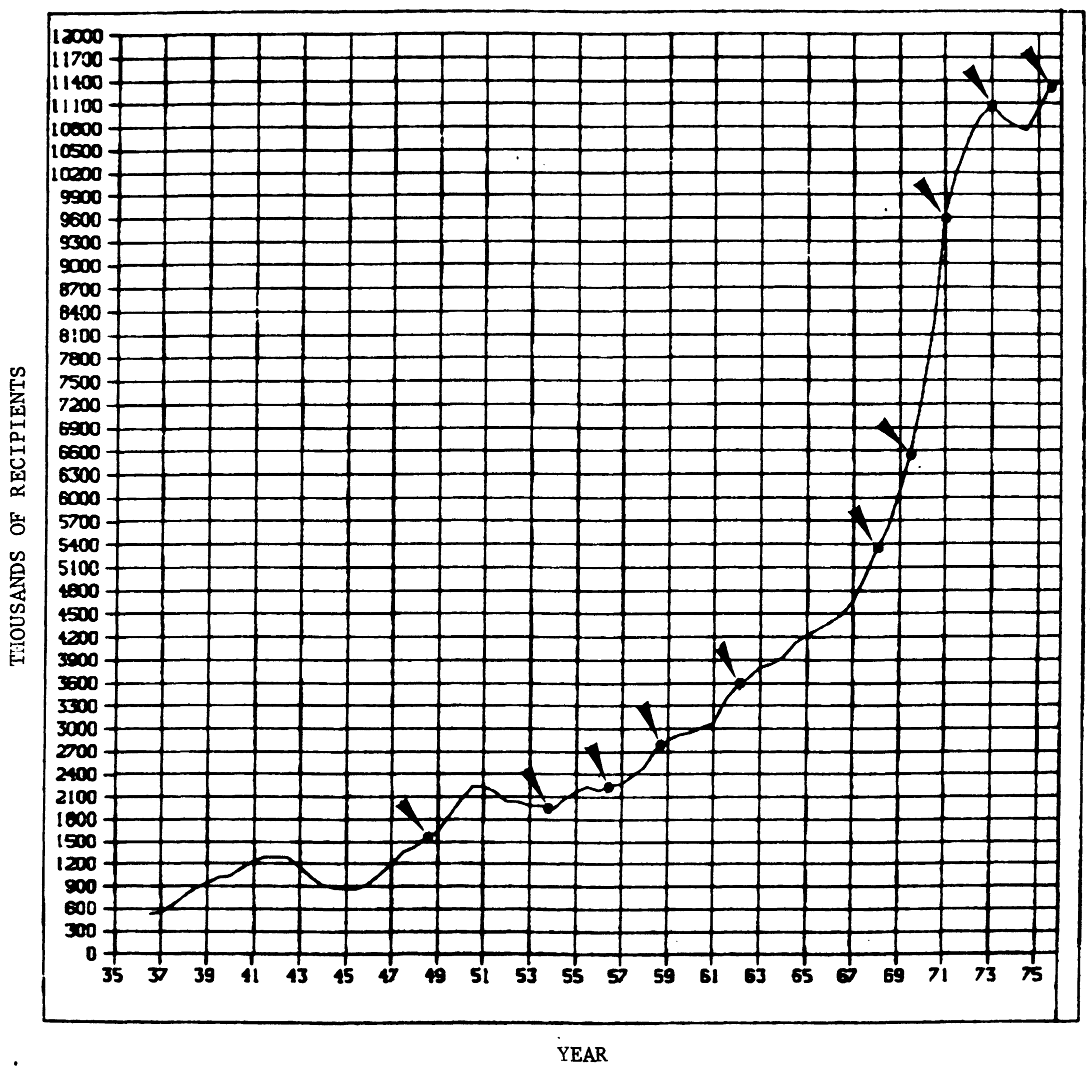


TABLE I

AFDC Characteristics Studies

Study	Number of AFDC Families	Sample: Minimum requirement	Sample: Number	Number of States Published
1948 (June)	449,202	**	**	50
1953 (Nov.)	541,353	500**	**	52
1956 (Jan.-Apr.)	610,070*	**	**	52
1958 (Oct.-Dec.)	746,401*	**	**	51
1961 (Dec.)	915,559	.01 or 500	**	51
1967 (Nov.-Dec.)	1,285,040*	.03 or 500	67,000	53
1969 (May)	1,652,306	.01	16,000	13
1971 (Jan.)	2,558,094	.01	25,000	21
1973 (Jan.)	3,142,860	.005	35,000	33
1975 (May)	3,481,363	.005	32,000	29
1976 (Jan.)	3,573,038			

* Average for the study months.

** Limited information is available on sample size for early studies. For example, the only reference for the 1953 study reads: "Reports for most States were based on samples of the caseload including a minimum of 500 families."

SOME PRELIMINARY RESULTS FROM A VALIDATION STUDY OF THE ESTATE MULTIPLIER PROCEDURE

H. Lock Oh and Fritz Scheuren*
Social Security Administration

The estate multiplier procedure has long been employed by economists and statisticians to estimate personal wealth [e.g.,1]. While there is a widespread recognition of the fact that the technique has potentially serious flaws, repeated calls for a vigorous evaluation of the method [e.g.,2] have generally been unavailing. A number of projects, however, are now underway which offer some promise in this regard [3,4]. The present paper is a report on one such effort that was recently undertaken using data from Social Security's Continuous Work History Sample.

ESTATE MULTIPLIER PROCEDURE

Before discussing this validation study and the results that were obtained, it might be useful to review some of the salient features of the multiplier procedure. The estate multiplier technique employs information from the records of decedents (usually financial records) to describe the characteristics of living individuals. To construct estimates, for each decedent, the data are weighted by the inverse of the mortality rate thought to be appropriate for the demographic category to which the individual belonged.

Illustration.--Consider the estimation of average house value for white males, 35 to 44 years of age, living in the State of Massachusetts in 1973. Assume we have at hand data on the number of 1973 decedents in this category, their house values, as obtained from some source (such as probate records) and their respective mortality rates. The first three rows of figure 1 provide some hypothetical data on these decedents. In particular, for the 35 to 39 year age group, there are 80 decedents, and their average house value is $40,000. For the 40 to 44 year age group, there are 100 decedents who had total house values of $4.5 million, or $45,000, on the average.

Figure 1.--Illustration of the Estate Multiplier Calculations

Item	Age (in years)		
	35-39	40-44	35-44
DECEDENT DATA			
Decedent Sample Size....	80	100	180
Total House Value (in millions of dollars)....	3.2	4.5	7.7
Average House Value (in dollars)...........	40,000	45,000	42,728
MORTALITY EXPERIENCE			
Mortality Rate..........	2/1000	4/1000	-
Estate Multiplier.......	500	250	-
ESTIMATES USING ESTATE MULTIPLIER			
Total Population Count..	40,000 (80x500)	25,000 (100x250)	65,000
Estimated Total House Value (in millions of dollars)..............	1,600	1,125	2,725
Estimated Average House Value (in dollars).....	40,000	45,000	41,923

The next two rows deal with mortality experience. By definition, the mortality rate for a given demographic category is the fraction of the persons in that category who died. If one assumes that the occurrence of death for any individual in such a group is equal to that for any other member, then, the mortality rate is equivalent to the sampling fraction in a stratum of a stratified survey design. This analogy to probability sampling is generally used [5,6] to "justify" the estimation procedure employed in the estate multiplier technique. In any case, given the mortality experience shown in figure 1, the estate multiplier procedure would weight the information for the 35 to 39 year-olds twice as heavily as for those 40 to 44. The balance of the illustration shows how this is done.

Advantages and Disadvantages of Method.--There are several advantages that the multiplier enjoys which make it an attractive alternative to probability samples employing household survey techniques. First, the information is generally obtained from records that have been very carefully prepared. Second, if the records are for estate tax or probate purposes, legal sanctions exist which would tend to further reduce misreporting problems. Coverage errors and errors arising from nonresponse are also lessened considerably because of the routine compliance procedures associated with the administration of the law. Questions of ownership and valuation do arise, of course; but, on the whole, at least in the case of the U.S. Federal estate tax returns, content errors are believed to be quite small [7].

The main disadvantage of the technique is that the estimates are not based on probability samples. In particular, the "randomization" is not under the control of the analyst, and, hence, he must guess about its nature. Thus, a subjective element is introduced which might have a crucial impact on the results.

As usually carried out, the multiplier technique stands, or falls, depending on whether or not the following assumptions hold:

1. the characteristics to be measured have not been distorted by the sampling process (i.e., by the occurrence of death);
2. the average mortality experience of the population about which inferences are being made has been adequately accounted for in the rates being used; and, finally,
3. the extent to which an individual's probability of death differs from the average for his demographic group is not related to the information one wishes to estimate.

Of course, it has generally not been possible to adequately check these assumptions. Even in the validation study we are about to describe, we have only been able to isolate the net effect of failures in these assumptions.

DESCRIPTION OF VALIDATION STUDY

One way to validate the multiplier procedure is to compare it with another measurement technique in which we have more (or complete) faith. This turns out to be a very formidable problem. In the context in which the multiplier procedures are usually conducted, only very limited success has been achieved so far.

It is possible, however, to examine the multiplier method in situations where it would not normally be used because other estimators are available. While not quite in the needed context, such situations do afford us a test of the procedure. Of course, the nature of the situation one studies naturally limits one's inferences about the validity of the procedure, but this should (and did) not deter us.

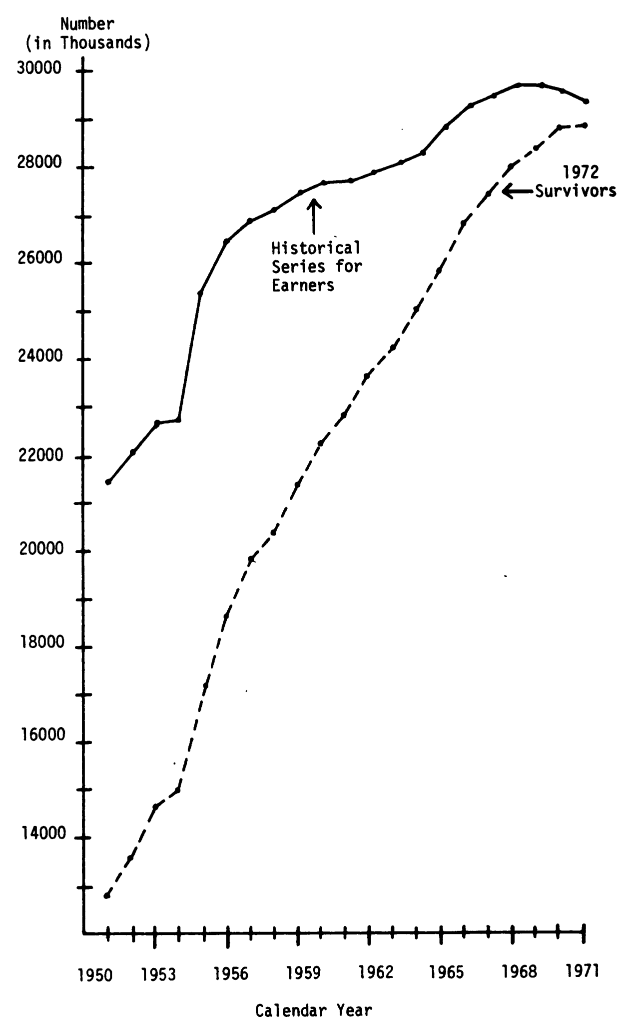
Use of Social Security (SSA) Records for Validation Purposes.--SSA maintains longitudinal records on each worker's social security covered taxable earnings up to the taxable maximum for any given year. Annual summaries from these administrative files are published in the Statistical Supplement to the Social Security Bulletin [8].

The basic study design called for a comparison between this overall historical information for the period 1951-1971 and estimates obtained by tabulating an appropriately weighted sample of persons in the Social Security system who were identified as having died in 1972. 1/ If the multiplier procedure were valid, it should be possible to estimate the earnings distribution in some prior year (say 1960), by employing the earnings histories obtained from the longitudinal files for 1972 decedents. However, a difficulty arises which must be faced. It has to do with the fact that the 1972 sample could only be expected to estimate the earnings distributions of individuals who survived to 1972. In particular, individuals who died prior to 1972, but who had earnings in, say, 1960, cannot be estimated using just the 1972 decedents.

Survivor Estimates.--Now, if all deaths for all persons in social security covered employment were always reported to SSA, the published historical series could be adjusted directly, so that only earners who had survived to 1972 would be included. Unfortunately, there is no necessity for all deaths to be reported. In many cases, there is an economic incentive to notify Social Security (if the individual has worked long enough to be eligible for a lump sum death benefit), but, even then, the death is not always reported.

There were two consequences of this: one is that, although the reporting has improved over the years, the 1972 decedent sample's coverage was such that we could only look at a rather restricted universe--males 35 years or older in 1972. Even with this restriction, there was still some coverage error, with only about 95 percent of all 1972 deaths among males aged 35 or more being

Figure 2. --Number of Male SSA Earners, 35 Years or Older: Comparison Between Historical Series and 1972 Survivors, 1951-1971



reported to SSA. Five year age-race specific coverage adjustments could be made for 1972 decedents;2/ however, it was not possible to use a direct method of calculating 1972 survivors, since the coverage error not only was greater in earlier years, but it is cumulative.

To "get around" this problem, we had to synthetically derive cohort survival rates by single year of SSA age (race and sex) for the period 1951 through 1971. The details of how these survival rates were calculated are provided in [9], which is available on request.

Figure 2 compares the 1972 survivor counts obtained by this process to the corresponding distribution of the number of persons in SSA's historical series for the years 1951 through 1971. There are no surprises here. For example, of the 21.3 million male earners 35 or older in 1951, the chart shows that 12.7 million, or only about 60 percent, survived to 1972. On the other hand, as might also be expected, the fraction of 1969 earners who survived to 1972 was over 96 percent.

The 1972 survivor totals are, of course, subject to various errors. On the whole, though, they probably afford a reasonable standard for the initial check of the multiplier procedure given in the next section.

INITIAL RESULTS OF STUDY

For the comparisons we will be making in this paper, the demographic categories in which the multiplier estimates (and the survivor figures) were calculated include only age, race, and sex. No adjustment for differentials by earnings class (or, indeed, between earners and nonearners) has been made. This limitation should be kept in mind as we look at figures 3 and 4 below.

Overall 1972 Survivor Series and Multiplier Estimates.--Figure 3 compares the number of survivors shown in the previous chart to multiplier estimates based on decedents' records for 1972.

Notice that the multiplier estimates are generally quite similar to the survivor totals. For the

Figure 3. --Number of Male SSA Earners, 35 Years or Older: Comparisons Between 1972 Survivors and Multiplier Estimates, 1951-1971

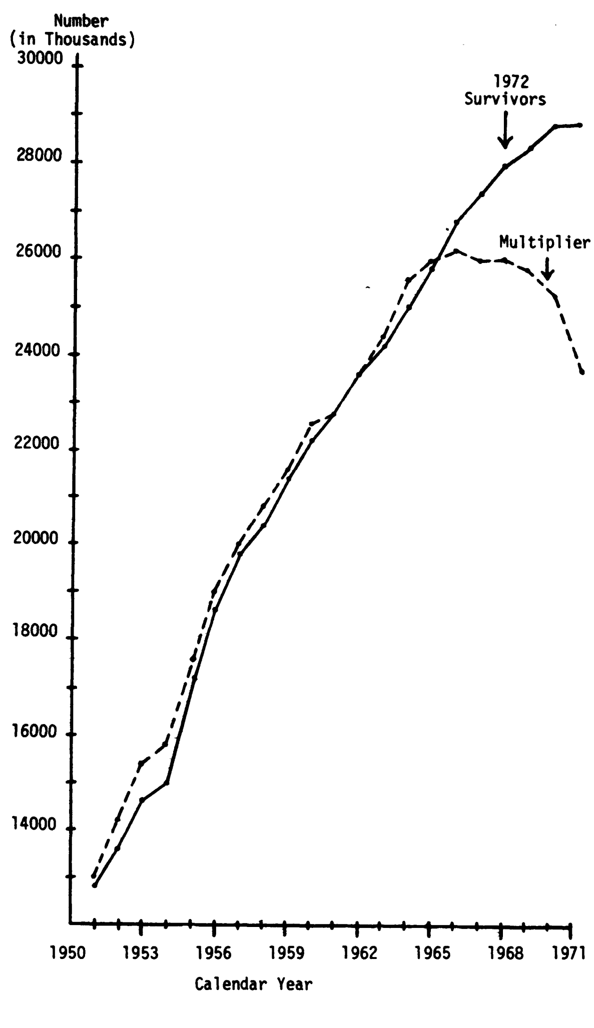
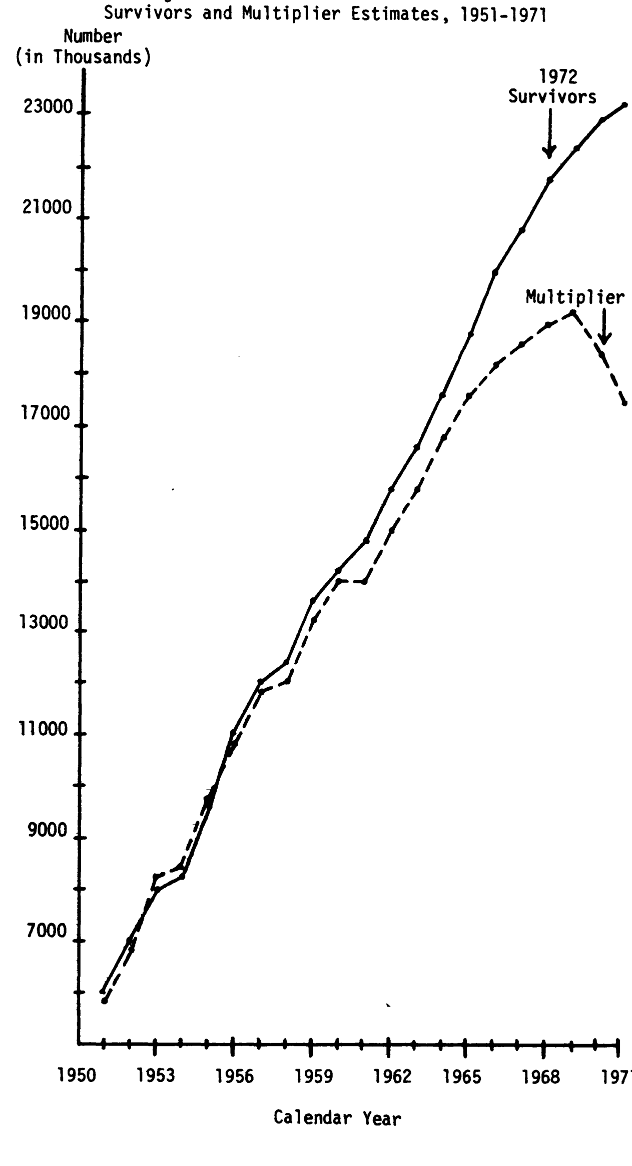


Figure 4. --Number of Male SSA Earners, 35 Years or Older with Earnings of $3600 or More: Comparison Between 1972 Survivors and Multiplier Estimates, 1951-1971



period before 1966, the two graphs track each other very closely. In fact, for each year, they are within sampling error.3/ However, the multiplier procedure fails to provide reasonable estimates after 1967. Apparently, as early as six years before death, the onset of health problems causes a decrease in earnings. We expected this behavior for 1971 and, to a lesser extent, for 1970. It was somewhat surprising to us, though, to see that the phenomenon begins so long before death and that its effect is so marked.

1972 Survivor Series and Multiplier Estimates for "High" Earners.--Figure 4 examines the multiplier estimates for persons with earnings of $3,600 or more.4/ Here, the multiplier appears to do well only until about 1960, while in figure 3, the estimates were in rough agreement as late as 1966. We speculate that not only are health problems a factor, but that there may also be a favorable differential at work in the mortality of these "high" earners relative to all social security account number holders. In this connection, it should be pointed out that, if there were such a differential, then, the survivor totals being used as a standard would be too low; hence, the gap shown in figure 4 would be an understatement.

FURTHER RESULTS AND SOME CONCLUSIONS

While this presentation has probably not emphasized it sufficiently, there are a number of "methods" issues imbedded in the approach we have taken. It is for this reason that we have labelled the findings as preliminary. Nonetheless, we do think some generalization may be warranted. Clearly, the estate multiplier procedure cannot be said to have been "validated" by the tests offered in this paper. However, proponents of the technique should not despair. For one thing, the impact of "health problems" would probably not be as severe for wealth variables as they are for earnings. On the other hand, the potential effect of differentials in mortality rates could be much greater in estimating wealth.

This paper has only touched on some of the overall results in the complete study. An extensive appendix is available to anyone interested in pursuing the matter further.

AN AFTER WORD

When we prepared this paper for the Proceedings, it struck us that we had not adequately stressed that, in our opinion, the multiplier technique has been overused. In particular, we feel the technique is unsafe to employ alone or without introducing external (and internal) checks. (See [4].) It is one thing to examine the method as an intellectual curiosity (as was done in this paper). It is entirely another matter to rely on it in situations where important decisions have to be made.

FOOTNOTES

* The authors would like to thank Wendy Alvey, Faye Aziz, Beth Kilss, and Bob Yuskavage for the computational and other assistance they provided. Helpful editorial comments were given by Keith Gilmour. The text was typed by Catherine Murphy.

1/ The decedent sample used was the 1% Continuous Work History Sample (CWHS) for the period 1937-1972. Information was also tabulated on all earners for each year from 1937 to 1971 using the 1937-71 0.1% CWHS. These tables were needed to augment the published historical data. (A full set of the tabulations will be made available upon request.)

2/ The assumption was made that, for a given race-age group, the earnings histories of covered and uncovered 1972 decedents were roughly the same. This assumption is undoubtedly false, but, because the uncovered group is so small, the differences which exist are not believed to have materially affected the overall outcomes of the present test. A study matching death certificates to social security records is planned to examine the characteristics of the uncovered group.

3/ For the "survivor" totals, the variance was fairly easy to obtain, since, at least at the national level, for the statistics examined, the CWHS (upon which the figures were based) can be treated as roughly a simple random sample. (See [10], table 1.) To estimate variances for the multiplier totals, we had to resort to a super-population model. Under this model we assumed that:

1. each calendar month's deaths were drawn independently of every other month's deaths;

2. the monthly samples can be divided into six strata--

A. January and December	D. May and October
B. February and March	E. June and July
C. April and November	F. August and September; and

3. within each of these 6 strata, the probability of an individual's death during a month is equal to some constant which is the same for both months. [12]

One advantage of this formulation is that it does not introduce any assumption about the nature of the bias in our estimator. It also takes some account of the seasonality which exists in the death rates. The approach is a "good bit" better than simply assuming that we are engaged in stratified sampling where the strata are race-age groups. However, it must be admitted, we are somewhat dissatisfied with this method, and, in the future work, we expect to be using decedent samples based on more than one year's deaths, so as to explore what might be more appropriate models.

In any case, standard errors can be calculated according to the above model (using the pseudo-replicate balanced half-sample procedures of McCarthy [11]). While the formulation gives us six degrees of freedom, we could only afford to use one pair of "replicates" (the cost per pseudo-replicate was about $750); therefore, the standard errors had to be based on an estimator with only one degree of freedom.

4/ The social security taxable maximum was $3,600 in 1951; by 1971, however, the maximum had risen to $7,800.

REFERENCES

[1] Atkinson, A. "The Distribution of Wealth in Britain in the 1960's--the Estate Duty Method Reexamined," Studies in Income and Wealth, National Bureau of Economic Research, no.39, 1975, pp. 277-319.

[2] Langley, K. "Discussion: Wealth in Britain--the Estate Duty Method Reexamined," Studies in Income and Wealth, National Bureau of Economic Research, no. 39, 1975, pp. 321-328.

[3] Smith, J. and Franklin, S. "Expost Estimations of Weights in a Nonprobability Sample," paper presented at the May 5, 1976, Washington Statistical Society Methodology Session on Validating the Estate Multiplier Technique.

[4] Scheuren, F. "Historical Perspectives on IRS Wealth Estimates with Prospectives for Improvements," paper presented at the May 2, 1975, National Bureau of Economic Research Workshop on Estate Multiplier Estimates.

[5] Internal Revenue Service. Statistics of Income--1972, Personal Wealth Estimated from Estate Tax Returns, Washington, D.C., U.S. Government Printing Office, 1976.

[6] Natrella, V. "Wealth of Top Wealthholders--1972," 1975 American Statistical Association Proceedings, Business and Economic Statistics Section.

[7] Smith, J. and Calvert, S. "Estimating the Wealth of Top Wealth-holders from Estate Tax Returns," 1965 American Statistical Association Proceedings, Business and Economics Statistics Section.

[8] 1974 Social Security Bulletin Annual Statistical Supplement, pp. 69-77.

[9] Scheuren, F. and Oh, H.L. "Some Preliminary Synthetic SSA Cohort Survival Rates, 1950-1971," Unpublished SSA working paper, April 1976.

[10] Finch, R. "Examples of Sampling Variability in the Continuous Work History Sample," paper presented at the May 13, 1976, CWHS Users Conference.

[11] McCarthy, P. "Replication: an Approach to the Analysis of Data from Complex Surveys," Vital and Health Statistics, National Center for Health Statistics, series 2, nos. 14 (1966) and 31 (1969).

[12] Rosenwaike, I. Seasonal Variation of Deaths in the United States, 1951-1960, Journal of the American Statistical Association, vol. 61, pp. 706-719.

THE OPTIMIZATION OF EDUCATIONAL VALUES IN NAVY CURRICULUM DESIGN

Ellis B. Page, University of Connecticut, Storrs

Summary

Curriculum theory has been recently shown to be analyzable with the quantitative tools of dynamic programming (DP), yet such applications are practically non-existent. What is required is (a) a list of candidate units or topics possible for inclusion; (b) for each candidate unit, at each considered level of instruction, a cost, usually expressed in time; (c) for each such unit, a value; and (d) some overall maximum acceptable cost. The most difficult problem is the fair and functional establishment of values, by instructors or other judges. Methods of apportionment and scaling were developed for a major technical Navy course, and found to have acceptable reliability. But a general phenomenon of "averaging" in judgment works against the more costly units when the DP algorithm is used, and compensations therefore need to be applied. Several statistical approaches are outlined for overcoming such a bias and making the technique usable in course design.

INTRODUCTION

We may say that a "curriculum" is a selection of materials for instruction, and an allocation to each material of some limited resource, usually time. A curriculum "problem" may be said to exist when when the candidate materials for such instruction exceed the available resource. Then we are forced to choose among the candidate materials, and to arrange the resource allocation such that we do not exceed the time available. Furthermore, if we wish to make the best possible curriculum, we will try to optimize the value of the materials included.

Recent theoretical work (Page, Jarjoura & Konopka, 1976) has treated this curriculum problem as a problem in operations research, and especially as a problem in dynamic programming (DP). We may picture it simply as if it were a college textbook, and we wished to move through the text assigning each chapter to some level of instruction (including the zero level, which would be omission of that chapter). Each level for each chapter has some estimated cost (in terms of hours) and some estimated value (in units to be discussed). And for the book as a whole, we begin with some maximum total cost (in hours assignable). The solution to the problem, then, consists in assigning each textbook chapter to some level of study, such that the total hours consumed do not exceed the maximum allowable, and such that the estimated value is as high as possible.

The theoretical analysis treated this textbook problem as if it were brand-new, and as if the "levels" for each chapter might be chosen (such as "read chapter once lightly") quite independently of the assignment of costs and values. In the applied situation herein studied, there were discovered problems which required adjustments in the approach, which considerably affect the results of the DP algorithm. Most particularly troublesome were certain behaviors in the assignments of value to the units, and in the effects of these behaviors on the computer solutions.

AN APPLIED NAVY ANALYSIS

In practical terms, the course Sonar Electronics Intermediate may be one of the most important courses in the Navy. Taught in San Diego, it had a well-experienced and able instructional staff, familiar with the course, with the fleet requirements, and with the trainees. Some of these instructors served as judges for the values of the larger units within the course, and of the topics within each unit.

The specification of costs

Navy courses are commonly taught, like certain civilian instruction, from a thoroughly worked-out curriculum guide. Such a guide will specify the content of each portion, the problems to be presented and solved, and the time to be allotted to each. In general, the materials of such a guide are explicitly allotted, then, a certain "cost" for our curriculum analysis.

After some experimentation, the most useful way of describing the costs for proposed changes, then, was in terms of the current costs. And certain modifications of that current cost were essayed, as follows: No hours assigned, 40% of current costs, 80%, 100%, 120%, and 240%. The guiding principle was that there would be five possible levels for a unit other than its present level. Two of these would be quite conservative in their change (20% to either side of the present level), and two of these would be more radical (halving or doubling these more conservative changes), and one would represent the possibility of excluding that unit from the Navy course.

The stimulus that resulted from these calculations, then, may be seen in Figure 1.

PROPORTION	0	40	80	100	120	240
HOUR COST	0	35*	70*	87	100*	200*
ESTIMATED VALUE	0			100		

Figure 1. The stimulus presented for scaling value in a Navy course for one unit.

In Figure 1, it is seen that the proportions of cost are scaled from 100%, which is the current program, and which uses 87 hours of instructional time. The hour costs marked by asterisk (*) are scaled to match the percent of 87 hours, as indicated in the top row. (Thus 35* hours is rounded from .40 x 87 = 34.80.) These cells, therefore, become hypothetical costs, amounts of time which might be considered by the judges of the course.

The specification of values

As noted above, the current costs for a Navy course are as prescribed by the established outlines. And the hypothetical costs are arbitrary, scaled for a spread of radical and conservative deviations from the current costs. On the other hand, the specification of values is performed by the expert judges themselves.

There are two acts of judgment performed by the experts in a curriculum. These may be done by the same judges, or by different judges; and they may be done at one sitting, or at two. The acts are:

a) Assigning values to units of the current curriculum. Suppose there are five units under consideration. Then one is given 100 tokens (poker chips, all of same value), and asked to "spend" these tokens in accordance with his beliefs about the values of the units as currently taught. He will spend all 100, so that his registered evaluation will be expressed as v_{ij}, where he is the ith judge, appraising the jth unit, and where

$$\sum_{j=1}^{m} v_{ij} = 100.$$

b) Assigning proportional values to hypothetical levels of cost. In a separate operation, a judge is asked to fill in the four blank cells in Figure 1. He is asked to assign a percent of value to a reduced, or an increased, effort dedicated to the unit in question. The first task, above, compared each unit with each other unit in assigning value. The present task scales the hypothetical value of certain fixed amounts of hypothetical effort. The judge's scaling, then, will be expressed as R_{ijk}, where he is the ith judge, appraising the kth level of the jth unit. These values must be non-decreasing, as k rises from 0 to 6.

RELIABILITIES

Reliabilities of the token strategy

For this pilot research, there were five Navy instructors who served as judges of the values. All were widely experienced in Sonar Electronics, both as instructors of the course and as technicians in the Fleet. Independently of each other, they all assigned token weights to the five units of the course, and to the topics of three of these units. A measure of concordance could then be taken by the correlation ratio,

$$\eta = \sqrt{\frac{SS_{bet.units}}{SS_{total}}}.$$

For the course as a whole, such analysis yielded a correlation ratio of .66. And for Units 1, 4, and 5, the corresponding statistics were .74, .76, and .76. Despite the small number of judges, these agreements were all statistically significant.

Therefore, it may be inferred that, among these instructors, there was some concordance of viewpoint, which renders their mean judgments of some importance.

Reliabilities of the level scaling technique

As noted above, each judge would scale the hypothetical value of each unit for each level (these levels being 40%, 80%, 120%, or 240% of the current expenditure). In order to appraise the concordance of judgment, and to avoid the problems of monotonicity, each level would be considered independently. For the overall level scaling of the course, across five units and five judges, at the 40% level of effort, the data are shown in Table 1.

THE 40% LEVEL FOR ALL UNITS

Judges	Units 1	2	3	4	5
A	80	100	40	80	50
B	50	40	10	10	60
C	40	30	10	5	20
D	90	35	25	10	20
E	25	70	10	30	25

η = .56

From sixteen such tables as Table 1, it was possible to calculate the correlation ratios for all the investigated levels, for the course as a whole, and for three units of that course. The results are as seen in Table 2, and the levels of significance are as indicated.

TABLE 2

CORRELATION RATIOS FOR FOUR LEVELS OF EFFORT FOR THE COURSE AND THREE UNITS

	Levels of Effort 40%	80%	120%	240%
COURSE (5 units)	.56*	.65*	.27	.35
UNIT 1 (20 topics)	.45	.72**	.49	.63**
UNIT 4 (19 topics)	.61**	.54*	.44	.45
UNIT 5 (9 topics)	.43	.62*	.54	.43

* Sig. at .05 level
** Sig. at .01 level

From Table 2, it is apparent that the agreement concerning levels of hypothetical value is enough to work with, when taken as an average. Furthermore, it was observed that instructors had no apparent difficulty, no great hesitation, in assigning these values. Familiarity with the course had given them fairly strong opinions about the units of the course, and about the topics of the units.

OPTIMIZING THE ASSIGNMENT OF HOURS

From such measures of value of the units, and of the hypothetical levels of effort, it is possible to compute an objective function for the dynamic program algorithm. First, each $\underline{v}_{ij}$ and each $\underline{R}_{ijk}$ is averaged across judges. Then the product of these averages becomes the (usually non-linear) objective function for the algorithm.

For example, in Figure 1, scaling percentages for the four empty cells were, from left to right, 57, 93, 107, and 110. The mean value of the unit, taken as a whole, was just .12 of the course's total value. Therefore, a new line to Figure 1, of "adjusted value," may be calculated, and these values will be 0, 6.84, 11.16, 12.00, 12,84, and 13.20. The value is thus seen to be curvilinear, with much of the current value (12.00) achieved with 20% less cost (11.16), and little to be gained by increasing the hourly expenditure for the unit.

The mathematical model is that of a "distribution of effort" problem, in standard works (cf. Wagner, 1970, pp. 254-257). The assignment of each unit or topic to a specific level of effort is made, such that the total expenditure of hours does not exceed some fixed quantity, and such that the sum of values of those assignments is maximized. And Moonan (1976) recently provided a useful algorithm for the solution of such problems.

Comparing the current and optimal assignments

From such dynamic programming, what is of interest is a comparison of the solutions with the values of the current strategy. What is also interesting is the hypothetical values which could be achieved with even <u>less</u> expenditure of hours than is currently given. The result of such comparison is seen in Table 3.

TABLE 3

VALUE ACHIEVED FOR FOUR TEST CASES AT THREE ALLOCATION LEVELS THROUGH DYNAMIC PROGRAMMING

	Allocation level compared w. present		
Case	100%	90%	75%
ALL UNITS	107	99	92
UNIT 1	129	124	118
UNIT 4	115	108	97
UNIT 5	109	103	89

Obviously, the algorithm was highly successful in finding an assignment which would increase the estimated value of the course. For the apportionment of time, at the current level, to the course as a whole, the increases over the current value amount to 7%, and for the three units examined, the gains are 29%, 15%, and 9%. In fact, even with the allotted time reduced, it is possible to increase the estimated value, as is evidenced in the two other columns of data in Table 3. With a 10% reduction in time-cost, it is possible to maintain or surpass the current value. And even with a one-quarter reduction in time-cost, the estimated values are sustained not too much below the current values.

THE NATURE OF THE SOLUTIONS

When the optimal assignment tables are produced (Page and Canfield, 1975, App. F), it is possible to study the relation of the optimal assignment in comparison with the current assignment. We may consider Figure 1 to contain six levels of effort, 0, 1, 2, 3, 4, and 5, corresponding with the columns, with level 3 representing the current solution. A "conservative" solution would be one in which level 3 was frequently the optimal level. A "radical" solution, on the other hand, would be one in which level 3 seldom appeared in the new assignment table.

By such a criterion, the solutions for this Navy course were decidedly "radical." Level 3 appeared only once of 15 possible decisions for the three tables of the course as a whole. It appeared only six times of a possible 60 for Unit 1; only eight times of a possible 57 for Unit 4; and only 4 times of a possible 27 for Unit 5. Given that this is a major Navy course, developed over years of intensive effort by many able people, such radical changes must be cause for very careful study -- not wrong per se but requir- great caution in interpretation. What is the nature of the solutions, and why are they so radical?

Part of the answer is that, in the solutions, the zero level appears a total of 24 times. That is, there are 24 occasions in the solution tables where a unit or topic is reduced to nothing. Especially, it appeared that this reduction to nothing sometimes occurred in the largest, most costly units or topics. To analyze this possibility, Table 4 was generated.

Table 4 reveals an apparent psychological phenomenon of the valuing strategy: a rather weak relation between the token value of a unit, and its current time cost. The token values are generally related <u>positively</u> to the current costs,

TABLE 4

CORRELATIONS BETWEEN CURRENT TIME-COSTS AND TWO MEASURES OF VALUE

CASE		Correlation of Time-Cost With	
		Token Value	Level Assign.
ALL UNITS	(5 units)	.41	-.63
UNIT 1	(20 topics)	-.45	-.65
UNIT 4	(19 topics)	.17	-.10
UNIT 5	(9 topics)	.81	-.03

but weakly. There is, in tokening, an apparent averaging which goes on, and which means that the variance in token proportion is much lower than the variance in cost proportion.

The effect of such averaging is seen in the last column of Table 4, in a strong negative bias against the larger units or topics. Subsequent debriefing of the judges revealed that it was not their intent to eliminate these large units; rather, it reflected that tendency to average their values, to lower the between-unit variance. There was, in fact, something of the feeling, though not verbalized during the rating, that they were evaluating the worth of a unit in terms of the "value per hour." But the effect of such valuing is to lower the estimated value for the unit (or topic) as a whole.

Yet the DP algorithm spends the total available hours in the most cost-effective way possible, which means that more effort is expended on those units where the value per hour is higher. And the algorithm computes the "value per hour" by dividing the total value by the number of hours currently allotted. Let us suppose two units, one of 50 hours and one of 10. If we appraise the first to be worth 20 tokens, and the second to be worth 10, then there is a correlation between cost and value. Yet the value per hour of the first is 20/50 = .4, while the value per hour of the second is 10/10 = 1.0. Thus in any DP algorithm, the second, shorter unit will commonly be emphasized. This phenomenon, then, is a feature of judge behavior, which must somehow be adjusted to render the the DP algorithm feasible in applied settings.

DISCUSSION

In a good many applications of OR, any difficulty is caused by an insecure fit of the model to human behavior. Most especially may this be true in educational applications, which have long suffered from inadequate measures of value (Page, 1972, 1974; Page and Breen, 1973). In the present application, there is here revealed a serious ambiguity, in the behavior of the judges, about what constitutes the "value" of a unit, whether their judgment reflects the worth of the unit as a whole (as it properly should, for the DP algorithm to function), or whether their judgment also reflects a factor of the worth of the unit per hour. The problem should call forth the ingenuity of workers in psychological scaling, and in operations research.

The following suggestions, therefore, are for future research in the field. They are broken into two sorts of strategy, one aimed at behavioral adjustments to this averaging tendency, and the other at statistical adjustments:

Behavioral adjustments

1) Begin the token strategy by placing before the judge a stack already proportional to the current time costs. This might serve to remind him of the overall nature of the evaluation, and to counteract the averaging tendency.

2) Or the judge might be explicitly told that it is the "value per hour" that he is estimating. Then, to use the DP algorithm, the weight of the unit as a whole would be the judge's estimate multiplied by the number of hours.

Statistical adjustments

3) The tokening might be skipped, and judges used only for the relative scaling of value at the various levels. To some extent, important information about the unit value is contained in such scaling. The statistical adjustment, for purposes of the DP algorithm, would be simply to insert a pseudo-token value equivalent to the proportion of current time costs.

4) The token judgments, once collected, might be adjusted so that the variance of their proportions would be made equivalent to the variance of the proportions of current time costs. This technique would leave the relative values of the judges undisturbed, but would fan them out in a way to eliminate the effects of the "averaging tendency."

Research along these lines should be successful in rendering the DP algorithm useful in the modification of curriculum of many different kinds.

ACKNOWLEDGEMENTS

This research and development was performed under NPRDC Code 306, Test and Application of Training Technologies and Systems (cf. Page and Canfield, 1975). The specific application was stimulated by the need to evaluate the course, Sonar Electronics Intermediate, taught by the Fleet Anti-Submarine Warfare School, San Diego. Earlier work with the same course was done by Stern and Aiken (1969).

During this work, valuable advice and assistance were given by E. C. Aiken, E. F. Alf, A. G. Archibald, J. G. Balaban, R. Boller, J. D. Carter, Jan D. Dodson, T. M. Duffy, J. D. Fletcher, E. I.

Jones, K. O. Leland, Carolyn L. McLandrich, W. J. Moonan, F. A. Muckler, J. J. Regan, B. Rimland, D. L. VanKekerix, H. L. Williams, and J. H. Wolfe, all of the Navy Personnel R & D Center, San Diego. Particular help was also given by P. V. Asa-Dorian and the instructors of the Curriculum Development and Evaluation Center, Fleet ASW School. Computer assistance was received from G. Grandon and C. Konopka, of the University of Connecticut, Storrs. Thanks are also due to P. Hawkes and A. Lubin. The most constant and direct collaboration was given by Dr. James Canfield, without whom this report would not have been possible.

REFERENCES

Moonan, W. J. DYNA: Dynamic programming analysis and the distribution of effort. Paper presented at the Military Testing Association Conference, October 1976, Pensacola, Florida.

Page, E. B. Seeking a measure of general educational advancement: The bentee. Journal of Educational Measurement, 1972, 9, 33-43.

Page, E. B. 'Top-down' trees of educational values. Educational and Psychological Measurement, 1974, 34(3), 573-584.

Page, E. B., and Breen, T. F. III. Educational values for measurement technology: Some theory and data. In W. E. Coffman (Ed.), Frontiers of educational measurement and information systems. Boston: Houghton-Mifflin, 1973.

Page, E. B., and Canfield, J. Design of Navy course structure through a dynamic programming algorithm. Consultant report at the Navy Personnel Research and Development Center, San Diego, June 1975. Pp. 69.

Stern, H. W., and Aiken, E. G. Training feedback study of the Sonar Technician Intermediate Electronics Course. Navy Personnel R & D Center, SRR 70-2, September 1969.

Wagner, H. M. Principles of management science: With applications to executive decisions. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

ACCURACY AND PRECISION OF FERTILITY PROJECTIONS MADE WITH TIME SERIES MODELS

Jeffrey S. Passel, U. S. Bureau of the Census

Demographers for many years have attempted to predict future trends in fertility and population, but have generally failed. Unfortunately, even as the science of demography has matured and its methodology become more sophisticated, the accuracy of fertility and population projections has not improved. Indeed, many modern projections are substantially worse than early projections made with sometimes crude techniques. This lack of accuracy coupled with the increased use of population projections in planning processes which result in the expenditure of substantial public and private funds suggests that improvements in projection methodology should be thoroughly investigated and that the error in projections should be quantified. The inclusion of distributional information with projections, either in the form of subjective probability estimates or explicit variance estimates, would immediately tell a user how much confidence to place in a projection and would make the projection more useful.

In developed countries, the component of population growth which has proved hardest to predict—fertility—has also been the major determinant of population growth. The accuracy of early projections can be attributed to the fact that fertility changes were very slow and gradual; however, this is no longer true as there seem to be swings of fairly large amplitude in fertility.

The problem of predicting long term trends in fertility for modern post-transition populations may not be as difficult as it has been for the past 50 years, however. It is possible to argue that in populations which have completed the demographic transition, there are no secular trends in fertility. With the completion of the adjustment of fertility to declining mortality, fertility in the future is likely to fluctuate around replacement levels.

Even if a model which included fertility fluctuating around replacement levels were accepted and brought into use, demographers would need some knowledge of the frequency and amplitude of the cycles in order for the model to be helpful for prediction purpose. The existence of such fertility cycles has been attributed to factors such as age structure which are internal to demographic systems (e. g. Easterlin and Condran, 1975) and to outside disturbances such as economic influences (e. g. Easterlin, 1973) mixed with environmental constraints (Lee, 1974b). The presence of such cycles has been empirically verified; and they could aid greatly in predicting fertility.

This conceptualization of fertility as fluctuating around a fixed level (replacement level) suggests that time series techniques could be profitably employed in studying fertility processes. Time series methods may not solve the problem of accurate trend prediction but the methodology can aid in quantifying the error process. Time series methods are well-suited for solving the problem of how much of a deviation from trend can be regarded as a fluctuation, not necessarily invalidating the entire forecast. With the addition of models for the variance of projections, the distributional information on population projections required for improved planning can be supplied in the form of confidence intervals.

1. Methods and Data

The projections reported in this paper were made with the classical discrete model of population dynamics, $\underline{X}_{t+1} = \underline{A}_t \underline{X}_t$. In this model, $\underline{X}_t$ is a column vector with each element representing the number of individuals in an age group at time t. The matrix $\underline{A}_t$ represents the survival of population members from one age group to the next and the entrance (birth) of new members into the first age group X_{t1}.

The variance calculations are based on the random transition matrix model of Sykes (1969) which assumes that the transition matrices, the $\underline{A}_t$'s, form a sequence of random variables:

$$\underline{X}_{t+1} = (\underline{A}_t + \underline{\Delta}_t)\underline{X}_t \ ; \ t = 0,1,\ldots \qquad (1.1)$$

where $\underline{\Delta}_t$ is a square matrix of random variables with the properties

$$E[\underline{\Delta}_t] = \underline{0} \quad ; \quad E[(\Delta_{iks}\,\Delta_{jlt})] = \underline{\Sigma}_{st}$$

and $\underline{\Sigma}_t$ is a singular variance-covariance matrix. With conditional arguments, it can be shown that the mean of the random process, $\underline{\mu}_t$, is identical to the deterministic process. The variance, $\underline{V}_t$, is then the sum of the one-step innovation variance, $\underline{M}_t$, and the weighted sum of the variance from the previous steps

$$\underline{V}_t = \underline{M}_{t-1} + \underline{A}_{t-1}\underline{V}_{t-1}\underline{A}_{t-1} \qquad (1.2)$$

where

$$\{m_{tij}\} = \{\sum_\alpha\sum_\beta \mathrm{Cov}(\Delta_{ti\alpha},\Delta_{tj\beta})(v_{t\alpha\beta} + \mu_{t\alpha}\mu_{t\beta})\}$$

The confidence interval for births can be shown to be (Schweder, 1971):

$$|X_{t1} - \mu_{t1}| \le \sqrt{V_{t11}} \qquad (1.3)$$

where is the percentage point in the χ^2-distribution with degrees of freedom equal to the rank of $\underline{V}_t$.

These basic equations are used to compute the projections but projected values for the parameters must be obtained by other means. The methods of time series analysis, in which the behavior of any variable over time is characterized by relationships to past values of the variable itself, can be used to predict values of the birth rates in $\underline{A}_t$, the b_j's, and the structure of the residuals can be used for $\underline{\Sigma}_t$.

The data used to fit the time series models are 36 series of annual central age-specific birth rates, ages 14-49 for 55 years, 1917-1971 (Whelpton and Campbell, 1960 and annual volumes of U. S. Vital Statistics from 1964). From the projected values of the annual central birth rate, f_j, the values of b_j are obtained by

$b_j = ({}_1L_0/2\ell_0)\ (f_j + s_j\ f_{j+1})$.

Because the major emphasis of this study is fertility, the values for survival rates were computed from actual life tables for 1950-1971 and from Census Bureau projections for the years 1972 and later. Annual immigration of 400,000 was assumed. For the initial population vectors, $\underline{X}_0$, data from the 1950 and 1960 Census were used; see Passel (1976) for more detail concerning data sources.

2. Time Series Methods and Results

The time series methods of Box and Jenkins (1970) were used to fit autoregressive, integrated, moving average (ARIMA) models to the birth rate series. Following their notation, the general form for a p-th order autoregressive, d-th order integrated, q-th order moving average model, also referred to as a (p,d,q) ARIMA model, is

$$\phi_p(B)\nabla^d f_t = \theta_0 + \theta_q(B)\epsilon_t \qquad (2.1)$$

The identification, selection, and estimation of appropriate time series models for the series of age-specific birth rates was accomplished following the methods suggested by Box and Jenkins (1970, Chapter 6).

The selection of an appropriate model for a given series depends on two major factors: 1) the uses of the model and 2) the nature of the series itself. All of the ARIMA models selected for fitting to the birth rates series included first or second differences. No unintegrated (i.e. d=0) models were selected for two reasons. First, examination of the autocorrelation functions for the series of age-specific and age-parity-specific birth rates suggests that practically all of the series, particularly those not at the older reproductive ages, are non-stationary. Second, and more important, the behavior of projections made using unintegrated models does not adequately portray the behavior of the birth rates. Projections made from undifferenced time series models decay exponentially to the mean of the series. There is no reason to impose this type behavior on the birth rate or probability projections. Furthermore, in every case, at least one integrated model produced a better fit than any unintegrated model. Differencing of higher than second order was not necessary to achieve apparent stationarity for any of the series.

The best fitting model for each age-specific birth rate was determined by the so-called portmanteau test, Q (Box and Jenkins, 1970:291), residual variances, and parsimony. Generally, for each birth rate series, the model with the smallest value of Q was chosen for the projections. If there were two or more models with small and approximately equal values of Q, the one with the smallest value of the residual variance, σ_ϵ^2, was chosen.

Of the 36 series of age-specific rates, all but 7 were fitted with models which yielded values of Q that had probabilities of occurrence of 0.70 or larger; none had a value smaller than 0.24. Box and Jenkins (1970:292-3) treat values of Q with probabilities of occurrence in the range 0.25 to 0.10 as indicative of some model inadequacy. Thus, all of the models presented in Table 1 fit the data quite well.

The fitted models are quite parsimonious. Only 8 of the series required second differencing to produce an adequate fit; only one of these (age 30) was at an age which contributes a substantial amount to overall fertility. Only 5 series (ages 30, 37, 39, 42, and 43) required four autoregressive terms, two degrees of differencing and two degrees in the AR coefficients. Seventeen series required moving average terms to provide an adequate fit but only one (age 35) required a second order MA term.

A further indication of the adequacy of the selected models is provided by the similarity of form and content shown by models for nearby ages. It is possible that other more complex models may provide a better fit for some of the series but this seems unlikely in view of the general similarity in form and the tendency for over-fitted models to produce worse results. Thus, the models presented in these tables, i.e. the parameter values, residuals, and residual variances, are the ones used in the population projections presented in the next section.

3. Results of the Projections

Two basic sets of projections are reported here: 1) starting in 1950 and continuing through 1975; and 2) from 1960 through 1975. The same time series models (Table 1) were used in each case. Thus, the "projection" models are based on data which include values from the projection period. This should lead to improved projections but, as will be seen, the overall quality of the projections is generally poor in spite of this bias toward "good" projections.

The behavior of the projections of the individual series of age-specific birth rates can be characterized by two main features: 1) each series levels off very quickly and 2) the amount of change from the initial value to the final value is relatively small. In order to assess the overall adequacy of the time series models the total fertility rate (TFR) and the projected births will be examined.

3.1 Fertility Projections: 1950-1975

TFR. The behavior of the TFR as projected from the 1950 base year is consistent with the behavior of the individual age-specific birth rates as just described. The projected TFR was essentially constant, staying roughly between 2,900 and 3,000; the oscillations are the result of minor deviations in fertility, mortality, and age structure. The projected value for the first year, 1951, was too low by 9 percent (2,968 vs. the actual value of 3,267) and the projected amount of change from 1950, -123, was in error by 262 percent. (See Figure A and Table 2.) The projection missed the 1950's "baby boom" entirely. By 1957, the peak year for TFR, the projected TFR was 811 points or 22 percent too low (3,760 vs. 2,949). Actual fertility fell after 1957 so that by 1965, the actual and projected TFR differed by only 47 points or 2 percent (2,928 vs. 2,881). However, the projections again did not predict another major fertility movement — the post-1960 decline in fertility. By 1975, the actual TFR was far below the projected value; the actual TFR (1,800) was 1,114 points less than the projected value (2,914) which was 62 percent above the actual TFR.

Births. The accuracy of birth projections made from base year 1950 with age-specific rates is comparable to the accuracy of the TFR projections for the first 15 or so years. Even the projection for the first year, 1951, is substantially in error. The difference between the projection (3,481,000) and the actual births (3,771,000) of 290,000 represents an error of 7.7 percent, a substantial amount for a single year forecast. In fact, the number of births fell outside the 50 percent confidence limits which were 3,481,000 ± 192,000 or a range from 3,289,000 to 3,673,000. (See Figure B and Table 2.)

The birth projections remain nearly constant between 3,400,000 and 3,500,000 per year through 1961. Then, the projected number of births begins to increase exponentially. This trend is, unfortunately, almost just the opposite of what actually occurred — an increase in the number of births from 1951 to 1961 followed by a general decline through 1975 (with the exception of 1969-1971). Although the curves of projected and actual annual births cross about 1966, the percentage errors increase again because of the different trend directions. By 1974, the difference between the actual number of births (3,115,000) and the projected number (4,676,000) reached 50.1 percent of the actual births!

The overall inaccuracy of the projected number of births can be seen quite clearly by examining the 50 percent confidence interval, which is ±192,000 for the first year of the projection and shows an increase of about ± 100,000 for each year of the projection. After only ten years, the 50 percent confidence limits reached ± 965,000 with the total width of the interval (1,930,000) being over 55 percent of the projected number of births. Even with these very wide limits which increased substantially each year, the increasing trend in the number of births is great enough to keep the actual value above the confidence interval for the first 8 years of the projection; only when the annual number of births levelled off in 1959 and 1960 did it fall within the 50 percent confidence limits.

3.2 Fertility Projections: 1960-1975

The projections from base year 1960 exhibit the same general tendencies as those from 1950. The projected TFR changes little from the initial value and levels off slightly above 3500 thus missing the decline of the 1960's. (See Figure A and Figure B.)

4. Discussion and Conclusions

The results indicate that the various projections are neither very accurate nor precise. Projections of births over periods of time as short as 1-5 years miss the actual births by substantial amounts. Over longer periods of time, 10-25 years, the projections bear practically no relationship to the actual course of events. The variance estimates appear to be quite large, too large in fact to provide useful confidence intervals for users of projections over short, medium, or long range. However, the inaccuracy of the projections, the large variances, and the failure of the actual value to fall within quite large fifty percent confidence intervals do not invalidate the variance models. The size of the estimated variance depends directly on the variances of the birth rates as determined from the projection models for the rates. Likewise, the accuracy of the projections is determined by the birth rate models. Thus, it is possible, even in light of the poor performance of the projections, to draw some meaningful conclusions about the precision of the projections, the variance models, and the method of computing confidence intervals.

The variance of the annual births is primarily a function of the variances of the projected birth rates. Projections of birth rates from time series models have variances which increase as the number of intervals (steps) from the origin increases. For the projections in this paper, the 50 percent confidence limits of annual births increase by about ± 50,000 to ± 150,000 per year. This means that by ten years into a projection, the 50 percent confidence limits for births are about ± 1,000,000 when annual births in this country generally have fallen in the range of 3,000,000-4,000,000. Furthermore, if the projection is carried far enough so that the projected birth cohorts with their large variances reach the childbearing ages, the variances of the next generation of births grow at an explosive rate. After 25 years or less, the 50 percent confidence limits for births could easily include no births. This property of increasing variance with time is logical and should be expected; the ability of demographers to forecast the number of births ten years (or even two years) in the future has proved to be quite limited. However, the magnitude of the variance of births in these projections is too great for practical application.

More precise projections, i.e. smaller confidence intervals, would provide better ex post assessment of projections and better ex ante limits for utilizing population projections. So, what can be done to make the confidence intervals and variances smaller? First, it should be stressed that the large variances found in this research are meaningful in the context of the projection models. The variances represent the actual experience in the United States over a period of 55 years as modelled with time series. So obviously, one way to decrease the variance of the projected births is to use more precise models for projecting the birth rates.

The size of the variance of a projection is a function of the number and size of the variances of the elements of the projection matrix. Thus, the overall variance could be reduced if the number of birth rates in the matrix could be reduced. In a related manner, the variance at lead time t could be reduced if the number of intervals between the origin and t could be reduced. By using 5-year age groupings (7 age-groups instead of 36) and 5-year time intervals, the variances of the projections for middle and long range projections could be reduced substantially.

A reduction in the variance of a projection as a result of increasing the size of the projection intervals and age groups would not be a mere statistical artifact. Small annual fluctuations in birth rates and births are removed by the aggregation. For many purposes, such projections would prove to be very useful. However, in a sense, this would be sidestepping important issues about improving, not the precision but, the accuracy of population projections, particularly over the short run. If the variance would be reduced, the projection models for fertility would still not be adequate.

The blame for the inaccuracy of the projections must be laid on the time series models. The models were fitted to the entire data series, 1917-1971, and were used to "project" points within the limits of the data as used to fit the models yet they still did a very poor job of predicting future values of the birth rates. What, then, went wrong with the projections of fertility and what can be done to improve them? Very little variation was predicted for individual birth rates; the smooth and relatively constant trends can be attributed directly to the properties of the time series models. Most of the year-to-year variation in birth rates is assigned to the random shock or residual term. In the projections these terms become zero so that their effect in the long run (after initially determining the level of the rate) is virtually nonexistent. The actual birth rates, however, are not mainly the result of random variation. Thus, the time series models are not adequate for most prediction purposes and do not model the actual process. However, improved time series models could possibly be useful for one- or two-year projections which could be updated annually.

Over medium and long ranges, the time series models used in this research do not predict trends very well. Essentially, the projected birth rates remain at a constant level close to the original level after four or five years rather than experiencing a prolonged, gradual movement to some, possibly very different, ultimate level. Although the model is based on the whole series, the projected trend and level are based wholly on the last p + d (usually not more than 4) values. Thus, the long range trend is a function of very short range variation and may be substantially different from the actual trend and from trends predicted in other ways.

Over both the short and the long term, the time series models do not do a very good job of predicting fertility (or of estimating the variance). Clearly, more structure than merely the autocorrelation of a series of birth rates itself is necessary to provide reasonable forecasts. Furthermore, it is unlikely that any model based solely on demographic variables such as age, sex, interval since last birth, parity, etc. will yield accurate projections of fertility even in the short run. (Figures A and B show that projections made by adding parity to the models are not substantially different. See Passel, 1976 for more discussion of these results.) A great deal of research has been done which demonstrates clearly that social and economic factors exert major influences on fertility. What is needed for accurate fertility (and, consequently, population) projection is a model which incorporates social and economic factors in addition to relevant demographic variables. However, we should not be limited to the usual regression techniques which ignore or suppress the known autocovariance structure of fertility.

A great deal of work in the area of fertility has been done with regression techniques and some with time series analysis (Saboia, 1974; Lee, 1974a), but the techniques have not been used together. However, methods are available which incorporate models of ARIMA processes and simultaneous equation models in such a way as to use regression techniques and time series analysis in estimating and checking the model (Zellner, 1975). Such linear multiple time series models could provide the basis for an integrated system relating social, economic, and demographic variables for providing accurate forecasts of fertility and population as well as estimates of the projection error.

Predicting the future for any purpose is difficult; the future of fertility especially has not proved to be easy to foretell. The methods of time series analysis, although useful tools (particularly for computing confidence limits), have unfortunately not provided an immediate solution to the problem of fertility projection. This does not mean that demographers should abandon the methods of Box and Jenkins. Rather, by combining the work of social demographers (with respect to correlates of fertility), economists (such as Easterlin for theory and Zellner for methodology), and statisticians, the work reported in this paper could be extended so that it might be possible in the not too distant future to predict with a reasonable degree of accuracy not only the short range (1-5 years) course of fertility and population growth, but also the long term (20-50 years) possibilities.

REFERENCES

Box, George E. P. and Gwilym M. Jenkins, 1970. Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day.

Easterlin, Richard A., 1973. Relative economic status and the American fertility swing. Pp. 170-223 in Eleanor Bernert Sheldon (Ed.), Family Economic Behavior: Problems and Prospects. Philadelphia: J. B. Lippincott.

________, and Gretchen A. Condran, 1975. A note on the recent fertility swing in Australia, Canada, England and Wales, and the United States. Forthcoming in Migration, Foreign Capital, and Economic Development: Essays in Honor of Brinkley Thomas.

Keyfitz, Nathan, 1972. On future population. Journal of the American Statistical Association 67 (338, June):347-363.

Lee, Ronald D., 1974a. Forecasting births in post-transition populations: stochastic renewal with serially correlated fertility.

Journal of the American Statistical Association 69 (347, September):607-617.
____. 1974b. The formal dynamics of controlled population and the echo, the boom, and the bust. Demography 11 (4, November):563-586.

Passel, Jeffrey S., 1976. Population projections utilizing age-specific and age-parity-specific birth rates predicted with time series models: Projections with confidence limits. Unpublished Ph.D dissertation, Department of Social Relations, The Johns Hopkins University, Baltimore.

Saboia, Joao L. M., 1974. Modeling and forecasting populations with time series: The Swedish case. Demography 11 (3, August): 483-492.

Schweder, Tore, 1971. The precision of population projections studied by multiple precision methods. Demography 8(4, November):441-450

Sykes, Zenas M., 1969. Some stochastic versions of the matrix model for population dynamics. Journal of the American Statistical Association 64 (March):111-130.

Whelpton, Pascal K. and Arthur A. Campbell, 1960. Fertility Tables for Birth Cohorts of American Women. Part I. Annual and Cumulative Birth Rates, by Age, by Order of Birth for All Women in Cohorts of 1876 to 1943. Vital Statistics — Special Reports. Selected Studies 51 (1, January 29). United States Public Health Service, Washington, D. C.: Government Printing Office.

Zellner, Arnold, 1975. Time series analysis and econometric model construction. Pp. 373-398 in R. P. Gupta (editor), Applied Statistics. Chicago: North-Holland Publishing Company.

TABLE 1

"Best" Fitting ARIMA Models for Time Series of Age-Specific Birth Rates for Women Aged 14-49, 1917-1971

Age	Model (p,d,q)	Q-Value	Probability[1]	Residual Variance	Mean	Autoregressive Coefficient		Moving Average Coefficient	
						Lag 1	Lag 2	Lag 1	Lag 2
14	210	10.2	0.99	0.15	3.86	-0.359	-0.203		
15	110	14.6	0.93	0.80	10.66	-0.010			
16	210	9.5	0.99	4.58	29.30	0.205	-0.170		
17	110	13.1	0.96	19.08	59.02	0.125			
18	210	18.3	0.74	49.11	99.41	0.189	-0.174		
19	210	18.9	0.71	91.34	137.54	0.234	-0.127		
20	110	19.0	0.75	120.66	159.54	0.196			
21	110	14.5	0.94	140.68	172.34	0.218			
22	110	17.0	0.85	149.49	180.51	0.191			
23	111	23.2	0.45	131.96	185.52	-0.272		-0.613	
24	110	23.0	0.52	148.33	183.64	0.152			
25	111	16.9	0.81	67.72	170.42	-0.177		-0.615	
26	210	26.6	0.28	73.27	162.42	0.317	-0.250		
27	111	16.4	0.84	54.45	151.10	-0.300		-0.741	
28	210	22.7	0.48	39.57	146.39	0.471	-0.170		
29	110	18.1	0.80	37.43	131.62	0.268			
30	220	27.3	0.24	21.63	123.89	-0.351	-0.560		
31	111	18.1	0.75	17.60	101.92	0.891		0.658	
32	111	13.1	0.95	16.64	100.65	0.897		0.634	
33	210	21.9	0.53	12.88	88.63	0.424	0.127		
34	111	15.5	0.88	13.78	80.86	0.895		0.665	
35	112	10.8	0.98	7.67	74.49	0.886		0.414	0.108
36	111	10.3	0.99	6.29	65.97	0.900		0.554	
37	220	20.8	0.59	5.85	56.14	-0.763	-0.365		
38	211	10.8	0.98	3.63	53.65	0.573	0.345	0.423	
39	220	10.5	0.99	3.28	41.77	-0.854	-0.416		
40	111	17.9	0.76	1.32	33.98	0.965		0.639	
41	211	13.2	0.93	0.72	21.90	0.777	0.184	0.644	
42	220	17.8	0.77	0.97	18.85	-0.909	-0.345		
43	221	17.8	0.72	0.27	12.02	-0.603	-0.507	0.137	
44	121	13.2	0.95	0.21	7.17	-0.317		0.824	
45	121	14.8	0.90	0.08	4.89	-0.105		0.660	
46	211	15.3	0.85	0.03	2.32	0.400	0.572	0.686	
47	210	16.7	0.82	0.02	0.98	-0.041	0.285		
48	111	8.3	0.99+	0.01	0.48	0.185		-0.130	
49	121	15.7	0.87	0.01	0.34	-0.496		0.863	

[1] Degrees of freedom = 25-p-q

[2] Births per 1,000 women

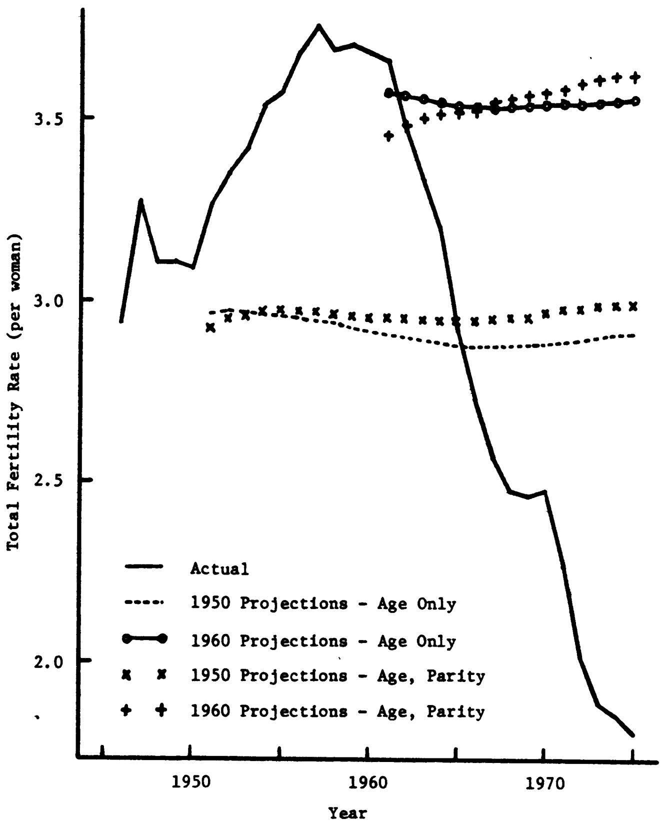


Figure A. Actual and Projected Total Fertility Rate, United States, 1946-1975.

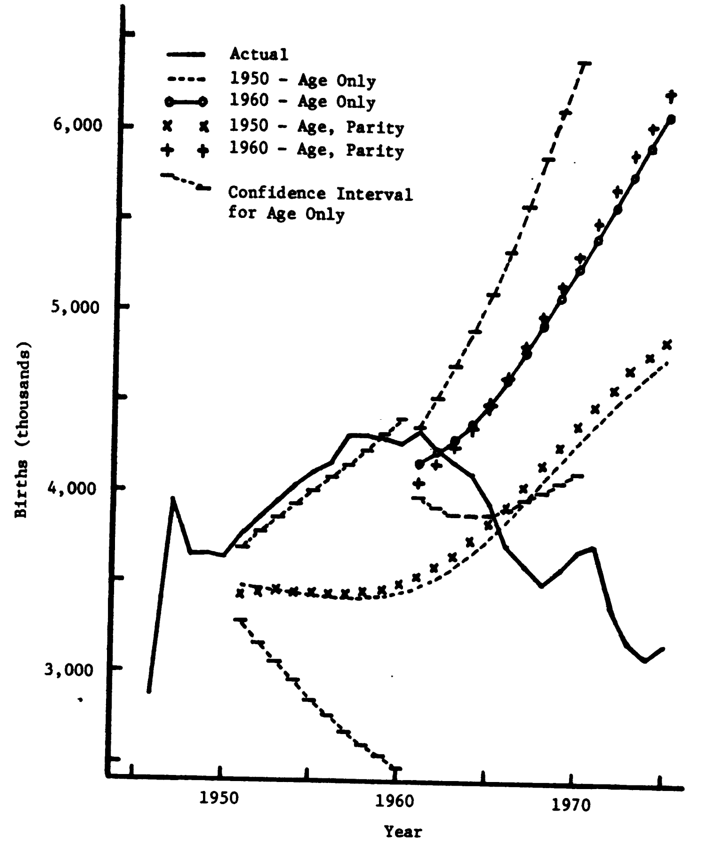


Figure B. Actual Annual Births, 1946-1975 and Projected Annual Births with Fifty Percent Confidence Intervals, 1951-1975 (Base Year 1950) and 1961-1975 (Base Year 1960), Projected with Age-Specific Birth Rates and Age-Parity-Specific Birth Probabilities.

TABLE 2

Total Population, Births, Confidence Intervals, and Annual Rates of Change -- Projections from Base Year 1950 Using Age-Specific Birth Rates: 1950 to 1975

(Populations in thousands. Based on Year of April 1 - March 31)

Year	Population[1]	50% Confidence Interval	Births	50% Confidence Interval	Total Fertility Rate[2]	Crude Birth Rate[3]	Crude Death Rate[3]	Percent Natural Increase
1950	150,216	(X)	(X)	(X)	(X)	(X)	(X)	(X)
1951	152,739	±199	3,481	±192	2,968	22.8	9.5	1.33
1952	155,230	839	3,471	305	2,970	22.4	9.5	1.29
1953	157,688	1,484	3,458	403	2,970	21.9	9.4	1.25
1954	160,110	2,274	3,443	492	2,966	21.5	9.4	1.21
1955	162,499	3,203	3,430	578	2,962	21.1	9.4	1.17
1956	164,857	4,260	3,420	657	2,957	20.8	9.4	1.14
1957	167,190	5,455	3,415	740	2,949	20.4	9.4	1.10
1958	169,507	6,350	3,418	814	2,940	20.2	9.4	1.08
1959	171,815	7,755	3,430	890	2,931	20.0	9.3	1.07
1960	174,124	9,285	3,453	965	2,922	19.8	9.3	1.05
1961	176,446	(X)	3,487	(X)	2,912	19.8	9.3	1.05
1962	178,836	(X)	3,535	(X)	2,906	19.8	9.1	1.07
1963	181,224	(X)	3,593	(X)	2,898	19.8	9.3	1.05
1964	183,629	(X)	3,664	(X)	2,889	20.0	9.5	1.05
1965	186,130	(X)	3,748	(X)	2,881	20.1	9.3	1.08
1966	188,692	(X)	3,849	(X)	2,878	20.4	9.4	1.10
1967	191,320	(X)	3,957	(X)	2,879	20.7	9.5	1.12
1968	194,063	(X)	4,069	(X)	2,884	21.0	9.3	1.17
1969	196,834	(X)	4,177	(X)	2,886	21.2	9.6	1.16
1970	199,716	(X)	4,284	(X)	2,888	21.5	9.4	1.21
1971	202,722	(X)	4,391	(X)	2,893	21.7	9.2	1.25
1972	205,843	(X)	4,496	(X)	2,899	21.8	9.0	1.28
1973	209,023	(X)	4,590	(X)	2,903	22.0	9.0	1.30
1974	212,253	(X)	4,676	(X)	2,910	22.0	9.1	1.29
1975	215,527	(X)	4,751	(X)	2,914	22.1	9.1	1.30

(X) Not available or not applicable

[1] April 1 population

[2] Per 1,000 women

[3] Per 1,000 mid-year population

POPULATION-ECONOMIC DATA ANALYSIS RELATIVE TO GEOTHERMAL FIELDS, IMPERIAL COUNTY, CALIFORNIA

James B. Pick, Charles Starnes, Tae Hwan Jung, and Edgar W. Butler
University of California, Riverside

1. INTRODUCTION

Social statistics aid in the understanding of current energy demands and in projecting energy utilization. Energy derived from nuclear, fossil fuel, and coal resources may be placed on an electric power grid, and for small increment costs may be conveyed to localities hundreds of miles away (Fowler, 1975). Such power grid relationships may alter according to season, regional supplies, alternate forms of energy, utility regulating decisions, international pricing developments, and so on. Thus the analyst may have difficulty in assessing demand.

Solar and geothermal energy sources differ from the others in predictability of locational supply-demand relationships. For solar, the current small capacity relative to major energy sources means that supply and demand locations are the same. Geothermal power stations are virtually nonexistent in the U.S. Regional use of geothermal power has the complication of changeable and distant dispersions by power grids, as well as plant capacity.

Regional relationships may be assessed, however, for nonelectrical uses of geothermal power. Geothermal energy production is based on subsurface hot water or steam, which in turn is heated by elevated magma layers. In geothermal power development, wells are drilled into the optimal three-dimensional geologic strata; hot water or steam is withdrawn and used to produce steam to turn generators which yield electricity. The regionally-fixed product of hot water may be used within about 60 km of the power plant periphery for heating, air conditioning, industrial plant processes, for agricultural purposes, and recreational use.

The purpose of this paper is to analyze the socio-economic characteristics of Imperial County, California, in relation to known locations of geothermal fields (KGRA's). Interpretation of these findings will be presented in a separate article by the authors. Socio-economic data are from the 1970 U.S. Census. Analytical methods used are graphic displays, demographic techniques and discriminant analysis. Before describing these it is important to examine relevant studies on energy development in the U.S., as well as past research on the population of Imperial County.

There have been few studies of socio-economic effects of energy development. A recent study (Vollintine and Weres, 1976) examined public opinion on geothermal energy development in Lake County, California, the major existing site of geothermal energy production in the U.S. Generally population variables were concluded to be uncorrelated with opinion on geothermal energy development. Kjos (1974) investigated the potential for industrialization in Calexico, the county's sister city of Mexicali. He drew essentially negative conclusions about the industrial development of Calexico due to lack of incentives for new businesses, a lack of governmental planning, local tensions in regard to border worker issues, and a lack of an industrial park, among other reasons.

2. POPULATION STATISTICS

Imperial County is located in the southeast corner of California; its central valley portion geologically is part of the Salton Trough. Before 1901, the valley consisted of Sonoran desert, with only several thousand residents in a land area of 4284 square miles. An increase in population was caused by the 1901 diversion of Colorado River water for irrigation. The northern drainage flow of residual irrigation water resulted in the formation of the Salton Sea.

As seen in Table 1, population rapidly increased to 60903 persons in 1930. At this time the population reached a plateau and in the next forty years increased only 22% compared to 251% for the state. Even with inclusion in 1970 of the 9000 estimated Mexican border commuters, population only increased 37%. A trend which paralleled total numbers was a decrease from a 1910 sex ratio of 1.9, 52% higher than the state, to a steady level 18-26% higher than the state's from 1930 to 1960. If international commuters (nearly all males) are included, the relative sex ratio continued 28% higher than the state's for 1970.

Since agriculture has consistently been the base industry of the county, the fundamental economic support for additional persons has not been significantly altered since 1930. The draw of large amounts of industry and commerce, which altered other formerly agricultural counties like Orange in this period, was probably precluded by labor and locational factors. The county's unusual sex ratio is interpreted as a lack of female occupations and amenities during the pioneering years of 1910-30, followed by an employment need for more males in agriculture since 1930.

A fundamental factor in Imperial's population is the presence of the border with Mexico and large numbers of Mexican Americans (henceforth abbreviated as MA). Although sparsely populated in the early 1900's, the northern region of Mexico has increased in population sharply since 1950 to a 1970 level of 857000, with 390000 in Mexicali. Table 2 shows comparative data on the MA population 1930-70. Comparison of the broadest definitions, "Mexican" for 1930 and "Spanish language" for 1970, again reveals greater equilibrium 1930-70 for the county. Imperial County's five-fold larger MA percentage than the state's in 1930 was likely due to border proximity and maximal Mexican emigration (Samora, 1971).

The stabilized nature of the county's agricultural economy set in 1930, has not offered an employment structure able to accommodate a skill mix of increased MA workers since that time. Noncitizen Spanish-origin persons have been present during this history. These are legal Mexican aliens (LMA's) and illegal Mexican aliens (IMA's). The detailed history of LMA and IMA trends and fluxes since 1900 is well-delineated in Samora (1971). LMA numbers have been influenced by economic demand for cheap labor which resulted in a federally-approved contract labor program in the 50's--the Bracero program. This program peaked with 400,000 contract workers in the U.S. during the late 50's. Since U.S. place of residence alone determines the census, a substantial number of Braceros were counted in the 1960 census for the county. With the ending of this program in the mid-60's, some 4500 to 6000 (U.S. Senate, 1971) farm workers established residence in Mexicali and commuted to jobs in the county daily, and are estimated to presently number 6,000-12,000. The Immigration and Naturalization Service estimated county LMA's at 8000 in 1973. Recent literature (Stoddard, 1976) has pointed out the extreme inaccuracy in estimates of IMA's.

Table 3 summarizes crude vital rates for the county from 1930-70. If we apply the average rate of natural increase (1930-70) of 1.65% we arrive at a compounded natural increase of 66% over the forty years (the process is assumed additive because the base population is only changing slightly). Under the assumption that border commuters should be counted as county residents in 1970, the forty year outmigration rate is 29%, or a compounded annual outmigration rate of .7%.

A standard demographic feature of MA population is higher fertility (Bradshaw, 1973). However, while the county's annual rate of natural increase is about 65% higher than the state's, this differential is only partly due to MA fertility. For completed cohort fertility in 1970 for ever-married women age 35-40, Imperial has a 49% additional increment for MA over non-MA fertility. When this is applied to the entire population, the fertility increment due to MA's is only 23%. The 42% differential is readily explained by the young age distribution of the county, especially of the MA population, since it is well-known that such youthful age distributions tend to inflate standardized fertility and deflate standardized mortality. Table 3 also shows life expectancies comparable to California's.

3. ENERGY CAPACITY AND CONSUMPTION

This steady level of total population of the county is presently supported by an ample local energy capability, which has evolved since the mid-1930's. Since the first energy installation in 1936 of a 2.2 mW diesel generator, installed capacity increased nearly linearly (at a smoothed linear rate of 3.4 mW/year) to 47.9 mW in 1950. From 1950-75, power plant installation again increased about linearly (at a smoothed linear rate of 12.0 mW/year) to 347.3 mW capacity in 1975. From 1940 to 1970, U.S. electrical energy production increased by a factor of 8.8 and per capita electrical production increased by 5.3 times (Fowler, 1975). By comparison, these factors for Imperial County energy capacity are respectively 12.9 and 10.4. Thus, while the county's population growth has lagged, its growth in energy capacity has greatly exceeded national rates. In fuel mix, the county has changed from steam-diesel-hydro proportions of 0:100:0 in 1940 to 52:7:41 in 1960 and to 68:11:21 in 1975. This latter compares to an estimated 1975 national mix of 76% petroleum and gas, 4% hydro, and 20% in other forms of energy (Penner et al., 1974). Thus the most unusual present feature of the county fuel mix is an enlarged component of hydroelectric power.

Per capita energy consumption was estimated by discounting national consumption in 1973 and county consumption in 1975 to 1970 by the national energy consumption growth rate of 5.6% per year (Fowler, 1975). Per capita estimates are 8272 kW hrs for the nation and 7325 kW hrs for the county (6530 kW hrs if border commuters are included).

Lower comparative per capita consumption contrasts with higher installed capacity per capita. Whereas the 1970 total and peak demand capacities are 1.67 kW/capita and 1.35 kW/capita for the U.S. (Fowler, 1975), they are 3.59 kW/capita and 1.89 kW/capita for the county (3.20 and 1.69 with border commuters). A final crude energy consideration is the current use of county electrical generation. The 1975 ratios of residential to commercial to industrial are 42:39:14 compared to 1973 national figures of 34:24:41. The housing fuel mix of the county reveals a major difference from the statewide pattern--reduced use of utility gas (37% below the state level) and increased use of electricity (271% above the state). A possible explanation is the low price of electricity due to the relatively large component of hydroelectricity in the county.

4. AGE STRUCTURE AND AGE SPECIFIC MIGRATION

As far as can be determined, there have been no past relationships between crude county population and energy development. Age structures and age-specific migration for the county have had very distinctive patterns, but seem unaffected by energy resources. However, age structures likely will be important for geothermal energy by determining present and potential labor force.

Age group data were obtained from the U.S. Census (1910-73), and data on births and deaths were gathered from state sources (California Dept. of Public Health, 1930-75). These were studied by the PYRAMID demographic display program (Pick, 1974).

In comparison to the nation, the county is significantly younger. Forty-five percent of males and 44.3% of females are below age 20, compared to national figures of 37.6% for both sexes. However, this youthfulness reflects only the MA population. While the above respective figures are 35% and 34.5% for anglos--levels below national ones--they are 56.4% and 53.6% for the MA population. This extreme youthfulness exaggerates fertility; since it also influences population projections, it will be discussed subsequently in the context of geothermal energy.

To study age specific migration between 1950-70 it is convenient to aggregate ages by three larger categories: 0-29, 30-64, and 65+. Age structures for 1960 and 1970 are distorted by border commuters previously referred to. Since these workers were nearly all males, comparisons of female age structures are more accurate. In 1960 female migration rates (based on past projected population with direction indicated) were -14.3, -11.5 and +10.0 for aggregated ages. In 1970 these rates were -10.3, +.4, and +9.2. Thus female children and young women have been leaving the county, while retirement age women have been returning. The middle age segment has changed from a strong outmigrating tendency in 1960 to neutrality in 1970. Because of the youthful age pyramid, numerically most migrants (counting in- and outmigrants equally) are in the 0-29 age category (65.2% in 1960 and 89.9% in 1970). The greatest female outmigration was for those ages of 20-24, with rates of -.257 and -.261 in 1960 and 1970. For the young, the draw of employment, educational, and life style forces in more urban, sophisticated regions of the West appear to account for such significant departure. Because of possible data unreliability, males' rates were done for ages 0-20 and 65-79. In 1960 rates were -6.5 and +12.0 for them; in 1970 rates were -4.6 and +9.7. Hence male migratory patterns also show the departure of the young and return of the old.

5. SOCIO-ECONOMIC CHARACTERISTICS

The economy of Imperial County, primarily based on agriculture, will be a critical element in possible geothermal development. The income distribution of the county was studied from 1950-70 by inflating all incomes by the ratio of the mean incomes by ten-year periods. To achieve a somewhat larger economic area for these inflators, Imperial's inflators were averaged with those for Los Angeles County. Incomes were inflated by 87% for 1950-60 and by 53% for 1960-70. Table 4 shows comparable income classes after multiplication by the above weights. The percent of top family incomes (above $28,611 in 1970) appears to be rather stable over twenty years. However, the second highest and lowest income brackets consistently decreased over the period, whereas the middle groups increased by 10%. This trend toward middle-class categories also appears for employment; there was an increase for 1950-70 of 11% in clerical and professional-managerial employment, compared to an increase of only 4% for California. In 1970, poverty levels were 35% higher than for California, and male and female unemployment were 7.1% and 6.4%, 69% and 129% above the state level.

Industrial classifications were examined for Imperial County, Kern County, and California. These can be best analyzed by separating occupational divisions by sex, and by eliminating the agricultural component entirely. In Table 5 are shown the nonagricultural industrial distributions by sex for 1950-70. In Imperial County, the most significant change in industrial classification for males is an 8% rise in public administration and government education (versus a 2.8% increase in these for California and a 3.7% increase for Kern). This increase reflects growth in the middle class. For males, a significant difference is a county manufacturing component only 57% smaller than the state, but a greatly increased utility and sanitary category (304% larger than the state). For females in Imperial, the most important longitudinal trend is an increase of 4% in government education workers. On the female side, compared to California, there is an even more sharply reduced manufacturing component (77% less than the state). Again the utility work force is larger. In summary, eliminating the effects of agriculture and sex ratio, two agricultural counties, Kern and Imperial, appear highly similar in industrial distribution but contrast with the state.

The above analysis controls for the influence of the labor force sex ratio. For the state, this ratio fell from 2.40 in 1950 to 2.03 in 1960 and finally to 1.64 in 1970. For Imperial County the figures were 4.12, 3.37, and 1.90 respectively. It is assumed that the labor force sex ratio will continue to decrease. Thus in considering labor market potential for geothermal-related industrialization, an important source of labor, presently only partially utilized, is females. The large gap with the state in female manufacturing employment would appear particularly important.

6. REGIONAL SOCIO-ECONOMIC COMPARISONS

In order to elucidate important county trends, socio-economic characteristics relevant to energy were studied by computer mapping and then by discriminant analysis between energy regions. Data for county enumeration districts (henceforth called ED's) were obtained from the fifth count of the 1970 Census (U.S. Bureau of the Census, 1974). These data were displayed using Automap II, a standard computer mapping program (Environmental Systems Research Institute). Five levels of data values were chosen, with extremes approximately eighths or sixteenths. Since 58.6% of the county's population resides in the towns of Brawley, Calexico, and El Centro, and another 35.7% lives in the central irrigated valley, the central valley and the largest three towns were mapped separately.

Geothermal energy is legally located underneath surface areas called Known Geothermal Resource Areas (KGRA's). Figure 1--a map of percent Spanish-speaking for the central valley--has superimposed on it a map of the three presently reported central valley KGRA's. For all analyses an ED was identified as geothermal if more than 50% of its surface area was contained within a KGRA. In addition, because of their size and location just outside the Heber KGRA border, ED's in the towns of Calexico and El Centro were identified as geothermal. Although not mapped, an ED in the lower East Mesa contains a promising KGRA which is being actively drilled by Republic Geothermal Corporation. However, this ED contains extensive land area and few of its 259 residents are likely to be affected; hence it was not identified as geothermal.

Using computer mapping technology, regional and geothermal results are presented below for the following characteristics: Spanish language, dependency ratio, labor force availability, mobility, income, and rental market over $100. Spanish-language residents show a concentration in Calexico and periphery (31% of MA's in county) and in Brawley and periphery (22.9%). Thus in the Heber KGRA and Brawley-Salton Sea KGRA's, the Spanish-speaking are of particular importance. Within towns, a directional axis of MA composition is apparent: southeast (high) to northwest (low) for Brawley; northeast (high) to southwest (low) for El Centro; and south (high) to north (low) for Calexico. In drilling and geothermal plant locations near these towns, such well-defined axes may be a consideration.

A dependency ratio was calculated and mapped (persons aged under 18 plus persons 65 and over/persons 18 to 64). This measure is ethnically biased because of the very large portion of children and adolescent MA population. Thus dependency closely corresponds to percent MA (r = .44), with a greater than average concentration in KGRA's. Directional axes for the towns are not apparent for dependency, but areas of highest concentration are present in east-central Brawley, northeast El Centro, and north Calexico. The opposite locations of high dependency and high percent MA for Calexico is due to the large percentage of young family housing for both ethnic groups in the north.

Male labor force availability is defined as the number of males 17-64/total number of males. By definition, it is inversely related to dependency, although the latter includes both sexes. For the northern combined KGRA, male labor force availability is very low in peripheral areas, but high in Brawley. For the southern KGRA, a medium level is apparent with El Centro medium and Calexico low. Brawley, which contains the preponderance of persons in the northern geothermal areas, clearly offers in addition an age structure favoring male work availability. Within the towns, this characteristic is highest in the two northern corners and west central part of Brawley, the northeast corner of El Centro, and the southwest corner of Calexico.

Female labor force availability is very highly correlated ($r = .58$) with that of males. Major differences are the reduction in female relative to male level in the highest income areas of southwest Brawley, El Centro and north Calexico. A similar reduction in the southwest part of Calexico, an area of farm labor activity, may be due to a reduced availability of married housing.

Regional mobility patterns were studied by 1965 residence in a different house (household mobility), different county (county mobility), and different state (state mobility). These data are for in-movement rather than out-movement. Hence for a county in which out-movement has predominated, the bulk of total moves is ignored. However, since town populations were very stable between 1960-70, it may be assumed that regional out-movement levels are highly correlated with in-movement levels as measured. These data show that mobility up until now does not correspond to KGRA boundaries. Household mobility has high rates in rural and peripheral areas in the eastern third and southern third of the central valley, except for the southeast corner; relatively high mobility in the northeast third; and medium to low levels in major towns. County mobility in nonurban areas in the northwest third is low in level, with moderate levels for the remainder; the cities of Calexico and Brawley had moderate rates while El Centro was high.

Since interregional mobility has been found to be an important correlate of other social variables including vital rates and prospective mobility (Butler, 1969; Pick, 1975), planners concerned with the labor force in El Centro should note the high mobility, with a decreasing gradient from west to east. Since interstate inmigrants compose only 30% of intercounty inmigrants, they are of less importance. The most notable feature, again, is elevated levels for El Centro and adjacent areas to the east and west.

Regional income differentials are very sharply defined, with higher incomes in El Centro, western rural El Centro, Holtville, and rural Calexico to the east. Rural parts of the Salton Sea and Brawley KGRA's and the town of Heber are at the lowest level of income. A general county-wide gradient for rural areas runs from north (high) to south (low). Clear income axes exist in Brawley (west-high to east-low), El Centro (southwest-high to northeast-low), and in Calexico (north-high to south-low). Regionally, the location of rentals in the rental market above $100 corresponds very closely to income--r (percent rentals $100-149 in the rental market, percent family income $9-25,000) = -.48, while r (percent rentals $200-249, percent family income $9-25,000) = .4.

Discriminant Analysis

As discussed earlier, for pollution by-products of non-electrical applications of geothermal energy, eventual effects from various field locations may be related to differences in the socio-economic characteristics of the surface areas above the fields. To attempt to establish what characteristics differentiate fields and relate these to geological features, a linear step-wise discriminant analysis was performed, using ED data. A standard step-wise program was used (Dixon, 1973), with an F criterion for inclusion of .01. In addition to the variables discussed in Section 6, the following variables were included for possible selection (all except the first three are in percent for the population considered): white fertility ratio (persons under 5/females 15-45), Spanish fertility ratio, persons now married, international mobility (persons age 5+ abroad in 1965), unemployment (white), unemployment (Spanish), white collar workers in the labor force (professional, managers, clerical, and sales), blue collar workers (craftsmen, operatives, service and laborers), farm laborers-farm foremen, families with income <$3,000, rentals $150-199 in the $100+ rental market, crowding factor of 3 or less persons in a unit, density of 1- persons/rm for the entire population and for the Spanish-speaking, house heating by utility gas, by bottled gas, and by electricity.

As seen in the KGRA base map, there are three major KGRA's in the central valley: the Salton Sea, Brawley, and Heber. Since only one ED is classified as geothermal for the Salton Sea field, it was decided to combine the Brawley and Salton Sea KGRA's, forming a larger area henceforth referred to as the Salton Sea consolidated KGRA. An inconsistency involves the three major towns, all of which touch KGRA boundaries--Brawley on the inside of the Salton consolidated KGRA, and El Centro and Calexico on the outside of the Heber KGRA. This irregularity was resolved by performing a discriminant analysis including the three major towns and another analysis including only rural ED's.

Results are summarized in Table 7. The first analysis grouped all geothermal ED's together; "non-geothermal" refers to the remainder of the county. The variable utility gas for house heating was by far the most important in separating geothermal from non-geothermal areas. Utility gas was also the best discriminant when the geothermal areas were classified into two groups. However, although the non-geothermal area was significantly separated, the Salton Sea consolidated and Heber fields were not significantly differentiated. Variables of less importance are percent Spanish-speaking, and percent of higher priced rentals ($200-249).

The importance of utility gas, higher rentals, and ethnicity is possibly due to a combination of geologic and human settlement reasons. The geothermal fields are contained in the Salton Trough, which extends from the Gulf of California to north of the Salton Sea, and are more likely to be on the central part of the trough. Likewise the irrigated valley is centrally located above the trough; and from a transportation and labor force standpoint, the county's towns were historically more likely to be centrally settled in the valley. Hence the significance of the utility gas and rental variables stems from the concentration of utility gas and higher rents in urban areas. MA preference for urban residence (Grebler et al., 1970) leads to the importance of the MA variable.

When a two-area analysis was performed for the rural portion of the county, Spanish-speaking was dominant--the above reasoning is still valid because of MA preference for areas near towns over more outlying areas.

7. POPULATION PROJECTIONS

Population projection was performed for the county, with special emphasis on geothermal energy. A set of four projections was made utilizing modified standard projection routines. The first two fifty-year projections (Table 6) were performed in a standard manner as follows: (I) fertility and mortality rates assumed constant at 1970 values; migration rates for both sexes assumed at the average rates 1950-70, with the sexes pooled and 1970 male rates for age group 20-65 assuming the values for females in 1970 to adjust for border commuters; and (II) fertility and mortality rates assumed constant at 1970 values; no migration allowed.

The potential stimulus of the county's present youthful age structure is clearly demonstrated by year 2020 projections of 71195 persons for (I) versus 175081 for (II). On a crude rate basis, the 1% difference in 1970 crude migration rates between (I) and (II) would account on a constant 1970 age structure for only about a 60% difference in 2020 totals. The actual 146% difference in these totals is attributable to a combination of the youthful age structure, high fertility, and the fertility dampening due to the high rates of outmigration prior to age 25 assumed in (I).

Possibilities for geothermal developments were incorporated into projections (III) and (IV). The key assumption for these is a linear increase in crude net migration rates calculated on the 1970 stationary age distribution (henceforth abbreviated as CNSMR = crude net stationary migration rates) until the time point of final installed capacity is achieved, with that time point's rate assumed thereafter. Geothermal development time scales were estimated from the only significant U.S. historical case--Lake County, California--and from known tentative corporate plans for geothermal development in Imperial County. Lake County's initial 11 mW of geothermal capacity has been added since 1971 at the approximate rate of 100 mW/year. Assuming in Lake County a total field capacity of 1500 mW and a constant future trend, one arrives at an estimated time scale of 25 years for complete development. Davis (1976) details present corporate plans for installation of power plant capacity running from 1978 to 1995. The spans of these two development scenarios were averaged to obtain a capacity installation span of 21 years for projection purposes.

Thus in projections (III) and (IV), CNSMR was increased linearly from 1970 to 1995 so that migration change is a function of power plant capacity. Year 1995 CNSMR was assumed at 20% for (III) and 30% for (IV) of the average crude migration rates 1950-70 for the five fastest growing California counties (these respective rates are 1% and 1.42% annually). The age distribution of inmigrants was a weighted average of 1) the county pattern 1950-70 used for (I)--weighted .3; and 2) the U.S. age structure for intercounty mobility (Bogue, 1959)--weighted .7. Results of these geothermal projections again reveal the county's high potential for rapid increase.

ACKNOWLEDGMENTS

The authors thank Mike Pasqueletti, Mildred Pagelow, Peter Force, and Manuel Sanchez for helpful suggestions. Funding was provided by NSF Grant AER75-08793.

REFERENCES

Bogue, D. J. 1959. The Population of the U.S. Glencoe, Ill.: Free Press.

Bradshaw, Benjamin and Frank Bean. 1973. Trends in the Fertility of Mexican Americans: 1950-70. Soc. Sci. Quarterly (March): 688-696.

Butler, Edgar W., F. Stuart Chapin, George C. Hemmens, Edward J. Kaiser, Michael A. Stegman and Shirley F. Weiss. 1969. Moving Behavior and Residential Choice: A National Survey. Nat. Cooperative Highway Research Program Report 81, Highway Research Board. Washington, D.C.: Nat. Academy of Sciences.

Davis, Cal. 1976. Preliminary Scenario, Imperial Valley Region, State of Calif. Pasadena: Jet Propulsion Laboratory.

Dixon, W. J. (ed.). 1973. Biomedical Computer Programs. Berkeley: University of California.

Environmental Systems Research Institute. Automap II User's Manual. Redlands, Calif.: Environmental Systems Research Institute.

Fowler, John M. 1975. Energy and the Environment. New York: McGraw-Hill.

Grebler, Leo, Joan Moore and Ralph C. Guzman. 1970. The Mexican-American People. New York: Free Press.

Penner, S. S. and L. Icerman. 1974. Energy, Vol. I. Reading, Mass.: Addison-Wesley.

Pick, James B. 1974. Computer Display of Population Age Structure. Demography 11(4): 673-682.

Pick, J. B. 1975. Correlates of Fertility and Mortality in Low Migration SMSA's. Proc. Am. Stat. Assoc., 588-595.

Samora, Julian. 1971. Los Mojados: The Wetback Story. Notre Dame, Ind.: U. of Notre Dame Press.

State of Calif., Dept. of Public Health. 1930-75. Vital Statistics of Calif., various vols. Sacramento, Calif.

Stoddard, Ellwyn R. 1976. Illegal Mexican Labor in the Borderlands. Pac. Soc. Review 19(2): 175-210.

U.S. Bureau of the Census. 1910-73. U.S. Census of Population. Various vols. Washington, D.C.: Government Printing Office.

U.S. Bureau of the Census. 1974. U.S. Census of Population, 1970. Fifth Count for ZIP Codes, Counties, and Smaller Areas.

Vollintine, Larry and Oleh Weres. 1976. Public Opinion Concerning Geothermal Development in Lake County, California. Report 4447, Lawrence Berkeley Laboratory.

Table 1. Population and Sex Ratios for California and Imperial County for 1910 to 1970

	Imper.	SR1	Calif.	SR2	SR1/SR2	Mexico
1910	13591	1.90	2377549	1.25	1.52	15 160 369
1920	43453	1.48	3426861	1.12	1.32	14 334 780
1930	60903	1.36	5677251	1.08	1.26	16 552 722
1940	59740	1.23	6907387	1.04	1.18	19 653 552
1950	62975	1.19	10586223	1.00	1.19	25 791 017
1960	72105	1.24	15720860	.99	1.25	34 923 129
1970	74492	.98	19953134	.97	1.28	48 313 438
1970	83492*	1.24				

* Includes border commuters
SR1 = Imperial County sex ratio
SR2 = California sex ratio
Source: U.S. Bureau of the Census, 1910-1970

Table 2. Spanish-Speaking Population in Imperial County 1950-1970

Spanish-speaking definition		Imperial	Kern	California
1930:	Mexican	.355	.086	.065
1950:	Mexican Birth	.135	.015	.015
1960:	Mexican Origin	.272	.050	.044
1970:	Mexican Birth			.020
	Mexican Origin	.304	.071	.055
	Spanish Language	.449	.155	.137

Source: U.S. Bureau of the Census, 1950-1970

Table 3. Measures of Fertility and Mortality for Imperial County from 1950 to 1974

Year	Births Imper.	Birth Rate Imper.	Birth Rate Ca.	Total Fertility Rate Imper.	Total Fertility Rate Ca.	Deaths Imper.	Death Rate Imper.	Death Rate Ca.	Imper. Life Expec. Male	Imper. Life Expec. Female
1950	1878	29.8	23.1	3836	3007	446	7.0	9.3	62.8	69.0
1960	1885	26.1	23.7	4347	3622	498	6.9	8.6	74.2	73.4
1970	1662	22.3	18.2	3211	2352	582	7.8	8.3	66.3	75.2
1974	1536	18.7	14.9			615	7.4	8.1		

Note: Imper. stands for Imperial County; Ca. stands for California; Life Expec. stands for life expectancy from age 0.

Source: State of California, various publications, 1930-1975

Table 4. Income Distribution for Imperial County 1950-1970 With Classification Subject to Inflation

Income Class 1950	Percent	Income Class 1960	Percent	Income Class 1970	Percent
10000+	3.1	18700+	3.7	28611+	2.9
5000-9999	21.2	9350-18699	19.3	14305-28610	15.3
3500-4999	15.0	7479-9349	14.8	10013-14304	19.6
1000-3499	44.8	1870-6544	50.5	2861-10012	50.8
-999	15.7	-1869	11.4	-2860	10.6

Note: Inflation of classifications calculated by average inflation of mean incomes for Imperial and Los Angeles Counties.

Source: U.S. Bureau of the Census, 1950-1970

Table 5. Non-Agricultural Industrial Distributions for Imperial County, Kern County, and California, 1950-1970

	1950 M	1950 F	1960 M	1960 F	1970 M	1970 F
Imperial County						
Construction-Mining	11.7	.7	8.4	.4	10.0	1.1
Manufacturing	12.5	2.2	12.8	2.0	11.3	3.5
Utilities-Sanitary	8.5	2.2	7.0*	4.4*	9.3	3.8
Wholesale-Retail	31.4	36.1	28.6	30.2	29.8	31.5
Public Admin.	6.9	5.4	8.8	5.9	10.7	6.0
Educ. (Gov.)	1.8	9.5	3.3	11.4	5.6	13.8
Other	27.6	44.6	29.0	63.2	22.2	40.2
Kern County						
Construction-Mining	26.7	2.0	20.0	1.6	21.9	2.4
Manufacturing	12.2	3.3	15.1	4.1	11.6	3.5
Utilities-Sanitary	2.4	.65	2.0*	1.4*	2.4	1.4
Wholesale-Retail	21.0	31.3	19.5	26.8	23.0	27.6
Public Admin.	10.8	9.5	12.0	8.4	12.3	7.2
Educ. (Gov.)	2.3	12.3	3.8	12.7	4.5	14.67
Other	39.1	49.0	36.1	47.3	24.5	43.3
California						
Construction-Mining	12.9	1.0	10.1	1.3	9.1	1.1
Manufacturing	23.9	15.3	29.3	17.4	26.5	15.3
Utilities-Sanitary	2.3	.7	1.6*	1.0*	2.3	.7
Wholesale-Retail	23.1	26.6	18.8	21.3	21.5	22.0
Public Admin.	7.3	6.7	6.9	5.5	7.5	5.2
Educ. (Gov.)	1.8	5.8	2.9	7.8	4.2	9.7
Other	28.7	43.8	28.9	46.0	28.9	46.1

*Sex division estimated from 1970 data.

Source: U.S. Bureau of the Census, 1950-1970

Table 6. Population Projections

Series	1970	1980	1990	2000	2010	2020
I	74492	75466	77448	76783	74270	71195
II	74492	88568	106194	124603	148124	175081
III	74492	76803	91387	119335	165981	233964
IV	74492	77381	95456	132525	198698	301989

Table 7. Discriminant Analysis of Geothermal Regions by Socio-Economic Characteristics

Grouping	Variables Included (F to remove in paren.)	F Matrix
Entire County [80] 2 groups	Utility gas (27.8÷÷) Rents $200-249 (11.5÷) Spanish-speaking (11.1÷)	Geo. [48] Non-geo. [32] 21.1÷÷
Entire County [80] 3 groups	Utility gas (15.4÷÷) Spanish-speaking (3.6*)	Salton [18] Heber [30] Heber [30] 2.1 Non-geo. [32] 16.7÷÷ 16.8÷÷ (Approx. F = 11.2÷÷)
Rural Portion [31] 2 groups	Spanish-speaking (12.9÷÷)	Geo. [6] Non-geo. [25] 12.9÷÷
Rural Portion [31] 2 groups Spanish-speaking variable excluded	Income $9-25000 (8.5÷) Bottled gas (5.6**)	Geo. [6] Non-geo. [25] 5.7÷

[ ] = number of ED's   * Signif. at .95   ** Signif. at .975   ÷ Signif. at .99   ÷÷ Signif. at .9995

Apparent Error Rate

Geo. .167
Non-geo. .281

Salton .389
Heber .467
Non-geo. .344

Geo. .333
Non-geo. .120

Geo. .333
Non-geo. .240

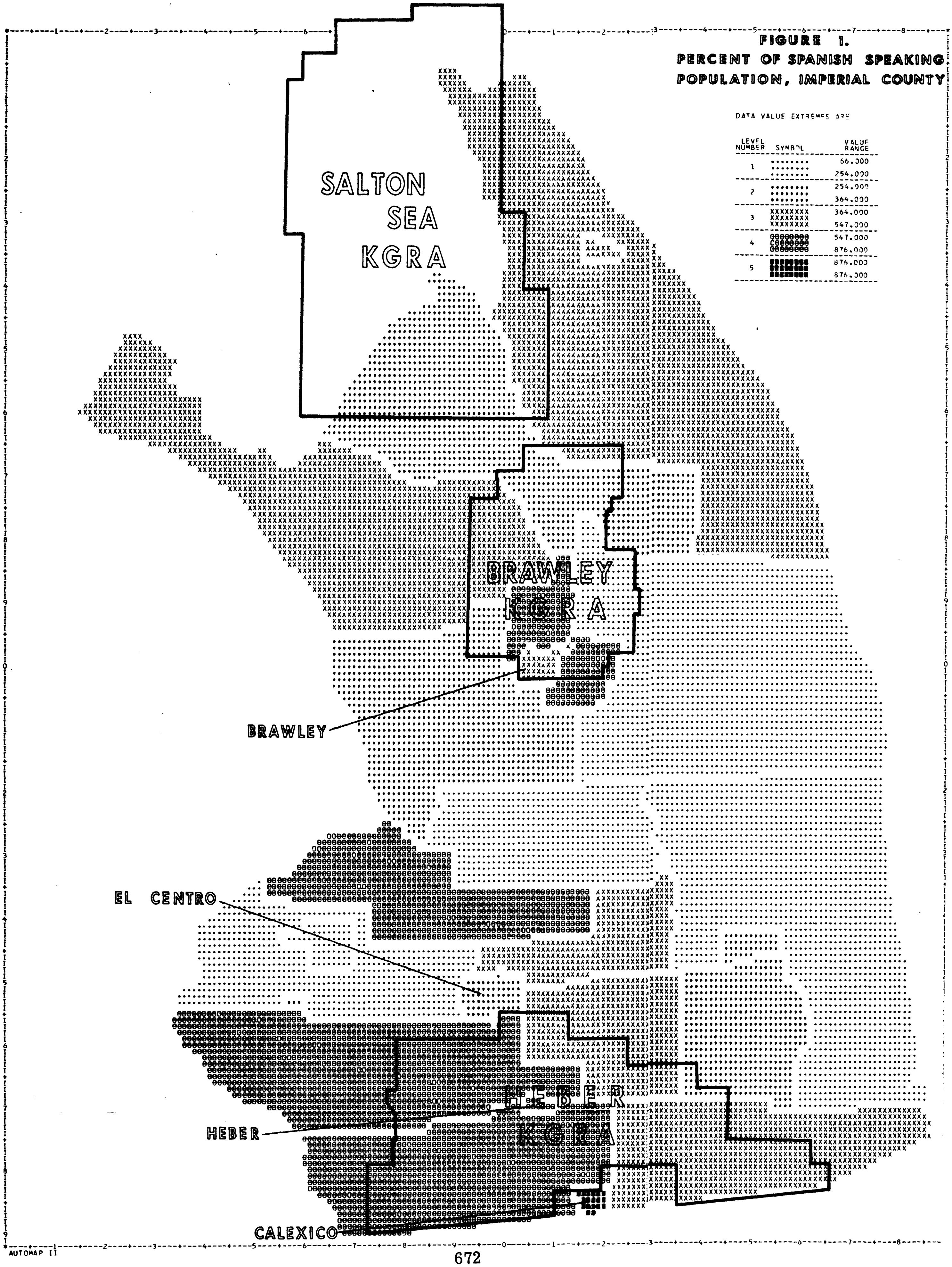


FIGURE 1.
PERCENT OF SPANISH SPEAKING
POPULATION, IMPERIAL COUNTY
DATA VALUE EXTREMES ARE
LEVEL NUMBER
SYMBOL
VALUE RANGE
1
66.000
254.000
2
254.000
364.000
3
364.000
547.000
4
547.000
876.000
5
876.000
876.000
SALTON SEA KGRA
BRAWLEY KGRA
BRAWLEY
EL CENTRO
HEBER KGRA
HEBER
CALEXICO
AUTOMAP II

ANALYSIS OF DESIGN EFFECTS AND VARIANCE COMPONENTS IN MULTI-STAGE SAMPLE SURVEYS

R. Platek and G.B. Gray, Statistics Canada

1. INTRODUCTION

a) General

Survey sample techniques have been in use for many years, and among probability samples, there are many types. They vary from simple random sampling to a complex multi-stage stratified sample design and it is the latter type which is the subject of this paper.

b) Need for Multi-Stage Sample Design

A particular survey design is determined by a variety of factors such as objectives of the survey, methods of enumeration, availability of census data, reliability of estimates required, and frequency of survey. The availability of funds usually plays an important, practical part in determining a survey design.

Any sample design, whether it uses personal interviews solely, or in conjunction with telephone or mail, requires clustering of the sample to some degree. Even mail surveys with personal follow-up of non-respondents may require clustering to a certain extent. Although clustering usually reduces the survey cost, the sampling variance of estimates of most statistics is increased. In order to introduce clustering in a survey design, a certain amount of information about the population is required. Census data may be used, for example, to delineate strata and primary sampling units, while more up-to-date information on areas within primary sampling units may be used to delineate sub-sampling units in subsequent stages.

The periodicity or frequency of a particular survey affects the type of survey design and the degree of clustering. For example, a continuous survey usually requires a permanent organization and permanent interviewers who must be trained and controlled.

c) Need for Components of Variance

For the sole purpose of obtaining measures of reliability of estimates, usually confined to estimates of sampling variability, it is selfom necessary to split up the variance into components and to estimate each component.

In a multi-stage design however, when total variances are studied, questions frequently arise as to the contribution of various stages to the total variance. Examination of the relative magnitude of variance components throws some light on the alternative sample design through changes in the sizes of the units, or changes in the allocation of sample units that might be adopted to decrease the sampling variance for a given budget, or decrease the cost for a given sampling variance. Clearly, the answer to either of these requires a cost analysis in conjunction with a variance analysis and we are aware that by providing variance analysis without cost analysis the study is incomplete. However, the magnitude of the components at certain stages may determine the desirability or necessity of isolating cost components at the same stages. In some instances, a variance component may be unacceptably high regardless of the corresponding cost component.

Total sampling variances and design effects [7] enable us to evaluate the gains or losses as a result of stratification, clustering and ratio estimation compared with simple random sampling and simple estimation. However, in order to analyze the design effect more fully, components of variances are required. For example, low design effects may mean that clustering plays no part in increasing the sample variance so that components of variance other than those between ultimate units may be quite low, while high design effects may indicate an extensive clustering effect which would lean toward high components of variance in at least one other stage besides the between ultimate unit component. Also, the size of the sampling unit at a given stage of sampling usually affects the variance component, since the population variance of which the variance component is a function, almost always increases with the size of a unit faster than by a linear relationship with the size. The reason for the faster than linear increase is a positive intraclass correlation [5]. Design effects by stages of sampling may also be calculated (See Definition in Section 4b).

2. VARIANCE COMPONENTS AND ANALYSIS

Variance components were undertaken for about 40 characteristics in June, 1973, for the Canadian Labour Force Survey, using Yates-Grundy estimation formulas [9], and further developed by Gray [3] and [4]. More details on the sample design of the Canadian Labour Force Survey may be obtained from [2] and [8], and the details related to variance component estimation may be obtained from [4].

The study of variance components based on the results from the LFS lead to two basic types of analysis, (a) a study of the variance components as percentages of the total variance and their relationship to design effects [7], and (b) a study of a variance function in terms of average weights and numbers of units at various stages and the population variances for which individual design effects may be obtained.

3. DESIGN EFFECTS AND CATEGORY OF CHARACTERISTICS

On the above broad definition of the types of characteristics as well as the observed design effects, the 19 characteristics were grouped into three categories, maintaining as much as possible the common category between urban and rural areas.

There is a rather high correlation of .860 between the design effects of the characteristics for urban and rural areas and without either "Finance and Insurance" or "Non-Agriculture Employed", the correlation increases to .907. The design effects in the urban and rural areas have also been ranked in order to show more clearly the correlation between the two sets of observations. In general, the clustering effect as seen by the magnitude of the design effect is more significant in the rural areas than in the urban areas, the average design effects being 1.604 and 2.020 in the urban and rural areas respectively. Although observations pertaining to category I are lacking in this study, monthly variance estimates of other categories using the Keyfitz formula [6] have revealed a similar phenomenon of low design effects among characteristics pertaining to specific age-sex categories.

4. ANALYSIS OF COMPONENTS OF VARIANCE

a) General

Each stage of sampling results in some contribution to the total sampling variance. The magnitude of the components vary from stage to stage and by characteristic within each stage of sampling. These have been calculated using methods developed in [4] for many Labour Force Survey characteristics in small urban and rural areas across Canada where four stage sampling was applied. The results are given in Table 2 for Urban and Rural. The percentages of the components of variance are studied in relation to the average weights and sizes of sampling units at the various stages of selection.

(b) Observation from Table 2 Dealing With Variance Components

i) The average weight $\bar{W}_r$ (i.e. the total number of rth stage units within all sampled (r-1)th stage units divided by the sampled number of units within all sampled (r-1)th stage units, "0" being taken as the stratum level) for each stage of sampling.

ii) The average size of rth stage unit $\bar{P}_r$. and averaged over the characteristics of each category.

iii) The average percentage contribution for each component of variance 100 V''_r/V'' by stage of sampling.

iv) The average design effect F_r, defined by $V''_r/(\sum_{s=1}^{r} \bar{W}_s - 1)\, \bar{P}pq$, and with some further approximation may be shown to equal $\bar{\sigma}^2_{r:r-1}/\bar{P}pq$.

v) The average measure of homogeneity, $\bar{\delta}_{4:r}$, referring to the measure of similarity in characteristics for any pair of households within rth stage units of a stratum as compared with any pair of households within the stratum, averaged over all strata across Canada.

The items above are tabulated for Canada urban and Canada rural as of June, 1973.

$\bar{\sigma}^2_{r:r-1}$ is the average population variance between rth stage units within (r-1)th stage units (not shown in these tables) averaged over all (r-1)th stage units. All averages mentioned above are estimates based on the sample.

In a particular (r-1)th stage unit $\underline{i}_{r-1}$,

$$N^2_{r|\underline{i}_{r-1}}\ \sigma^2_{r|\underline{i}_{r-1}} = \sum_{i_r} p_{i_r|\underline{i}_{r-1}} (X_{i_r|\underline{i}_{r-1}}/p_{i_r|\underline{i}_{r-1}} - X_{\underline{i}_{r-1}})^2;$$

where $N_{r|\underline{i}_{r-1}}$ is the number of rth stage units in $\underline{i}_{r-1}$, $p_{i_r|\underline{i}_{r-1}}$ is the relative size of rth stage unit i_r, and $X_{i_r|\underline{i}_{r-1}}$ is the characteristic value of unit i_r in $\underline{i}_{r-1}$ such that $\sum_{i_r} X_{i_r|\underline{i}_{r-1}} = X_{\underline{i}_r}$.

In the above, i_o = h (stratum); i_1 = (PSU), i_2 = K (segment in type of area j); i_3 = c (cluster) and i_4 = household.

c) Analysis of Table 2

The percentage contribution to the total variance, design effect by stages, and measures of homogeneity all show a distinct pattern by the three categories of characteristics in both urban and rural areas.

Turning first to the percentage contribution, we find that the total variance for category I characteristics is almost entirely contributed by the between household component (80% in the urban area, and 59% in the rural area). For category II characteristics, the between household component of variance drops to 68% in the urban area and 47% in the rural area with slight changes to moderate increases in the other components. For category III characteristics, the between household component drops to 44% and 30% in the urban and rural areas respectively, but the between PSU component increases to about 1/3 in both types of areas.

The high between household components of variance is reflected in low design effects and low measures of homogeneity at all stages. However, looking at individual characteristics for the moment rather than the averages, one notices that high components of variance at a given stage are reflected in high design effects and high measures of homogeneity. Negative components of variance estimate and negative design effects may be interpreted as estimates of unknown values that are positive but close to zero.

If measures of homogeneity are calculated, design effects may be considered redundant. However, measures of homogeneity are difficult to calculate and apply in the formulae for variance functions and population variances must be calculated individually (not averaged over characteristics), since they are a function of the size of the characteristic total or mean. Design effects are useful both for individual characteristics and for sets of characteristics to arrive at estimated population variances $\bar{\sigma}^2_{r:r-1}$ in the components of variance functions.

5. VARIANCE COMPONENTS AND DESIGN EFFECTS IN ALLOCATION STUDIES

For a given design (i.e. strata, units with their sizes unchanged), the population variances remain unchanged and the variance function (#7 Appendix) will be a function of the average weights for sampling at the various stages only.

Or, algebraically, the rth stage component of variance is given by:

$$V_r = [\prod_{s=1}^{r-1} \bar{W}_s (\bar{W}_r - 1) / \prod_{s=1}^{r-1} \bar{W}''_s (W''_r - 1)] V''_r$$

where " denotes the present value of the parameters. For example,

$1/W''_1$ = present sampling fraction of PSUs

and

V''_1 = present between PSU component of variance.

For a 4-stage sample design, $V = \sum_{r=1}^{4} V_r$, and by substituting the current percentage variance components for different weights, one can readily see whether the altered weights would increase or decrease the variance according as V is greater or less than 100. Table 3 provides the variances that would occur under different sample allocations of weight changes while maintaining the same overall weight (product of the 4 weights kept the same). Fifteen different values of $100 \sum_{r=1}^{4} V_r / \sum_{r=1}^{4} V''_r$ are obtained for each of the three categories using different component weights such that $\prod_{s=1}^{4} \bar{W}_s = \prod_{s=1}^{4} W''_s$.

The weights are varied mostly by halving and doubling the number of selected units per unit of the next lower stage for two of four stages (indicated by .5* and .2*). One special case of a census of primary sampling units with a more scattered sample within is also considered. In Table 3, the variance as compared with the present are obtained for Canada Urban and Canada Rural.

The total cost of the survey will not be the same for all of the allocations presented in Table 6, since the households may be on an average more spread out or more clustered than at present, but it serves to illustrate one method of optimization on the assumption that the total survey costs may not differ appreciably if the total sample size is kept constant, whatever the allocation by stages.

Under the survey constraints of a fixed overall sample size and fixed strata and delineated sampling units, it turns out that in only one of the 15 cases examined in Table 3 will the total variance be reduced for all 3 categories of characteristics, although striking reductions may occur in any one of the categories. If every PSU is taken (i.e. each PSU a stratum), then the variance of category III variables would be substantially reduced, but 1/2 of the segments and 1/2 of the clusters and the appropriate proportion of households within clusters to maintain the same overall sampling ratios (see note at bottom of Table 3) would be required. There is a relatively small choice of allocations to examine. $\bar{W}_r$ cannot exceed $\bar{P}_{r-1}/\bar{P}_r$, for we would take fewer than one rth stage unit per (r-1)th stage unit. Also $\bar{W}_r$ cannot be less than one, for we would be taking more units in the sample than there exist in the population. It would be possible to take twice as many rural segments per PSU as we now do, but not twice as many urban segments, since in most PSUs we already take every urban segment. For this reason, the cases of 2* for segments were omitted from Table 3.

The change in the variance from the present under different allocations given in Table 3 (no changes in the sizes or delineations of the units at the various stages) is given by:

$$V - V'' = \sum_{r=1}^{4} [\prod_{s=1}^{r-1} a_s^{-1} \bar{W}''_s (a_r^{-1} \bar{W}''_r - 1) / \prod_{s=1}^{r-1} \bar{W}''_r - 1) - 1] V''_r ,$$

where $\bar{W}''_r$ denotes the present weight (inverse sampling ratio) for sampling rth, stage units, $a_r^{-1} W''_r$ denotes the altered weight according to the allocation strategy as of Table 3 (commonly a_r = .5 or .2), and V''_r denotes the present variance between rth stage units, expressed as a present total variance. By substituting the appropriate values of a_r^{-1} in the above formula, simple relationships between V and V'' may be derived for each allocation strategy and the conditions between V''_r's necessary to ensure a decrease in the variance readily derived. The simplified relationship may help explain peculiar results such as allocation #4 and #5, where it turns out that $V''_4 < 4.1\ V''_3$ in order to ensure a decrease in V from V'' in the urban area.

<u>Conclusions from Table 3</u>

i) Despite the existence of take-all clusters, the between household component of variance accounts for a significant portion of the total variance and in fact dominates the total variance for category I variables (e.g., 80% of the total variance in Canada urban). Consequently, for fixed total sample size, changes in the allocation will have a small effect on the variance for category I variables.

ii) If we select only one PSU instead of 2 as in allocations #6 and 7, the variances of category II and III variables would increase substantially, in some cases over

50%. While the variance is substantially increased, the field costs might be only marginally reduced since the workload in each PSU is doubled and we might require two interviewers as we would in the case of 2 selected PSUs. If, however, the interviewer's workload could be doubled without jeopardizing the quality of the interviews, then there would be some reductions in the cost with fewer interviewers to train and lower travel costs.

iii) If we select 3 or 4 PSUs as in the case of allocations 8 to 12 and 13 to 15 respectively, there are only small decreases or increases for Category II Urban but a greater tendency for reductions for Category II Rural. For Category III Urban, the tendency leans towards reduced variance up to 20% (exceptions being allocation 11 and 12) but for Category III Rural, the reductions vary from a negligible 3% in the case of allocation 11 and 12 to about 30%. The interview costs would likely increase substantially either through increased travel between more PSUs which would result in larger areas to cover or through an increase in the number of interviewers in smaller areas and smaller workloads for each, thus increasing the training and interviewer control costs.

iv) If, in the extreme case, we select every PSU in each stratum, the reductions in the variance for Category II Urban and Category III variables are large (e.g., 45% reduction for Category III in the rural area). The effect on Category I variables and Category II in the urban areas is minimal (less than 8%). If we had delineated strata the size of PSUs, there would perhaps have been a larger reduction than indicated in this paper. For the philosophy behind the delineated strata is to make them as distinct as possible while PSUs should be as much alike as possible.

v) Despite the rather severe constraint of a fixed overall sample size with varying weights by stages but with delineated strata and units unchanged, many interesting and important results were revealed. The data of Tables 2 and 4, and the variance function in terms of weights could be further utilized to determine the variance for the requirements of reducing or increasing the total sample size, examining various strategies of reduction or increase of the number of selected units at the different stages of selection.

6. VARIANCES OF CHARACTERISTICS AND SIZES OF UNITS

Just as in Section 5, we examined the changes in the variances as the allocation of the sample by stages (maintaining or fixed overall sample size and the same delineated units and strata), we can examine the changes in the variances as the average sizes of the units are altered (maintaining the same strata and sampling rates by stages). In a similar manner employed in Section 5, the average sizes of the units are halved or doubled in such a manner that a lower stage unit would never be smaller than a higher stage unit.

Using #2 in the methodology, $\underline{V}$ was obtained for different average sizes of $\bar{P}_r$'s for r = 1, 2, and 3 while $\bar{P}_0$ and $\bar{P}_4$ are fixed. The results are presented in Table 4.

Analysis of Table 4

It can be seen that only marginal improvements are observed for any strategy of altered sizes of units and they are largely confined to the case of changed segment and cluster sizes. When the PSU size is doubled, however, the rural estimates are shown to possess an increase in variance up to almost 20% for Category I variables and almost 40% for Category III variables. In the case of Category III variables with high measures of homogeneity for most sizes of units, any increase in the average size of units will tend to increase rather than decrease the variance even while maintaining the fixed overall sample size. From the observed large between PSU variance of rural Category III variables, one would expect some substantial reductions in the variance of these variables if smaller PSUs were delineated and twice as many of them were selected. The interview costs for smaller PSUs, however, may be greater on a per household basis because of the extra spread of the sample or the necessity to hire extra interviewers to avoid the travelling between PSUs.

General Conclusion

The analysis of the components of variance pertained strictly to the current LFS design. However, the methods can be easily adopted to any multi-stage sample design similar to the above, such as the revised LFS design. Only one survey's data was used in this article and it appears that the allocation of the sample by stages and the size of delineated units are near optimum values for the given sample size. However, the cost function for changes in the weights and/or sizes of units was not taken into account so that one cannot draw fully certain conclusions about the optimum properties of the sample design. To monitor the sample design for its efficiency in terms of allocation and size of units continuously in a continuing survey where growth and features are constantly changing it is recommended that components of variance be obtained four times per year across Canada with special runs for smaller areas where problems are occurring.

REFERENCES

[1] Fellegi, I.P., "Sampling With Varying Probabilities Without Replacement: Rotating and Non-Rotating Samples", Journal of the American Statistical Association, 58 (1963), 183-201.

[2] Fellegi, I.P., Gray, G.B., and Platek, R., "The New Design of the Canadian Labour Force Survey", Journal of the American Statistical Association, 62 (1967), 421-453.

[3] Gray, G.B., "Variance Components and Variance

Function", Proceedings of the Canadian Conference in Applied Statistics, Statistics '71 Canada, 1971, 119-126.

[4] Gray, G.B., "Components of Variance in Multi-Stage Stratified Samples", Household Surveys Development Staff Survey Methodology, Vol. 1, No. 1, (June 1975), 27-43.

[5] Hansen, M.H., Hurwitz, W.N., and Madow, W.G., "Sample Survey Methods and Theory", Vols. 1 and 2, New York: John Wiley & Sons (1953).

[6] Keyfitz, N., "Estimates of Sampling Variance Where Two Units Are Selected From Each Stratum", Journal of the American Statistical Association, 52 (1957), 503-510.

[7] Kish, L., "Survey Sampling", New York: John Wiley & Sons (1965), 257-259.

[8] Sampling and Survey Research Staff, Dominion Bureau of Statistics (now Household Surveys Development Staff, Statistics Canada): Methodology, Canadian Labour Force Survey", (1965).

[9] Yates, F. and Grundy, P.M., "Selection Without Replacement From Within Strata With Probability Proportional To Size", Journal of the Royal Statistical Society, Series B, 15 (1953), 253-261.

Table 1: Average Design Effects by Category (no. of characteristics in brackets*) for Canada Urban and Canada Rural

Category	Canada Urban	Canada Rural
1	.864 (2)	.989 (2)
2	1.141 (9)	1.529 (10)
3	2.310 (8)	3.018 (7)

* with only one exception, the category of the characteristic was the same in urban and rural areas

Methodology For The Study

As mentioned earlier, the variance components were estimated by the Yates-Grundy formulas [9], also quoted in Fellegi [1] and the approximate variance function in terms of weights and population variances, developed in more detail by Gray [3] and [4] is given by:

$$V_r = (P/\bar{P}_r)\cdot\bar{W}_1\bar{W}_2 \ldots \bar{W}_{r-1}(\bar{W}_r-1)\bar{\sigma}^2_{r:r-1}/(1 - \bar{P}_{r-1}/P_r) \qquad (1)$$

= rth stage component of variance, where
p = total population
p_r = average size of rth stage unit
$\bar{W}_s$ = average weight for sth stage units within (s-1)th stage units
$\bar{\sigma}^2_{r:r-1}$ was defined in the text.

If the strata and sampling units at all stages remain fixed as in Table 3, then $V_r \propto \prod_{s=1}^{r-1} \bar{W}_s\cdot(\bar{W}_r-1)$. Although the variance estimates were obtained using the appropriate individual and joint selection probabilities, the variance function was developed on the assumption that the average f.p.c. $r_{FP} = -1/(\bar{N}_{r:r-1} - 1)$, where $\bar{N}_{r:r-1} = \bar{P}_{r-1}/\bar{P}_r$ is the average number of rth stage units per (r-1)th stage unit.

To study the changes in the variances as the sizes of the units are changed, but maintaining fixed household and strata sizes, as in Table 4, the population variance function was derived below.

$$\bar{\sigma}^2_{r:r-1} = \bar{N}_{4:r}\,\sigma^2_{4:o}\{[1+(N_{4:r}-1)\,\bar{\delta}_{4:r}] - \bar{N}^{-1}_{r:r-1}[1+(\bar{N}_{4:r-1}-1)\bar{\delta}_{4r-1}]\} \qquad (2)$$

where $\bar{N}_{s:r}$ = average no. of sth stage units per rth stage unit.

Estimates of $\bar{\delta}_{4:r}$ were obtained from June, 1973 data and interpolated by the reciprocal of the average size of units as different sizes of units at various stages were considered.

Table 2: Average weight (wt.), size of unit; percent component of variance %, design effect F, and measure of homogeneity δ by stage of sampling for Canada by type of area and by Category of variable (June, 1973).

Type of Area, Stage of Sampling	Aver. Wt.	Aver. Size	Category I %	Category I F	Category I δ	Category II %	Category II F	Category II δ	Category III %	Category III F	Category III δ
Urban, 0	1.00	16,053									
1	6.52	1,232	6.08	1.385	.0013	10.49	3.736	.0044	33.57	21.71	.0249
2	1.24	492	1.87	0.958	.0044	3.81	2.818	.0129	8.80	13.70	.0646
3	4.36	34.9	11.12	0.562	-.0235	17.58	1.113	.0169	13.46	1.718	.0807
4	4.16	2.33	80.95	1.002	1.0000	68.07	1.061	1.0000	44.18	1.339	1.0000
Rural, 0	1.00	24,180									
1	5.98	2,020	11.10	3.035	.0037	14.44	6.801	.0056	37.17	29.75	.0214
2	2.77	332	5.63	0.673	.0036	20.79	4.084	.0242	19.17	7.616	.0488
3	3.53	17.9	24.49	0.859	-.0091	18.20	0.946	.0055	13.93	1.473	.0530
4	2.47	2.27	58.78	0.930	1.0000	46.57	1.036	1.0000	29.73	1.273	1.0000

Table 3: Variance/Present Variance using different allocation of weights while maintaining current overall weight

Alloc. No.	Sample Allocation PSUs/ Str	Seg/ PSU	Clus/ Seg	Hhld/ Clus	Category I Urban	Category I Rural	Category II Urban	Category II Rural	Category III Urban	Category III Rural
1)	*(2)	*	*	*	100.00	100.00	100.00	100.00	100.00	100.00
2)	*	.5*	2*	*	106.35	99.13	114.45	125.34	141.46	124.49
3)	*	.5*	*	2*	95.17	93.31	115.71	119.06	114.95	123.71
4)	*	*	.5*	2*	88.82	94.18	101.26	93.71	103.49	99.21
5)	*	*	2*	.5*	105.59	102.91	99.37	103.41	98.26	100.39
6)	.5*(1)	*	2*	*	105.74	109.28	110.97	130.94	144.45	158.30
7)	.5*	*	*	2*	94.56	103.46	112.22	124.65	147.93	157.51
8)	1.5*(3)	2/3*	*	*	100.20	96.62	101.17	98.13	99.04	88.71
9)	1.5*	*	2/3*	*	98.08	96.33	96.35	106.57	85.21	96.87
10)	1.5*	*	*	2/3*	101.81	98.84	95.93	91.78	84.05	80.81
11)	1.5*	.5*	4/3*	*	102.32	96.33	105.99	106.57	112.87	96.87
12)	1.5*	.5*	*	4/3*	98.59	94.39	106.41	104.48	114.04	96.61
13)	2.*(4)	.5*	*	*	100.31	94.93	101.74	97.20	98.51	83.10
14)	2.*	*	.5*	*	97.13	95.36	94.52	84.53	77.78	70.85
15)	2.*	*	*	.5*	102.72	98.27	93.89	87.68	76.03	71.25
16)	all	.5*	.5*	(a)	100.43	94.05	92.44	80.56	67.38	55.81

Number of PSUs selected per stratum in brackets

(a): $\bar{W}_4$ = 6.78(U); 3.69(R)

*: Means current sampling ratio of selected units within units at the next lower stage on an average.
all: Means that there is no sampling of units at a particular stage, i.e. all units are selected.

Table 4: Variance/Present Variance for different size of units by stages of sampling (strata unchanged) and the same weights applied at each stage as currently

Strategy Number	Sizes of Units Compared to Current PSU	Seg	Clus	Category I Urban	Category I Rural	Category II Urban	Category II Rural	Category III Urban	Category III Rural
1	1.*(2)	1.*	1.*	100.00	100.00	100.00	100.00	100.00	100.00
2	1.*	.5*	1.*	99.258	100.242	98.190	91.136	94.980	91.472
3	1.*	1.*	.5*	108.168	105.923	102.586	105.329	100.622	102.212
4	1.*	2.*	1.*	100.793	99.936	101.777	101.790	105.131	101.733
5	1.*	1.*	2.*	98.526	99.177	100.106	98.913	100.465	99.889
6	1.*	.5*	.5*	107.937	106.560	101.109	96.789	95.747	93.865
7	1.*	.5*	2.*	97.784	99.548	98.356	90.221	95.444	91.360
8	1.*	2.*	.5*	108.307	105.693	103.318	106.073	103.083	103.003
9	1.*	2.*	2.*	100.193	99.064	103.803	102.746	111.092	103.396
10	2.*(1)	.5*	1.*	98.568	112.633	104.214	99.607	98.400	127.782
11	2.*	1.*	.5*	107.393	118.124	108.082	113.193	103.488	137.958
12	2.*	2.*	1.*	100.017	111.678	106.289	109.131	107.997	137.029
13	2.*	1.*	2.*	97.750	111.376	105.617	106.971	103.327	135.636
14	2.*	.5*	.5*	102.809	118.950	107.132	105.259	99.173	130.170
15	2.*	.5*	2.*	97.094	111.809	104.378	98.712	98.861	127.750
16	2.*	2.*	.5*	107.531	117.436	107.873	113.415	105.949	138.300
17	2.*	2.*	2.*	99.417	110.806	108.372	110.088	113.849	138.693

A * means A times as large as current average size.
No. of selected PSUs per stratum in brackets

RELIABILITY OF PATIENT RECALL OF EVENTS IN A CANCER EPISODE

Lincoln Polissar, Linda Styron, Anita Francis
Fred Hutchinson Cancer Research Center, Seattle, Washington

1. INTRODUCTION

Multiple sources of information have been compared to determine the accuracy of patient interview responses to questions on events in episodes of cancer of the colon or rectum. In section 4 we summarize the types of discrepancies found between the patient interviews and the other sources. This report is a part of an ongoing study of the patterns and care received by cancer patients in the health care system (1.). Information has been gathered concerning the dates of onset of symptoms, first contact with a doctor, subsequent visits to medical providers, and surgery. Data have also been gathered on costs associated with the cancer episode.

2. SOURCES OF DATA

Interviews were conducted with a panel of persons residing in King County (including Seattle) who were diagnosed during 1975 as having cancer of the colon or rectum. The first interview occurred just after treatment. Followup interviews were held every three months up to one year's duration after treatment. The interviewers recorded the patient's responses to questions on the number of medical contacts and on the dates and doctors associated with each contact.

Some of the patients permitted us to see their medical bills and insurance reports relating to the illness. These bills and reports included dates of medical contacts (referred to hereafter as "contacts") and doctor's names. For other patients we received dates of contacts and doctors' names from the claim records of a large third-party insurer or from the medical records of a prepaid group practice. These sources served as an excellent source of comparison for the patients' recall of contacts as reported in the interviews. Another source of information, used mainly for checking demographic data supplied by the patients, was tumor registry records.

In all, data from these sources could be compared on a maximum of 72 patients. The number of patients and contacts included in calculations shown in the tables below varies according to the number of missing values for the variables involved.

3. METHODOLOGY

Information regarding contact dates and doctors were compared between medical bills and coded interviews for the period from first visit to a doctor up to discharge from the hospital after surgical treatment. We considered the medical bills to be the more reliable source, so that a difference between bills and interview report was counted as an error in the patient's recall of the event. It is also possible that coding and keypunching of the data may have introduced errors, but this appears to be minimal, as we shall point out later.

Each contact recorded on the medical bills was compared to contacts on the coded interviews. Any date or doctor differences between the two sources was noted. We also noted contacts which the patient failed to report, as well as extra, fictitious contacts reported by the patient. A reported contact could be considered fictitious only if the set of bills did not include the contact and the bills were known to be very complete for the period including the alleged contact.

A methodological problem in comparing the two contact records is that we could not link some contacts that perhaps should have been linked. An example will illustrate this. A patient reports seeing Dr. Early on February 4. A complete bill record shows that, by that date, the patient had not seen Dr. Early for a month, was now seeing Dr. Late, and there was no contact with Dr. Late on February 4. In reality, the patient may have been thinking of the events during a visit with Dr. Late on February 18, but gave the wrong name and was in error on the date by two weeks. Our coding system would label the February 4 contact as fictitious, and would code an unmatched bill contact on February 18 with Dr. Late as missed by the patient. From one point of view, this coding scheme acceptably represents the error. Alternative coding schemes, however, could be formulated.

In some cases, the bills covered only certain periods or certain doctors, so that comparisons could not be made between bills and interviews for some number of contacts. One of us evaluated the degree to which the bill record covered the period of interest for each patient (table 1.). 83% of the patients had bills with medium to good coverage.

TABLE 1

COVERAGE OF CONTACTS BY MEDICAL BILLS

Completeness of coverage	Number of patients	%
Good	41	57%
Medium	19	26%
Poor	12	17%
	72	100%

4. RESULTS

4.1 Accuracy in reporting numbers of contacts.

Bills and interviews could be compared for 72 patients who reported 478 contacts in the aggregate. Among the 478 contacts, 415 (87%) could be compared between bills and interview to find discrepancies. The remaining 63 contacts were reported by the patients for periods not covered by our bill records.

Table 2 summarizes the types of between-source discrepancies found by comparing bills and interviews for the 343 contacts with doctors. (Hospital admissions are not included in table 2).

The most serious error reflected in Table 2 is the proportion of contacts not reported by the patients. Among the 310 contacts which were listed on the bills and thus appear to be *bona fide*, 98 (32%) were not reported by the patients.

Almost half (45/98) of contacts not reported by the patients were referrals by a doctor for a single contact with an outside specialist (Table 3). Most often this specialist was a radiologist.

TABLE 2

RESULTS OF COMPARING BILLS AND INTERVIEWS FOR 343 CONTACTS WITH DOCTORS

	#	%
Patient and bills agree on date	74	22
Patient and bills disagree on date	121	35
Fact of contact verified by bills, patient did not know date	17	5
Patient did not report contact	98	29
Sub-total, contacts listed on bills	(310)	
Fictitious contact reported by patient (inconsistent with bills).	33	10
Total	343	101

Such contacts may be forgotten due to the brevity of the doctor-patient involvement and the fact that the patient usually hears the results of the visit at a subsequent visit with his own doctor.

TABLE 3

CONTACTS NOT REPORTED BY PATIENTS

		#	%
Single contact referrals		45	46
to radiologist	24		
to internist	12		
to GP or family Practitioner	4		
other	4		
	45		
Other contacts		53	54
Total		98	100%

Another possible reason for some of the large number of missed contacts may be the difficulty of matching some of the bill and interview contacts. As mentioned above, some contacts may have been reported by the patient with a wild date and incorrect doctor, leading us to code a single error as two errors: one missed contact (an unmatched contact on the bills) and one fictitious contact (an unmatched contact on the interview which is inconsistent with the bills). Therefore, if the 98 unreported contacts are reduced by 33 - the number of fictitious contacts - the remaining 65 "unreported" contacts are 21% - rather than 32% - of the 310 contacts listed on the bills.

For each patient we calculated the difference between the number of ficitiious contacts and the number of unreported contacts. A difference of zero indicates that the patient has recalled the correct total number of contacts. We classified the patients into those who did and those who did not report the correct number of contacts. The relationship between various factors and percent of patients reporting the correct number of contacts is shown in table 4. The correct number of contacts was reported more frequently in longer interviews, for patients whose responses the interviewers felt were more reliable, in cases where there were fewer contacts to report, and in interviews taking place outside of the hospital. (The interviewers had previously commented that the patients were more receptive and alert outside of the hospital, when they had recovered from medication and the trauma of the surgery.) Patients age, sex, education, and occupation did not affect the accuracy of reporting number of contacts. Table 4 also shows no difference between interviewers in per cent of patients reporting the correct number of contacts.

TABLE 4

EFFECT OF SELECTED FACTORS ON PERCENT OF PATIENTS REPORTING CORRECT NUMBER OF CONTACTS*

Factor	Category	N/N	%
Interview time	≤ 1 hr.	13/37	= 35%
	> 1 hr.	9/17	= 53%
Interviewer	A	9/23	= 39%
	B	13/33	= 39%
Interviewer's assessment of reliability	Good	15/30	= 50%
	Fair	5/19	= 26%
	Poor	2/7	= 29%
Total # contacts	1-4	6/9	= 67%
	5-8	12/31	= 39%
	9+	4/16	= 31%
Place of interview	Hospital	4/13	= 31%
	Home, other	18/43	= 42%
Total		22/56	= 39%

* 100 x (patients with correct number) ÷ (all patients in category)

4.2 Accuracy in reporting dates of contacts.

Another statistic used to measure the reliability of the interview data is the difference, in days, between the patient-reported date and the bill date for a particular contact. Table 5 shows that, while the date differences range widely, 86% of the differences are within one week. Among the contacts with non-zero differences, 77% have differences within one week. Note, also, that the differences are quite evenly divided about zero (the median difference is zero.)

TABLE 5

DISTRIBUTION OF TIME DIFFERENCES BETWEEN BILL RECORD AND PATIENT REPORT OF CONTACTS*

DIFFERENCE IN DAYS (Bills Minus Patient Report)	NUMBER OF CONTACTS	%	
-22 or more negative	5	3	
-15 to -21	3	2	
-8 to -14	7	4	
-1 to -7	44	23	86%
0	74	38	
1 to 7	48	25	
8 to 14	8	4	
15 to 21	2	1	
≥ 22	2	1	
Total	193	101%	

Maximum = 34
Minimum =-90
Median = 0

*Includes Hospital Admissions

Recall of dates for early events in the cancer episode was generally less accurate than for later events (Table 6). The standard deviation of date difference decreased from the earliest verifiable event, the first contact with a doctor, to later events. This is also reflected in the increase in the percent of differences of one week or less. It is worth noting, however, that a number of patients (12/72) could not recall their hospital discharge date. The dates reported by the patients for the events in Table 6 were dispersed evenly about the bill date. The median differences are all trivially different from zero.

TABLE 6

DIFFERENCES BETWEEN PATIENT REPORT AND BILL RECORD FOR SOME KEY CONTACTS IN THE CANCER EPISODE

	Median difference	Std. Dev.	% of differences ≤ one week	No. of differences
First visit to doctor	-.2 days	21.6	76	55
Referral to surgeon	.1	8.9	88	48
Hospital admission	.0	3.9	99	71
Surgery	.0	2.1	96	70
Hospital discharge	.0	1.9	100	59

4.3 Accuracy in identifying doctors.

We checked to see how accurately we could identify certain doctors from information supplied by the patient. The patients could sometimes supply only the doctor's last name or were uncertain of the spelling of the name. The coder would have to attempt to more fully identify the doctor from various physician directories. We compared the result of this process with the doctor's name given on the bills. For 19% (11/17) of the patients, the first doctor seen was not identified correctly. The error in identifying the surgeon - the doctor whom we usually contacted for permission to call the patient and whose name we knew accurately - was only 3% (2/67), and probably indicates the level of error in coding and keypunching.

4.4 Accuracy in reporting demographic data

We were able to check three demographic variables by comparing hospital records and the interview responses for marital status, race, and age. (Interviewer's estimate of race was used for coding race in the interview). For marital status and race, the differences were minor (Table 7). There were a large proportion (51%) of age differences between the two sources. However, only one difference was as much as two years and all of the remaining differences were one year. Age differences occurred more commonly among the elderly.

TABLE 7

AGREEMENT BETWEEN PATIENT/INTERVIEWER REPORT AND HOSPITAL RECORD FOR THREE DEMOGRAPHIC VARIABLES

VARIABLE	CATEGORIES USED IN COMPARING	% AGREEMENT
Marital Status	Married, single	57/58 = 98%
Race	White, Non-white	55/56 = 98%
Age*	Single years	29/59 = 49%
Age < 65	Single years	19/32 = 59%
Age ≥ 65	Single years	10/29 = 34%

4.5 Effect of probing vs. brief interview

The effect of style of interviewing on response could be compared for several items (Table 8). In one interview situation a number of topics were covered briefly, including cost of care. In the second situation, one of the authors interviewed the patients in depth about cost and factors related to it. In this second interview the patients usually provided bills which allowed an accurate assessment of cost items, but if bills were not available, the interviewer asked a number of probing questions about cost. The differences in responses between the interviews were, more often than not, due to the probing interview finding a less desirable situation than the brief interview. That is, the probing interview found more hospital admissions, more impact on family finances, and larger amounts for the hospital bill and out-of-pocket costs.

In many of the cases where there was agreement between the two interviews on two of the cost items (amount of hospital bill and surgeon's fee), the patients were referring to the same bill for their responses in both interviews. The low percentage of agreement (21%) between interviews on the out-of-pocket cost is probably due to the lack of a single bill which covered this cost and which the patient could refer to. In fact, of the 75 patients who were asked about out-of-pocket costs on both interviews, 22 (29%) would not hazard even a guess in response to this question. This under-reporting of cost is also reflected in the small denominators for the cost and finance items in table 8. Potentially, there were 75 respondents, and the attrition is due to "don't know" responses on the brief interview or to our being unable to obtain the probing interview.

TABLE 8

COMPARISON OF RESPONSES TO A BRIEF INTERVIEW AND TO A PROBING INTERVIEW

ITEM	COMPARISON OF RESPONSES	#	%
Number of hospital admissions	Probing more admissions than brief	11	15
	Same	60	85
	Brief more admissions than probing	0	-
		71	100%
Impact of cancer episode on family finances	Probing more impact than brief	6	14
	Same	35	81
	Brief more impact than probing	2	5
		43	100%
Amount of hospital bill	Probing > brief	14	35
	Probing = brief	19	48
	Brief < probing	7	18
		40	101%
Amount of surgeons fee	Probing > brief	5	16
	Probing = brief	22	69
	Probing < brief	5	16
		32	101%
Patient's out-of-pocket cost for cancer episode	Probing > brief	11	46
	Probing = brief	5	21
	Probing < brief	8	33
		24	100%

5. SUMMARY

Patient interviews and other sources of data have been compared to determine the reliability of the interview data. Patients tended to forget some of their contacts with their doctors and, to a lesset extent, to report fictitious contacts. Single-contact referrals to a specialist accounted for half of non-reported contacts. When patients did report events, they reported the dates of occurrence fairly accurately. Demographic characteristics of the patients did not affect the accuracy of the reports, while some of the interviewing-related characterisitcs did. Accuracy of reporting decreased as time between event and interview increased. Age, race, and marital status were reported with acceptable accuracy. A probing interview technique uncovered higher costs and impact of the cancer episode than a brief interview did. Under-reporting was a problem in obtaining cost data.

REFERENCES

1. Anita Francis, Lincoln Polissar, "The Social Epidemiology of Cancer Care: Patterns of Health Care for Colo-rectal Cancer." (In preparation)

SPURIOUS CORRELATION AND SOCIAL SCIENCE RESEARCH: THE EFFECT OF USING RATIO VARIABLES*

James E. Prather, Georgia State University

The theoretical and analytical treatment of what Pearson (1897) called "spurious correlation" has import in and of itself. However, to illustrate its impact upon actual research applications is the intent of this paper. Lacking empirical antecedents showing the practical implications of spurious correlation, the mathematics of spurious correlation is but an academic exercise. This paper presents examples which clearly illustrate that research models which contain deflated variables (whether called percentages, ratios or per capita measures) are not inconsequential or transparent data transformations. It will be shown that deflation does change the assumptions about the model an investigator is using.

The first section of the paper contains a brief introduction of the mathematics of the spurious correlation, of which a detailed treatment may be found in Prather (1975). The next section consists of a collection of empirical examples of how ratio variables have been dealt with in the social sciences. Schuessler (1974) describes the way ratio variables are used in social research, but this paper aims at looking beyond correlation of variables pairs, toward analyzing the model specification process, i.e., the role of ratio variables in the structure of the analytic approaches often employed by social scientists.

CALCULUS OF SPURIOUS CORRELATION

The "spurious" correlation question, as developed by Pearson (1897), was that, given

$a = X/Z$

$b = Y/Z$

and even if X, Y, and Z are stochastically independent, then,

$E(r_{ab}) \neq 0.$

This non-zero expected correlation has troubled observors in the decades since Pearson's exposition, and the practical import of spurious correlation has often been debated without satisfactory resolution.

It is the thesis of this paper that spurious correlation is a result of poor statistical model construction. The specification of linear models to include nonlinear, multiplicative variables expressed as ratios, percentages, or per capita measures is often not the product of theoretical demands, but rather a lack of appreciation for the purpose of the linear model. Thus, the remainder of this section focuses upon the deflation question from the least squares (regression) perspective.

The use of analysis models with deflated variables is found throughout social science literature. The models most frequently employed take a form similar to the following:

$$Y/Z = a + b_1 X_1/Z + \dots + b_n X_n/Z + e/Z \quad (1)$$

where Y, $X_1, \dots, X_n$ and Z are fixed measures and e is random. If equation (1) is expressed in its undeflated form, then:

$$Y = a Z + b_1 X_1 + \dots + b_n X_n + u. \quad (2)$$

It is noted that equation (2) lacks a constant term. If the aim is to correct for a disturbance term with a multiplicative factor, i.e., $e = u/Z$, then the deflation of (2) to yield (1) will result in best, linear unbiased estimators (BLUE), given that the other assumptions of the linear model are met (Johnston, 1972, p. 122). If the problem is that Var (u) = $\sigma^2 Z^2$ then the more reasonable deflated equation would be,

$$Y/Z = A/Z + b_1 X_1/Z + \dots + b_n X_n/Z + e/Z. \quad (3)$$

This specification is given by Johnston (1972, p. 216) to correct for heteroscedastic disturbances. Belsley (1974) has noted that the careless specification of the constant term, such as found in (1) can lead to inefficient, if not biased, estimators. Also, if the measures Y, X's and Z are stochastically independent, the b's will not have expected values of zero, i.e., $E(b_1), \dots, E(b_n) \neq 0$. However, in equation (2) the null values for the b's would be zero (Prather, 1975).

Deflation may be viewed as an application of generalized least squares (GLS) or, more specifically, the type of GLS labelled weighted least squares. Equation (3) is in the form of GLS, where the deflation is for purposes of developing efficient estimators for equation (2). Belsley (1972) argues that if Var (u) $\neq \sigma^2 Z^2$, then deflation is not appropriate in that it will result in inefficient estimators if Z is fixed, and if Z is random, then spurious (bias) correlation is likely to be introduced.

There is a debate in the social sciences as to whether equation (1) is a structural model (a causal model), or just "fitting" an equation. The structural equation debate about ratio measures is a long standing one that has generated more emotional reaction than enlightenment. (Freeman and Kronenfeld, 1973; Fugitt and Lieberson, 1974; Schuessler, 1974). It is hoped that the following examples will demonstrate that deflation used for theoretical reasons is not as straightforward as some have maintained.

EXAMPLES

Professor Neyman has, for several decades, warned researchers of the pitfalls of ratio variables:

> In more modern times, spurious methods of studying correlations were involved in a great variety of empirical research; in astronomy, in farm economics, in biology, the study of elasticity of demand, in the problems of drunkenness and crime, of railroad traffic, and of racial segregation. On occasion, they were used in arguments about public policy matters. This applies to the health-pollution literature (Neyman, 1973, p. 31).

The following examples are presented with the aim of noting how deflation impacts on model specification itself. The implicit goal of each of the following examples is to increase the cumulative knowledge of these fields, freeing them from methodological questions of spurious correlation which interfere with the substantive pursuits of the discipline.[1]

Interstate Commerce Commission and Railroad Cost Studies. A case where deflation had policy implications may be found in the railroad cost studies performed for the Interstate Commerce Commission (ICC). Neyman (1952) made note of these data describing them as a possible example of deflated variables that could be misspecified. Neyman critiqued (pp. 151-154) the use of constructing railroad cost models with miles-of-track as a deflator. He expressed amazement that no one objected to this model.

It was not until 1972 that Griliches published a critique of deflated equations in cost studies. This example emphasizes the importance of the deflator used in a railroad cost model. Griliches (p. 29) pointed out that "it is very difficult to proceed to a discussion of the correct measurement of 'percent variable' unless it is possible to agree on which percent variable one is interested in--which 'average,' for what railroads, and at what traffic density."

The Griliches analysis reviewed the Cost Section's model, which contained these components: total costs (C), total gross ton-miles (X), and miles of track (M), the deflator. Griliches commented that, "to use deflation to eliminate the size component, one must assume that miles of road are in fact the relevant size measure and that the cost relationship is homogeneous in output and size, i.e., that there are no costs which are independent of size" (p. 31). The cost section used this model:

$$C/M = a + bX/M. \tag{4}$$

Griliches noted, however, that this assumed the undeflated model to be

$$C = aM + bX. \tag{5}$$

But Griliches proposed this model:

$$C = aM + bX + c, \tag{6}$$

and by deflating:

$$C/M = a + bX/M + c/M. \tag{7}$$

"The assumptions underlying the Cost Section procedure imply that the coefficient $\underline{a}$ is 'significant' while the coefficient $\underline{c}$ is insignificantly different from zero" (p. 32).

Griliches then fitted the equations using data from 97 railroads with the variables which are averages for the 1957-1968 period. The standard errors of the estimators for M in equations (6) and (7) showed that since the estimator for M is "insignificant," there is no evidence that M belongs in the equation. This is because the total miles of road variable is a poor measure of the "size" of the railroad. Obviously, division by an irrelevant variable is unneeded. Griliches then considers the problem of using deflation for stabilizing transformations. "Since the argument for any deflation must be made on efficiency grounds, it is not unreasonable, other things being equal, to prefer that procedure which yields the highest precision (lowest standard error) in estimating the parameter of interest (percent variable)" (p. 34).

Crime and Arrest Data Expressed in Ratios. Spurious correlation has involved not just simple mathematical relationships, but it also has involved the foundations of some researchers' arduous work, combined with their personal perceptions. An illustration of conceptual needs exceeding a linear model is found in Logan (1971a, 1971b, 1972) who studied the relationship between crime (C), population (P), and arrests (A), focusing upon the correlation $r_{C/P,A/C}$. Logan spent a whole chapter (pp. 102-116) justifying his correlation of ratio variables. The correlation $r_{C/P(A/C.C)}$, the Pearson formula, and a data simulation were employed. Logan wrote: "in summary, the correlations obtained between certainty of imprisonment and crime rate cannot be explained as spurious, indexical artifacts" (1971a, p. 111). It is interesting that Logan's model is expressed:

$$C/P = a + bZ/C + e.$$

Note that the variable of prime interest--crime--cannot be expressed in this model in a manner where it does not appear on both sides of the equation. If the Logan Model were given as the formulation,

$$C/P = a/C + bA/C,$$

then it could be reexpressed as,

$$C^2 + aP + bAP.$$

This is certainly not what Logan had conceptualized nor would he be likely to be willing to accept such a peculiar model.

Crime and Population Density. The consequences of careless deflation can have policy implications contrary to what appropriate analyses might imply. An example of this possibility is a recent newspaper headline--"Crime Linked to Population Density"--(Thornton, 1974), which reported a study by Kvalseth (1974), to the effect that, "the negative relationships between crime rates and population density established in the present study are highly conclusive and statistically significant as determined by both the multivariate regression analysis and simple correlation analysis" (Kvalseth, 1974, p. 31). A look at the model used might lead us to doubt his conclusion:

$$R/P = a + bA + cA/P + b_1 X_1/P + \ldots + b_n X_n/P + e,$$

where robbery incidents (R), square acres (A), total population (P) and X's represents additional exogenous variables (all deflated) and their estimators. The estimator b is always negative, and the estimator c is always positive. The correlation $r_{A,A/P}$ would be expected to have a high value which introduces the problem of multicollinearity, as the author noted (p. 34). This leads to inefficient estimator variance and possibly unstable estimators (Deegan, 1972). This may account for the significant t-test of the estimator c. Another model used by Kvalseth (1974, p. 17), reveals yet another deflation problem in which acreage in commercial use (C) is found:

$$R/P = a + b\ (A/P) \quad (C/A) + e.$$

That expression (A/P) (C/A) equals C/P is obvious, simply showing the R^2 of .73 with robbery is indicative of commercial use being related to robberies. However, Kvalseth claims that the results ". . . indicate again the substantial influence exerted on the robbery rate in an area by its population density . . ." (p. 17).

The general problem of how to conceptualize the role of density in theory building was reviewed by Lawrence (1974, p. 712):

> While it is difficult to draw firm conclusions from the existing meager

> body of data, the balance of evidence appears not to support any simple causal relationship between density and socio-or psychopathology.

Deflation in Economic Models.

> Applied econometricians have long been careless in the way they introduce deflators into linear models (Belsley, 1972, p. 923).

The economics discipline is as advanced as any social science discipline in quantitative applications to its models. The economics literature supplies a number of examples of how many economists deal with the problem of deflated data.

A recent exchange of journal communications provides an example. This controversy over deflation was originated by Sato (1971) who criticized an article by Vanek and Studenmund (1968) for "spurious correlations [which] arise because the equation is divided by a variable whose value is much dispersed" (Sato, 1971, p. 625). Studenmund (1974, p. 497) rebutted that 'sound' theoretical and econometric reasons exist that show that the equational forms . . . are not subject to spurious correlations." As a further defense, Studenmund cited, from Sato's work, a model in the form of equation (3):

$$Y/Z = a/Z + Y/Z + e/Z.$$

Studenmund (p. 497) comments on the Sato model: "As can be clearly seen, Sato's equation is--according to his own technique--just as subject to spurious correlation as are those of our article because both sides of his equation are divided by [Z]." Studenmund uses as his reference the Kuh and Meyer (1955) work.

Sato (1974) replied to Studenmund (1974) by defending his own model specification and re-emphasizing that the Vanek and Studenmund (1968) effort was amiss because " . . . as the growth rate appears on both sides of the equation, it is subject to spurious correlations in a high degree" (p. 499). Further, Sato argues that the Incremental Capital-Output Ratio (referred to as Y/Z above "is a single variable in <u>theory</u>, [but] it is derived in <u>fact</u> as a ratio of the investment share to the growth rate" (p. 499). Sato (p. 500) also argues that Vanek and Studenmund have used a deflator that is "largely stochastic." Sato concluded that "I believe that my results do not suffer from spurious correlations in spite of the Studenmund argument to the contrary" (p. 502).

In summary, from this exchange, it is readily apparent that deflation <u>per se</u> was not causing spurious correlation. Rather, all the authors seemed to express little if any awareness of what deflation does to their models, although Sato cites Kuh and Meyer's (1955) observation that a random deflator is of more concern than one that is non-random.

Administrative Intensity. In the field of management, one area of study that has treated the problem of using ratios at length is the analysis of administrative intensity (Freeman and Kronenfeld, 1973, and Freeman, 1973, p. 761). The problem was first noted by Akers and Campbell (1970), and Freeman and Kronenfeld devoted an entire article to the problem, calling it one of "definitional dependency." The frequently used model is that of

$$A/T = a + bT + e,$$

with numbers of administrators (A) and total employees including administrators (T). Inflating by T, it becomes

$$A = aT + bT^2 + eT.$$

Akers and Campbell looked at the model as being

$$In(A) = a + b \ln(T) + e.$$

They used a sample of 197 national membership associations and compared the numbers of staff members with the numbers of total members in the associations. Their equation was (p. 246) as follows:

$$lN(A) = 1.974 + .784 \ln(T)$$

with $R^2 = .686$. The authors noted that the elasticity (slope) of .784, "fall[s] short of a perfect proportional relationship between membership and staff size. Staff size increases with organizational growth but at a slightly decreasing rate" (p. 247). The undeflated data was expressed in this equation:

$$A = 9.619 + .0012T \quad R^2 = .692.$$

The undeflated model was in need of correction for heteroscedasticity, whether through log or deflation transformations. Freeman and Kronenfeld (p. 119) reviewed the use of the logarithm model, but doubted its conceptual rationale.

The work on administrative intensity by Evers, Bohlen and Warren (1976) is an interesting case where many types of transformations and ratios are used and they result in unexpected and most unappealing specifications. However, Freeman and Hannan (1975) analyzed organizational structure using a model in the form of equation (3) with GLS.

CONCLUSIONS

If "spurious" correlation were a methodological or technical problem, the debate concerning the subject would have ended decades ago. Yet observers, e.g., Sockloff (1976), continue to pursue a technical "solution," hoping to adjust the Pearson formula of 1897 to yield an "exact" measure of the correlation of ratio variables. These efforts might be categorized as statistical exercises which really cannot answer the underlying question of what is the "true" correlation.

The correlation of ratio variables is not really a "bivariate" correlation. There are three variables in non-linear combinations. The constant term of equation (1) must be accounted for in one's theoretical and conceptual formulation. That is, the problem is normative or conceptual, not statistical. The introduction of deflation into a model transforms the model, and it is necessary for the researcher to be aware of the implications of this transformation.

There are numerous researchers who would strenuously argue that ratio variables are "structural" measures, having genuine meaning as ratios. Does <u>per capita income</u> really exist as a causal variable? The linear model "fits" linear relationships. The implied multiplicative functions of ratio variables should be introduced into linear models only with the greatest caution. When theory requires ratio variables, it is advised that the constant term be carefully specified and that goodness-of-fit as measured by R^2 be viewed with caution. Buse (1973) notes that the R^2 in GLS may be derived in several ways--all of which differ from the R^2 of OLS.

If one approach to data analysis is that of the "exploratory" mode (recommended by Tukey and others) where the objective is to "fit" equations using the linear model, then all variables formed by division or multiplication should be avoided for these reasons: (1) the constant term would not be altered; (2) each of the variables can be tested (estimated as to their functional relationships--using deflation assumes a proportional or non-linear relationship); and because (3) the coefficient of determination (R^2) is not subjected to the ambiguity resulting from correlating ratio variables.

If heteroscedasticity is thought to be present (or is tested for by Harvey's [1976] general procedure), then deflation may be a solution for achieving efficient estimators. However, the double-logarithm transformation deals with this source of estimation inefficiency, typically better than does deflation (Prather and Hutcheson, 1976).

* I wish to thank John D. Hutcheson, Jr., Georgia State University, for his helpful comments and criticism.

NOTES: 1) An interesting, but purely historical example from the psychology literature is found in the debate over the impact of spurious correlation on the IQ ratio. Relevant studies are Thomson and Pinter (1924), Douglass and Huffaker (1929), Jackson (1940), DuBois (1948), McNemar (1946, p. 136) and Guilford (1965, p. 351).

Examples of the ratio conundrum in applied natural science may be seen in Sutherland (1965) who appeared most confused on the question of ratio data in experimental research; and Katch and Katch (1974) attempted, erroneously, to use ratio variables as a form of partial correlation.

REFERENCES

Akers, R. L. and Campbell, F. L. 1970. "Size and Administrative Component in Occupational Association." Pacific Sociological Review. 13(Fall): 241-251.

Belsley, D. A. 1972. "Specification with Deflated Variables and Specious Spurious Correlation." 40(September): 923-927.

Belsley, D. A. 1974. "The Constant Term and Deviations About the Mean." American Economist. 18(Fall): 109-112.

Buse, A. 1973. "Goodness of Fit in Generalized Least Squares Estimation." American Statistician. 27(June): 106-108.

Deegan, J. 1972. The Effects of Multicollinearity and Specification Error on Models of Political Behavior. Unpublished Ph.D. Dissertation, University of Michigan.

Douglass, H. R. and Huffaker, C. L. 1929. "Correlation Between Intelligence Quotient and Accomplishment Quotient." Journal of Applied Psychology. 13: 76-80.

DuBois, P. H. 1948. "On the Statistics of Ratios." Paper presented at the Annual Meeting of the Psychometric Society.

Evers, F. T., Bohlen, J. M., and Warren, R. D., 1976. "The Relationship of Selected Size and Structure Indicators in Economic Organizations." Administrative Science Quarterly. 21(June): 326-342.

Freeman, J. H. 1973. "Environment, Technology, and Administrative Intensity of Manufacturing Organizations." American Sociological Review. 38(December): 750-763.

Freeman, J. H. and Hannan, M. T. 1975. "Growth and Decline Processes in Organizations." American Sociological Review. 40(April): 215-228.

Freeman, J. H. and Kronenfeld, J. E. 1973. "Problems of Definitional Dependency: The Case of Administrative Intensity." Social Forces. 52(September): 108-121.

Fuguitt, C. V. and Lieberson, S. 1974. "Correlation of Ratios or Difference Scores Having Common Terms." Sociological Methodology, 1973-1974, edited by H. Costner. San Francisco: Jossey-Bass.

Griliches, Z. 1972. "Cost Allocation in Railroad Regulation." Bell Journal of Economics and Management Science. 3(Spring): 26-41.

Guilford, J. P. 1965. Fundamental Statistics in Psychology and Education. (4th ed.) New York: McGraw-Hill.

Harvey, A. C. 1976. "Estimating Regression Models with Multiplicative Heteroscedasticity." Econometrica. 44(May): 461-465.

Jackson, R. W. B. 1940. "Some Pitfalls in the Statistical Analysis of Data Expressed in the Form of IQ Scores." Journal of Educational Psychology. 31: 677-685.

Johnston, J. 1972. Econometric Methods (2nd ed.). New York: McGraw-Hill.

Katch, V. L. and Katch, F. I. 1974. "Use of Weight-adjusted Oxygen Uptake Scores that Avoid Spurious Correlations." Research Quarterly. 45(December): 447-451.

Kuh, E. and Meyer, J. R. 1955. "Correlation and Regression Estimates When the Data are Ratios." Econometrica. 23(October): 400-416.

Kvalseth, T. O. 1974. "Statistical Models of Urban Crime." Paper at Joint National Meeting of the Operations Research Society of America and the Institute of Management Sciences, October. San Juan, Puerto Rico.

Logan, C. H. 1971a. Legal Sanctions and Deterrence from Crime. Ph.D. Dissertation, Indiana University.

Logan, C. H. 1971b. "On Punishment and Crime (Chiricos and Waodo, 1970): Some Methodological Commentary." Social Problems. 19(Winter): 280-284.

Logan, C. H. 1972. "General Deterrent Effects of Imprisonment." Social Forces. 51(September): 64-73.

McNemar, Q. 1949. Psychological Statistics. New York: Wiley.

Neyman, J. 1952. Lectures and Conferences on Mathematical Statistics and Probability (2nd Ed.). Washington, D. C.: U. S. Department of Agriculture.

Neyman, J. 1973. "Epilogue of the Health-Pollution Conference." Bulletin of the Atomic Scientists. 29(September): 25-34.
Pearson, K. 1897. "Mathematical Contributions to the Theory of Evolution--On a Form of Spurious Correlation Which May Arise When Indices are Used in the Measurement of Organs." Proceedings of the Royal Society of London. 70: 489-497.
Prather, J. E. 1975. The Methodology of Spurious Correlation: Implications for Public Policy Analysis. Unpublished Ph.D. Dissertation, Political Science Department, University of Georgia, Athens.
Prather, J. E. and Hutcheson, J. D. 1976. "A Critique of Aggregate-Level Urban Policy Research." Paper presented to the Annual Conference of the Public Choice Society, Roanoke, Virginia.
Sato, K. 1971. "International Variations in the Incremental Capital-Output Ratio." Economic Development and Cultural Change. 19(July): 621-640.
Sato, K. 1974. "Spurious Correlation and Incremental Capital-Output Ratio: Reply." Economic Development and Cultural Change. 22(April): 499-502.
Schuessler, K. 1974. "Analysis of Ratio Variables: Opportunities and Pitfalls." American Journal of Sociology. 80(September): 379-396.
Sockloff, A. L. 1976. "The Analysis of Nonlinearity Via Linear Regression with Polynominal and Product Variables: An Examination." Review of Educational Research. 46(Spring): 267-291.
Studenmund, A. H. 1974. "Spurious Correlation and Incremental Capital-Output Ratio." Economic Development and Cultural Change. 22(April): 496-498.
Sutherland, T. M. 1965. "The Correlation Between Feed Efficiency and Rate of Gain, A Ratio and Its Denominator." Biometrics. 21(September): 739-749.
Thomson, C. H. and Pinter, R. 1924. "Spurious Correlation and Relationship Between Tests." Journal of Educational Psychology. 15(October): 433-444.
Thornton, L. 1974. "Crime Linked to Population Density." Atlanta Constitution. (August, 26): 7-A.
Vanek, J. and Studenmund, A. H. 1968. "Towards a Better Understanding of the Incremental Capital-Output Ratio." Quarterly Journal of Economics. 82(August): 435-451.

A RESPECIFICATION OF VARIABLES IN COHORT ANALYSIS

Daniel O. Price, University of Texas at Austin

Some of the basic problems in cohort analysis were delineated and a suggested approach was presented in Mason, Mason, Winsborough and Poole (1973). Some of the same problems were dealt with in a somewhat different framework by Riley, Johnson, and Foner (1972). The present article attempts to deal with the problem of separating Age, Cohort, and Period effects in cohort analysis by respecifying the variables, then presents an example of the proposed approach.

Cohort Analysis. A cohort is a group of individuals, areas, families, or other units of observation that were similar on some specified characteristic at some point in time. The most frequent usage of the term is as a birth, or age, cohort -- those individuals born in a particular year or time period. However, one might have a marriage cohort made up of all individuals who were first married during a certain year, a labor force cohort made up of all individuals who first sought work during a certain year or time period, or even a census tract cohort made up of census tracts that first exhibited some specified characteristic at a particular point in time. In cohort analysis this group, homogeneous with respect to some characteristic at a specified point in time, is "followed" over the years subsequent to the time when it was identified as a cohort. Other cohorts that exhibited the relevant characteristic at different points in time are also "followed," and comparisons are made within cohorts, between cohorts, and at particular points in time.

Utility of Cohort Analysis. Norman Ryder (1964, 1965, 1968) has made a sound case for the value of examining social change by cohorts. This leads to a better understanding of the phenomena than does the examination in cross-section. In fact examination by cross-section may be misleading. If, for example, one looks at median income by age one sees the medians increasing with increasing age up to about ages 40 or 45 and then decreasing with further increases in age. This has led to the conclusion that people reach their maximum earning capacity sometime in their forties and are likely to have declining earnings after that. An examination of median earnings by cohorts indicates that median earnings continue increasing throughout the labor force history or at least until beyond age 60. At a particular point in time people aged 55 may have lower median earnings than those aged 45, but their earnings at age 55 are higher than they were when they themselves were 45, ten years earlier. This can be seen only with examination of the data by cohorts.

Previous Specification of Relevant Variables in Cohort Analysis. Most works in cohort analysis have accepted the specification of the relevant variables as Age, Period, and Cohort effects. The Age Effect is the change in the relevant variable with increasing age. The Period Effect is the change in the relevant variable across all ages at a particular point in time associated with some societal event such as a recession, sudden inflation, a war, etc. The Cohort Effect is the differences in level of the relevant variable from one cohort to the next.

These sources of variation seem to be conceptually distinct but, as ordinarily defined, are not mathematically independent because if any two of these variables are known the third is determined. One of the best analyses of this situation and one of the best proposals for separating these three effects was that made by Mason, et. al. (1973). They point out that the technique they suggest cannot be expected to yield meaningful results unless the investigator also utilizes relatively strong hypotheses about the nature of these effects. The results obtained by investigators using the method suggested by Mason et. al. (1973) have, in general, not been too useful because this point has been ignored. (See Glenn, 1975.) A respecification of the relevant variables seems to be due, and one proposal for this is the basis of the present article.

RESPECIFICATION OF RELEVANT VARIABLES

If further progress is to be made in cohort analysis it would seem that the relevant variables must be more precisely defined. Ideally the specification would make possible the separation of Age, Period, and Cohort Effects but should also be based on an underlying reasonableness. The following identification of these variables not only seems to build on an underlying reasonableness but gives mathematical specifications for the variables.

The approach followed is to separate, by definition, the variables from their effects. Age, Cohort and Period are inextricably confounded since Age plus Cohort equals Period. However, it seems possible to define the effects of these variables in such a way as to be independent, especially if we specify a certain priority among them.

Age Effect. The Age Effect not only seems to be clear conceptually but would seem to occupy a certain primacy among the various effects because it is so omnipresent. Increasing age is the essential characteristic of a cohort, and virtually all dependent variables change with increases in age. Riley et. al. (1972) documents the basic importance of age in our society, not only as a variable, but as a basis for stratification. The age effect is therefore defined in traditional fashion as the changes in the dependent variable associated with increases in age of a cohort.

In virtually no dependent variable is there any reason for assuming that the age effect is constant at all ages. In general the age effects should be assumed to be nonlinear, the exact form dependent on the dependent variable under consideration. Age effects exist within cohorts and should not be assumed to be the same from one cohort to another.

The observations of a dependent variable made on the members of a cohort at increasing ages form a time series of observations. These are not independent observations, and if the time interval (the different ages) at which observations are made is sufficiently reduced, the relationship forms a smooth curve. Thus the age effect is defined mathematically as the slope of the curve fitted to the set of observations made at different points in time (at different ages) for a specified cohort. If we take

$$Y = f(A|C)$$

which indicates that the dependent variable is a function of age for a specified cohort, then the age effect can be defined as

$$\text{Age Effect, } E_A = f'(A|C)$$

which in words says that the Age Effect is the derivative, or slope, of the mathematical function relating the dependent variable, Y, to age for a specified cohort.

Important assumptions in the above definition are that the function is <u>continuous</u> and reflects the <u>pattern</u> of changes in Y with increasing age. It is possible to develop "pathologic" situations in which the function is discontinuous, but these are rare or nonexistent in real life. If every person were required to retire at precisely age 65, then a discontinuity would result at that point; however, even then we might not observe a discontinuity if observations were being made on age groups.

It is also important that the above definition assumes that the function, $f(A|C)$, reflects the <u>pattern</u> of changes in Y with increasing age. This is always a problem in the analysis of time series -- to extract the trends without allowing perturbations to disturb the trend but also without ignoring perturbations that might be reflecting "real" effects or changes in trend. This problem is recognized along with the subjectivity which is involved in efforts at its solution. As Mason et. al. (1973) points out, the investigator must utilize relatively strong hypotheses about the substantive material under study in order to come out with meaningful findings. We will not discuss this problem further at this point but will come back to it later because we assume that the period effects which we are trying to identify are the cause of some of these perturbations or deviations from pattern.

We have defined Age Effect as the first derivative of the function relating the dependent variable to Age in a specified cohort, and have pointed out that this function should be taken as nonlinear. The exact nature of the function should ideally be developed from theoretical notions about the relationship between the dependent variable and age. The development of functions which can be theoretically justified is an important area of research and should not be looked upon as simple empirical curve-fitting. A good fit to any curve can be obtained with the use of Fourier Series, but theoretical sense can be made of the result only in certain areas of electronics. The use of regression equations is "curve-fitting", but rarely is any thought given to the appropriateness of the usual assumption of linearity. Efforts must be made to associate mathematical form with substantive theory.

There are situations, however, where a good fitting mathematical curve is useful even when it has no basis in substantive theory. If the mathematical equation fits the observations very closely, it may be used as a convenient description of the data, and information about the data may be abstracted from the mathématical equation more easily than from the observations. In such situations it should be kept in mind that the mathematical equation is a description of the observations and it would be dubious to use it for extrapolation beyond the range of the observations. In the example given in the latter part of this paper, the mathematical equation is taken as a description of the observations from which information may be abstracted. No attempt is made in this example to develop a theoretically justifiable mathematical form.

<u>Cohort Effects</u>. Cohort Effects are usually referred to as the differences between cohorts, and in Mason et. al. (1973) are assumed to be a constant level of difference between cohorts in the dependent variable. Following the manner in which age effects were defined, we define cohort effects as the change in the dependent variable as one moves from one cohort to the next <u>at a specified age</u>. In symbols, if

$$Y = f(C|A),$$

then the Cohort Effect is ΔY for two successive cohorts at a specified age.

At this point it is not assumed that the function, $f(C|A)$, is necessarily continuous. We assumed that $f(A|C)$ was continuous because observations in a cohort at successive ages are not independent and may theoretically be taken at infinitesimally close intervals. A similar line of reasoning may be utilized if we consider theoretically narrow cohorts, say only one day in width rather than the usual 5 or 10 years. Given the stability of the social structure and the social interaction of individuals of similar ages, it is reasonable to assume that a characteristic of individuals who are some specified age today, say 34, will not differ significantly from the same characteristic of individuals who were 34 yesterday or those who will be 34 tomorrow. Thus, we now assume continuity of the function $f(C|A)$ for all ages. If

$$Y = f(C|A)$$

is a continuous function for all ages, then the Cohort Effect can be defined as

$$\text{Cohort Effect, } E_C = f'(C|A).$$

In words, the Cohort Effect is the derivative, or slope, of the function relating the dependent variable to Cohort for a specified age.

This definition does not imply that the cohort effect between two cohorts is constant for all ages, nor does it imply that the cohort effect is constant between all cohorts at any given age. Just as thought needs to be given to the nature of the association between the dependent variable and age, so thought needs to be given to the nature of the relationship between the dependent variable and cohorts. This is the way a particular characteristic is changing over time for a specified age group in society. In fact, this definition of Cohort Effect can be considered an operational definition of one aspect of social change.[1] The mathematical form describing this relationship should be developed on the basis of our theories of social change. Again, the substantive implications of the mathematical form should be examined for reasonableness. If the mathematical form chosen is a polynomial of second degree with respect to cohort, this implies that the cohort effect will change between successive pairs of cohorts but that the change will be by a constant amount. For example, school enrollment of individuals aged 20-24 may well be expected to be higher from one cohort to the next, but do we expect the increase to be by a constant amount or do we expect the proportion enrolled to increase but by decreasing amounts as the level approaches 100 percent? It is important to consider the substantive implications of the mathematical form and to derive the mathematical form from theoretical considerations if possible. For purely descriptive purposes, however, an empirically fitted mathematical function can be useful.

Age and Cohort Interaction Effects. If $f(C|A)$ is continuous for all Ages, and if $f(A|C)$ is continuous for all Cohorts, as assumed above, then it follows that

$$Y = f(A,C)$$

is a continuous function. This function may be visualized geometrically as a nonlinear surface above an Age by Cohort base in which the height of the surface above the base at a particular point is the value of the dependent, or relevant, variable for that particular Age and Cohort combination. The slope of this surface in any plane perpendicular to the Age axis is the Cohort effect at that point, and the slope of the surface in any plane perpendicular to the Cohort Axis is the Age effect at that point. In symbols we can define the age and cohort effects on the basis of the function.

$$Y = f(A,C)$$

$$\text{Age Effect} = E_A = \frac{\partial Y}{\partial A}$$

$$\text{Cohort Effect} = E_C = \frac{\partial Y}{\partial C}$$

Another important aspect which can be examined within this formulation is the change in age effect from one cohort to another. This is the same as the change in cohort effect with increasing age but is perhaps easier conceptualized as the changing age effect over time. Is the rate at which white females are leaving the labor force between the ages of 25 and 26 increasing or decreasing? The answer to this question is called the Age and Cohort Interaction Effect and is defined as

$$\text{Age and Cohort Interaction Effect} = E_{AxC} = \frac{\partial Y}{\partial A}\frac{\partial Y}{\partial C}$$

The changes in Age Effect from one time period to another can be taken as the measure of another aspect of social change.

The identification of this interaction between age and cohort effects provides additional information in cohort analysis. If the direction of this interaction effect follows a pattern for all ages at a particular point in time, then it might be identified as a period effect, because it would seem reasonable to attribute it to some social event happening at that time. The social event has upset the regular pattern of age and cohort effects and thus has caused the cohorts to change with respect to the variable being studied. These Age by Cohort Interaction Effects are important in their own right and may or may not be identified as one aspect of Period Effects, which are discussed next.

Period Effects. In studying social change the effects associated with aging, and the historical effects over time associated with cohort effects, are felt to be of primal importance. The Period Effects are defined as those changes in the dependent variable caused by some event or events at a particular period in time that cause transitory departures at all ages from the established age and cohort patterns. It is not specified that the effect is the same at all ages.[2]

It has been indicated above that if the interaction effect follows a pattern for all ages at a particular point in time, it may be identified as a period effect since Period Effects are defined as departures at all ages from the established Age and Cohort patterns.

Period effects may take a different form. They may occur as consistent deviations from the Age and Cohort patterns established by the function $Y = f(A,C)$. The identification of Period Effects may be made from an examination of the deviations from the surface described by $Y = f(A,C)$, at a particular point in time. On the basis of substantive knowledge of the dependent variable and on the basis of the deviations, one begins to formulate ideas as to the nature of the Period Effects one is trying to identify.

There are several sorts of Period Effects that one can identify theoretically, and the search procedures will vary depending on the nature of the Period Effect. For example, if the dependent variable is income, one might imagine a recession causing a drop (or slackening in increase) of median incomes at all ages, and thus a trough effect. If the mathematical form fitted to the surface is quite complex, it may fit into this trough and deviations from the surface will not help in identifying it. However, if the mathematical form is not too complex and the longer term, more stable, pattern is followed by the surface, the deviations for this point in time will all be negative. A peak effect might be similarly identified with positive deviations.

If the Period Effect is in the nature of rapid inflation, one may have a cliff effect where the median incomes rise sharply at that point in time for all ages and then continue on their previous trend but at a higher level. This type of period effect will be more difficult to identify from deviations because it will change the overall slope of the basic pattern, causing negative deviations on one side of the "cliff" and positive deviations on the other side.

If the surface, f(A,C), were not fitted by ordinary least squares but by the least absolute value approach or a related technique (see Armstrong and Frome, 1976 and in press, and Tukey, 1977), the Period Effects as defined here would be more easily identified. A case could also be made that the Age and Cohort Effects, as defined, would be more meaningful because of the surface's conformity to "pattern". Some of the possibilities in this direction are being explored.

A computer plot of the observed data over an Age-by-Cohort base may be useful in identifying Period Effects. Since period effects may be of several different sorts, there is probably no one best technique for identifying them. This is consistent with our general view of reality. Events happening at a particular point in time are of a wide variety of sorts and can produce a variety of consequences so that no one particular type of pattern can be expected as a Period Effect.

One might object to the proposed definition of Period Effects on the basis that all Period Effects aren't apparent at the time of the causative factor. The Depression, for example, had effects on children growing up at that time that were not exhibited until later in life. In the system proposed here, such effects become identified as cohort effects or age by cohort interaction effects. Period effects as defined here are the consequences of those events which have a relative transitory effect across all ages.

ILLUSTRATIVE EXAMPLE

The dependent variable examined in this illustration is the labor force participation of white females as measured by percent in the labor force. The data are for the period 1940 through 1970. The cohorts are five year age groups identified by the year in which they were 20-24 years of age. The group 45-49 in 1940 is defined as cohort 15 since they were 20-24 years old in 1915. Age is taken as the midpoint of the five year age interval. The basic data are shown in Table 1.

Since the purpose of this illustration is to demonstrate the sorts of descriptive information that can be extracted from a set of observations by use of the approach suggested above, no effort is made to develop a mathematical expression on the basis of substantive theory. So long as the expression for f(A,C) fits the observations closely, the derived information can be taken as descriptive of the original data. In this illustration f(A,C) is taken as

$$f(A,C) = a+b_1C + b_2A + b_3A^2 + b_4A^3 + b_5C^2 + b_6AC + b_7A^2C + b_8A^3C + b_9AC^2 + b_{10}A^2C^2 + b_{11}A^3C^2$$

This expression was fitted to the observations of Table 1 by ordinary least squares with a resulting multiple correlation of .984. This high correlation indicates that we have achieved a fairly accurate description of the original observations and can place some confidence in the derived descriptive information. The high multicollinearity and the lack of substantive theory used in developing the polynomial form indicate that we should not attempt to attach meanings to the specific values of the regression coefficients. Partial derivatives of the expression f(A,C) were evaluated at five year Age and Cohort intervals to produce the values of Age, Cohort, and Age by Cohort Interaction Effects shown in Tables 2, 4 and 5. If f(A,C) had been developed on the basis of substantive theory, we might have computed valued for years yet to come and a comparison of these projections with actual observations would become a test of the theory underlying the model.

Age Effects. The data of Table 2 show that recent cohorts of white females start reducing their labor force participation about age 55 and that this age has been declining. The cohort of 1910 for example didn't start reducing its labor force participation until about age 60. Also the rate of withdrawal from the labor force after these ages has been increasing in more recent cohorts; the cohort of 1910, for example, at age 62.5 was reducing its labor force participation about half a percentage point with each increase of one year in age while the cohort of 1930, at the same age, was withdrawing from the labor force by nearly three percentage points with each year increase in age. This is doubtless due to Social Security and the increasing number and quality of retirement plans.

Given the mathematical formulation it is possible to make estimates of the age at which each white female cohort begins reducing its labor force participation. This is also, of course, the point of maximum participation. For selected years between 1940 and 1970, it was determined which cohort was at its maximum point at the specified year, the age of the cohort at that point, and the level of labor force participation. (Cohorts one year in width were utilized in this exercise.) The results are shown in Table 3. The data of this table suggest that the age of maximum labor force participation for white female cohorts has been declining but seems to be stablizing at about age 54.

The age effects in Table 2 show that among white females the age effects are negative at younger ages and then become positive somewhere before age 35. This is the "labor force drop-out" of white females during childbearing ages. Table 2 indicates that the age at which this drop-out stops, a point of minimum labor force participation, has been declining. The cohort which had reached its minimum point in each of selected years was determined and the age and labor force participation at that point was computed. The results are shown in Table 3 along with the maximum points. This estimated age of minimum labor force participation has declined from 37 in 1940 to a little over 25 in 1970, and if the pattern continues future cohorts of white females will not show a "labor force drop-out". This changing pattern is the result of several factors -- declining fertility, increasing childlessness, and increasing pressure for development of maternity policies that do not require females to withdraw from the labor force during the child bearing years.

Cohort Effects. Table 4 shows us that the cohorts of 1940, 1945, and 1950 had negative cohort effects at age 22.5. Thus we can say that the proportion of white females in the labor force at age 22.5 declined from 1940 through 1950 but started increasing between 1950 and 1955 and by 1970 was increasing by approximately one and three fourths percentage points a year. Table 4 shows us that except for younger females in earlier cohorts the labor force participation of females at all ages has been increasing between 1940 and 1970. This table also shows that below age 40 this increase has been at an increasing rate but above age 40 the amount of increase each year has been declining. This suggests that there might be a plateau for level of white female labor force participation at various ages. At the oldest age, 62.5, the increase in labor force participation has been by approximately two thirds of a percentage point a year and increasing very slightly.

Age by Cohort Interaction Effects. Table 5 shows the change in cohort effect per year increase in age, or what is the same thing, the change in age effect at a given age from one cohort to the next. At all ages under 30 these interaction effects are positive which means that the change in white female labor force participation per year increase in age is changing in a positive direction from one year to the next. Those white females who were 22.5 in 1945 were dropping out of the labor force by nearly three percentage points with each year increase in age, but this positive interaction term tells us that this rate of change was becoming less negative (more positive) with each succeeding year. The change in labor force participation with increasing age has been becoming more positive with each succeeding year for all white females under 30 during the period 1940 through 1970.

If we look at ages over 50 we find the interaction terms all negative. The rate at which cohorts over 50 are changing their labor force participation with increasing age has become less positive with each succeeding year. Women at 52.5 are still increasing their proportion in the labor force with each increasing year of age, but from year to year in time the amount of increase is less. Since females over 55 are dropping out of the labor force with increasing age, this negative interaction effect means that each succeeding year in time they are withdrawing from the labor force at increasing rates.

Period Effects. Table 6 shows the deviations between the observed points and the surface described by the regression polynomial. These deviations are arranged by time periods because the most likely sort of period effect would result in nearly all positive (or negative) deviations at some point in time. The data of Table 6 show no indication of a period effect although it is possible that if some method other than least squares had been used to fit the regression surface, a period effect might have shown up.

Limitations. The utility of the method outlined and illustrated is still to be fully explored but it would appear to be fruitful for variables such as labor force participation, income, etc. The development of mathematical forms with substantive meaning will doubtless pose problems but may stimulate more rigorous thinking on the nature of age and cohort effects. Two different mathematical forms, each of which correlates with the observations as closely as the example above, will probably yield similar interpretations, though one may provide more meaningful projections. This approach would not seem particularly applicable to a variable such as fertility where for most individuals the relevant value of the variable is its cumulative value. However, a cumulative measure of fertility might be used with a generalized Pearl-Reed Growth curve as the mathematical form. (An analogous situation would be trying to study mean or median income in a cohort when the variable that was most relevant to each individual was the total earned since birth, with individuals stopping work when a certain grand total had been earned.) A careful examination of a computer drawn perspective plot of the original observations or of the surface described by the basic regression equation might yield many of the same interpretations made from Tables 2-6, but the actual magnitude of the changes taking place would not be known.

Conclusions. Definitions of age effects, cohort effects, and age by cohort interaction effects have been developed in mathematical form for use in cohort analysis. Period effects have been defined independently of these and suggestions have been made for identifying period effects depending on their nature. The method has been illustrated utilizing the labor force participation of white females for the period 1940 through 1970. The approach suggested seems to offer real possibilities for use in cohort analysis. The approach also points up the need for continuing examination of the relationship between substantive theory and mathematical form.

FOOTNOTES

[1]This suggested operational definition of social change is subsumed under that suggested by Ryder (1964). He says "The definition of social change prompted by these considerations is the modification of processual parameters from cohort to cohort." (p. 461) His discussion indicates that he includes "short-run change," which he identifies with a "period-specific event," along with "long-run change," which "is characterized by differences in functional form from cohort to cohort other than those betraying the characteristic age pattern." The definition of social change suggested by the present author is not identifical to Ryder's "long-run change" but is closely related.

[2]This point of view is similar to that of Ryder (1964) in his discussion of short-run changes. ". . . the manifestation of a period-specific event or situation that 'marks' the successive cohort functions at the same time, and thus at successive ages of the cohorts in question. Frequently such manifestations take the form of fluctuations, in the sense that a counteracting movement occurs subsequently, which erases the impact of the situation in the eventual summary for the cohort." (p. 462)

REFERENCES

Armstrong, R. D. and E. L. Frome. "Least absolute value estimators for one-day and two-way tables." Center for Cybernetic Studies Research Report CCS 266. Austin: The University of Texas, 1976.

______. "A comparison of two algorithms for absolute deviation curve fitting." Journal of the American Statistical Association. Accepted for publication.

Glenn, Norval D. "Some cautions concerning statistical attempts to separate age, period, and cohort effects." Paper given at American Political Science Association meeting, San Francisco, Sept. 2-5, 1975.

Mason, Karen Oppenheim, William M. Mason, H.H. Winsborough and W. Kenneth Poole. "Some methodological issues in cohort analysis of archival data." American Sociological Review 38 (April, 1973) 242-258.

Riley, Matilda White, Marilyn Johnson and Anne Foner. Aging and Society, Vol. III: A Sociology of Age Stratification. New York: Russell Sage Foundation, 1972.

Ryder, Norman B. "Notes on the concept of a population." The American Journal of Sociology 59 (March, 1964) 447-463. Reprinted in Riley, et. al., 1972.

______. "The cohort as a concept in the study of social change." American Sociological Review 30 (December, 1965) 843-861.

______. "Cohort analysis." The International Encyclopedia of the Social Sciences. Vol. 2. New York: The Macmillan Company and the Free Press, 1968. 546-550.

Tukey, J. W. Exploratory Data Analysis. Reading, Mass.: Addison-Wesley, 1977 (Limited preliminary edition, 1970).

Table 1. Percent of White Females in the Labor Force by Age and Cohort, 1940 to 1970. (Cohorts identified by year in which members were 20-44.)

Cohort	Age 20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64
1970	56.4								
1965		55.9							
1960	44.8		43.7						
1955		33.5		42.1					
1950	43.6		33.5		49.0				
1945		31.3		38.5		52.2			
1940	45.7		29.1		43.9		51.5		
1935		34.2		32.1		46.4		47.1	
1930			29.1		34.9		45.1		35.9
1925				26.1		33.6		39.1	
1920					24.0		29.8		29.0
1915						21.9		25.2	
1910							19.8		20.0
1905								17.4	
1900									13.9

Table 2. Age Effects by Age and Cohort for Labor Force Participation of White Females, 1940 to 1970. (Age effect is the percentage points change in labor force participation per year increase in age.)

Cohort	Age 22.5	27.5	32.5	32.5	42.5	47.5	52.5	57.5	62.5
1970	.02								
1965	- .42	.17							
1960	- .94	.06	.74						
1955	-1.53	-.16	.78	1.31					
1950	-2.21	-.50	.71	1.40	1.58				
1945	-2.97	-.94	.50	1.37	1.64	1.34			
1940	-3.81	-1.50	.18	1.21	1.61	1.37	.49		
1935		-2.16	-.27	.94	1.48	1.34	.52	-.97	
1930			-.85	.55	1.24	1.24	.55	-.85	-2.94
1925				.03	.91	1.09	.57	-.66	-2.58
1920					.48	.87	.58	-.40	-2.06
1915						.59	.59	-.07	-1.38
1010							.59	.32	- .55
1905								.78	.44
1900									1.59

Table 3. Estimated Ages of Minimum and Maximum Labor Force Participation Between Ages 22 and 65 for White Female Cohorts, and Percent in Labor Force at those Points for the Period 1940-1970

Year	Minimum point Age	Minimum point Percent	Maximum point Age	Maximum point Percent
1940	37.0	26.7		
1945	33.8	28.3		
1950	31.5	29.9	56.8	28.5
1955	29.8	32.4	55.5	36.2
1960	28.2	37.1	54.8	42.6
1965	26.8	44.5	54.5	47.8
1970	25.4	53.5	54.4	52.2

Note: The maxima and minima are for cohorts, not for years. The cohort of white females that entered the labor force about 1916 reached maximum labor force participation at age 56.8 in 1950, their maximum being lower than the minimum labor force participation of the cohort that entered the labor force about 1941 and reached a minimum at age 31.5 in 1950.

Tables 4-6 are available from author: Dept. of Sociology, University of Texas, Austin, Tx. 78712.

TESTING FOR CONTEXTUAL EFFECT

C. H. Proctor, North Carolina State University

1. Introduction

In a pair of recent papers appearing in the journal Sociological Methods and Research (Farkas, 1974 and Hauser, 1974) the two authors disagreed on the attribution of belief in a, so called, contextual effect for a given set of data. Farkas decided it was there, and Hauser maintained it was not established. They said much more and, in fact, the two papers constitute an excellent review of contextual effect analysis. Hauser admitted that "...the effects are statistically significant at the usual α-levels, since the sample has about 5,300 cases," (p.368) and then went on to argue that statistical significance is not practical significance. However, that t-value was around 10, and I feel that such statistical significance as $t = 10$ exhibits (even with 5,300 cases) almost has to be of practical significance as well. The largeness of this t-value, however, may be due to their use of an inappropriate model equation. In short, although there were 5,300 students, they came from only 99 colleges and estimation of a context effect as a contrast on college means may better be viewed as based on a sample size of 99 than of 5,300.

The following discussion is an attempt to summarize some current methods available for testing for contextual effect. The first two are truly rough and ready approaches. The first one struggles to obtain balanced data by deleting and duplicating at random, while the other is a version of the celebrated Tukey Jackknife. The next two use variance component models and exhibit the rather involved considerations that seem to arise when estimating components of variance with unbalanced data along with testing a regression coefficient in a mixed model. One of these latter two estimates the variance components by equating mean squares to their expectations and then tests by use of generalized least squares, while the final method is simply likelihood all the way. The situation is an instructive example of a data handling problem that is quite common in social survey analysis and although no new methods are here proposed, it may be useful to see what already-existing statistical practice seems to suggest can be done.

2. Model Equations

In a concrete example treated in the two papers, scores were given for his answers on a questionnaire to each student on his attitude toward drinking (the independent or X-variable) and on the extent of his drinking behavior (the dependent or Y-variable). As one would expect, these two measures were rather closely related. The question posed was whether the between colleges regression slope of behavior on attitude was equal to the within colleges slope. There is suspicion that the between colleges slope may be steeper due to some cumulative, reasonance interpersonal phenomena as described by (Farkas,1974, p.339ff). This is called contextual effect. A difference in slopes may be due to enumerator or sample selection biases or to omitted variable effects or to many other causes but the data handling methods would still be called testing for contextual effect.

It is by no means uncommon to do regression analysis in the presence of a cluster or nested sample design. One simply removes the cluster or "block" effects (schools, counties, area segments or whatever) and operates upon the within cluster cross products with perhaps a pious expression of hope that there are no interactions between slopes and clusters (at any rate I have done so in S-63, 1974, p.85). In the present case of being interested in the contextual effect there is no way to avoid the possibly upsetting presence of confounded college effects. To see how they enter let's consider model equations.

Regression calculations would be governed by a model equation such as·

$$(1)\quad y_{ij} = \mu' + x_{ij}\beta + \bar{x}_{i.}\delta + e_{ij}, \text{or (redefining } \mu)$$

$$= \mu + (x_{ij} - \bar{x}_{..})\beta + (\bar{x}_{i.} - \bar{x}_{..})\delta + e_{ij}\ ,$$

where y_{ij} is the score on his drinking behavior reported by the j^{th} student among n_i at the i^{th} college where $i = 1,2,\ldots,t$. His attitude score is written x_{ij}; the college mean is written $\bar{x}_{i.}$ and the grand mean is denoted $\bar{x}_{..}$. The e_{ij} quantities ($n = \sum^{t} n_i$ of them) are taken to be independent, identically distributed with variance σ_e^2 . The objective of the analysis is to test H_0: $\delta = 0$.

I suspect that including a term to be written u, where the u_i's $i = 1,2,\ldots,t$ are independent, identically distributed with variance σ_u^2 , would allow the model to fit more closely to the actual data. This issue should be decided by examining the data although one's judgement based on past experiences with similar situations should enter also. The model equation then becomes

$$(2)\quad y_{ij} = \mu + (x_{ij} - \bar{x}_{..})\beta + (\bar{x}_{i.} - \bar{x}_{..})\delta + u_i + e_{ij}\ .$$

Suppose that the data have already been cleaned (of outliers, for example) and possibly some transformations made to assure linearity so that least squares estimation is not too far from optimal. There are two, somewhat idealistic, circumstances in which the test of $\delta = 0$ becomes relatively simple to compute. We consider these first and show how they can become the bases for two, more realistic although approximate, corresponding methods of analysis.

3. Two Rough and Ready Methods

One of the ideal cases is of balanced data. That is, if all colleges furnished the same number of students then the computational methods of the analysis of covariance (Fisher, 1958) become available. For example, Snedecor and Cochran's textbook (1967,pp.436-438) includes formulas for estimating σ_u^2 (called σ_B^2 there) and for testing equality of the between classes slope to the within classes slope. Although the present data set does not have equal sample sizes in every college it would likely be possible to consider a number of colleges having almost equal (say the ratio of largest to smallest is less than 3 to 1) sample sizes. Then one may randomly delete some cases in the schools with larger sample sizes and perhaps randomly duplicate some in the others and so end with a balanced data set without having damaged to any appreciable extent the information on the context effect. This possibility is discussed at some length by Searle (1971,p.364). Before recommending widespread use of this procedure one should investigate, at least through empirical sampling, the effects of varying amounts of deleting and duplicating.

The other ideal case involves the independent replication of the entire sample design in both the sample selection as well as the measurement and tabulation phases. Such designs have been advocated by Deming (1960, Part II, Replicated Sampling Designs) and are in common use, for example, for sampling business accounts whenever there is great, perhaps legal, need of an unbiased estimate of sampling variance. In such a case one would compute separate estimates of the parameter δ on each sub-sample and then compute a t-statistic to test that the population mean of these estimates is zero.

In practice, replicated designs may be deemed expensive and cumbersome to carry out or may not have been used because variance estimation was given low priority. In such cases the data can still be broken into pseudo-replicates (McCarthy, 1966) or portions and then one can use the above-described t-test procedure or some other method. One of these other methods is the Tukey Jackknife (Mosteller and Tukey, 1968) that proceeds as follows. One portion of the data is deleted, an estimate of the context effect is computed from the remainder and a pseudo-value is then formed. This portion is returned to the sample, another one deleted, and again a pseudo-value is computed. This is done until we have as many pseudo-values as portions. The pseudo-values are then used to compute a t-value as a test statistic for the hypothesis that the population average pseudo-value is zero. It is important to define "portions" so as to reflect the principal source of sampling variance in the application considered. In the present example a school is a portion. Each pseudo-value equals the-number-of-schools times the-estimate-of-context-effect-from-all-schools minus the-number-of-schools-less-one times the-estimate-based-on-all-schools-except-one.

4. Regression Approach

We return now to a more direct attack guided by model equation (2) and first using methods of estimating variance components described by Searle (1971) in conjunction with a generalized least squares procedure to estimate δ and to test that $\delta = 0$. First, however, we might ask how the test based on the least squares regression calculation that was employed in the cited papers performs when model equation (2) holds. If we adopt a more concise notation of $d_{ij} = x_{ij} - \bar{x}_i$ and $c_i = \bar{x}_i - \bar{x}_{..}$ then model equation (2) can be written in terms of deviates as:

$$(3) \quad y_{ij} = \mu + \beta\, d_{ij} + \delta\, c_i + u_i + e_{ij} ,$$

with $\sum\sum_{i,j} d_{ij} = 0$ and $\sum\sum_{i,j} c_i = 0 = \Sigma n_i c_i$. Notice that $\sum_{j=1}^{n_i} d_{ij}/n_i = c_i$.

The regression computations proceed from the 4 by 4 matrix of sums of squares and cross products. This is shown in partitioned form just prior to removing the β and μ effects as:

$$(4) \quad \begin{array}{c} \text{Effects:} \\ \\ A = \end{array} \begin{array}{cc:cc} y & \delta & \beta & \mu \\ TYY & BXY & TXY & y_{++} \\ BXY & BXX & BXX & 0 \\ \hdashline TXY & BXX & TXX & 0 \\ y_{++} & 0 & 0 & n \end{array} = \begin{array}{cc} A_{11} & A_{12} \\ A_{21} & A_{22} \end{array}$$

Partitioning A into 2 by 2 sub-matrices as shown in (4) and sweeping out the μ and β effects corresponding to the last two rows and columns, leaves the following residual cross products matrix.

$$(5) \quad A_{11\cdot2} = A_{11} - A_{12}A_{22}^{-1}A_{21}$$

$$= \begin{array}{cc} TYY - y_{++}^2/n - (TXY)^2/TXX & BXY - (TXY)(BXX)/TXX \\ BXY - (TXY)(BXX)/TXX & BXX - (BXX)^2/TXX \end{array} .$$

Next sweeping out the now-adjusted δ effect from $A_{11\cdot2}$ yields an estimate of δ as:

$$\hat{\delta} = \frac{[BXY - (TXY)(BXX)/TXX]}{[BXX - (BXX)^2/TXX]}$$, and the following

regression sum of squares (call it DSS) for a test of $\delta = 0$:

$$(6) \quad DSS = [BXY - (TXY)(BXX)/TXX]^2/(BXX - BXX^2/TXX).$$

In the absence of the u_i terms this quantity, DSS, would be taken to have a single degree of freedom chi-square distribution and could be divided by the error mean square, say $RSS/(n-3) = RMS$, to furnish an F-ratio test (or, since $t = \sqrt{F}$ when numerator degrees of freedom are one, a t-test) statistic for the hypothesis $\delta = 0$. This is likely to have been the calculation that led to the t = 10 value in Farkas (1974), and will be denoted $t_1 = \sqrt{F_1} = \sqrt{DSS/RMS}$.

In case that model equation (2) holds with $\sigma_u^2 \neq 0$ we find E(DSS) as follows. First, express DSS as:

$$(7) \quad DSS = (TXX\ BSY - TXY\ BXX)^2/BXX(TXX-BXX)TXX .$$

Writing $WXX = TXX - BXX$, introducing the expectation, and substituting for y_{ij} from model equation (2) leads to:

$$(8) \quad BXX\ WXX\ TXX\ E(DSS) = (BXX\ WXX)^2\delta^2 + WXX^2\Sigma n_i^2 c_i^2 \sigma_u^2 + BXX\ WXX\ TXX\ \sigma_e^2 .$$

The final result is:

$$(9)\quad E(DSS)=\frac{BXX\ WXX}{TXX}\,\delta^2+\frac{\Sigma n_i^2 c_i^2 WXX}{BXX\ TXX}\,\sigma_u^2+\sigma_e^2 .$$

On the other hand as we will see below in (15):

$$(9a)\quad E(RMS)=\sigma_e^2+[n-tr(Z'X(X'X)^{-1}X'Z)]\sigma_u^2/(n-3),$$

where Z and X in this formula refer to data matrices. The coefficient of σ_u^2 in E(DSS) is generally larger than that of σ_u^2 in E(RMS), and thus the ratio F_1 is biased upward from 1 .

It is thus evident that only if $\sigma_u^2 = 0$ (i.e., if model equation (1) holds) will the regression approach be correct. If $\sigma_u^2 \neq 0$ then to form an F-ratio test statistic one needs to search elsewhere than the error mean square (RMS) for a denominator. If one examines the expected value of the residual sum of Y-squares after removal of DSS, namely of:

$$(10)\quad RSS = TYY-y_{++}^2/n - (TXY)^2/TXX - DSS ,$$

it is found that both σ_u^2 and σ_e^2 appear, but no fixed effects do. By next sweeping out the college effects as well, one can obtain an even more refined error sum of squares, (ESS,say) and thus calculate an unbiased estimate of σ_e^2 alone. Using this, along with the coefficients of σ_u^2 and σ_e^2 in E(RSS-ESS), allows one to get an estimate of σ_u^2. It bears emphasizing that one should examine the F-ratio in which the mean square (RSS-ESS)/(t-2) is divided by the error mean square, as a test of $\sigma_u^2 = 0$. If rejection of the hypothesis $\sigma_u^2 = 0$ is not possible even at the, say, 20% level, then one may proceed to act as if model equation (1) held and perform a t-test using $\hat{\delta}$ and the error mean square EMS=ESS/n-t-1 . I would be fairly reluctant to suppose that $\sigma_u^2 = 0$ and so would not recommend this test. At any rate it may be denoted $t_2 = \sqrt{DSS/EMS}$.

5. Estimating Variance Components for the Test

A denominator for a F-ratio statistic with DSS as numerator can be constructed using method of moments estimates based on ESS and RSS of σ_u^2 and σ_e^2 or more elaborate estimates of σ_u^2 and σ_e^2. The notation and most of the methods are to be found in Searle's textbook (1971), and we will first move toward that notation.

Model equation (2) can be written in matrix form as (see p.465 in Searle,1971)

$$(11)\quad \underset{\sim}{y} = X\underset{\sim}{b} + Z\underset{\sim}{u} + \underset{\sim}{e} ,$$

which in the present case becomes:

$$(12)\quad \begin{bmatrix} y_{11} \\ y_{12} \\ \cdot \\ \cdot \\ \cdot \\ y_{t,n_t} \end{bmatrix} = \begin{bmatrix} 1 & d_{11} & c_1 \\ 1 & d_{12} & c_1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & d_{t,n_t} & c_t \end{bmatrix} \begin{bmatrix} \mu \\ \beta \\ \delta \end{bmatrix} + \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & & 0 \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \cdot \\ \cdot \\ \cdot \\ u_t \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ \cdot \\ \cdot \\ \cdot \\ e_{t,n_t} \end{bmatrix}$$

Searle assumes that the only linear relationship of columns of X to those of Z is by way of the first column of X equal to the sum of the columns of Z. In our case there is yet another relationship so that we must generalize his derivation as follows. Let r be the rank of X and t (the number of PSU's) is, of course, the rank of Z. Then define λ_1 to be

$$(13)\quad \lambda_1 = r + t - r[X\ Z] ,$$

where r[X Z] is the rank of the combined matrix [X Z]. In our example $\lambda_1 = 2$, but we might as well be prepared for dealing with numerous independent variables since the contextual effect may be multi-componential.

What had previously been denoted as TYY is now written as $\underset{\sim}{y}'\underset{\sim}{y}$. The quantity R(b) will refer to the reduction in sum of y-squares due to fitting the model:

$$(14)\quad \underset{\sim}{y} = Xb + e ,$$

while R(b,u) will refer to the reduction achieved by fitting the full model of rank r[X Z]. The additional reduction may be written R(u|b) = R(b,u)-R(b) and upon using formula (79) in Searle (1971,p.445) its expectation is found to be:

$$(15)\quad E(R(u|b))=\sigma_u^2(n-tr[Z'X(X'X)^{-}X'Z])+\sigma_e^2(t-\lambda_1) ,$$

while

$$(16)\quad E(\underset{\sim}{y}'\underset{\sim}{y})-R(b,u) = [N-(r + t - \lambda_1)]\,\sigma_e^2 .$$

In our example, X is of full rank and upon extending the method to cover more than one independent variable when the observed values are obtained by survey methods (rather than from a balanced experiment), it will continue to be of full rank so that we could have written $(X'X)^{-1}$ rather than just $(X'X)^{-}$, the generalized inverse, as appears in the textbook.

To test $\sigma_u^2 = 0$ one computes the ratio

$$(17)\quad F = \frac{R(u|b)}{t-\lambda_1} \Big/ \frac{(\underset{\sim}{y}'\underset{\sim}{y} - R(b,u))}{N-(r + t-\lambda_1)}$$

and refers it to a table of the $F(t-\lambda_1, N-(r + t-\lambda_1))$ distribution.

The estimates can be written explicitly as:

$$(18)\quad \hat{\sigma}_e^2=(\underset{\sim}{y}'\underset{\sim}{y}-R(b,u))/[N-(r + t - \lambda_1)]$$

$$(19)\quad \hat{\sigma}_u^2=(R(u|b)-(t-\lambda_1)\hat{\sigma}_e^2)/(n-tr[Z'X(X'X)^{-1}X'Z]).$$

At this point one could refer to expression (9) and using the values obtained for $\hat{\sigma}_e^2$ and $\hat{\sigma}_u^2$, compute $[\Sigma n_i^2 c_i^2\ WXX\ \hat{\sigma}_u^2/BXX\ TXX + \hat{\sigma}_e^2]$ which is then divided into DSS to get an F-ratio test statistic.

In practice, one could perhaps improve the estimates of σ_e^2 and σ_u^2 in accord with Thompson (1969), by first computing the ratio

$\hat\lambda = \hat\sigma_e^2/\hat\sigma_u^2$ and then new value of $R(b,u)$ and $R(u|b)$ as $R^*(b,u)$ and $R^*(u|b)$ based on the maximum likelihood equations for b (Searle,1971, section 11.7c). Upon iteration the estimators may be denoted with an upper squiggle as $\tilde\sigma_u^2$ and $\tilde\sigma_e^2$. They can be written explicitly as:

$$(18a)\quad \tilde\sigma_e^2 = (\underset{\sim}{y}'\underset{\sim}{y} - R^*(b,u))/(N - r)$$

$$(19a)\quad \tilde\sigma_u^2 = R^*(u|b)/c \ ,$$

where c is the same denominator as for $\hat\sigma_u^2$ in equation (19).

We can now use generalized least squares to test H_0. The maximum likelihood equations for estimating the fixed effects are:

$$(20)\quad \begin{matrix} X'X & X'Z \\ Z'X & P \end{matrix} \quad \begin{matrix} \underset{\sim}{b}^* \\ \underset{\sim}{u}^* \end{matrix} = \begin{matrix} X'y \\ Z'y \end{matrix}$$

where $P = Z'Z + \lambda I$ and $\lambda = \sigma_e^2/\sigma_u^2$. The appearance of λI is the new ingredient. From

$$(21)\quad X'X\,\underset{\sim}{b}^* = X'(\underset{\sim}{y} - Z\,\underset{\sim}{u}^*) \text{ and } P\underset{\sim}{u}^* = Z'(\underset{\sim}{y} - X\,\underset{\sim}{b}^*)$$

we get

$$(22)\quad X'X\,\underset{\sim}{b}^* = X'(\underset{\sim}{y} - Z\,P^{-1}Z'\underset{\sim}{y} + X'Z\,P^{-1}Z'X\,\underset{\sim}{b}^*$$

or

$$(23)\quad X'(I - Z\,P^{-1}Z')X\,\underset{\sim}{b}^* = X'(I - Z\,P^{-1}Z')\underset{\sim}{y}$$ or finally:

$$(24)\quad \underset{\sim}{b}^* = (X'T\,X)^- X'T\,\underset{\sim}{y} \text{ where}$$
$$T = (I - Z\,P^{-1}Z')'.$$

The matrix $\underset{\sim}{b}^*$ contains estimates of the parameters μ, δ and β (call them μ^*, δ^* and β^*) and its covariance matrix is estimated by $\hat\sigma_u^{-2}(X'T\,X)^{-1}$. Therefore a test of $\delta = 0$ can be based on the ratio of δ^* and its standard error obtained from $\hat\sigma_u^{-2}(X'T\,X)^{-1}$.

6. Maximum Likelihood

It was most helpful to have someone do the differentiations for a maximum likelihood solution as Jennirch and Sampson (1976) have provided. It may be that the method of maximum likelihood is justly criticized for its small sample properties. In the present case of variance estimation it provides negatively biased estimators since division is by n rather than n-1 . It does however, furnish a lot of useful results in such a situation as the present one where nuisance parameters (σ_e^2 and σ_u^2 as well as μ and β) abound. Their (Jennirch and Sampson's) paper (and previous ones by Hartley and others that they cite) are complete and explicit so that I would only show my tenuous grasp of the material by paraphrasing it.

7. Discussion

There are a number of general observations on data handling that this problem of testing for contextual effect prompts. The first is the recognition of some difference in point of view or, better, of methods between the pseudo-replicate or jackknife methods and the variance component estimation methods. There are many sources of this difference. One is more data-based and the other is more model-based, although both have models and both worry about data collection and processing. One proceeds from the survey sampling side and the other from the analysis of experimental data.

My own bias is toward favoring the components of variance approach and this paper (in its relative neglect of the literature on pseudo-replication) reflects that bias. I do believe that in the face of the complexities of sample design and measurement techniques, not to mention the population distributions, of the variables included in most social surveys the pseudo-replicate approach is the only contender in practice. The method does demand a considerable amount of judgement and experience to form replicates which will faithfully represent the uncertainty in the data. One variation of the method that I confess I do use in practice is to run regression on individual subjects and then correct t-values and chi-squares with the ratio of actual sample size to effective sample size. I use the results of Kish and Frankel (1974) that showed such ratios of 1.4 or so for regression coefficients.

The reason that I favor the variance component approach is its explicit attention to the model and particularly to its distributional assumptions not just the systematic part. There are, I suspect, a good many more components of variance floating around that should be included both for making honest tests and also for improving on future surveys. The controversy over tests of significance that bubbled in the sociological literature a few years back may have been fueled in part from misjudgements about the modeling of variances. Given the level of complexity in variances that was mentioned, it is no wonder to me that there have been serious mistakes in calculating levels of significance and I'm sure I've contributed to them. The solution is certainly not to scrap the tests, but rather to study variances harder.

A brief point needs to be made about computing requirements of the methods. The balanced data case is the least demanding and should be chosen by the researcher who wants quick and cheap, but honest results. I think that the jackknife also can be done by an investigator operating with limited computing resources. In doing the calculations for maximum likelihood and the regression-type computations I used a procedure, PROC MATRIX, from the software package known as SAS (Barr, Goodnight, Sall, Helwig,1976), and I can report that it was only moderately painful to write, while recognizing that my skills at programming are quite limited.

Returning to the problem as an exercise in data handling, it bears remarking again that no new methods were developed. The model is the mixed one. It has the, perhaps novel, feature of an additional linear dependence between the X matrix and the Z matrix (or U matrix as written by Hartley, et.al.). That is, the sum of the columns of Z equaling the constant column of X is the usual dependency, while the context effect (being a linear combination of group means) is a second dependency.

REFERENCES

Barr, A. J., Goodnight, J. H., Sall, J. P. and Helwig, J. H. (1976) A User's Guide to SAS76. Sparks Press, Raleigh, N. C.

Deming, W. E. (1960) Sample Design in Business Research. New York: Wiley.

Farkas, G. (1974) "Specification, residuals and contextual effects", Sociological Methods and Research, 2 (Feb.): 333-363.

Fisher, R. A. (1958) Statistical Methods for Research Workers. 13th edition, New York: Hafner.

Hauser, R. M. (1974) "Contextual analysis revisited", Sociological Methods and Research, 2 (Feb.): 365-375.

Kish, L. and Frankel, M. R. (1974) "Inferences from Complex Samples" Journal of Royal Statistical Society, A, 137: 1 - 37.

Mosteller, F. and Tukey, J. W. (1968) "Data Analysis" in Lindsey, G. and Aronson, E.,Eds., The Handbook of Social Psychology, Vol. 2, second edition. New York: Addison-Wesley.

Searle, S. R. (1971) Linear Models, New York: Wiley.

Snedecor, G. W. and Cochran, W. G. (1967) Statistical Methods, 6th edition, Ames, Iowa: Iowa State University Press.

Southern Region Research Project S-63, (1974), Influences on Occupational Goals of Young People in Three Southern Sub-cultures, Mimeo, Agricultural Experiment Stations of Alabama, Kentucky, Mississippi, North Carolina, South Carolina, Tennessee, and Virginia.

Thompson, R. (1969) "Iterative estimation of variance components for non-orthogonal data: Biometrics, 25: 767 - 773.

McCarthy, P. J. (1966) "Replication: An approach to the analysis of data from complex surveys", Vital and Health Statistic. PHS Publication, No. 1000, Series 2, No. 14, Washington, D. C.

An expanded version of the paper with an appendix containing examples of the computations is available from the author upon request.

THREE-STAGE SAMPLING ON SUCCESSIVE OCCASIONS

Dharam S. Rana, Old Dominion University
and R. P. Chakrabarty, Jackson State University

Introduction -- To maintain an up-to-date study of a time-dependent population, technique of sampling over successive occasions is used. In the present paper, the theory of sampling on successive occasions for a three-stage sampling design has been developed. In the case of three-stage successive sampling, there are about twelve sampling procedures available to alter the composition of the current sample. The estimates of population mean on the second occasion, the change between two consecutive occasions and the overall mean for two occasions are obtained for four selected procedures namely (5), (7), (8) and (11). The limited space did not permit the inclusion of the results for the overall mean.

2.1 Consider a population Π consisting of N primary stage units (PSU's) and each PSU consisting of M second-stage units (SSU's) and each SSU containing K third-stage units (TSU's). In the present study it is assumed that the population units are fixed, the variances on the two consecutive occasions are equal, the sample sizes remain same on each occasion and N, M and K are large so that the finite population correction factors at all three stages are negligible. The discussion is confined to two occasions only although the results obtained can be extended to more than two occasions. It is further assumed that the selection of the first sample, which consists of n PSU's, m SSU's within each of the n PSU's and k TSU's within each of the nm SSU's, is carried out by the method of simple random sampling without replacement and this applies to all the four sampling plans considered in this paper.

2.2 Estimate of the Mean by Procedure (5). On the second occasion, retain all the PSU's of the first sample but retain only a fraction r of the SSU's with their samples of TSU's in each of the PSU retained and select afresh a fraction s of SSU's such that r + s = 1. Let $y_{hij\ell}$ denote the value of the response variable y for the ℓ-th tertiary unit in the j-th second-stage unit within the i-th first-stage unit on the h-th occasion. A general linear unbiased estimator of $\bar{Y}_2$ the population on the second occasion may be written as $\bar{y}_{2(5)} = a\,[\bar{y}'_{1(5)} - \bar{y}^*_{1(5)}] + c\,\bar{y}'_{2(5)} + (1-c)\,\bar{y}^*_{2(5)}$ where $\bar{y}'_{h(5)} = \frac{1}{nrmk}\sum_{i=1}^{n}\sum_{j=1}^{rm}\sum_{\ell=1}^{k} y_{hij\ell}$ $\quad h = 1,2 \quad$ $\bar{y}^*_{h(5)} = \frac{1}{nsmk}\sum_{i=1}^{n}\sum_{j=1}^{sm}\sum_{\ell=1}^{k} y_{hij\ell}$, h = 1, 2. The variance of $\bar{y}_{2(5)}$ is given by $\mathrm{Var}\,[\bar{y}_{2(5)}] = a^2\,[\frac{Sw^2}{nrm} + \frac{St^2}{nrmk} + \frac{Sw^2}{nsm} + \frac{St^2}{nsmk}] + c^2\,[\frac{Sb^2}{n} + \frac{Sw^2}{nrm} + \frac{St^2}{nrmk}] + (1-c)^2\,[\frac{Sb^2}{n} + \frac{Sw^2}{nsm} + \frac{St^2}{nsmk}] + 2ac\,[\rho_b\frac{Sb^2}{n} + \rho_w\frac{Sw^2}{nrm} + \rho_t\frac{St^2}{nrmk} - \rho_b\frac{Sb^2}{n}] + 2\,c\,(1-c)\,\frac{Sb^2}{n}$ where $Sb^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\bar{Y}i.. - \bar{Y}...)^2$ = true variance among PSU means, $Sw^2 = \frac{1}{N(M-1)}\sum_{i=1}^{N}\sum_{j=1}^{M}(\bar{Y}ij - \bar{Y}i..)^2$ = true variance among SSU means, $St^2 = \frac{1}{NM(K-1)}\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{\ell=1}^{K}(y_{ij\ell} - \bar{Y}ij.)^2$ = true variance among TSU's. $\rho_b\,Sb^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\bar{Y}_{1i}.. - \bar{Y}_1...)(\bar{Y}_{2i}.. - \bar{Y}_2...)$ = true Covariance among PSU means, $P_w\,Sw^2 = \frac{1}{N(M-1)}\sum_{i=1}^{N}\sum_{j=1}^{M}(\bar{Y}_{1ij.} - \bar{Y}_{1i..})(\bar{Y}_{2ij.} - \bar{Y}_{2i..})$ = true covariance among SSU means. $\rho_t\,St^2 = \frac{1}{NM(K-1)}\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{\ell=1}^{K}(y_{1ij\ell} - \bar{Y}_{1ij.})(y_{2ij\ell} - \bar{Y}_{2ij.})$ = true covariance among TSU's.

In the above formulae, ρ_b, ρ_w and ρ_t denote the true correlation coefficients among PSU means, SSU means and TSU's respectively and $\bar{Y}_h...$, $\bar{Y}_{hi}..$ and $\bar{Y}_{hij}.$ represent true mean of the Psu means true mean of the i-th PSU and the true mean of the j-th SSU in the i-th PSU on the h-th occasion (h = 1,2) respectively. The optimum weights a_o and c_o that will minimize the variance of $\bar{y}_{2(5)}$ are

$$a_o = \frac{-rs\,\alpha_o\,\beta_o}{\alpha_o^2 - s^2\beta_o^2} \quad \text{and} \quad c_o = \frac{r\,\alpha_o^2}{\alpha_o^2 - s^2\beta_o^2} \quad \text{where}$$

$$\alpha_o = Sw^2 + \frac{St^2}{k} \text{ and } \beta_o = \rho_w\,Sw^2 + \rho_t\,\frac{St^2}{k} \qquad (2.2.1)$$

It should be noted here that both $\bar{y}_{2(5)}$ and its variance are independent of ρ_b. With optimum weights, the variance of $\bar{y}_{2(5)}$ is given by

$$\mathrm{Var}\,[\bar{y}_{2(5)}] = \frac{1}{n}\,[Sb^2 + \frac{\alpha_o}{m}\,\frac{(\alpha_o^2 - s\,B_o^2)}{(\alpha_o^2 - s^2\beta_o^2)}] \qquad (2.2.2)$$

The optimum replacement fraction s_o that would minimize this variance is

$$s_o = 1/[1 + (1 - \frac{\beta_o^2}{\alpha_o^2})^{1/2}]$$

It may be easily shown that $s_o \geqslant 1/2$. There are two special cases of interest here e.g., (i) s = 0 and (ii) s = 1. In either case, it follows from (2.2.2) that $\mathrm{Var}[\bar{y}_{2(5)}] = \frac{1}{n}[Sb^2 + \frac{\alpha_o}{m}]$. Thus it is clear that a complete retention or complete replacement of SSU's within the PSU's from the first sample on the second occasion does not help to improve on the estimate of the current population mean.

2.3 Estimate of the Mean by Procedure (8)

From the first sample retain only a fraction p of the PSU's along with their samples of SSU's and TSU's on the second occasion. Replace the remaining fraction q (such that p + q = 1) of the PSU's by a fresh random selection of PSU's on the second occasion.

A general linear unbiased estimator of $\bar{Y}_2$ may be

$$\bar{y}_{2(8)} = a\,[\bar{y}'_{1(8)} - \bar{y}''_{1(8)}] + c\,\bar{y}'_{2(8)} + (1-c)\,\bar{y}''_{2(8)}$$

and its variance is given by

$$\mathrm{Var}\,[\bar{y}_{2(8)}] = a^2[\frac{\alpha}{np} + \frac{\alpha}{nq}] + c^2\frac{\alpha}{np} + (1-c)^2\frac{\alpha}{nq} + 2ac\frac{\delta}{np}.$$

where $\bar{y}'_{h(8)} = \frac{1}{npmk}\sum_{i=1}^{np}\sum_{j=1}^{m}\sum_{\ell=1}^{k} y_{hij\ell}$, $h = 1, 2$ $\quad \bar{y}''_{h(8)} = \frac{1}{nqmk}\sum_{i=1}^{nq}\sum_{j=1}^{m}\sum_{\ell=1}^{k} y_{hij\ell}$, $h = 1, 2$ $\quad \alpha = Sb^2 + \frac{Sw^2}{m} + \frac{St^2}{mk}$ and

$$\delta = \rho_b\,Sb^2 + \rho_w\frac{Sw^2}{m} + \rho_t\frac{St^2}{mk} \qquad (2.3.1)$$

The optimum weights are; $a_o = -\frac{pq\,\alpha\,\delta}{\alpha^2 - q^2\delta^2}$ and $c_o = \frac{p\,\alpha^2}{\alpha^2 - q^2\delta^2}$. The variance of $\bar{y}_{2(8)}$ with optimum weights is

$$\mathrm{Var}\,[\bar{y}_{2(8)}] = \frac{\alpha}{n}\,\frac{(\alpha^2 - q\delta^2)}{(\alpha^2 - q^2\delta^2)} \qquad (2.3.2)$$

The optimum replacement fraction q_o that will minimize the variance in (2.3.2) is

$$q_o = \frac{\alpha^2 - \alpha\sqrt{\alpha^2 - \delta^2}}{\delta^2} \text{ and } \mathrm{Var}\,[\bar{y}_{2(8)}]_{opt} = \frac{\delta^2}{2n[\alpha - \sqrt{\alpha^2 - \delta^2}]}$$

2.4 Estimate of the Mean by Procedure (7):

Retain all the PSU's from the first sample. In each of the PSU's retained, further retain only a fraction r of the SSU's and select a fraction s of the new SSU's such that r + s = 1. In each of the matched SSU's, retain only a fraction t of TSU's and select fresh a fraction u of TSU's such that t + u = 1.

A general linear unbiased estimator of $\bar{Y}_2$, may be

$$\bar{y}_{2(7)} = a\,\bar{y}'_{1(7)} + b\,\bar{y}^{**}_{1(7)} - (a+b)\,\bar{y}^{*}_{1(7)} + d\,\bar{y}'_{2(7)} + e\,\bar{y}^{**}_{2(7)} + (1-d-e)\,\bar{y}^{*}_{2(7)}$$

where

$$\bar{y}'_{h(7)} = \frac{1}{nrmtk}\sum_{i=1}^{n}\sum_{j=1}^{rm}\sum_{\ell=1}^{tk} y_{hij\ell},\ h = 1,2$$

$$\bar{y}^{**}_{h(7)} = \frac{1}{nrmuk}\sum_{i=1}^{n}\sum_{j=1}^{rm}\sum_{\ell=1}^{uk} y_{hij\ell} \text{ and } \bar{y}^{*}_{h(7)} = \frac{1}{nsmk}\sum_{i=1}^{n}\sum_{j=1}^{sm}\sum_{\ell=1}^{k} y_{hij\ell}.$$

The variance of $\bar{y}_{2(7)}$ is given by

$$\mathrm{Var}\,[\bar{y}_{2(7)}] = (a^2 + d^2)\,\alpha' + (b^2 + e^2)\,\beta' + [(a+b)^2 + (1-d-e)^2]\gamma' + 2(ab + de)\alpha^* + 2(ae + bd + be)\delta^* + 2ad\delta' - 2(a+b)(d+e)\,\rho_b\frac{\delta b^2}{n} + 2[(d+e)(1-d-e) - (a+b)^2]\frac{Sb^2}{n}$$

where

$$\alpha' = \frac{Sb^2}{n} + \frac{Sw^2}{nrm} + \frac{St^2}{nrmtk},\ \beta' = \frac{Sb^2}{n} + \frac{Sw^2}{nrm} + \frac{St^2}{nrmuk}$$

$$\gamma' = \frac{Sb^2}{n} + \frac{Sw^2}{nsm} + \frac{St^2}{nsmk},\quad \delta' = \rho_b\frac{Sb^2}{n} + \rho_w\frac{Sw^2}{nrm} + \rho_t\frac{St^2}{nrmtk},\quad \alpha^* = \frac{Sb^2}{n} + \frac{Sw^2}{nrm} \text{ and } \delta^* + \rho_b\frac{Sb^2}{n} + \rho_w\frac{Sw^2}{nrm}.$$

2.5 Estimate of the Mean by Procedure (11)

Retain a fraction p of the PSU's from the first sample and select a fraction q (such that p + q = 1) of new PSU's on the second occasion. Further, retain a fraction r of the SSU's with their samples of TSU's in each of the PSU's retained and select anew the remaining fraction s of SSU's such that r + s = 1.

A general linear unbiased estimator of $\bar{Y}_2$ is

$$\bar{y}_{2(11)} = a\,\bar{y}'_{1(11)} + b\,\bar{y}^{*}_{1(11)} + (1-d-e)\,\bar{y}''_{1(11)} + d\,\bar{y}'_{2(11)} + e\,\bar{y}^{*}_{2(11)} + (1-d-e)\,\bar{y}''_{2(11)}$$

$$\bar{y}'_{h(11)} = \frac{1}{nprmk}\sum_{i=1}^{np}\sum_{j=1}^{rm}\sum_{\ell=1}^{k} y_{hij\ell},\ \bar{y}^{*}_{h(11)} = \frac{1}{npsmk}\sum_{i=1}^{np}\sum_{j=1}^{sm}\sum_{\ell=1}^{k} y_{hij\ell}$$

and

$$\bar{y}''_{h(11)} = \frac{1}{nqmk}\sum_{i=1}^{nq}\sum_{j=1}^{m}\sum_{\ell=1}^{k} y_{hij\ell},\quad h = 1, 2$$

The variance of $\bar{y}_{2(11)}$ is

$$\mathrm{Var}\,[\bar{y}_{2(11)}] = \frac{a^2}{np}[Sb^2 + \frac{\alpha_o}{rm}] + \frac{b^2}{np}[Sb^2 + \frac{\alpha_o}{sm}] + \frac{(a+b)^2}{nq}\alpha + \frac{d^2}{np}[Sb^2 + \frac{\alpha_o}{rm}] + \frac{e^2}{np}[Sb^2 + \frac{\alpha_o}{sm}] + \frac{(1-d-e)^2}{nq}\alpha + 2ab\frac{Sb^2}{np} + 2\frac{ad}{np}[Sb^2 + \frac{\beta_o}{rm}] + \frac{2\,Sb^2}{np}[a\,e\,\rho_b + bd\,\rho_b + be\,\rho_b + de].$$

The optimum weights are given by

$$a_o = -\Omega\,pr\,\alpha_o\,[s\,Sb^2\,\beta_o + q\,\rho_b\,Sb^2\,\alpha_o + \frac{(q+ps)}{m}\alpha_o\,\beta_o]$$

$$b_o = \Omega\,ps\,[Sb^2(qs\,\rho_b\,\beta_o^2 - q\,\rho_b\,\alpha_o^2 + r\,\alpha_o\,\beta_o) + \frac{pr}{m}\alpha_o\,\beta_o]$$

$$d_o = \Omega\ pr\ \alpha_o\ [Sb^2\ (\alpha_o + qs\ \rho_b\ \beta_o) + \frac{\alpha_o^2}{m}]$$

and

$$e_o = \Omega\ ps\ [Sb^2(\alpha_o^2 - qr\ \rho_b\ \alpha_o\ \beta_o - s\ \beta_o^2) + \frac{\alpha_o}{m}$$
$$(\alpha_o^2 - (q + ps)\ \beta_o^2)]$$

where

$$\Omega = \alpha/[(qs\ \rho_b\ \beta_o + Sb^2\ \alpha_o + \frac{\alpha_o^2}{m})^2 - (q\ \rho_b\ Sb\ \alpha_o +$$
$$s\ Sb^2\ \beta_o + \frac{(q + ps)}{m}\ \beta_o\ \alpha_o)^2]$$ and α_o, β_o and α are the same as defined in section 2.2 and 2.3.

2.6 Relative Efficiency

In a design problem, it is important to study the relative performance of the different estimators. To study the relative efficiency of the four sampling procedures, it is necessary to have the same overall replacement fraction. Denoting the overall replacement fraction by q^*, it may be shown that for the Procedure (7), $q^* = s + u - s \times u$ and for the Procedure (11), $q^* = q + s - q \times s$. Let RM85 = Var $[\bar{y}_{2(5)}]$/ Var $[\bar{y}_{2(8)}]$, RM87 = Var $[\bar{y}_{2(7)}]$/Var $[\bar{y}_{2(8)}]$ and RM811 = Var $[\bar{y}_{2(11)}]$/Var $[\bar{y}_{2(8)}]$.

The relative efficiencies are computed for some selected values of the parameters and the design quantities. Limited space permits to present only a small fraction of the results in table 2.1. In most cases, procedure (8) is more efficient than the other procedures. The effects of various quantities on the relative efficiencies are discussed in details in the technical report. In tables 2.1 and 3.1, the symbols $\phi = Sw^2/Sb^2$, $\psi = St^2/Sb^2$, m = 16, k = 8, q = u = .5, s = .1 and $q^* = .55$.

3.1 Estimate of the Change by Procedure (5)

A general linear unbiased estimator of $(\bar{Y}_2 - \bar{Y}_1)$ the change, may be written as

$$\Delta_{(5)} = a\ \bar{y}'_{1(5)} - (1 + a)\ \bar{y}^*_{1(5)} + c\ \bar{y}'_{2(5)} +$$
$$(1 - c)\ \bar{y}^*_{2(5)}$$

and its variance is given by

$$\text{Var}\ [\Delta_{(5)}] = a^2[\frac{Sb^2}{n} + \frac{\alpha_o}{nrm}] + (1 + a)^2[\frac{Sb^2}{n} + \frac{\alpha_o}{nsm}]$$
$$+ c^2\ [\frac{Sb^2}{n} + \frac{\alpha_o}{nrm}] + (1 - c)^2\ [\frac{Sb^2}{n} + \frac{\alpha_o}{nsm}] + 2ac$$
$$[\rho_b\ \frac{Sb^2}{n} + \frac{\beta_o}{nrm}] + 2[a(1 - c) - c(1 + a) - (1 + a)$$
$$(1 - c)] \times \rho_b\ \frac{Sb^2}{n} - 2[a(1 + a) - c(1 - c)]\ \frac{Sb^2}{n}$$

For $\bar{y}'_{h(5)}$, $\bar{y}^*_{h(5)}$, α_o and β_o see Section 2.2.

The optimum weights are

$$a_o = \frac{-r\ \alpha_o}{(\alpha_o - s\ \beta_o)} \quad \text{and} \quad c_o = \frac{r\ \alpha_o}{(\alpha_o - s\ \beta_o)}$$

The variance of $\Delta_{(5)}$ with optimum weights is

$$\text{Var}\ [\Delta_{(5)}] = \frac{2}{n}\ [(1 - \rho_b)Sb^2 + \frac{\alpha_o}{m}\ \frac{(\alpha_o - \beta_o)}{(\alpha_o - s\beta_o)}] \tag{3.1.1}$$

Special Cases: It is easily seen from (3.1.1) that if s = 0,

$$\text{Var}\ [\Delta_{(5)}] = \frac{2}{n}\ [(1 - \rho_b)Sb^2 + (1 - \rho_w)\frac{Sw^2}{m} + (1 - \rho_t)\frac{St^2}{mk}] \tag{3.1.2}$$

and if s = 1, we obtain

$$\text{Var}\ [\Delta_{(5)}] = \frac{2}{n}\ [(1 - \rho_b)\ Sb^2 + \frac{Sw^2}{m} + \frac{St^2}{mk}] \tag{3.1.3}$$

From (3.1.1) and (3.1.3) it is clear that for positive correlations, it is advantageous to retain a fraction of SSU's from the first sample to estimate the change.

3.2 Estimate of Change by Procedure (8)

A general linear unbiased estimator of $(\bar{Y}_2 - \bar{Y}_1)$ is

$$\Delta_{(8)} = a\ \bar{y}'_{1(8)} - (1 + a)\ \bar{y}''_{1(8)} + c\ \bar{y}'_{2(8)} +$$
$$(1 - c)\ \bar{y}''_{2(8)}$$

and its variance is given by

$$\text{Var}\ [\Delta_{(8)}] = a^2\ \frac{\alpha}{np} + (1 + a)^2\ \frac{\alpha}{nq} + c^2\ \frac{\alpha}{np} +$$
$$(1 - c)^2\ \frac{\alpha}{nq} + 2ac\ \frac{\delta}{np}$$

where $\bar{y}'_{h(8)}$, $\bar{y}''_{h(8)}$, α and δ are the same as defined in Section 2.3. The optimum weights are

$$a_o = \frac{-p\ \alpha}{\alpha - q\delta} \quad \text{and} \quad c_o = \frac{p\ \alpha}{\alpha - q\delta}$$

The variance of $\Delta_{(8)}$ with optimum weights is

$$\text{Var}\ [\Delta_{(8)}] = \frac{2}{n}\ \alpha\ \frac{(\alpha - \delta)}{(\alpha - q\delta)}$$

3.3 Estimate of the Change by Procedure (7)

One possible linear unbiased estimator of the change may be of the form

$$\Delta_{(7)} = a[\bar{y}'_{2(7)} - \bar{y}'_{1(7)}] + b[\bar{y}^{**}_{2(7)} - \bar{y}^{**}_{1(7)}] +$$
$$(1 - a - b)[\bar{y}^{**}_{2(7)} - \bar{y}^*_{1(7)}]$$

and its variance is

$$\text{Var}\ [\Delta_{(7)}] = \frac{2}{n}[a^2(\lambda_1 - \lambda_4 + \lambda_3) + b^2(\lambda_2 - \rho_w\ \frac{Sw^2}{rm}$$

$+ \lambda_3) + (1 - \rho_b)Sb^2 + \lambda_3 + 2ab \{(1 - \rho_w) \frac{Sw^2}{rm} + \lambda_3\} - 2 \lambda_3(a + b)]$

where

$$\lambda_1 = \frac{Sw^2}{rm} + \frac{St^2}{rmtk}, \quad \lambda_3 = \frac{Sw^2}{sm} + \frac{St^2}{smk},$$

$$\lambda_2 = \frac{Sw^2}{rm} + \frac{St^2}{rmuk}, \quad \lambda_4 = \rho_w \frac{Sw^2}{rm} + \rho_t \frac{St^2}{rmtk}$$

The optimum weights are given by

$$a_o = rt\alpha_o [(1 - u\rho_t)(1 - s\rho_w)Sw^2 + (1 - s\rho_t - ru\rho_t)\frac{St^2}{k}]^{-1}$$

$$b_o = ru(1 - \rho_t)\alpha_o [(1 - u\rho_t)(1 - s\rho_w) Sw^2 + (1 - s\rho_t - ru\rho_t) \frac{St^2}{k}]^{-1}$$

3.4 Estimate of the Change by Procedure (11)

One possible linear unbiased estimator of the change is

$$\Delta_{(11)} = a[\bar{y}'_{2(11)} - \bar{y}'_{1(11)}] + b[\bar{y}^*_{2(11)} - \bar{y}^*_{1(11)}] + (1 - a - b)[\bar{y}''_{2(11)} - \bar{y}''_{1(11)}]$$

and its variance is

$$\text{Var } [\Delta_{(11)}] = 2a^2[\frac{Sb^2}{np} + \frac{\alpha_o}{nprm} - (\rho_b \frac{Sb^2}{np} + \frac{\beta_o}{nprm})] + 2b^2[\frac{Sb^2}{np} + \frac{\alpha_o}{npsm} - \rho_b \frac{Sb^2}{np}] + 2(1 - a - b)^2 \frac{\alpha}{nq} + 4ab(1 - \rho_b)\frac{Sb^2}{np}$$

The optimum weights are

$a_o = pr \alpha_o \alpha/D'$ and $b_o = ps(\alpha_o - \beta_o)\alpha/D'$

where

$$D' = Sb^2(1 - q\rho_b)[\alpha_o - s\beta_o] + \frac{\alpha_o}{m} [\alpha_o - (q + ps)\beta_o]$$

The variance of $\Delta_{(11)}$ with optimum weights is

$$\text{Var } [\Delta_{(11)}] = \frac{2}{n} (Sb^2 + \frac{\alpha_o}{m})[Sb^2(1 - \rho_b)(\alpha_o - s\beta_o) + \frac{\alpha_o}{m} (\alpha_o - \beta_o)] \times [Sb^2(1 - q\rho_b)(\alpha_o - s\beta_o) + \frac{\alpha_o}{m} \{\alpha_o - (q + ps)\beta_o\}]^{-1}$$

3.5 Relative Efficiency

Let RC75 = Var $[\Delta_{(5)}]$/Var $[\Delta_{(7)}]$, denote the relative efficiency of Procedure (7) with respect to Procedure (5). Symbols RC78 and RC711 have similar meanings. These relative efficiencies are studied numerically and some of the results are presented in Table 3.1. The Procedure (7) provides the most efficient estimate of the change.

4.1 Sample Allocation

In most applications of sampling, cost is an important factor since the resources for sample surveys are always limited. Therefore, it is important to study the optimum allocation of the sample subject to a given cost. The optimum distribution of the sample to estimate the current population mean by Procedures (5) and (8) in two-stage successive sampling is considered here. It is assumed that the travel cost between units is unimportant.

4.2 Allocation of the Sample in Procedure (5)

A simple cost function for two occasions may be

$$c(1) = c_1n + c_2nm; \; c(2) = c'_2nrm + c''_2 nsm$$

where c_1 is the cost of preparing frame and c_2 the cost of enumeration on the first occasion. c'_2 and c''_2 are the costs of enumeration on the matched and unmatched parts of the sample on the second occasion. The total cost for two occasion is

$$c = c_1n + (c_2 + c'_2r + c''_2 s)nm \qquad (4.2.1)$$

where

$c = c(1) + c(2)$. The variance of $\bar{y}_{2(5)}$ in two-stage, successive sampling is

$$\text{Var } [\bar{y}_{2(5)}] = \frac{Sb^2}{n} + (\frac{1 - s\rho_w^2}{1 - s^2\rho_w^2}) \frac{Sw^2}{nm} \qquad (4.2.2)$$

From (4.2.1) and (4.2.2) it may be shown that the optimum values of m and n are

$$m_o = \{ c_1\phi(1-s\rho_w^2)/[(1-s^2\rho_w^2)(c_2+c'_2r+c''_2s)] \}^{1/2}$$

$$n_o = \frac{[c(1-s^2\rho_w^2)]^{1/2}}{c_1(1-s^2\rho_w^2)^{1/2} + \{c_1\phi(c_2+c'_2r+c''_2s)(1-s\rho_w^2)\}^{1/2}}$$

4.3 Allocation of the Sample in Procedure (8)

The total cost for two occasions in procedure (8) is

$$c = (c_1+c'_1q)n + (c_2+c_2p+c''_2q)nm \qquad (4.3.1)$$

It is noted here that c'_1q is the additional cost of frame due to new selection of a fraction q of PSU's on the second occasion. The variance of $\bar{y}_{2(8)}$ for two-stage successive sampling is

$$\text{Var } [\bar{y}_{2(8)}] = \frac{1}{n}(Sb^2 + \frac{Sw^2}{m}) \times$$

$$\frac{[(Sb^2 + \frac{Sw^2}{m})^2 - q(\rho_b Sb^2 + \rho_w \frac{Sw^2}{m})^2]}{[(Sb^2 + \frac{Sw^2}{m})^2 - q^2(\rho_b Sb + \rho_w \frac{Sw^2}{m})^2]} \qquad (4.3.2)$$

Eliminating n from (4.3.1) and (4.3.2) we obtain

$$Var[\bar{y}_{2(8)}] = \frac{1}{c}[c_1+\tfrac{1}{2}c_1'+\frac{m}{2}(2c_2+c_2'+c_2'')](Sb^2+\frac{Sw^2}{m}) \quad x$$

$$\frac{[(Sb^2 + \frac{Sw^2}{m})^2 - \frac{1}{2}(\rho_b Sb^2 + \rho_w \frac{Sw^2}{m})^2]}{[(Sb^2 + \frac{Sw^2}{m})^2 - \frac{1}{4}(\rho_b Sb^2 + \rho_w \frac{Sw^2}{m})^2]} \qquad (4.3.3)$$

Now $\frac{\partial}{\partial m}$ [Var $(\bar{y}_{2(8)}]$ = 0 provides a sixth degree equation in m which is solved for the optimum m by the method of successive approximations. From (4.3.1), optimum n is obtained.

4.4 Relative Efficiency

Let REMC58 = Var $[\bar{y}_{2(8)}]$/Var $[\bar{y}_{2(5)}]$ represent the relative efficiency of Procedure (5) with respect to Procedure (8). On the basis of numerical study (some of the results presented in Table 4.1) made, it is observed that the Procedure (5) is more efficient than the Procedure (8) to estimate the current population mean.

Conclusions: It is observed from the extensive numerical study of the relative efficiencies that if the sampling is inexpensive and the precision of the estimates is of major interest, the sampling Procedure (8) is more efficient than the other procedures in most cases. However, the gains of Procedure (8) over Procedure (5) are modest in most cases. If cost is taken into consideration, Procedure (5) is more efficient than Procedure (8) to estimate the current population mean.

References

[1] Chakrabarty, R. P. and D. S. Rana (1974). "Multi-Stage Sampling with Partial Replacement of the Sample on Successive Occasions". Proceedings of the ASA (Social Statistics Section) 1974. pp. 262-268.

[2] Eckler, A. R. (1955). "Rotation Sampling". Ann. Math. Stat., 26, pp. 665-685.

[3] Singh, D. (1968). "Estimates in Successive Sampling Using a Multi-Stage Design". Jour. Amer. Stat. Assoc., 63, pp. 99-112.

[4] Singh, D. and Kathuria, O. P. (1969). "On Two-Stage Successive Sampling". The Australian Journal of Statistics, 11, No. 2, pp. 59-66.

[5] Chakrabarty, R. P. and D. S. Rana, "Three-Stage Sampling on Successive Occasions". Technical Report. Department of Stat. and Comp. Sci., University of Georgia. August, 1976.

Table 2.1 Values of RM85, RM87, RM811

ρ_b	ρ_w	ρ_t	ϕ=.5, ψ=2 x	y	z	ϕ=5, ψ=2 x	y	z
	.5	.5	100	107	100	100	107	100
.5	.7	.7	92	107	100	93	108	100
	.9	.9	79	107	100	82	109	100
	.5	.5	108	116	100	106	113	100
.7	.7	.7	100	116	100	100	115	100
	.9	.9	86	116	100	88	118	100
	.5	.5	125	133	101	116	124	101
.9	.7	.7	116	134	102	111	128	101
	.9	.9	100	135	102	100	133	101

x = RM85 y = RM87 z = RM811

Table 3.1 Values of RC75, RC711, RC78

ρ_b	ρ_w	ρ_t	ϕ=.5, ψ=2 x	y	z	ϕ=5, ψ=2 x	y	z
	.5	.5	101	133	137	108	132	136
.5	.7	.7	101	133	138	108	137	141
	.9	.9	101	135	139	105	142	148
	.5	.5	102	151	159	111	147	152
.7	.7	.7	102	152	161	113	152	159
	.9	.9	101	154	163	108	159	168
	.5	.5	104	175	188	119	165	171
.9	.7	.7	104	176	190	125	171	179
	.9	.9	103	179	195	120	180	191

x = RC75 y = RC711 z = RC78

Table 4.1 Values of REMC 58

ρ_b	ϕ	q^* = .5 ρ_w=.5	.7	.9	q^* = .7 ρ_w=.5	.7	.9
	.5	129	129	133	141	142	147
.5	.5	117	118	121	123	124	130
	10	114	115	118	119	120	126
	.5	122	122	124	133	133	136
.7	5	113	113	115	119	119	122
	10	111	111	113	116	116	119
	.5	113	111	111	121	118	117
.9	.5	109	107	107	114	112	110
	10	108	107	106	113	110	109

c = 10550 c_1' = 70.50 c_2' = 12.50

c_1 = 64.40 c_2 = 14.75 c_2'' = 16.25

THE MULTIPURPOSE SURVEY OF SAUDI ARABIA - AN EXPERIMENT IN COMPRESSION

Ali Rashid, Central Department of Statistics
Kingdom of Saudi Arabia
and
John Rumford, United States Bureau of the Census/U.S. Representation
United States - Saudi Arabian Joint Commission on Economic Cooperation

In accordance with the current Five Year Plan of Saudi Arabia, the Central Department of Statistics (CDS) is committed to implement three major household sample surveys. Included are surveys of vital events and migrations, labor force activities and family consumption expenditures. To collect this information as soon as possible and at the same time work within the constraints of the critical shortage of statistical manpower that is prevalent in Saudi Arabia today, the three surveys have been compressed into a single unitized system. The design of any one component of this multipurpose survey is not particularly unique, however, the integration of three very different types of information into one survey vehicle, the multiplicity of data that can be generated, and several of the evaluation and control techniques introduced are considered unique and may prove useful to other countries or areas where similar data are required in a short period of time and under conditions where statistical manpower is extremely difficult to acquire.

The Survey Vehicle

One of the most important elements in designing the multipurpose survey was the configuration of the vehicle itself. This involved considerations of sample size, sample dispersion, and sample periodicity. Two factors were instrumental in determining the sample size finally selected. The first was the geographic scope of the survey and the second was the precision of the estimates that were considered to be of paramount importance.

With respect to the scope of the survey, the vast differences between the urban and rural areas of the Kingdom in cultural heritage, living conditions, life style and facilities, made it necessary that separate estimates be obtained for each area.

Concerning the precision requirements of the survey, the most nationally significant estimates that will be generated are the birth and death rates. Current estimates of vital rates derived from the latest United Nations information available for the Kingdom 17/18/ suggests that the birth rate is approximately 50 per 1,000 population and the death rate between 20 and 23 per 1,000. Based on results from similar surveys in other countries 6/7/8/9/15/16/, a sample of between 50,000 and 75,000 persons should yield relative sampling errors that will be within 1 to 3 percent for the birth rate and about 4 to 7 percent for the death rate.

With regard to sample dispersion, though a sample size of under 75,000 persons is not particularly large, if the sample units are scattered throughout the Kingdom, the logistical and control problems necessary to service them would be excessive. To avoid widespread dispersion, a multistage clustered sample was selected. The population within each last stage, sample unit in a cluster is approximately 500 persons. A unit of this size is compatible with the village settlement pattern in Saudi Arabia and, as has been demonstrated in other countries, relatively efficient as far as the casefinding of births and deaths is concerned 12/.

Regarding periodicity, the subject matter of the survey demands that the enumeration be conducted more than once each year. Labor force activities are seasonal and month-to-month variations must be monitored. Price levels, and family expenditures also vary throughout the year and estimates must realistically reflect how the current state of the economy affects the budget of Saudi citizens. Moreover, the measurement of births and deaths require a continuing enumeration schedule to promote accurate reporting.

The most critical periodicity component is the timely recording of births, deaths, and migrations. These events occur throughout the year and must be recorded continuously or they will be lost through memory decay or intentional misreporting. Moreover, vital rates are expressed as incidence measurements and are time specific. One of the more successful enumeration periods found to maximize vital event and migration reporting is by a regular monthly household visit 1/6/10/11/14/.

In the case of the family consumption component, it was necessary to take into consideration the seasonal patterns affecting production and consumption. Saudi Arabia has two distinct seasons, a mild winter and a very hot, dry summer. A semiannual enumeration, therefore, was scheduled in the midseason periods of January and July.

Labor force activities are also affected to some degree by the same climatic extremes and the main survey efforts are again conducted semiannually in the months of November and May. With labor force activities, however, it was also necessary to monitor changes in the employment status on a continuing basis, therefore, like vital events, a supplementary monthly cycle is used.

Implementation of the Vital Event Component

As indicated previously, the most critical component of the multipurpose survey is the casefinding of vital events and migrations, particularly the former. Because of this, the major implementation, enumeration and control decisions were oriented toward relieving the problems associated with the capture of these events. The occurrence of a birth or death is a very personal experience in any culture. To record these events successfully, a carefully trained and controlled network of enumerators atuned to the sensibilities of the respondents is required. Because of the paramount importance of enumerator—respondent rapport, the permanent resident enumerator concept was adopted. This method has proven successful in many countries where people are reluctant to confide in persons who do not come from their immediate cultural subgroup 5/10/11/.

To implement the resident enumeration system, elementary school teachers are to be recruited as part-time enumerators. In most instances, these teachers teach school in or very close to the survey sample unit.

The initiation of the monthly system is relatively uncomplicated. For the first round, beginning October 8, 1976, a date chosen because it marks the beginning of the school year, the enumerator conducts a complete initial canvass of every household in his sample unit. He fills out in triplicate all the basic census information required for all household members. Nothing is asked at this time about vital events, migrations, labor force activities or consumption expenditures. One copy of the questionnaire is retained by the enumerator, one is transferred to the Regional Headquarters and one is filed in the Central Office in Riyadh. At the end of every month, the enumerator visits each houshold, taking along the initial questionnaire. He asks if any births, deaths, migrations, or labor force changes took place since his last visit. If any of these events or changes occurred, they are recorded on three new questionnaires and the household census information is adjusted to reflect the changes brought about by the event. Again, one copy is kept by the enumerator, one is transmitted to the regional office and one is filed at the Central Office. Each monthly report for a household is cumulative. The latest questionnaire not only contains the events and household composition of the current month, but also includes all the events that may have occurred since the survey began. Considering the low probability of the occurrence of a given event in a population of less than 100 households in any one month, most rounds are completed within a few hours.

On the surface it would seem that if the monthly enumerators made their rounds religiously and that household respondents dutifully reported all events that occurred, at the close of the year the computation of exact birth, death or migration rates would be a straightforward matter. Unfortunately, this is not the case. Even the most conscientious enumerators neglect monthly rounds. Respondents often do not bother to recall a migration or, through the heartbreak of tragedy, fail to mention the death of a loved one. Because of these problems, it has been demonstrated that one of the most successful methods of obtaining relatively complete reports of events is by using a dual estimation procedure 1/2/3/5/10/11/. The basis of the system is to record events using two independent data collection systems covering the same population and time period. The results of the two systems are compared or "matched", differences are reconciled and the results are accumulated.

In linking records, each report is classified by the enumeration system that found the event. The three classifications are (a) the same event was recorded by both systems, (b) the event was recorded by the first system but missed by the second system, and (c) the event was recorded by the second system but missed by the first. The matching process used for record linking is the singular or "one-way" method. Singular matching is employed because of the relative simplicity and speed associated with the technique and because it is particularly appropriate where geographic out-of-scope problems related to indistinct enumeration area boundaries are common.

There are several methods used to implement this dual enumeration process. One of the better techniques is to link a monthly system to an independent enumeration conducted on an annual, semiannual or quarterly basis 4/5/10/11/12/13/. Because of the shortage of trained cadre of full-time enumerators in Saudi Arabia, the annual enumeration system was selected to complement the continuing monthly report. This annual survey is implemented by a completely different team of enumerators at the end of the survey year and uses a recall period dating back to the beginning of the school year.

Although the dual system has many provisions for control through self-evaluation, it is effective only after the fact in most cases. To rectify this problem, a unique control procedure is implemented that not only serves a control function, but also provides a separate, independent, and complementary vital event coverage estimate.

To apply this system, the sample units in each region are divided into three sections. At the end of the first three months, a regional enumeration team visits one of these sections and records all births and deaths that occurred since the survey's beginning. These results are matched against the accumulative monthly reports and coverage estimates are generated for the three month period. This procedure is repeated every three months using a sliding recall period. This type of active control insures that every sample unit is reinterviewed twice in a survey year. Moreover, the additional coverage estimate provides a running account of enumeration quality.

Although these enumeration and control procedures are mainly oriented toward improving the coverage of vital events, they also improve the implementation, coverage and quality of the remaining two components of the multipurpose survey - - the semiannual labor force study and the family consumption expenditure report.

Implementation of the Labor Force Component

The major purpose of the labor force component is to provide an accurate and current estimate of the domestic and foreign working population in the country. This estimate includes both an inventory of occupational skills by business and industry and selected information about the characteristics of the workers themselves.

As indicated previously, the primary labor force survey enumerations are to be conducted in November and May with a short supplemental monthly canvass. These particular months were selected because they represent periods of relative stability in the work force. Due to the extremely hot weather extending from June through September, many businesses and industries curtail operations. This is the period of extended vacation and leave taking. In contrast to this exodus, a large influx of Saudi students temporarily enter the labor market during the summer recess and create a distorting picture. The November/May canvass measures the normal pattern; the monthly canvass, the seasonal cycles.

The semiannual enumerations are conducted by the same, part-time resident enumerators who do the vital event rounds. They are issued a separate questionnaire containing only nine questions. The enumeration is restricted to persons age 12 years and over. Because the first labor force round is conducted only one month after the initial census round, it is not necessary to ask such questions as name, age and sex. These and other items are linked automatically to the initial questionnaire. The second semiannual round conducted in May is designed to measure any changes in occupations and incomes that may have occurred during the six month period. The monthly inquiry consists of only one question, this pertains to changes in the employment status of the eligible labor force population. It requires a simple, self-coded answer and very little enumeration time is expended.

Implementation of the Family Consumption Expenditures Component

A general multipurpose survey system is not the vehicle to use for the execution of a traditional, intensive family consumption expenditure survey. Inquiries of this type require the techniques of multiple visits over short periods of time, long lists of expenditure items and full-time enumerators. No attempt was made to interject a consumption survey of this type into the multipurpose system. Instead two surveys are implemented. One, is a component of the multipurpose system, confined to the rural areas and designed only to provide consumption inputs into National Accounts. The other is a traditional expenditure survey executed exclusively and independently in the urban areas and designed to provide information for a Current Price Index.

Because data from both surveys will eventually be used to create Kingdomwide SNA statements, a standard classification scheme is used for both at the publication level.

As indicated previously, the rural expenditure survey is conducted semiannually in January and July. This again is necessitated by the seasonal extremes that dramatically affect the availability and quality of food stuffs and services. This is particularly true in the rural areas where "own product" commodities and services are used. Like the labor force survey, the same part-time enumerators conduct the enumeration. They use a separate questionnaire for each household. The expenditure recall period extends to the past week for frequently purchased items and, to the past month for such items as rent, fuel and most durables.

Multiplicity of Information Through Data Integration

The integrated multipurpose survey approach as implemented in Saudi Arabia presents a total of information that is much greater than the sum of the individual parts. It is possible for example to examine fertility, mortality and migration differentials by labor force activities and consumption patterns. These added dimensions are provided without any increase in enumeration and supervisory personnel or additions in logistical support.

The multipurpose survey design will probably not make specialized subject matter technicians completely happy. Obviously, the training of enumerators on any one component cannot be especially thorough. Part-time enumerators usually owe allegiance to their main source of income and not the after hours survey. The respondents themselves may be subject to confusion at times over exactly what the overall study is about. However, considering the economy of operation in personnel and time, and the multiplicity of information available, the multipurpose survey approach incorporating some of the enumeration and control procedures used in Saudi Arabia may offer nations with a scarcity of statistical manpower a method of collecting different types of information rapidly; in-control and with the administrative and logistical advantages inherent in a unitized survey system.

References

1) Ahmed, N., and Krotki, K.: Simultaneous Estimations of the Population Growth: The Pakistan Experiment. Pakistan Development Rev 3: 37-65 (1063).

2) Cantrelle, P.: Repeated Demographic Observation in a Rural Area in Senegal: Method and First Results. Document No. B.6 6/V/F/207. World Population Conference, Belgrade, 1965.

3) Chandrasekar, C., and Deming, W.E.: On a Method of Estimating Birth and Death Rates and the Extent of Registration. J. Am. Stat. Assoc. 44:101-115, March 1949.

4) Marks, E.S., Seltzer, W. and Krotki, K.: Population Growth Estimation: A Handbook of Vital Statistics Measurement. New York. The Population Council, 1972.

5) National Statistical Office: Malawi Population Change Survey: February 1970 - January 1972. Zomba, Malawi: Government Printer 1973.

6) Republic of Liberia: Patterns of Natality, 1970. Series, N-1, Ministry of Planning and Economic Affairs, 1971.

7) Republic of Liberia: Patterns of Mortality, 1970. Series D-1, Ministry of Planning and Economic Affairs, 1971.

8) Republic of Liberia: Natality Profiles, 1971. Series N-2, Ministry of Planning and Economic Affairs, 1972.

9) Republic of Liberia: Mortality Profiles, 1971. Series D-2, Ministry of Planning and Economic Affairs, 1972.

10) Rumford, J.C., Heperkan, Y., and Fincancioglu, N.: The Principles and Preliminary Results of the Turkish Demographic Survey.

11) Rumford, J.C.: Use of the Chandrasekar-Deming Technique in the Liberian Fertility Survey. Public Health Rep. 85:965-974, November 1970.

12) Rumford, J.C.: Factors that Affect Case-finding in the Liberian Fertility Survey. Health Service Reports, 87:247-262, March 1972.

13) Rumford, J.C.: Factors Influencing the Casefinding of Migrations in the Liberian Fertility Survey. Demography 9:431-441 (1972).

14) Seltzer, W.: Some Results from Asian Population Growth Studies. Population Studies 23:395-406, November 1969.

15) Turkish Ministry of Health and Social Welfare: Vital Statistics from the Turkish Demographic Survey. School of Public Health, Ankara, 1967.

16) Turkish Ministry of Health and Social Welfare: Vital Statistics From the Turkish Demographic Survey. School of Public Health, Ankara, 1970.

17) United Nations Population and Vital Statistics Report. Series A, V. 26, No. 4 U/W. New York, New York, 1974.

18) United States Bureau of the Census: Recent Demographic Estimates for Countries and Regions of the World. In S.P.C. May 1974.

PREDICTING THE OUTCOME OF FOOTBALL GAMES OR CAN YOU MAKE A LIVING WORKING ONE DAY A WEEK

Larry T. Revo, Meharry Medical College

It seems that every few years some individual comes along who claims to have supernatural powers of prognostication. The latest and, perhaps best, prognosticator currently performing is applying his specialized talents to the prediction of football games. His name is Danny Sheridan and he lives in Mobile, Alabama. This paper will present the results of Mr. Sheridan's selections in 1974 and 1975 and compute the probabilities associated with his performance under the assumption that his chances of picking the winner are 50-50 (i.e., the null hypothesis of no predictive power).

What Mr. Sheridan does so well is predict the outcome of football games "against the spread". The spread is the number of points that the favorite has to give to the underdog to make the game even (i.e., to make betting a supposedly 50-50 proposition). If the favorite wins the game by more than the point spread, the favorite is considered to have beat or "covered" the spread. Conversely, if the favorite wins by less than the spread or loses the game outright, the underdog is considered to have beaten the spread. If the favorite beats the underdog by an amount exactly equal to the spread, the outcome is a tie.

In September of 1974, Sheridan started sending his weekly selections to Bill Sellers, a Mobile Press Register reporter. He usually sent them on Thursday and Sellers opened them on the Monday following the games. By November, he had interested Sports Illustrated writer, Harold Peterson, in his forecasting talent and started sending Peterson his weekly selections. In 1975, he continued sending Sellers his weekly selections and started sending same to Donald Friedlander, a Mobile lawyer. In response to increasing requests, Sheridan started a phone and "football newsletter" service to paying customers in September of 1975. Thus, for all of 1974 and 1975, his weekly selections can be fully documented and it can be verified that he made these selections prior to game time. In addition, I have checked the outcome of Sheridan's selections myself. As far as can be determined, the results presented in this paper are reliable and verifiable.

The following table summarizes the results of Mr. Sheridan's selections in 1974 and 1975 versus the spread, with the selections separated into the categories of "favorite" or "underdog" and College or Professional:

		1974 Won	1974 Lost	1974 Total	1975 Won	1975 Lost	1975 Total
College	Favorite	51 (.89)	6	57	39 (.76)	12	51
	Underdog	123 (.87)	19	142	87 (.77)	26	113
	Total	174 (.87)	25	199	126 (.77)	38	164
Pro	Favorite	4 (1.0)	0	4	14 (.74)	5	19
	Underdog	15 (.94)	1	16	46 (.66)	24	70
	Total	19 (.95)	1	20	60 (.67)	29	89

Thus, for instance, there were 57 college games in 1974 in which Sheridan picked the favorite; the favorite won (beat the spread) 51 of these games for an astounding 89% correct. Similarly, there were 70 pro games in 1975 in which Sheridan picked the underdog; the underdog won 46 of these games for a winning percentage of 66%, still hard to believe, since each game is supposed to be a 50-50 proposition.

Under the null hypothesis that Mr. Sheridan has a 50% probability of correctly predicting the winner of a single game (versus the spread), the number of correct predictions, X, is a binomial variate with parameters p = .5 and n, the total number of predictions. This model assumes that the outcomes of Sheridan's selections are independent - that is, that the outcome of one game vis-a-vis his prediction has no effect on the outcome of any other of his predictions. This is clearly a reasonable assumption since the outcomes of any two games played on the same day are independent for all practical purposes.

The probability of correctly predicting X or more games in n attempts is given by

$$\sum_{i=X}^{n}\binom{n}{i}p^{i}(1-p)^{n-i} = \sum_{i=X}^{n}\binom{n}{i}(.5)^{n} \quad \text{with } p = .5$$

under the null hypothesis.

This quantity was computed for the various entries in Table 1, with the following exact results:

Table 2

Exact Probabilities (P Values) Under Ho for the number of Successes (Wins) in Table 1

		1974	1975
	Favorite	2.84 E-10	9.90 E-05
College	Underdog	3.86 E-20	3.56 E-09
	Total	5.74 E-29	1.61 E-12
	Favorite	6.25 E-02	3.18 E-02
Pro	Underdog	2.59 E-04	5.76 E-03
	Total	2.00 E-05	6.68 E-04

All probabilities are given in scientific notation; the number after the E (exponent) specifies the exponent of 10 which the number preceeding the E is to be multiplied by. Thus, 2.84 E-10 = 2.84×10^{-10} which is equivalent to .000000000284 in decimal notation.

Clearly, there is overwhelming evidence against the null hypothesis of no predictive power, especially in 1974. However, 1974 was a year in which college underdogs did much better than 50% versus the spread. In fact, college underdogs won 174 of 297 games, a winning percentage of .596. Thus, had a prognosticator simply selected all college underdogs in 1974, he would have beaten the spread in almost 60% of those games.

A glance at Table 1 indicates that Mr. Sheridan concentrated on underdogs in 1974; 71% of his college selections were underdogs. Thus, a reasonable question is - did Mr. Sheridan do so well simply because he selected so many underdogs, or did he in fact do substantially better than someone could have done by selecting every 1974 college underdog, i.e., did he do substantially better than win 59.6% of his selections? Thus H_o: $p = .596$

H_a: $p > .596$

The exact probability of having 174 winners in 199 selections under H_o is 5.64 E-18. Thus, we can not attribute Mr. Sheridan's 1974 college record to simply selecting underdogs indiscriminantly.

Summary

The binomial model has been used to test the hypothesis that Danny Sheridan has no powers of prognostication. The model was applied to his 1974 and 1975 football selections. The results emphatically reject the null hypothesis, and argue that his powers, at least during 1974 and 1975, were of a supernatural order.

HEROIN EPIDEMICS - DECLINE AND FALL OF EPIDEMIOLOGIC RESEARCH

Alex Richman, Beth Israel Medical Center
Helen Abbey, Johns Hopkins School of Hygiene

"But, of course, having regard to the data it would have been more than genius, it would have been magic."

Greenwood, M. 1948

INTRODUCTION

Smooth, unimodal, skewed distributions of year of onset of heroin use, derived from groups of treated heroin addicts, are called "epidemic-like" curves and are generally alleged to reflect incidence.

This paper shows how "epidemic-like" curves can be produced from treatment data without associated changes in incidence, ascertainment or duration of disorder. These findings are relevant to other types of non-infectious disorders where the number of patients ascertained is a non-linear function of time since onset.

BACKGROUND

Identification of "new cases" of heroin addiction in the community is difficult since voluntary reporting is unlikely and other forms of ascertainment are unreliable. Community surveys have difficulties in constructing suitable frames for probability sampling; establishing reliable and valid instrumentation; inferring the characteristics of those non-interviewed; relating the quantity, frequency, duration and recency of heroin use to the "need for treatment" and assessing the characteristics of the relatively small proportion of the population who admit heroin use. O'Donnell et al's probability survey of Selective Service registrants elicited 148 heroin users out of the 2,510 men interviewed (471 were not interviewed). Half of those who admitted heroin use had "used" heroin less than 10 times.

In contrast to the relatively low yield of heroin users in community surveys is the large amount of data available for heroin addicts entering treatment programs. These treatment data have been considered to be "valid" indicators of the year of onset of heroin use or heroin addiction in the community.

Graphs of these data show a rise and fall in the year of onset of heroin use among admissions; the rise and fall is considered to reflect experience in the community. Although this interpretation seems rational, it is the purpose of this paper to show how such "epidemic-like" curves can occur independently of changes in onset in the community. "Epidemic-like" curves of heroin use do not necessarily indicate a common time for exposure; the results of person-to-person spread; the removal of susceptibles; or changes in virulence.

"Epidemic-like" curves of year of onset of heroin use in treated populations do not prove that there were corresponding peaks of heroin use in the community. Factors other than changes in incidence can produce "epidemic-like" distributions of the year of onset of heroin use.

PEAK YEAR OF ONSET - Assumptions

Much attention has been directed to the year of onset of heroin use (or addiction) among patients entering treatment programs. This distribution has been used to infer changes in morbidity in the community, to predict future needs for treatment, to assess the effectiveness of intervention programs, and for surveillance and monitoring of national trends. (Hunt and Chambers; Greene)

Various assumptions are implied when changes in morbidity in the community are inferred from persons entering treatment. These assumptions include:

i. The ratio of admissions to onsets is assumed to be constant. Changes in the number and type of admissions are assumed to represent proportionate changes in the number and types of persons in the community in need of treatment for the first time.

We can only use treatment data to reflect changes in morbidity if we know that the ratio of admissions to onsets has not changed over time. Since the probability of admission may vary over time, among different demographic subgroups and between places, assessments of community morbidity from admissions to treatment must be based on knowledge of the proportion of onsets who are admitted. If that proportion were known, treatment data would not be needed to infer changes in morbidity. A doubling in the number of new patients admitted to treatment doesn't demonstrate a doubling in the number of persons in the community in need of treatment for the first time.

ii. Smooth frequency distributions are assumed to represent homogeneous sub-populations. The greater the number of random and independent variables which are combined, the more likely the output will result in a normal distribution function with a "peak". (King)

The sum $x_1 + \ldots + x_n$ of independent random variables, regardless of their individual distributions, has approximately the normal distribution under very general conditions (Ljapunov Theorem). Addition of data from populations with diverse distributions of year of onset will result in a smooth distribution with a peak. (Richman and Richman, 1975)

iii. Changes in percentage distributions are assumed to be equivalent to changes in population based rates.

The number of admissions of long-term addicts varies from time to time, and fluctuations in the number of long-term addicts will affect the percentage of short-term addicts among the total number admitted even without any changes in the absolute number of short-term addicts. Changes in onset must be inferred from population based rates.

iv. The probability distribution of admission for addicts in the community (specific for time since onset) is assumed to be stable over calendar time.

HUNT'S ESTIMATION OF "LAG" IN ENTERING TREATMENT

The delay between onset of heroin use and subsequent entry into treatment is referred to as "lag". Recently "lag" data have been used to project future admissions to treatment for a given program. Hunt (1975) asserts that "lag" is stable from time-to-time, consistent from place-to-place and can be estimated from onset cohorts.

Hunt has published data on the distribution of lag intervals to be used in projecting future admissions. He assumes probability of admission is the same for addicts of specific duration of addiction regardless of year of onset, clinical correlates or complications, availability of treatment in previous years, type of current treatment or demographic characteristics. Hunt assumes that the duration of addiction or prospect of remission, or death are fixed, and do not change.

The distribution of year of onset of heroin use (or addiction) resembles graphs of time of exposure or onset of contagious disorders to such an extent that their assumptions and implications for heroin addiction have not been adequately tested. The following model has been developed for testing assumptions, implications and programmatic relevance of statistical approaches to inferring changes in incidence from distributions of year of onset among treated populations.

- input variables - various number of years for which data collated.
  - time-specific probabilities of admission among new onsets in the community.
- output variables - distribution of time since onset among admissions for treatment.
- status variables - conditions which are kept constant throughout operation of the model are incidence, duration of disorder, remission, mortality, treatment capacity, and perceptions of treatment needs.

To begin, let's construct a model with stable incidence, ratio of admissions to onsets, and duration - specific probabilities of admission from year to year. (FIGURE 1)

Later, we plan to extend this model to include other cells of Figure 1 and to consider the effect of changes on some of the factors previously listed as status variables. We will also defer consideration of the differential effect of readmissions to long standing programs being called first admissions in newly established programs.

What type of distribution of years of onset will result from the steady-state situation in Figure 1? Can we get "epidemic-like" curves without any change in incidence?

TIME SPECIFIC PROBABILITIES OF ADMISSION AMONG NEW ONSETS IN THE COMMUNITY

Let us use Hunt's projection of the time between onset and admission for those addicts entering treatment. (Later, we will use other estimates for this distribution). Hunt projected that 12% of addicts would enter treatment within 1 year, 22% within 1-2 years, 26% within 2-3 years, 26% within 3-4 years, 6% within 4-5 years, 3% within 5-6 years and 3% within 6-7 years. (FIGURE 2)

FIGURE 3 shows Hunt's projection with the number of dots representing the value of the percentage distribution. Each pattern depicts the number of patients entering treatment in a specific year following onset.

DEVELOPMENT OF THE GRAPHIC MODEL

Figure 4 shows the result of using Figure 3 (an onset-year cohort with admissions in successive years) as a module to build up the distribution of patients admitted during 1965-1975 with onset of heroin use in 1965-1975. A stable state is assumed with constant incidence of heroin use, a constant rate of treatment entry and no changes in duration of the disorder, perceptions of treatment, or treatment capacity. The onset-year cohorts are shown horizontally, the admission groups are represented vertically. The group admitted during 1971 is flagged; it can be seen that 12 of the patients began heroin use in 1971, 22 in 1970,...and 3 began in 1965. With no change in incidence of heroin use in the community, what is the distribution of year of onset of heroin use among the 1971 group of admissions?

Figure 4 was rotated and the 1971 admission group separated in Figure 5. What was the time since onset of heroin use among patients admitted in 1971? Among the 1971 admissions the onset of heroin use had peaked 3-4 years earlier. Figure 6 converts the abscissa from years since onset to calendar year of onset. "Epidemic-like" curves occur with distributions of years between onset and admission other than Hunt's projection. (Fig.7)

DISCUSSION

The model is based on assumptions of a stable state with:

- no changes in onset in the community;
- constant probability of eventual admission for addicts in the community;
- unchanging time-specific probabilities of admission;
- uniform duration of the disorder,or rate of mortality;
- fixed capacity for admitting new patients.

Such a stable state has the properties of a stationary population in a life table. In a stationary population the numbers of births and deaths are constant and the age distribution of deaths in the cohort and the age distribution in the life table population are identical (Lotka). In our demographic model for heroin addiction, the number of onsets (births) and the number of admissions (deaths) are constant over time; the distribution of time since onset of heroin use in onset cohorts is identical with the distribution of time since onset of heroin use in a group of admissions. Figure 6 is a mirror image of Figure 2.

In a non-truncated version of Figure 4 any column and any row would be identical, e.g., the distribution of time since onset of heroin use is identical for the cohort with onset of heroin use in 1970 and the group admitted in 1970.

Therefore, a log-normal distribution of time since onset of heroin use will appear in a group of admissions if that is the distribution among the onset cohort. This log-normal distribution is an "epidemic-like" curve.

This phenomenon of apparent clustering in time is not restricted to log-normal distributions, but also applies to normal distributions and other non-linear functions.

Pseudo-epidemics (or clustering in time without change in incidence) can occur in other conditions where there is a log-normal distribution of time between exposure and diagnosis. These conditions include, in addition to bacterial and viral diseases (Sartwell), post-radiation leukemia, iatrogenic blood dyscrasias and bladder tumors in dye stuff workers (Armenian and Lilienfeld). If data on the year of exposure were graphed for groups of patients diagnosed with chronic conditions,we would get a clustering in year of exposure, a rise and fall in the distribution of patients with the condition even if there had been no change in the extent of environmental exposure. It is essential to be able to differentiate this pseudo-epidemic from situations where there have been true changes in exposure and incidence in the community.

Further work is necessary to assess the effect on the model of other factors. What is the effect on the nature and characteristics of the "epidemic'like" curve when there are variations in the following individual factors:

ascertainment - probability of eventual admission; time specific rates of admission for those who are admitted; availability, accessibility of different types of treatment.

natural history - probability of becoming addicted after initial trial of heroin; interval between initial use of heroin to onset of addiction among those addicted; demographic characteristics of addicts (ethnic, sex, age); remission of addiction, or mortality.

social or community factors - short term changes in law enforcement, availability of heroin, methadone on the street, social sanctions.

An epidemiologist has been defined as an expert from out of town who slides to glory on the descending limb of the epidemic curve. (Fox, Holland, Elveback) Let us strive to differentiate the descending limbs of onsets in the community from the descending limb of pseudo-epidemics.

* * * *

REFERENCES

Armenian HK and Lilienfeld AM: The distribution of incubation periods of neoplastic diseases, Am. J. Epidemiol. 99: 92-100, 1974.

Greene MH, Kozel NJ, Hunt LG and Appletree RL: An assessment of the diffusion of heroin abuse to medium sized American Cities, Washington: USGPO, 1974.

Greenwood M: Medical statistics from Graunt to Farr, Cambridge University Press, 1948.

Hajek J: Probability theory (Ljapunov theorem) in Survey of Applicable Mathematics (K. Rektorys ed.) Cambridge, Mass: The M.I.T. Press, 1969.

Hunt LG: Admission Records. A tool for treatment program planning, Washington: Drug Abuse Council, 1975.

Hunt LG and Chambers CD: The Heroin Epidemics, New York: Spectrum Publications, 1968.

King JR: Probability Charts for Decision Making New York: Industrial Press, 1971.

Lotka AJ: Elements of Mathematical Biology, New York: Dover, 1956.

O'Donnell JA, Voss HL, Clayton RC, Slatin GT and Room RGW: Young Men and Drugs-A nationwide survey, Washington: USGPO, 1976.

Richman A and Richman VV: Heroin addict forecasting-statistical fallacies and disputed facts, Proceedings of the American Statistical Association, Social Statistics Section, Annual Meeting, 1975.

Sartwell PE: The distribution of incubation periods of infectious disease. Amer. J.Hyg. 51:310-318, 1950.

FIG.1

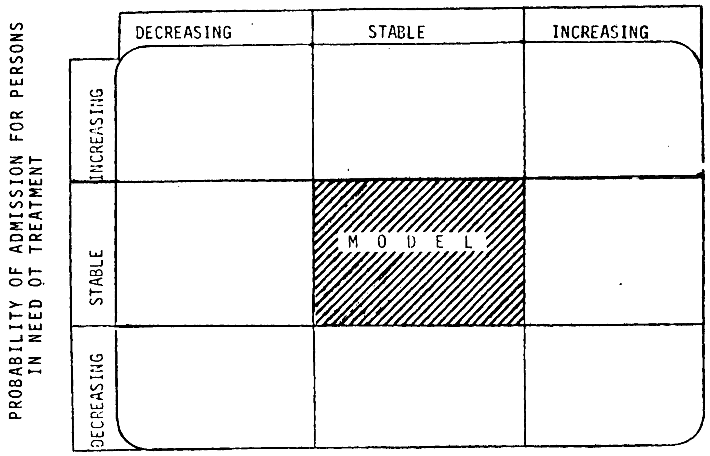
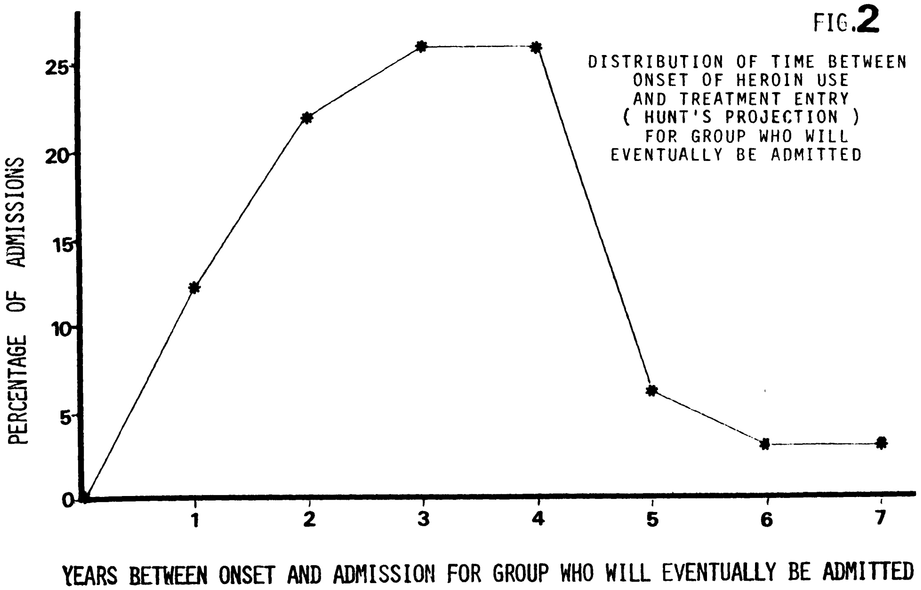


FIG.2

DISTRIBUTION OF TIME BETWEEN ONSET OF HEROIN USE AND TREATMENT ENTRY (HUNT'S PROJECTION) FOR GROUP WHO WILL EVENTUALLY BE ADMITTED



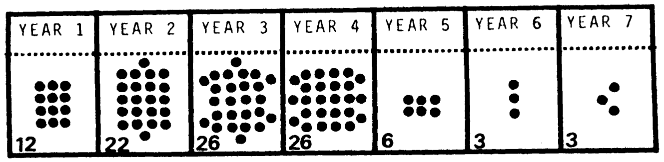


FIG.3

YEARS BETWEEN ONSET AND ADMISSION FOR GROUP WHO WILL EVENTUALLY BE ADMITTED

(NOTE: THESE ARE THE SAME DATA AS IN FIG. 2)

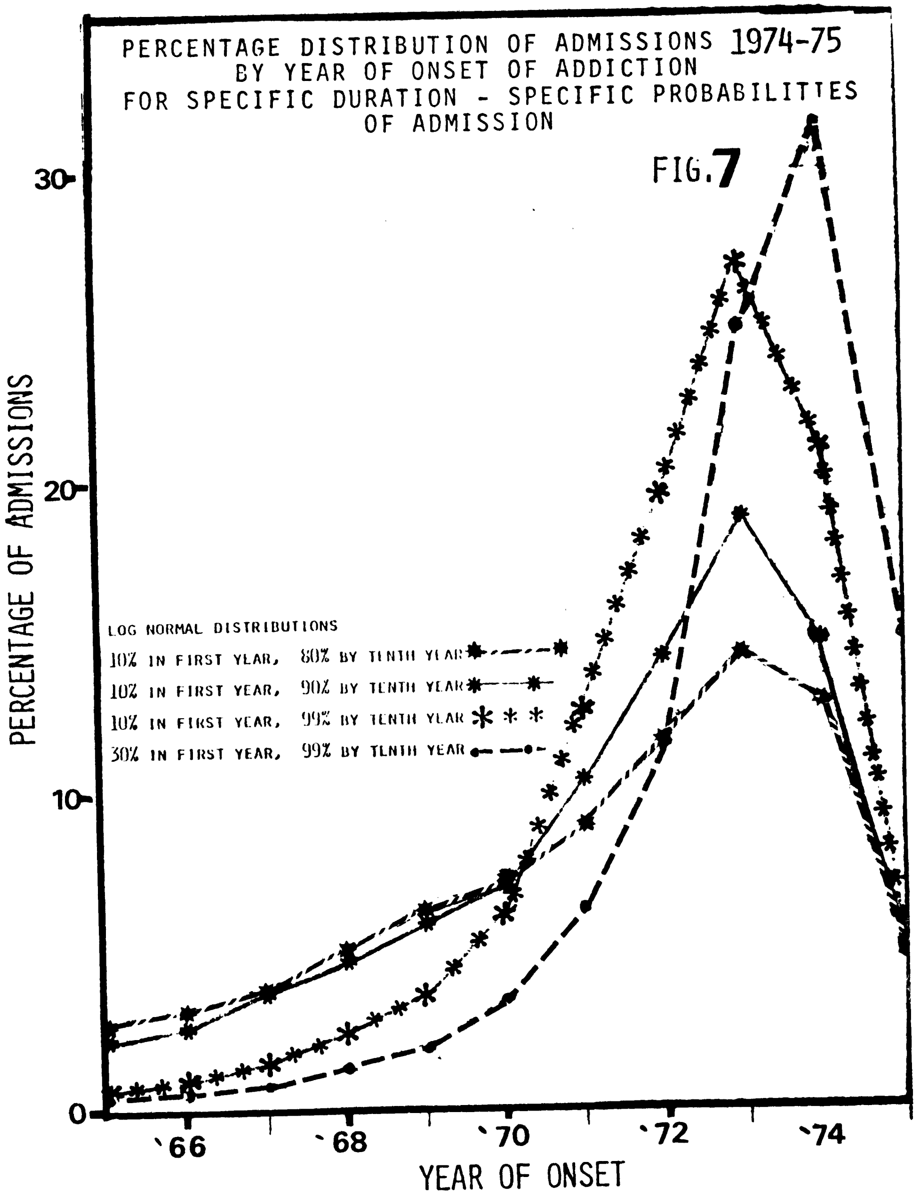
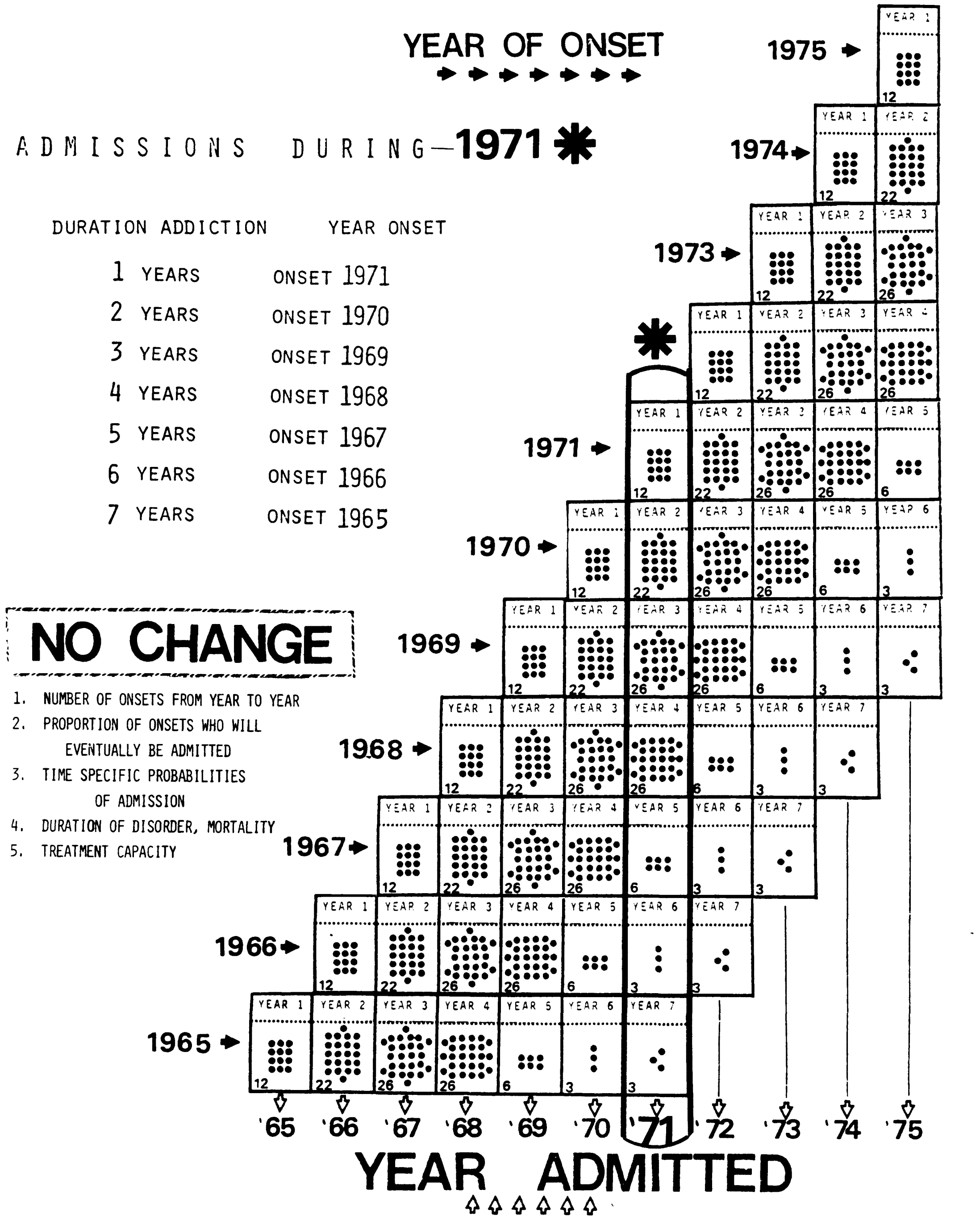


FIG.7

FIG.4



YEAR OF ONSET
ADMISSIONS DURING—1971
DURATION ADDICTION YEAR ONSET
1 YEARS ONSET 1971
2 YEARS ONSET 1970
3 YEARS ONSET 1969
4 YEARS ONSET 1968
5 YEARS ONSET 1967
6 YEARS ONSET 1966
7 YEARS ONSET 1965
NO CHANGE
1. NUMBER OF ONSETS FROM YEAR TO YEAR
2. PROPORTION OF ONSETS WHO WILL EVENTUALLY BE ADMITTED
3. TIME SPECIFIC PROBABILITIES OF ADMISSION
4. DURATION OF DISORDER, MORTALITY
5. TREATMENT CAPACITY
1965
1966
1967
1968
1969
1970
1971
1973
1974
1975
'65 '66 '67 '68 '69 '70 '71 '72 '73 '74 '75
YEAR ADMITTED

REARRANGEMENT OF FIGURE 4
TO EMPHASIZE
ADMISSIONS DURING 1971

FIG.5

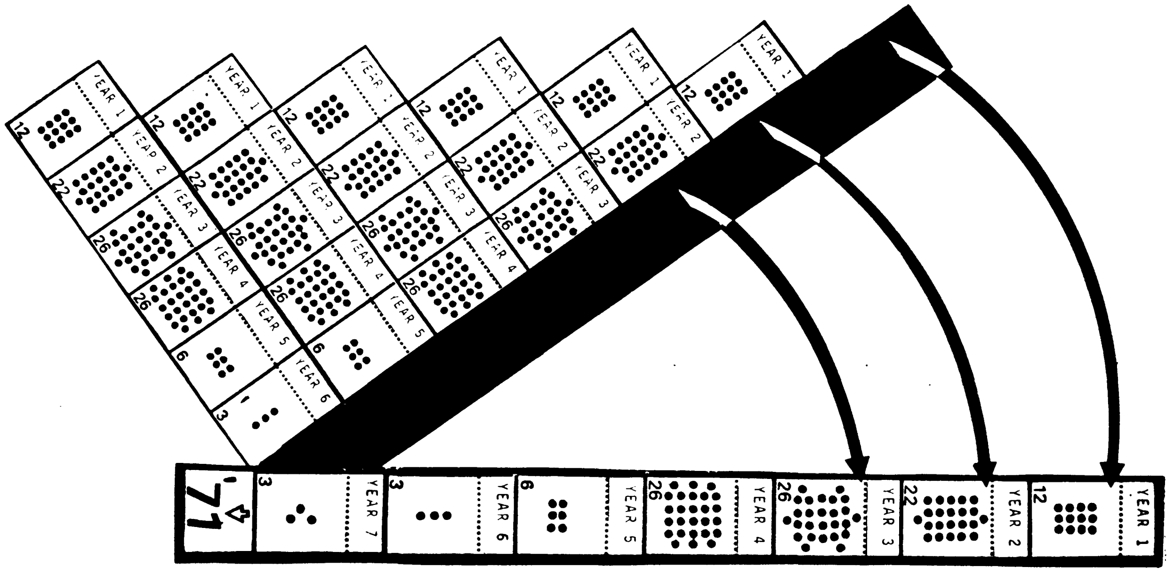
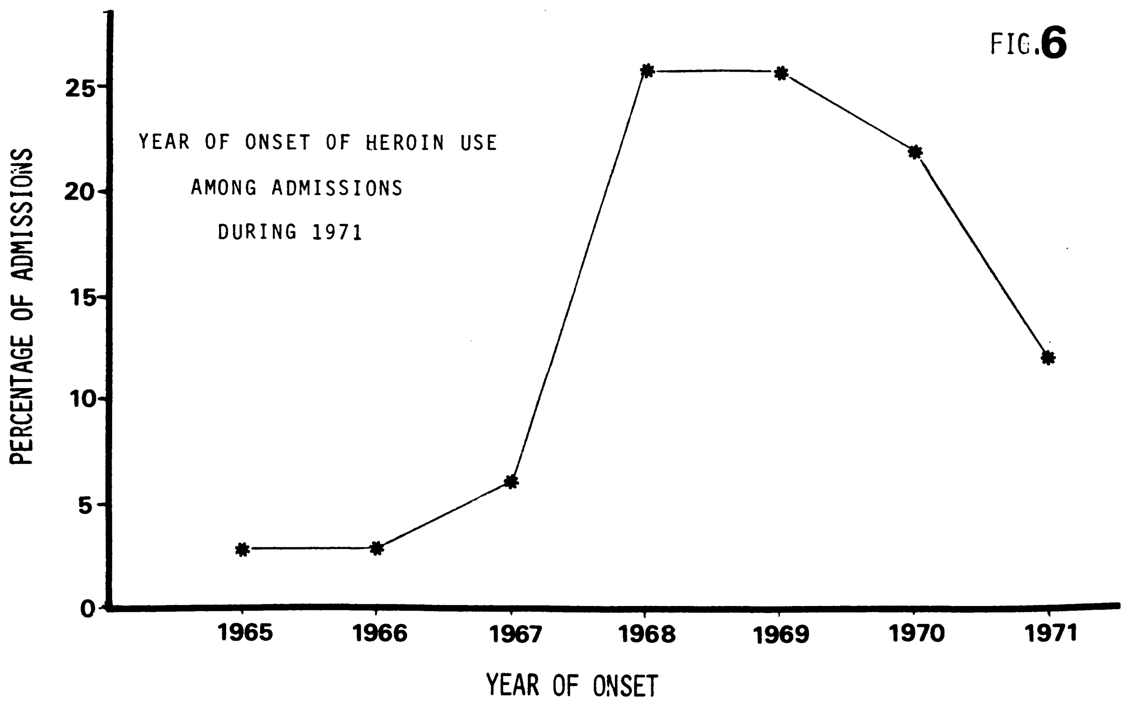


FIG.6



The assistance of Ms. T. Dubrow and Ms. P. Zappel, and the comments of: Drs. P. Kohn, P.V. Lemkau, B. Locke, F. Lundin, R.W. Morgan, J.A. Newmeyer, J.O'Donnell, P. Person, S.B. Sells, D.T. Wigle and K.W. Deuschle are acknowledged.

This research is supported by Grant Number 00666-03 from the National Institute on Drug Abuse.

307 Second Avenue
New York, N.Y. 10003

A MODIFIED HOUSING-UNIT METHOD FOR SMALL-AREA POPULATION ESTIMATION

Norfleet W. Rives, Jr., University of Delaware

1. INTRODUCTION

The College of Urban Affairs and Public Policy at the University of Delaware maintains an extended survey capability to support various planning and evaluation projects conducted by State and local public agencies. A central element of this survey capability is the New Castle County household address file. The University is located in New Castle County, the largest county in the State. The estimated 1975 population of the county is 395,000. The household population is distributed among approximately 130,000 housing units. The county household address file contains information on physical location for each house, apartment, or other structure occupied or intended for occupancy as separate living quarters.

The ability of the household address file to support different programs of survey research in New Castle County is well documented in numerous reports published by the College of Urban Affairs and Public Policy. These reports typically summarize the findings of household surveys conducted on a large scale, but, more recently, and with further improvement in the accuracy and completeness of the basic file, several projects on a smaller scale have been undertaken. One project is concerned with the estimation of local population.

The procedure involved in using the county address file to estimate population is based on the same principle as the housing-unit method for estimating the postcensal population of urban areas (Shryock and Siegel, 1971). The housing-unit method produces a figure for total resident population by combining information on (1) the stock of housing units, (2) average household size, (3) the vacancy rate, and (4) the population in group quarters. The estimating equation is:

$$P = bH(1-w)/(1-g), \qquad (1)$$

where: P = total resident population,

H = stock of housing units,

b = average household size,

w = vacancy rate, and

g = proportion of population in group quarters.

The usual method for estimating H is given by the expression:

$$H = H_c + H_a - H_d, \qquad (2)$$

where: H_c = stock of housing units in the most recent census,

H_a = housing units added to stock during the postcensal period, and

H_d = housing units demolished between census date and the estimation date.

Various sources of information may be used to obtain numerical values for b, w, and g.

The accuracy of the conventional housing-unit method clearly depends upon the accuracy with which the components of equation 1 can be estimated. The usual procedure in many areas is to derive estimates for b, w, and g from the most recent census. This procedure is simple and inexpensive, even though the dated usefulness of census information is well known. The usual procedure for estimating H is to update the census figure (H_c) with information collected locally on housing-unit completions and demolitions. According to several recent studies on the accuracy of different postcensal population estimation methods, the quality of completion and demolition certificates is subject to considerable variation at the local level, and this is a principle reason for the frequently poor performance of the housing-unit method (Morrison, 1971; Starsinic and Zitter, 1968).

The use of a household address file to estimate the current stock of housing units will certainly not overcome all of the problems associated with the use of census statistics in combination with unit completion and demolition certificates, but the former approach will not be any less effective, and, depending upon how the file is updated, it may actually be more effective. Segments of the New Castle County file, corresponding to specific geographic areas, are periodically re-field-listed in connection with particular survey projects, but this is not the primary method for continuous monitoring. The general file is routinely updated using (1) completion and demolition certificates, (2) public utility records on the connection and disconnection of electric meters, and (3) regular reports from survey personnel on the actual status of housing units encountered in the course of conducting interviews. This particular combination of information has been found to facilitate the maintenance of the address file, and, therefore, it should facilitate the estimation of the total number of county housing units on a given date.

If one can assume that a household address file provides a reasonably accurate and complete specification of a particular stock of housing units, then one is in the rather fortunate position of being able to construct a useful population estimate with a relatively simple statistical procedure. Population estimates produced by this procedure are designated modified housing-unit estimates, because the procedure is mechanically similar to the conventional housing-unit method. The following section presents the theoretical foundations of the proposed method, and the final section illustrates its application to a particular city in the State of Delaware.

2. THEORETICAL FRAMEWORK

Let H be the total number of housing units in a perfectly specified household address file for a given area, and let X_i be the number of persons who usually reside in the ith unit; vacant units will have X_i equal to zero. Suppose now that one samples this file, selecting n household addresses at random. The sample may be either simple random or systematic, depending upon whether one is willing to assume that the ordered series $(X_1,\ldots,X_H)$ forms a random sequence. In practice, this will usually be a tolerable assumption, primarily because systematic samples are less expensive to select and routinely produce only slightly larger variance terms than those obtained from simple random samples (Kish, 1965).

Information on the number of persons who usually reside at each household address in the sample group may be obtained either by mail or by personal interview. The mail approach is less expensive, but more difficult to control, and it may produce a troublesome level of nonresponse, even though one can hardly expect respondents to become disgusted and impatient with a simple prepaid postcard containing a pleasant explanatory note, a single non-threatening question, and possibly some official emblem such as a university seal. The interview approach is more expensive, but it is less difficult to control and can usually be relied upon to produce the desired information with relatively few problems. If only a single question is involved, however, as in the present case, then an interview survey can hardly be justified, unless the interviewing process is virtually costless. Conversely, if the single question on household size can be appended to a larger survey which has the proper sample design and would have been conducted in any event, then the marginal cost of interviewing to obtain household size data for the sample group is relatively small.

When the household address file has been sampled and the household survey completed, one will be left with the statistical series $(X_1,\ldots,X_n)$. Average household size (b) can then be estimated from the equation:

$$b = (X_1+\ldots+X_{n-k})/(n-k) \ , \tag{3}$$

where k is the number of vacant units in the sample. The estimated vacancy rate is given by the expression:

$$w = k/n \ . \tag{4}$$

A point estimate of the total population in households (P_h) can now be obtained from the equation:

$$P_h = bH(1-w) \ . \tag{5}$$

This estimate, combined with an independent estimate of the population in group quarters, can be used to produce a final point estimate of total resident population. If g is the proportion of the area population in group quarters, then total resident population (P) is given by the expression:

$$P = P_h/(1-g) \ . \tag{6}$$

To provide for the direct estimation of P from the sample data, equation 6 can be rewritten in the form:

$$P = bH(1-w)/(1-g) \ . \tag{7}$$

A useful source of information on the proportion g is the most recent census. If the local group-quarters population is small in relation to the total resident population, as is typically the case, then the census proportion can be safely applied in equation 7. In areas where the population in group quarters is proportionately large, this population will almost always have a primary source, such as an institution of higher education or a correctional facility. Information on the size of these populations can normally be obtained from the appropriate institutional officials, but, in the case of student populations, one should be careful to select only students who are area residents and not otherwise subject to the risk of having been included in the household survey. If the area under study contains a military installation, one should obtain a separate estimate of the total resident population of the installation from appropriate installation personnel. Information of this type is routinely available for official purposes, with the approval of the installation commander.

An important methodological issue in the estimation of average household size (b) and the vacancy rate (w) is the determination of sample size. The number of housing units surveyed will directly affect the cost of the project, the time required to complete it, and the statistical precision of the population estimate.

Assume that an estimate of the true vacancy rate (R) is desired accurate to within a fixed proportion E, with statistical precision given by the standard normal variate z. Under these circumstances, the minimum sample size (n_w) can be obtained from the expression:

$$n_w = \frac{Nz^2R(1-R)}{(N-1)E^2 + z^2R(1-R)} \ . \tag{8}$$

This expression is the familiar equation for the determination of sample size when the sampling frame is finite (Lapin, 1975).

Equation 8 cannot be solved without a value for R, and this puts one in the rather awkward position of needing the value of a parameter to find the value. Actually, the value for R in equation 8 is an intermediate value in the calculation procedure, and one can afford to select a value in a rather casual manner. If some estimate of the vacancy rate cannot be obtained locally, then the most recent census may be a useful source of information. A less desirable solution is to use the maximum value of the product R(1-R) in equation 8. Since R is a fraction, the product R(1-R) will reach a maxima when R equals 0.5, but it is quite unlikely that a local vacancy rate for all structures combined would ever reach 50 percent.

The minimum sample size required to estimate average household size can be found using an expression similar to equation 8. Let $V(X)$ be the variance of household size, and assume that an estimate of true average household size is desired accurate to within F persons per unit, with statistical precision again given by the standard normal variate z. The sample size in question is the solution to the equation:

$$n_x = \frac{Nz^2 V(X)}{(N-1)F^2 + z^2 V(X)} . \qquad (9)$$

This equation, like equation 8, is the usual expression for the determination of sample size when the sampling frame is finite (Lapin, 1975).

The variance term $V(X)$ in equation 9 should ideally be the population variance, but the actual parameter is almost always difficult to obtain. Census tabulations on household size could be used to construct a reasonably good approximation to $V(X)$, but this approach will be subject to one rather important limitation: Since the most recent census in 1970, there has been a significant increase in the rate of formation of primary (single-person) households in the United States (Kobrin, 1976). This means that the variance of household size has almost certainly declined during the postcensal period. The extent of any such reduction will obviously depend upon the particular local population being studied. In general, the estimated variance of household size obtained from the 1970 Census should be considered a maximum value, and, indeed, one may even want to use a smaller figure in the calculation for sample size.

The minimum sample size required to estimate the vacancy rate with a given statistical precision will rarely equal the minimum sample size required to estimate average household size with the same precision. Since the household survey operates on a single sample, the estimates from equations 8 and 9 must be reconciled. The most defensible solution to this problem is to select the larger of the two sample sizes. This will insure a certain minimum level of statistical precision for one of the parameter estimates and a somewhat higher-than-desired level for the other. Increasing the number of housing units to be surveyed beyond some lower limit should not be a cause for concern, because the marginal cost of sampling is usually quite small, especially when the survey is conducted by mail.

3. EMPIRICAL APPLICATION

The modified housing-unit method to estimate local population was originally evaluated using census and survey statistics for the City of Newark, Delaware. The Newark household address file was the first component of the New Castle County file, with other areas being added to the system between 1968 and 1973. The procedure for updating the Newark file is well established, and the quality of the information used in this process has been subjected to rigorous testing on more than one occasion. The most recent estimate of the total resident population of Newark was prepared for the midyear date, July 1, 1975. The following discussion summarizes the estimation procedure.

The Newark household address file contained 6,509 housing unit locations on July 9, 1975, the day the survey sample to estimate local population was actually drawn. Since the Newark file had been updated only five weeks before this date, no minor adjustments were made for file specification errors. Under other circumstances, it might have been necessary, or at least advisable, to adjust the file for known biases.

Prior to selecting the sample, it was decided that the vacancy rate should be estimated to within 1 percent and average household size to within 0.1 persons per unit, both with 80 percent statistical significance. According to an analysis of census tabulations, the approximate variance of household size in Newark in 1970 was 1.72. To allow for the effect of an increase in the proportion of primary households on the variance of household size, the census figure was reduced to 1.50; this adjustment is clearly arbitrary. The 1970 Census vacancy rate for the Newark area was less than 2.5 percent, but a 1974 estimate prepared locally put the rate closer to 3.0 percent. After some deliberation, it was decided to accept the more recent figure.

The minimum number of housing units required to estimate the vacancy rate with the desired precision was computed using equation 8. The solution value for n_w was 445. The corresponding solution value for the minimum sample size necessary to estimate average household size with the desired precision, following equation 9, was 237. According to the decision rule previously established, the minimum sample size for the survey should be 445, since 445 is obviously larger than 237. The actual survey sample consisted of 500 housing units, however, to allow for the effect of survey nonresponse. This adjustment was purposely smaller than the usual adjustment, because the survey, which was conducted by mail, contained only a single, relatively insensitive question. The usual over-sampling rate for mail surveys conducted by the College of Urban Affairs and Public Policy is closer to 20 percent.

Each household receiving a survey postcard was asked to provide a simple count of the number of persons who usually resided in the unit on July 1, 1975. The survey period was set at four weeks from the date of mailing, and, at the end of this period, 462 responses had been obtained. The Newark Post Office indicated that 15 units were vacant, and, according to other sources, 2 units had been demolished. The remaining 21 units were simply classified as legitimate nonresponses, although 3 postcards were received several months after the survey had technically been completed. No attempt was made to identify nonrespondent units.

The survey data were processed manually to produce estimates for the vacancy rate (w) and average household size (b). The estimated value of b was 2.94, and the estimated value of w was 3.1 percent. When these estimates are substitued into equation 5, one obtains an estimated household population of 18,553.

The final estimate of the total resident population requires only an estimate of the population in group quarters. The principal source of the group-quarters population in Newark is the University of Delaware. According to University enrollment records, the number of students to be added to the resident population of Newark was approximately 6,800. This figure was compiled from data for the preceding academic year and the first session of summer school.

If 6,800 were added to the population in households, then the estimated total resident population of the City of Newark on July 1, 1975, would be 25,353. The 1970 Census population was 21,078. This would make the average annual growth rate during the postcensal period 3.5 percent, an entirely plausible figure for a small metropolitan community in New Castle County.

REFERENCES

Kish, L. SURVEY SAMPLING. New York: Wiley, 1965.

Kobrin, F. "The Fall of Household Size and the Rise of the Primary Individual in the United States." DEMOGRAPHY 13 (1976): 127-138.

Lapin, L. STATISTICS: MEANING AND METHOD. New York: Harcourt, Brace, Jovanovich, 1975.

Morrison, P. DEMOGRAPHIC INFORMATION FOR CITIES: A MANUAL FOR ESTIMATING AND PROJECTING LOCAL POPULATION CHARACTERISTICS. R-618-HUD. Santa Monica: Rand, 1971.

Shryock, H. and J. Siegel. THE METHODS AND MATERIALS OF DEMOGRAPHY. 2 Vols. Washington: U. S. Bureau of the Census, 1971.

Starsinic, D. and M. Zitter. "Accuracy of the Housing Unit Method in Preparing Population Estimates for Cities." DEMOGRAPHY 5 (1968): 475-484.

STATE DEMOGRAPHIC CENTERS: THEIR CURRENT STATUS

Harry M. Rosenberg
The University of North Carolina

George C. Myers
Duke University

An element common to the statistical planning needs of many State and local programs is demographic information, that is, statistical data on the size, characteristics, and geographic distribution of the population. Such information serves a variety of governmental purposes below the national level. These include, for example, establishing program targeting needs in human services (health, education, social services, employment training), estimating population-related demand for capital improvements, allocating revenues under the many statutory formulae that include a population size factor, and conducting social and economic research for States and local areas.

Associated with the growing use of State and local demographic information has been the emergence in recent years of identifiable centers, within State government, capable of processing, analyzing, and disseminating population information directly relevant to the needs of specific State and local programs. Because development of such "State demographic centers" has proceeded almost autonomously, that is, largely in the absence of Federal guidance, standards, and resources, little comparative information has heretofore been available on the nature of the centers--on the scope of their activities, on the adequacy of their resources to meet the public's need, and on the way in which their activities articulate with those Federal statistical agencies whose capabilities they assist in extending at the state and local levels.

Recognizing the importance of strengthening the capability of States to make the most effective use of Census information, particularly as we approach the 1980 Census of Population, a survey of states was undertaken under the auspices of the Southern Regional Demographic Group,[1] to collect comparative information on the way in which state governments are presently organized to respond to State and local needs for demographic services.

The Survey

The survey, which covered 16 Southern states and the District of Columbia,[2] was designed to provide information on the nature of the administrative environment within which demographic research and services are provided by State government in the South, on the personnel and budgetary aspects of these activities, and on the nature of the actual services rendered. While the survey focussed on the Southern Region of the United States, it is our impression--after reviewing results of the survey with State officials outside the Region and with staff of the U.S. Bureau of the Census--that the general characteristics of the State operations and the issues surrounding them are not unique to the South. Indeed, because of the rapid modernization of government in the South, State demographic activities in this Region may be at the forefront of developments elsewhere in the Nation.

The survey entailed sending questionnaires to each State in the Region and to the District of Columbia, directed to those government officials thought to have the most comprehensive perspective and knowledge of state statistical activities. These were Officers of State Budget, State Planning officials, and technical persons identified by the U.S. Bureau of the Census as State Liaison for coordinating county population estimation activities.[3] Completed questionnaires were received from all the areas.

Results of the survey provide a source of comparative multi-state information on the way in which States provide statistical information on population to their constituents; on the way in which State governments extend the data use capabilities of the Census Bureau; and on how State demographic centers constitute an unheralded but important element of the larger national statistical system.

Administrative Setting

In many different agencies of State government there exists an impressive capability to use population statistics, particularly information from the Census of Population, in support program needs. For example, State health departments use population data in the analysis of vital statistics; Employment Security Commissions and State Departments of Labor use demographic information and methods to carry out manpower and labor force analyses; and the use of Census data and projection methods is often important in State economic development activities. More recently, population work has come to be associated with yet another administrative setting within State government, namely Offices of State Planning, which are emerging as the State foci for planning coordination among line departments.[4] These offices, which often work closely with the Governors' Offices and which provide program analysis support to the State Budget function, are now found in virtually every Southern State.[5]

Presented at the 1976 Annual Meeting of the American Statistical Association, Boston, Massachusetts, August 1976. This is to acknowledge Conrad Taeuber, Center for Population Research, Georgetown University and Meyer Zitter, Population Division, U.S. Bureau of the Census, who assisted in developing the questionnaire and made useful comments on the survey results.

The ubiquity of Census data and demographic capabilities in State government has had two important consequences over the years. The first is that many departments prepared their own current population estimates and population projections for the State and its geopolitical subdivisions long before the Census Bureau initiated similar activities at the National level. A second consequence of the dispersion of demographic work among State programs, was confusion among users, both local and State agencies, of a 'best' place to obtain demographic technical assistance and population data. This was complicated particularly with respect to identifying an 'official' set of population estimates and population projections which could be regarded as methodologically reliable and officially sanctioned for, say, planning or State revenue distribution purposes.

From a situation in which many State agencies have played a role in demographic work, there developed in many States the recognition of a need for a 'lead agency' that could carry the burden of providing official, authoritative figures to the public; that could consolidate costly data resources; that could coordinate liaison with local and Federal officials with respect to Census data; and that could respond to or coordinate responses to population queries directed at State government. This movement toward 'lead agency' designation was given impetus during the 1960's by two developments: the emerging planning coordination role of Offices of State Planning and the initiation of the Census Bureau's program, in 1967, to cooperate with States in the production of county population estimates.[6] A key feature of the Census Bureau's program was a requirement that the Governor designate a State official who would serve as State Liaison to the U.S. Bureau of the Census for population estimates. In his role, the State Liaison would provide the Census Bureau with State-based data for the Census production of his State's figures; and he would review the Census Bureau's results for his State. Gubernatorial designation of the State Liaison was tantamount to identifying the State lead agency for demographic activities from among the many engaged in demographic work. These lead agencies hereinafter are referred to as "State demographic centers," although we recognize that important demographic work, in support of State programs, is carried out in many other State agencies; and State-related population research is often carried out by State universities as well.

The distribution of agencies with lead responsibility for State demographic work is shown in Table 1. In the South, the majority of such centers are situated in State agencies, the balance in State universities. Among those located in agencies of State government, the predominant administrative setting is in Offices of State Planning. Seven of the ten State agency centers are in these offices; and the survey indicates that there is increasing movement of the demographic function toward this central location in State government. In response to a question on where these functions were previously located in State government, two States reporting such an interagency transfer specified a movement from the State Department of Health to the Office of State Planning.

In States where the Governor had designated the State University to act as Liaison to the Census Bureau for county population estimation, centers tended to be concentrated in the University bureaus of business and economic research. These bureaus have long served a function of providing extramural technical assistance in business, economics, and statistics to the State and local business community.[7]

Outside the Southern Region, Table 1 shows that State demographic centers are more often situated in State government agencies than in State universities; and among State agencies, State Health Departments and State Planning Offices play the role of lead demographic agency with equal frequency.

Functions

All the State demographic centers, by Gubernatorial designation, serve as State Liaison to the U.S. Bureau of the Census program for county population estimation. In that role, the State Centers accumulate necessary data for the Census estimation program; and they also review the final estimates prepared by the Bureau prior to official release. State centers' demographic activities usually go considerably beyond that. Many centers prepare sets of State population projections for counties and for smaller areas since these are not available from the U.S. Census of the Bureau. In all the Southern states, centers maintain files of published census information and have access to tapes of unpublished Census information. Two of the centers, in addition, are official Census Tape Processing Centers.

All centers report providing both statistical information and analytical technical assistance to other public agencies, and, within resource constraints, to private requestors. The volume of technical assistance services was not recorded on a uniform basis among the States; however, some respondents did provide some such information. One area, predominantly urban in character, reported responding to about 2,000 requests for demographic information each year; a similar figure was provided by a center in a predominantly rural State.

Personnel

In the South, the average personnel complement of the State demographic centers was between three and four persons including a programmer and at least one junior level person who served as a statistical clerk. The size of the units ranged, however, from as small as a single professional staff member to centers with as many as six staff members.

Questions about the length of professional service were asked in the survey. For seven of the 16 responding areas, results showed that the senior professional had been working in the center for less than five years; in one-fourth for less than one year. Such a record of short average tenure suggests that the centers experience considerable staff turnover; and indeed this has been identified independently by Census bureau staff as a considerable impediment to the success of the local population estimation program in which the States participate cooperatively with the Bureau. Census staff noted that their contacts with State demographers in the cooperative population estimation program had changed by 25-percent in a recent six months period.

Budgets

Average annual outlays in FY1975 for State-sponsored demographic activities were about $54,000 per center (Table 2). However, budgets varied considerably among States, from less than $30,000 per year to almost $100,000. Only three states supplemented State revenues with Federal grants to support central demographic activities. These Federal matching supplements, known as HUD-701 grants are available for general 'state planning' purposes;[8] they do not constitute a stable fiscal base with which to support sustained statistical programs at the State level.

Table 2 also indicates that State budgets expended for demographic activities are allocated mainly to build the State agency staff function, rather than to purchasing consulting services. Of the 17 reporting areas, only three used consultants. Average expenditures for consulting were small in comparison with total program outlays, about $12,000 annually.

To determine if there were any systematic variation in State demographic outlays, we related total expenditures to two variables, population size and per capita income for the States. Population size was viewed as a proxy for demand for State services, while income was seen as a measure of potential State resources. The analysis indicated that size of State is unrelated to expenditures for demographic services, but that there is a reasonably strong and statistically significant relation between outlays and State per capita income. Large income differences among Southern States, ranging from $2,600 to $5,300 in 1970, were associated with the variation in outlays for State demographic centers shown in Table 2.

Budget constraints constitute one of the most pressing and oft-cited problems of State demographic centers, even as they are for other Federal-State statistical activities.[9] In attempting to identify issues and problems that impede the effective provision of State demographic services, we asked open-ended questions about possible resource constraints. Seven areas provided comments focussing on budget problems; they are reproduced in full below:

- Budget is inadequate to provide more than minimal service.
- Budget limitations have prevented expansion of research into related areas, curtailed survey and field work, and exacerbated retention of staff.
- The State has decided that it needs demographic expertise at constant funding.
- The few demographic services [available in the State] are provided at the initiative of individual State agencies and must be incorporated into their separate operating budgets.
- The best that can be said [for this State's demographic budget] is that with a relatively good memory and graduate assistants who are replaced every other year, and with inadequate facilities to meet our needs, I have been able to hold things together. I hope we have been of some use to those who are working with State problems with needs for demographic inputs. [No budget data available for this university-based Center.]
- Small budget!
- We get numerous requests for demographic services from Federal, State, and local agencies, none of which are accompanied by offers of financial assistance. All of the funding for the Demographic Unit comes from State funds and, as such, priorities usually are arranged accordingly. It would be helpful to get some Federal funding, especially when massive Federal requests are made.

  We believe that the addition of age, race, and sex to the current estimates program would be an asset to all users of demographic data. However, at this time, we do not have sufficient funds to expand into this area.

Reports and Publications

All the reporting areas provided information on publication activities during the 1970-75 period, as summarized in Table 3.[10] A total of over 100 reports were issued, according to the survey results. Of these, less than half were publications of population estimates and population projections. Most were methodological and analytical studies, focussing on such subjects as migration, the analysis of demographic change, and general reports on the socio-economic characteristics of the population of the State and smaller areas.

Coordination of State Demographic Activities

That many State agencies have developed demographic expertise in support of their own programs has given rise to a considerable diffusion of these statistical resources, and has sometimes led to problems of statistical coordination. Dispersal of these capabilities, moreover, has often been at the expense of developing a strong, fiscally viable, professionally-staffed function that could serve many agencies and the public more generally as well. While emergence of the State planning function and Gubernatorial designation of a State liaison to the Census Bureau for population estimation have promoted consolidation of the State function, the survey results suggest that many States still bear the imprint of dispersed agency involvement in population statistics.

This expresses itself in two ways: one is that respondents in one-fourth of the areas did not know that there had been a Gubernatorially-designated liaison with Census Bureau staff for population estimation. Another related manifestation is continuing redundancy in the production of population estimates and population projections for substate areas such as municipalities. In one reporting State, three agencies currently prepare alternative population projections for counties. In another State, similar services are provided by the State Planning Office, the State Health Department, and the State university. In a third State, population estimates are prepared by one branch of the State university; projections by another.

The survey showed that the problems of coordination and redundancy of estimates were viewed as serious by State officials. An in-depth study by Rosenberg in North Carolina identified this as a major concern of State data users as well. One user noted,[11]

> We are in need of population bases which meet the highest standards of reliability. We are presently considering the use of the postcensal estimates prepared by one agency, but to date have no evidence that they are any better than other estimates of like specificity prepared by two other agencies.

Evans, commenting on his experience in South Carolina, writes,[12]

> We discovered in our work with State agencies that there were eight persons in seven different agencies making county population estimates and projections. In most cases there was a lack of methodology, or the person performing the work did not have sufficient qualifications to know whether the methodology was good or bad.

Fragmentation of State demographic services, and statistical services more generally, should be understood as partly reflecting the disjoint, episodic way in which Federal statistical activities are initiated in support of larger programs.[13] It also reflects an absence of statistical coordination at the Federal level due to a paucity of resources provided to the Statistical Policy Division of the Office of Management and Budget for this purpose.[14] Accomplishments in statistical coordination and consolidation of demographic services at the State level must be viewed as a tribute to State initiatives rather than as emulation of a Federal model.[15]

Cooperative Programs and Resources

State demographic centers, as an element of the national statistical system, can be viewed within the broad perspective of Federal-State statistical activities, in which partnerships have been forged between the Federal government and States for the collection, processing, and use of statistics in a number of areas. Such joint Federal-State statistical partnerships have evolved over the past 50 years, beginning in agriculture and labor statistics and now covering many subject areas. Cavanaugh, in a recent review, identified "cooperative statistical programs" in areas that include vital statistics, crime statistics, law enforcement, manpower and employment projection, income occupation, and labor force.[16] Duncan and Wallman have described cooperative programs in additional areas.[17]

The cooperative statistical programs strive to engage active State participation to upgrade statistical quality, improve statistical comparability among areas, and enhance usefulness of data to the local areas. Federal involvement assures uniform standards, provision of technical assistance and training, and financial assistance to reduce inequities in statistical program resources at the State and local levels. Wallman's recent analysis of cooperative programs stresses that these activities are highly variegated with respect to administrative arrangements, geographic coverage, program scope, resource and personnel configurations, and the respective roles of the Federal government and participating States.[18]

So central to the success of these joint statistical endeavors, are Federal resources that Wallman proposes, as a guideline, matching funds for all cooperative statistical activities between States and the Federal government.[19] The provision of Federal financial assistance to States participating in cooperative statistical activities recognizes the important Federal role for reducing geographic inequities in public program resources and for reimbursing States for those costs incurred in statistical reporting and analysis activities required by the Federal government. Many state demographic activities are in compliance with Federal requirements.[20]

Compared with other cooperative activities, Wallman has noted that Federal investments in State demographic activities have been very limited.[21] While budgets of cooperative programs are not comparable due to the varying nature of State statistical responsibilities, the level of Federal commitment to these activities is an instructive benchmark. Thus the U.S. Department of Agriculture's Statistical Reporting Service budget, expended for 400 field-based employees, is about $17 million annually.[22] The U.S. Bureau of Labor Statistics allocates some $3 million annually to support regional and State operations.[23] The Cooperative Health Statistics System spent nearly $7 million in FY1975 to support research and development activities that will eventually lead to an operational national health statistics system.[24] Finally, in the area of Law Enforcement, some $21 million has been made available to States in the form of grants for statistical activities.[25] Currently, no Federal resources are allocated to States to strengthen their Census data use capabilities, or to participate in the Census Bureau's cooperative program for population estimation.

Federal and State Roles
In Cooperative Demographic Activities

Both Federal and State action can contribute to improving the quality of U.S. demographic services at all levels of government. State initiatives are important particularly in the areas of coordination and consolidation of services. These can reduce redundant acquisition of costly data and duplication of population estimates and population projections. Consolidation of activities might also make available additional resources to strengthen a central capability.

The survey suggests that that is a need for increased Federal commitment to cooperative demographic activities. Areas in which joint activities might be initiated or bolstered include, for example: fostering information exchange on State programs; implementing a broad training program for State personnel on the use of Census data and related statistical information available from Federal agencies; developing methodological guidelines for local population estimation and population projection; and formulating guidelines for quality control. In each of these areas, there are opportunities for joint Federal-State participation.

The survey leaves no doubt that the greatest single requirement for strengthening cooperative demographic activities is provision of matching Federal dollars to supplement existing State investments for this purpose. The provision of resources can serve not only to improve and extend capabilities; it can also serve the important symbolic purpose of recognizing a State role in the national statistical system for Census data use.

Footnotes

1. The Southern Regional Demographic Group is devoted to promoting research and teaching in demography, and to improving the use and quality of demographic information and services. The Association is sponsored by the Oak Ridge Associated Universities, a consortium of 43 colleges and universities in the South.

Papers related to State demographic activities were presented at the Annual Meeting of the Southern Regional Demographic Group, Atlanta, Georgia, October 1976:

- Frederick J. Cavanaugh, "The Perspective of the Federal Government on the Role of State Government in Demographic Activities: A Joint Governmental Effort."
- Thomas P. Evans, "The Perspective of State Government on Demographic Activity."
- Forrest H. Pollard, "Issues in Relating to State and Federal Agencies."
- Harry M. Rosenberg, "State Demographic Activities--A View from the South."
- William J. Serow, "State and Local Population Estimates: Issues on Relating to Local Government."

2. Alabama, Arkansas, Delaware, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia.

3. State liaison are listed in U.S. Bureau of the Census, Current Population Reports, "Federal-State Cooperative Program for Local Population Estimates--Status Report: January 1975", Series P-26, No. 118, July 1975.

4. Leonard U. Wilson and L.V. Watkins, "How the States Plan," Challenge 18(6), January-February 1976, pp. 43-51.

5. Council of State Planning Agencies, Council of State Governments, 1970. James A. Catanese, "Testing of An Emerging Model of State Planning: A Report Card," Atlanta, Georgia: Georgia Institute of Technology, May 1972.

6. U.S. Bureau of the Census, op. cit., July 1975, p. 1.

7. E.g., Pollard, op. cit., pp. 1-2.

8. HUD-701 refers to state and local planning grants from the U.S. Department of Housing and Urban Development authorized under provisions of Section 701 of the Housing Act of 1954, as amended.

9. Joseph W. Duncan and Katherine K. Wallman, "Regional Statistics and Federal-State Cooperation," keynote address presented at the annual meeting of the Association for University Business and Economic Research, Williamsberg, Virginia, October 20, 1975, p. 11.

10. The list of publications is available, on request, from Harry M. Rosenberg.

11. Harry M. Rosenberg, "North Carolina Demographic Data Needs and Capabilities in Proceedings of the North Carolina Demographic Data Workshop, Chapel Hill: Carolina Population Center, University of North Carolina, 1975, p. 34.

12. Communication with author.

13. Harry M. Rosenberg, "Demographic Data and the Public Planning Mandate," in Proceedings of the North Carolina Demographic Data Workshop, Chapel Hill: Carolina Population Center, University of North Carolina, 1975, p. 11.

Joseph W. Duncan, "Developing Plans and Setting Priorities in Statistical Systems, "Statistical Reporter, No. 76-14, August 1976, p. 281.

14. Summary of "Report of the Joint Ad Hoc Committee on Government Statistics," in

Newsletter of the Federal Statistics Users' Conference, 17(8), August 25, 1976, p. 3.

15. Harry M. Rosenberg, "State Initiatives in Improving Demographic Data," Public Data Use, 3(2), April 1975, pp. 16-21.

16. Cavanaugh, op. cit.

17. Duncan and Wallman, op. cit.

18. Katherine K. Wallman, " 'Getting It All Together': The Development of Appropriate Relationships Between Federal and State Governments for Statistical Programs," paper presented at the Annual Meeting, American Statistical Association, Boston, August 1976, p. 13.

19. Ibid., p. 19.

20. Ibid., p. 13.

21. Wallman, op. cit., p. 16.

22. Duncan and Wallman, op. cit., p. 5.

23. Ibid.

24. Ibid., p. 8.

25. Ibid., p. 6.

TABLE 1. STATE AGENCY WITH PRIMARY RESPONSIBILITY FOR CURRENT POPULATION ESTIMATES UNITED STATES, 1975

Agency	South a/		Rest of Nation b/	
	Number	Percent	Number	Percent
Total	18 c/	100.0	34	100.0
University	8	44.4	6	17.7
Center or Institute	2	11.1	0	0.0
Business & Economic Research	4	22.1	4	11.8
Sociology Department	1	5.6	2	5.9
Extension Service	1	5.6	0	0.0
State Government	10	55.6	28	82.3
State Planning	7	38.8	9	26.5
Economic Development	0	0.0	4	11.7
Employment Security	1	5.6	2	5.9
Health	1	5.6	9	26.5
Finance/Budget Control	1	5.6	3	8.8
Labor	0	0.0	1	2.9

a/ Includes District of Columbia

b/ Excludes Puerto Rico

c/ In one of the 17 reporting areas, there are two officially-recognized centers, one in a state agency (State Planning), the other in the state university (Center for Business and Economic Research).

TABLE 2. DISTRIBUTION OF SOUTHERN STATES
BY SIZE OF BUDGET FOR STATE-SPONSORED
DEMOGRAPHIC ACTIVITIES
BY PRESENCE OF FEDERAL SUPPORT
AND BY USE OF CONSULTANTS, 1975

FY 1975 Budget	Number of States	Type and Amount of Federal Support[a/]	Use of Consultants
$ 15,000 - 29,999	4		
30,000 - 49,999	2	One state, HUD 701	One state
50,000 - 69,999	3	One state, HUD 701	
70,000 - 89,999	2		Two states
90,000 - 100,000	3	One state, HUD 701	
Average	$ 53,900	$ 28,300 [b/]	$ 12,000[b/]

a/ HUD 701 refers to planning grants from the U. S. Department of Housing and Urban Development under the provisions of Section 701 of the Housing Act of 1954, as amended.

b/ Average for these areas.

TABLE 3. STATE-SPONSORED DEMOGRAPHIC PUBLICATIONS
BY TYPE, FOR THE SOUTHERN STATES
1970-1975

Type of Publications	Number	Percent
Population estimates	22	19.3
Population projections	23	20.2
Socio-economic analyses	9	7.9
Migration	12	10.5
Methodology	8	7.0
Demographic analysis	36	31.6
Housing analysis	4	3.5
TOTAL	114	100.0
Average per area	6.7	--

Note: Includes 16 Southern states and the District of Columbia. See text for explanation of number of reporting areas.

DRUG ABUSE: CURRENT ESTIMATE OF THE COSTS TO SOCIETY[1/]

Brent L. Rufener, Alvin M. Cruze, J. Valley Rachal, Research Triangle Institute

INTRODUCTION

This paper provides an estimate of the total economic costs experienced by the U.S. society due to the abuse of drugs. This is part of the information needed by society in its policymaking regarding drug abuse in order that alternative efforts to prevent and treat this serious social problem may be analyzed and the most effective and efficient selected for implementation.

The total costs are computed by estimating the individual components of the total. The reference time period is fiscal year 1975. An important characteristic of the study is that only secondary data were available for this effort. Thus, while we used what we considered the best and most current information of which we were aware, numerous conceptual, definitional and measurement shortcomings are evident in many of the data used. We attempt to point these out as appropriate.

This paper is based in part on a study conducted at Johns Hopkins University [1], hereinafter referred to as the Hopkins study. The methods and organization of the Hopkins study were used as we think appropriate in our effort to revise and update the economic costs. The Hopkins study was a useful point of departure for this effort and was an advance over previous work that had been done [2,3,4].

This paper does not purport to be the definitive answer regarding the economic costs to society of drug abuse. In some cases it was necessary to use data of questionable reliability where nothing better was available. In other cases it was necessary to use data from small, limited studies to estimate totals for the entire U.S. For some cost components, too little is known about the related behavioral phenomena to provide definite conclusions regarding the impact of drug abuse. For example, the amount of non-drug criminal activity "caused" by drug abuse is an important question about which definite answers are not yet available. Subject to the qualifications given, we feel, however, that this study presents a useful national estimate of the economic costs to society of drug abuse in fiscal year 1975.

METHODOLOGY

General

Our basic approach to developing these estimates of the costs of drug abuse to society is to treat all costs in terms of foregone opportunities for members of society to use resources to produce goods and services. This basic approach focuses on the real costs of drug abuse to society and does not include the effects of only changes in relative prices. These latter effects, termed pecuniary effects in the benefit cost and program evaluation literature, merely serve to redistribute existing goods among members of society and do not alter the production capacity of society.

Given this basic approach, economic costs may be distinguished as being either explicit or implicit. An explicit or direct cost is one in which resources are used and a formal payment is made in cash or in kind (i.e., through the direct provision of some commodity or service). When resources are used to treat the medical consequences of drug abuse, the labor is paid for in wages, the materials used are paid for and the capital used receives a return. On the other hand, an implicit or indirect cost occurs whenever resources are used for which no formal payment is made, that is, these resources are not priced by any market mechanism. When a drug abuser is treated in a hospital his time is being used, but no formal payment is made for its use. In order to place a value on this implicit cost one looks at the value of what the individual gave up, or the opportunity cost. In other words, when an abuser is treated in a hospital he is unavailable for work and the implicit cost of his time is the goods and services he could have produced during the treatment. In assessing the total costs to society of drug abuse it is necessary to include these implicit costs, since they represent foregone production that could be used to improve the total welfare of society. These implicit costs are easier to overlook than explicit costs in enumerating the entire range of cost elements, so care must be taken to be sure all implicit costs are included.

In conducting an analysis of costs it is necessary to use a common unit of measure in which to express the costs so that they can be added together. The most common unit of measure used is the currency of the country for which the analysis is being conducted, in our case dollars. When expressing costs in monetary units it is quickly obvious that some costs cannot be so expressed. Here we must make the distinction between tangible and intangible costs. Tangible costs are those that can be priced, while intangible costs are those that cannot be priced. When one says that a cost cannot be valued it should be made clear whether the cost cannot logically be valued or cannot be valued without an excessive use of resources. Often it is the latter that is true. It is important to note that in this paper we present only those costs that can be valued with a reasonable use of resources, that is the tangible costs. The term intangible will be used to refer to those costs that cannot either logically or reasonably be valued, such as the costs of "pain and suffering." Costs of such elements as the displeasure people feel when viewing abusers and the conditions drug abuse creates, the disruption of neighborhoods, destruction of families, and the corruption of law enforcement officials are additional examples of intangible costs that are excluded from our study.

A final point is important regarding the economic costs of drug abuse. This paper estimates the economic costs to society of drug abuse for only one year. Some costs, however, extend for more than one year. For example, if a drug abuser dies in fiscal year 1975 not only is an economic cost present in that year but it extends into the future as well. The cost would be the loss of the goods and services the individual would have produced throughout his or her working lifetime. If costs in the period beyond fiscal year 1975 were included it would be necessary to discount the future costs to express them on a basis comparable to those present in fiscal year 1975.

Principal Cost Elements

In developing these cost estimates, the first step that was undertaken was an enumeration of all of the major items of costs that will be borne by society. This initial step can be thought of as developing a cost framework or model and is undertaken in order to ensure the inclusion of all relevant costs. The major cost items that were identified for this study are listed below, categorized by the major classification of either direct or indirect costs.

Direct Costs	Indirect Costs
Increases in:	Increases in:
Medical treatment	Unemployability
Law enforcement	Emergency room treatment
Judicial system use	Inpatient hospitalization
Corrections	Mental hospitalization
Nondrug crime	Drug-related deaths
Drug traffic control	Absenteeism
Drug abuse prevention	Incarceration
Housing stock loss	Work lost due to treatment

A variety of secondary data sources and estimation procedures were used to estimate the magnitude of each of these cost elements for fiscal 1975. Details of these procedures, together with a complete description of the scope of each of the cost elements, are available in a separate report [5]. Highlights of particularly critical methodological considerations or assumptions that were used in developing these estimates are summarized in the following paragraphs.

Special Issues

Within the perspective of the cost framework above, we note again that this report estimates the economic costs that were present in fiscal year 1975. The number of abusers and their activities were a historical fact at this point in time. While is is reasonable to assume that the number of abusers and the economic costs vary directly, for many cost components we simply measured the costs incurred; it was unnecessary to know the number of addicts or abusers. This method of estimating costs might be called the "aggregate" approach; only total cost figures are used. For other cost components, however, we were unable to obtain reasonable cost estimates using the aggregate approach and thus reverted to what we term the "summing approach." This approach computes the cost per abuser and then sums the cost across estimates of the number of abusers. The nondrug crime element of the direct costs and the unemployability element of the indirect costs are computed using this second method. The estimates of these cost elements thus depend on the estimates of the number of abusers assumed, even though actual costs and number of abusers are indeed historical facts.

In a related research effort, statistical procedures for estimating the number of heroin addicts were critically reviewed [6]. These methods include the capture-recapture method, widely used to estimate the size of a fish or wildlife population, analyses of prevalence data from various secondary data sources, and indirect verification of incomplete prevalence data on narcotics addiction from one source with independent data from another source. The overall conclusion of this review was that the statistical methodologies used to develop the estimates were sound but that the results were based on a number of tenuous assumptions that were difficult, if not impossible, to verify and the uncertain quality of available data.

Due to the uncertainties in estimating the size of this population, we decided to employ a range of estimates for purposes of developing these cost estimates. This range includes three values--250,000, 500,000 and 750,000--each of which has been cited in the literature as an estimate of the number of heroin addicts in the U.S. The results presented in the following section have been developed for these three assumptions which are labelled low, medium, and high, respectively.

A second major problem with respect to estimating the social costs of drug abuse is to measure the extent to which nondrug crime committed by drug abusers is caused by drug abuse. Previous studies of this type have been based on the assumption that all of the nondrug crime committed by drug abusers is caused by drug abuse. This assumption clearly produces upwardly biased estimates of this cost element, since evidence indicates that many drug abusers would have engaged in criminal activities even in the absence of their drug abusing behavior [7]. In the absence of an acceptable measure of the strength of the relationship between drug use and criminal behavior, the following procedure was used in developing our estimates.

First, we assumed that only the abuse of heroin causes significant amounts of nondrug crime. It was also assumed that 60 percent of the criminal behavior of heroin abusers is caused by the abuse. This figure is based on a previous study [8] in which it was found that 50 percent of a heroin abuser's funds come from income-producing crime. Based on interviews with addicts treated in Phoenix House, it was estimated that a one dollar increase in habit cost will lead to income-producing crime that nets the addict \$.30. Thus .30/.50 or 60 percent of nondrug criminal activity by heroin abusers is assumed to be caused by the heroin abuse.

Another study [9] indicated that of those arrested for the crimes of robbery, burglary, larceny-theft, stolen property violations, weapons violations and commercialized vice, 30.9 percent were current heroin users, as determined through a urine sample or questionnaire response. Although the categories of crime are not precisely the same as those used to determine the percent of all arrests that were for crimes of which heroin abuse causes more frequent violation, the categories seem close enough so that no adjustment is necessary. With 30.9 percent of the arrests associated with heroin abuse, 60 percent of this amount is assumed to be caused by the abuse. Thus, 18.5 percent (60% x 30.9%) of all arrests in the above categories are assumed to be caused by heroin abuse. This percentage figure was used for appropriate cost elements in order to estimate the proportion of total costs for that element that was caused by drug abuse.

The final special issue of our methodology that will be highlighted in this paper is the procedure we used to estimate the costs of nondrug crime associated with drug abuse. Becker [10] and Tullock [11] indicate that the total real income loss to society from crime is the sum of (a) the labor and capital inputs into criminal activity and (b) the costs of controlling crime. These latter costs, which consist of enforcement, judicial system, and correction costs, were estimated from various secondary data sources in a relatively straightforward manner in our study. The labor and capital inputs into criminal activity represent a cost to society in that the resources used for criminal activities are unavailable for other uses. The amount of labor and capital used in criminal activities is a direct cost because as this resource is used it is paid with stolen goods (payment in kind) and/or stolen cash. Several analyses of the social costs of drug abuse refer to the value of stolen property as the direct cost of drug abuse. In fact, however, it is only an indirect measure of the labor and capital inputs to criminal activity. The value of the stolen goods contains an involuntary transfer element which is not a cost to society. Including this involuntary transfer as part of the costs of nondrug crime (as many estimates have) grossly overestimates the social cost due to this element.

In estimating this cost element we first assumed that only heroin abuse results in significant amounts of nondrug criminal activity, since other drugs are available at sufficiently low prices so that abusers do not have to commit crimes in order to support their abusive behavior. Next we estimated the mean earnings of a heroin abuser. Using data on the age, race and sex distribution of heroin abusers furnished by NIDA and data on earnings of full-time workers by age, race and sex from the U.S. Department of Commerce [12], we estimated the annual earnings of the average heroin addict if he were employed full-time, as $8,896 in 1975.

The next step in calculating the cost of nondrug criminal activity is to determine what proportion of the earnings of a heroin abuser is obtained from nondrug criminal activity caused by heroin abuse. This proportion is then multiplied times the mean earnings figure to give the cost per abuser. By using this procedure the assumption is being made that a heroin abuser spends the same amount of time "working" in criminal behavior and other legitimate income-producing activities as the average full-time employee spends in legitimate activity. It may be the case that heroin abusers spend more total time in these two activities because of the need to secure drugs. Because of this possibility the estimates of the cost of nondrug criminal activity produced by our procedure may be an underestimate of this element.

Goldman estimated that for a one dollar increase in habit cost an abuser will commit crimes that will bring him thirty cents, which indicates that 30 percent of an abuser's income used for drugs comes from drug-caused criminal activities [8]. Goldman did not offer any evidence regarding abusers' living expenses. It will be assumed that 30 percent of the total income of an addict comes from drug-caused crime. Since we do not have any detailed data on the allocation of an addict's time, we assume that 30 percent of an abuser's "working" time is spent in drug-caused crime.

The percent income and percent time spent may not be the same if certain activities are more remunerative. At the margin the return per hour should be the same for various activities, but the percent time spent for an activity and the percent income received may not be equal. Multiplying 30 percent times $8896 gives a cost per abuser of $2669. The total estimated cost for this element was then obtained by multiplying this dollar figure by the three assumed numbers of abusers.

RESULTS

Based on the framework and assumptions developed in this study the economic costs to society of drug abuse for fiscal year 1975 are estimated to be between $8.4 and $12.2 billion, with a middle estimate of $10.3 billion (table 1). Approximately 48 percent of these costs are due to direct costs or expenditures for services or goods needed as a direct result of drug abuse. Details of these cost components are presented in table 2. The remaining portion of the total costs are indirect costs and result from decreased production of goods and services as a result of drug abuse. Details of these cost components are presented in table 3.

Given these cost estimates, it is possible to develop a more detailed examination of the various cost components and to indicate a number of general policy implications of the results of this study. These elements include a separation of these costs into those associated with heroin abuse and those associated with other drugs, an indication of who in society bears these costs, and a comparison of the costs of drug abuse with the costs to society of alcohol abuse and alcoholism. Each of these topics is briefly summarized in the remainder of this paper.

The total estimated costs of drug abuse reported in this study are the result of the abuse of many different types of drugs. One point of interest for readers of this paper is to have this total cost separated into the cost of heroin abuse and the cost of other drug abuse. Heroin abuse is considered by many to be the most serious abuse problem in the U.S. and has certainly been the target of substantial research, prevention and control efforts. Based on the details presented in [5] and the three assumed numbers of heroin addicts it is estimated that the economic

TABLE 1
The Economic Costs of Drug Abuse, Fiscal Year 1975 ($ Millions)

Cost Component	Assumed Number of Addicts* Low	Middle	High
Direct Costs	$4,265	$4,932	$5,599
Medical Expenses	494	494	494
Law Enforcement	1,342	1,342	1,342
Judicial System	296	296	296
Correction	294	294	294
Nondrug Crime	667	1,334	2,001
Drug Traffic Control	93	93	93
Drug Abuse Prevention	995	995	995
Housing Stock Loss	84	84	84
Indirect Costs	$4,167	$5,406	$6,644
Unemployability	1,239	2,478	3,716
Emergency Room Treatment	0.4	0.4	0.4
Inpatient Hospitalization	20	20	20
Mental Hospitalization	8	8	8
Drug-related Deaths	12.5	12.5	12.5
Absenteeism	1,594	1,594	1,594
Incarceration	1,205	1,205	1,205
Treatment Costs	88	88	88
TOTAL COSTS	$8,432	$10,338	$12,243

*The three assumed numbers of heroin addicts were: 250,000, 500,000, and 750,000.

TABLE 2
Direct Costs of Drug Abuse, Fiscal Year 1975 ($ Millions)

Cost Component	Assumed Number of Addicts* Low	Middle	High
Medical Expenses	$ 494	$ 494	$ 494
Law Enforcement	$1,342	$1,342	$1,342
Drug laws	667	667	667
Nondrug laws	675	675	675
Judicial System	$ 296	$ 296	$ 296
Drug laws	172	172	172
Nondrug laws	124	124	124
Correction	$ 294	$ 294	$ 294
Drug laws	158	158	158
Nondrug laws	136	136	136
Nondrug Crime	$ 667	$1,334	$2,001
Drug Traffic Control	93	93	93
Drug Abuse Prevention	$ 995	$ 995	$ 995
Housing Stock Loss	$ 84	$ 84	$ 84
TOTAL DIRECT COSTS	$4,265	$4,932	$5,599

*The three assumed numbers of heroin addicts were: 250,000, 500,000, and 750,000.

cost of heroin abuse in the U.S. in fiscal year 1975 was:

Low number of addicts: $4.49 billion
Middle number of addicts: $6.39 billion
High number of addicts: $8.30 billion

In addition, we have estimated that the cost of the abuse of other drugs in the U.S. in fiscal year 1975 was $3.95 billion. Thus, 62 percent of the total economic costs of drug abuse can be attributed to heroin abuse, using the middle assumption of the number of addicts.

One of the questions of interest in conducting a study of the economic costs of a social problem is to determine who bears these costs. Among the direct costs the medical expenses are borne by taxpayers in the form of increased local and state taxes to pay for hospitalization and increased local, state and federal taxes to pay for medical care through programs to help the indigent. The costs of law enforcement, judicial system use and corrections are borne by those who pay taxes, with the greatest part of the cost borne by those paying local taxes. The costs of nondrug crime are borne by all members of society in the form of reduced goods and services. The costs of control and prevention are borne by

TABLE 3
Indirect Costs of Drug Abuse, Fiscal Year 1975 ($ Millions)

Cost Component	Assumed Number of Addicts* Low	Middle	High
Unemployability	$1,239	$2,478	$3,716
Emergency Room Treatment	0.4	0.4	0.4
Inpatient Hospitalization	20	20	20
Mental Hospitalization	8	8	8
Drug-related Deaths	12.5	12.5	12.5
Absenteeism	1,594	1,594	1,594
Incarceration	$1,205	$1,205	$1,205
Drug laws	845	845	845
Nondrug laws	360	360	360
Treatment Costs	88	88	88
TOTAL INDIRECT COSTS	$4,167	$5,406	$6,644

*The three assumed numbers of heroin addicts were: 250,000, 500,000, and 750,000.

taxpayers, primarily through increased federal and state taxes. The cost of the housing stock loss is borne by those who pay insurance on housing in the form of higher premiums.

For the indirect costs it is primarily the abuser himself who suffers the loss of goods and services produced through his or her lost net income. Taxpayers also bear part of this burden as abusers pay less taxes on their income. For some indirect costs an abuser does not suffer an actual loss of income. An example would be where an abuser misses work due to some form of hospitalization but is still paid his salary. In this case society as a whole bears the burden in the form of reduced wages, salaries and profits to pay for the time off.

This study is an analysis of the economic costs of drug abuse to society as a whole. One of the costs to society is the labor used in nondrug criminal activity, which represents foregone opportunities to produce goods and services that benefit all members of society. As pointed out in other sections of this paper, the value of these stolen goods does not represent a net cost to society in the usual economic sense. This is due to the fact that these stolen goods represent an involuntary transfer among members of society and are not associated with foregoing production opportunities.

However, if one separates society into two groups, those from whom goods are or may be stolen and those who do or may receive stolen goods, then it may be of interest to examine the economic cost of stolen goods to the former group. In this case it is appropriate to include the involuntary transfer that takes place when goods are stolen. The cost of stolen goods directly associated with drug abuse in fiscal year 1975 is estimated to be $396 million. Details of the calculations and sources used in developing this estimate are presented in [5]. It should be noted that drug abusers are not among those who typically receive stolen goods. Abusers perform the transfer and usually sell the goods to others to obtain cash. It may be the case that this transfer is relatively efficient when compared to other forms of transferring income.

Finally, it may be useful to compare the economic costs of drug abuse to the costs of other social problems in order to gain some insight into the relative magnitude of the costs of these problems. Alcoholism and smoking are problems similar to drug abuse in that all involve the use of a substance in the body to produce physical and mental changes which yield undesirable consequences for health and behavior. The economic costs of alcohol abuse and alcoholism are detailed in table 4. We have not critically examined the cost data provided. Since the study estimated the costs in 1971 we are unable to precisely determine what the costs were in fiscal year 1975. However, we did adjust these costs using changes in the Consumer Price Index over this four year period with an estimate of $32.04 billion for fiscal 1975.

The economic costs of smoking were estimated in the Little study [2]. In 1971 the estimated total cost in the U.S. was $5.625 billion. The Little study used a study based on Canadian data, and we have not attempted to assess the appropriateness of using Canadian data for the U.S. Adjusting this figure to fiscal year 1975 prices using changes in the Consumer Price Index gives an estimated total cost to U.S. society from smoking of $6.735 billion for the 1975 fiscal year.

TABLE 4
Economic Costs of Alcohol Abuse and Alcoholism
($ Billions)

Cost Component	1971 Prices	FY 1975 Prices*
Lost Production	$ 9.35	$11.98
Health and Medical	8.29	10.34
Motor Vehicle Accidents	6.44	8.25
Alcohol Programs and Research	0.64	0.82
Criminal Justice System	0.51	0.65
TOTAL	$25.23	$32.04

*Adjusted using the Consumer Price Index. For health and medical care the medical care group was used; for the other cost components the entire CPI less medical care was used.

SOURCE: Reference [13].

Thus, using these sources and procedures, the costs to society of drug abuse in fiscal 1975 are approximately 60 percent larger than the costs of smoking but only about one third as large as the costs of alcohol abuse and alcoholism.

In summary, we believe the techniques used for developing the costs in this paper are economically sound. The basic shortcomings of the estimates developed are related to the unavailability or deficiencies of the data used and/or of knowledge about relationships between drug abuse and behavioral or other phenomena. This is particularly true and relevant with respect to the criminal behavior/drug abuse and unemployment/drug abuse relationships, as mentioned previously. Further, we did not undertake to develop new and perhaps more appropriate data sources. The data used were for the most part developed for other uses; more appropriate data for the purpose at hand could probably be developed for several cost categories. We should also like to remind the reader that we attempted as a matter of practicality to deal only with tangible costs in this volume. We, and others, recognize the intangible costs, including those under the rubric "pain and suffering," as well as others such as neighborhood deterioration, official corruption, tax base losses, etc. are probably extensive. Thus, in actuality the total economic costs due to drug abuse are probably underestimated. We think, however, that the data assembled and costs estimated can provide useful inputs for the development of alternative social policies with respect to drug abuse.

FOOTNOTES

1. The research from which the materials in this paper were developed was sponsored by the National Institute on Drug Abuse (NIDA), under terms of Contract No. 271-75-1016, Work Order No. 02. The findings and opinions stated in this paper do not necessarily represent the official position or policy of the National Institute on Drug Abuse.

REFERENCES

[1] Lemkau, Paul V.; Amsel, Zili; Sanders, Bruce; Amsel, Jonathan; and Seif, Thomas F. Social and Economic Costs of Drug Abuse, Contract No. 271-75-1021. Baltimore, Md.: Department of Mental Hygiene, School of Hygiene and Public Health, The Johns Hopkins University, May 1974.

[2] A.D. Little Co. Social Cost of Drug Abuse. Cambridge: A.D. Little Co., 1975. Draft copy.

[3] Fujii, Edwin T. "Public Investment in the Rehabilitation of Heroin Addicts." Social Science Quarterly 55 (June 1975): pp. 49-50.

[4] McGlothlin, William H., and Tabbush, Victor C. "Costs, Benefits, and Potential for Alternative Approaches to Opiate Addiction Control." In Drugs and the Criminal Justice System, pp. 77-124. Edited by James A. Inciardi and Carl D. Chambers. Beverly Hills: Sage Publications, 1974.

[5] Rufener, Brent L.; Rachal, J. Valley; and Cruze, Alvin M. Management Effectiveness Measures for NIDA Drug Abuse Treatment Programs, Volume II: Costs to Society of Drug Abuse, FR 23U-1149. Research Triangle Park, N.C.: Research Triangle Institute, May 1976.

[6] Glenn, William A., and Hartwell, Tyler D. Review of Methods of Estimating Number of Narcotic Addicts. Research Triangle Park, N.C.: Research Triangle Institute, August 1975.

[7] For a more complete discussion of the issues involved in this critical behavioral assumption see Drug Use and Crime, Rockville, Md.: National Institute on Drug Abuse, July 1976 Draft.

[8] Goldman, Fred. The Relationship Between Drug Addiction and Participation in Criminal Activities: An Econometric Analysis. New York: Center for Policy Research and Columbia University, June 1975.

[9] Eckerman, W.C.; Bates, J. Donald; Rachal, J.V.; and Poole, William K. Drug Usage and Arrest Charges. Washington, D.C.: Drug Enforcement Administration, 1971.

[10] Becker, Gary S. "Crime and Punishment: An Economic Approach." Journal of Political Economy 76 (March/April 1968); pp. 169-217.

[11] Tullock, Gordon. "The Welfare Costs of Tariffs, Monopolies, and Thefts." Western Economic Journal 5 (June 1967): pp. 224-233.

[12] U.S. Department of Commerce, Bureau of the Census. "Money Income in 1974 of Families and Persons in the U.S." Current Population Reports, Series P-60, No. 101. Washington: U.S. Government Printing Office, 1976, pp. 106-107.

[13] Berry, Ralph; Boland, James; Laxson, Joan; Haylor, Donal; and Sillman, Margery. The Economic Cost of Alcohol Abuse and Alcoholism. Boston: Policy Analysis, Inc., 1974.

USING TAX RETURNS TO STUDY WAGE AND TAXPAYING PATTERNS BY SEX, 1969 AND 1974

by
Peter J. Sailer*
Internal Revenue Service

The Internal Revenue Service annually publishes statistics based on data from individual income tax returns (Forms 1040 or 1040A). The major classifiers used in these reports are size and source of income, form of deduction, tax rate, and State of residence. Occasionally, by using the taxpayer's first name (and title --such as Mr., Mrs., etc. --when given), IRS also produces data by sex of taxpayer. Unfortunately, this method of classifying income by sex is only useful in examining "nonjoint" returns. In the case of married persons filing jointly, the incomes (and the subsequent taxes) of both the husband and wife are combined. 1/ Hence, it is not possible, using the tax return alone, to distinguish between the incomes and taxes which should be attributed to each spouse separately.

However, there is another document which taxpayers are required to file with IRS which does help in distinguishing between married taxpayers' incomes: the Form W-2 (or Wage and Tax Statement). Since there is a separate W-2 for each job, by linking these forms with the respective tax return, a distinction can be made between the wages of the husband and those earned by his wife. Furthermore, in contrast to the matching problems which have been discussed in other papers in these and earlier Proceedings (e.g.,[1-6]), the linkage of W-2's and tax returns for the same individual presents no major technical problems. This is primarily because taxpayers are supposed to file both sets of forms in the same envelope. All that the statistical editor has to do is use the W-2's to separate the wages of the husband from those of his wife, and add up each group separately.

Studies in which the Form W-2 was used to distinguish the salaries of husbands from those of their wives were conducted for tax year 1969 [7], and repeated for tax year 1974 [8]. The main purpose of this paper is to illustrate some of the kinds of statistics which will shortly be made available by the Internal Revenue Service for income year 1974. As might be expected, such data are bound to be of interest to persons concerned about sex discrimination in the labor force. Therefore, before beginning, it is necessary to make a disclaimer: since the tax return contains no data on the educational level achieved by the taxpayer, the number of hours worked, or the extent to which the taxpayer would have liked to work (i.e., fewer or more hours), this paper will have only slight bearing on the issue of job discrimination by sex. Still, it is hoped that it will shed some light on the economic situation of men and women, as reported on tax returns filed in 1975.

Since the financial circumstances of married persons filing joint returns differ markedly from those of taxpayers filing nonjoint returns, these two groups will be discussed separately. In the first section, the incomes of men and women filing nonjoint returns will be examined. Then, in section 2, the salaries of joint filers will be considered separately for husbands and wives. Section 3 presents some comparisons between joint and nonjoint returns. Next, there will be a brief look at the 1969 and 1974 studies to see if there are any changes in income patterns emerging by sex (section 4). Finally, in section 5, the Federal income taxes of married couples will be estimated based on the tax laws governing incomes of nonjoint filers.

1. INCOME DIFFERENCES BY SEX FOR NONJOINT TAXFILERS

"Nonjoint" returns for 1974 included 37 million returns filed by unmarried persons, as well as approximately 2 million returns of "married persons filing separately." Since some taxpayers in the latter group are not, in fact, married spouses living together, but, rather, separated persons who, because they are legally married, do not qualify as single people, there is a theoretical question as to whether they should actually be included in the "married" or "unmarried" category. There is no convenient way to differentiate between different types of "separate returns of husbands and wives." Therefore, for the purpose of this paper, they have all been included in the nonjoint category.

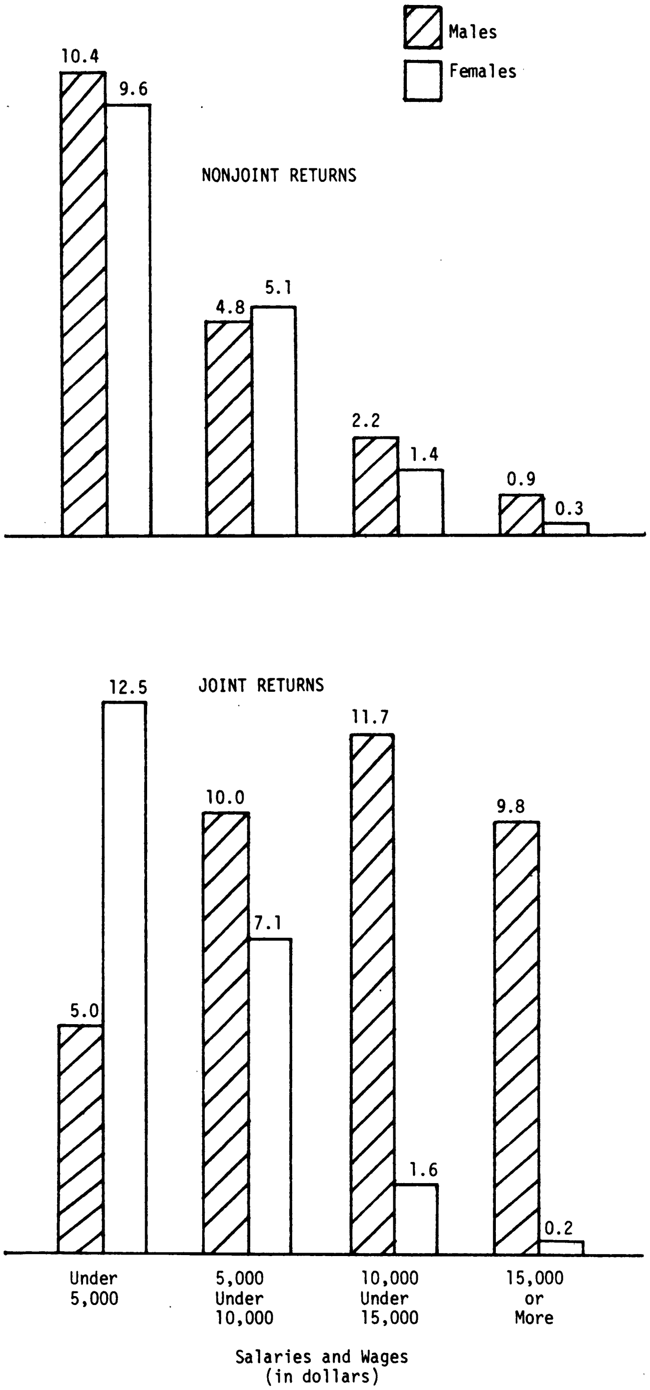
Average Income.--In 1974, the mean adjusted gross income (AGI) 2/ for men filing nonjointly was $5,888; the average for women was $5,460. In other words, there is only about an 8 percent difference between means. However, the disparity between average salaries is somewhat greater--$5,633 for men, $4,827 for women, or a difference of 14 percent. Part of the reason that there is a smaller gap in the case of adjusted gross income than is so for salaries may be due to the fact that women have a greater tendency to inherit men's wealth than vice versa [9]. For this and other reasons, salaries constitute a smaller proportion of adjusted gross income for females (76 percent) than they do for males (87 percent).

Income Distributions.--While differences in mean income and earnings are useful, they do not tell the whole story. There is also a marked difference between the male and female distributions of both these variables. Figure 1 compares the distributions of wages and salaries by sex. As can be seen from the top of the chart, nonjoint returns for women have a greater tendency to be concentrated around the mean ($4,827), while the frequencies for males are more widely distributed (from their average, $5,633).

2. INCOME DIFFERENCES BY SEX FOR JOINT TAXFILERS

Average Income.--The average adjusted gross income on joint returns was $15,449, or close to three

Figure 1. --1974 Returns by Sex and Size of Salaries and Wages for Nonjoint and Joint Filers
(Numbers in Millions)



times the level reported by nonjoint filers. As might be expected, the distribution of these returns by income class is also markedly different from that for nonjoints. In fact, 42 percent of all joint returns fall into the $15,000 or more AGI size class, while only 6 percent of those for nonjoints do. Furthermore, the mean salary for these married couples was $14,398 for returns in 1974--again, nearly three times the level of their counterparts filing singly. These, of course, are averages per return, representing the combined incomes of husbands and wives.

Salaries by Sex.--Using data from Forms W-2, the husbands' and wives' contributions to this combined income can be considered separately. Of the $560 billion of Form W-2 salaries associated with joint returns, 82 percent was earned by men. A brief glance at the lower portion of figure 1 shows one of the major reasons for this disparity: 58 percent of the wives fall into the "under $5,000" salary class, as opposed to only 14 percent of the husbands. Furthermore, the means of the distributions for husbands and wives are $12,495 and $4,760, respectively.

The second reason for the disparity between total salaries of husbands and wives is the lower proportion of labor force participation by wives. To better illustrate this, table 1 presents the earnings distributions for two years by number of wage earners on joint returns. As can be seen, in 1974, 45 percent of returns with salaries had no earnings by women; only 6 percent had no earnings by men.

Table 1.--Joint Returns with Salaries and Wages from Form W-2, 1969 and 1974

(Number in thousands)

Type of Return	Number of Returns		Percent of Returns	
	1969	1974	1969	1974
All Returns with Salaries and Wages......	37,516	38,978	100	100
Returns with No Female Wage Earners....	18,331	17,481	49	45
Returns with No Male Wage Earners.........	1,902	2,421	5	6
Returns with Male and Female Wage Earners..	17,283	19,076	46	49

In all, 19 million returns (nearly half the total of joint filers with salaries and wages) showed earnings by both the husband and his wife. Table 2 further analyzes these returns by cross-classifying the size of the husband's income by that of his wife, using $1,000 intervals up to $10,000 and somewhat larger classes thereafter. It shows that only 1.3 million of these returns (about 7 percent) are "on the main diagonal"--that is, the husband and his wife fall into the same wage class. Eighty-two percent of the returns fall "below the diagonal" (i.e., the husband's wages are greater than his wife's) and the remaining 11 percent are "above the diagonal" (the wife's salary exceeds her husband's).

TABLE 2.--JOINT RETURNS WITH TWO WAGE EARNERS: NUMBER OF RETURNS BY SIZE OF HUSBAND'S AND BY SIZE OF WIFE'S SALARIES AND WAGES, 1974

(NUMBERS IN THOUSANDS)

SIZE OF HUSBAND'S SALARIES AND WAGES (IN DOLLARS)	ROW TOTAL	SIZE OF WIFE'S SALARIES AND WAGES (IN DOLLARS)													DIA-GONAL TOTAL	CLASSES ABOVE MAIN DIAGONAL
		1 UNDER 1,000	1,000 UNDER 2,000	2,000 UNDER 3,000	3,000 UNDER 4,000	4,000 UNDER 5,000	5,000 UNDER 6,000	6,000 UNDER 7,000	7,000 UNDER 8,000	8,000 UNDER 9,000	9,000 UNDER 10,000	10,000 UNDER 15,000	15,000 UNDER 20,000	20,000 OR MORE		
COLUMN TOTAL..........	19,076.1	3,179.5	2,266.4	1,971.1	1,880.5	1,864.2	1,775.3	1,513.2	1,300.1	1,004.2	734.9	1,344.8	200.3	41.8	19,076.1	GRAND TOTAL
1 UNDER 1,000....	529.1	114.6	50.7	49.1	54.1	50.4	37.2	40.1	41.1	34.8	14.3	36.4	6.2	0.3	2,156.9	SUBTOTAL
1,000 UNDER 2,000....	475.8	95.3	68.1	60.4	55.0	35.9	32.1	42.7	19.1	15.1	20.7	27.5	3.3	0.4	0.3	12 CLASSES
2,000 UNDER 3,000....	491.9	103.4	63.6	62.5	59.1	49.5	51.4	17.6	20.4	25.0	12.2	25.0	1.9	0.4	6.5	11 CLASSES
3,000 UNDER 4,000....	472.2	92.3	65.4	70.4	50.8	34.9	37.2	28.7	34.6	16.0	15.1	23.3	1.6	2.0	40.5	10 CLASSES
4,000 UNDER 5,000....	691.3	137.0	98.5	89.5	71.9	84.0	63.2	50.0	26.4	16.5	11.4	36.4	6.4	0.1	45.6	9 CLASSES
5,000 UNDER 6,000....	858.0	133.5	118.1	104.2	107.2	110.9	93.4	63.6	36.3	43.6	12.1	29.3	5.6	0.1	82.1	8 CLASSES
6,000 UNDER 7,000....	1,002.1	141.6	125.6	128.8	129.6	130.1	112.5	88.7	45.2	40.1	24.0	35.7	0.7	0.6	98.1	7 CLASSES
7,000 UNDER 8,000....	1,250.7	136.9	154.6	131.3	132.4	147.2	153.1	106.9	83.5	61.5	45.2	44.9	2.9	0.3	141.8	6 CLASSES
8,000 UNDER 9,000....	1,332.5	194.9	160.7	110.5	168.7	142.6	146.9	138.9	104.3	67.7	40.4	52.5	3.8	0.5	157.9	5 CLASSES
9,000 UNDER 10,000...	1,440.6	204.5	149.7	153.5	143.4	169.7	162.6	122.7	103.5	92.3	43.5	83.0	10.6	1.5	202.3	4 CLASSES
10,000 UNDER 15,000...	6,429.1	979.8	731.1	618.5	577.9	601.0	590.2	549.4	521.6	382.9	304.9	500.6	63.9	7.3	314.3	3 CLASSES
15,000 UNDER 20,000...	2,677.9	485.8	307.7	264.2	214.4	206.7	211.7	184.3	192.4	151.9	137.4	267.1	45.4	8.9	432.9	2 CLASSES
20,000 OR MORE........	1,425.0	309.9	172.4	128.2	117.0	101.2	83.6	79.7	71.8	56.7	53.9	183.2	47.6	19.6	634.7	1 CLASS
DIAGONAL TOTAL........	19,076.1	15,596.9	309.9	658.2	1,415.8	1,315.9	1,278.7	1,369.2	1,442.6	1,575.2	1,620.2	1,567.8	1,594.7	1,447.7	1,322.3	NONE
CLASSES BELOW MAIN DIAGONAL	GRAND TOTAL	SUB-TOTAL	12 CLASS-ES	11 CLASS-ES	10 CLASS-ES	9 CLASS-ES	8 CLASS-ES	7 CLASS-ES	6 CLASS-ES	5 CLASS-ES	4 CLASS-ES	3 CLASS-ES	2 CLASS-ES	1 CLASS	NONE	DIAGONAL CLASSIFIER

Note: The diagonal totals were obtained by adding up the table entries in a diagonal direction. Consider, for example, the main diagonal, which has been underlined. If one sums these underlined figures, the diagonal total obtained is 1322.3 which is shown in the lower right-hand corner of the table. The figures identified as "subtotal" in the diagonal marginal are counts of the number of cases above and below the main diagonal and can be used to get an overall sense of the degree to which disagreements between the classifiers are offsetting. For more details on tabular displays such as this, see Scheuren, F.J. and Oh, H.L., Comm. in Stat., July 1975.

Source: Derived from Internal Revenue Service's Statistics of Income -- 1974, Individual Income Tax Returns (to be released in early 1977).

3. COMPARISON OF JOINT AND NONJOINT RETURNS

When the data on nonjoint returns discussed earlier are compared to the information on joint returns, it is found that the average salary of married women filing jointly is about one percent lower than the corresponding mean for women filing nonjoint returns ($4,760 versus $4,827, respectively). On the other hand, the average salary of married men filing jointly ($12,495) is well over twice the size of the average salary of men filing nonjoint returns ($5,633).

A comparison of the nonjoint and joint salary distributions shown in figure 1 reveals that 88 percent of the taxpayers reporting salaries of $15,000 or more were married men filing joint returns. Many factors, including age and educational level, played a role in this distribution. It should also be kept in mind that the statistics presented here include part-time and occasional workers, as well as full-time employees. According to data published by the Bureau of Labor Statistics [10], only 3 percent of married-men-with-spouse-present who were employed in non-agricultural industries were voluntarily working less than full-time in 1974. This compares to the 22 percent of unmarried men and 25 percent of women, regardless of marital status, who preferred part-time employment.

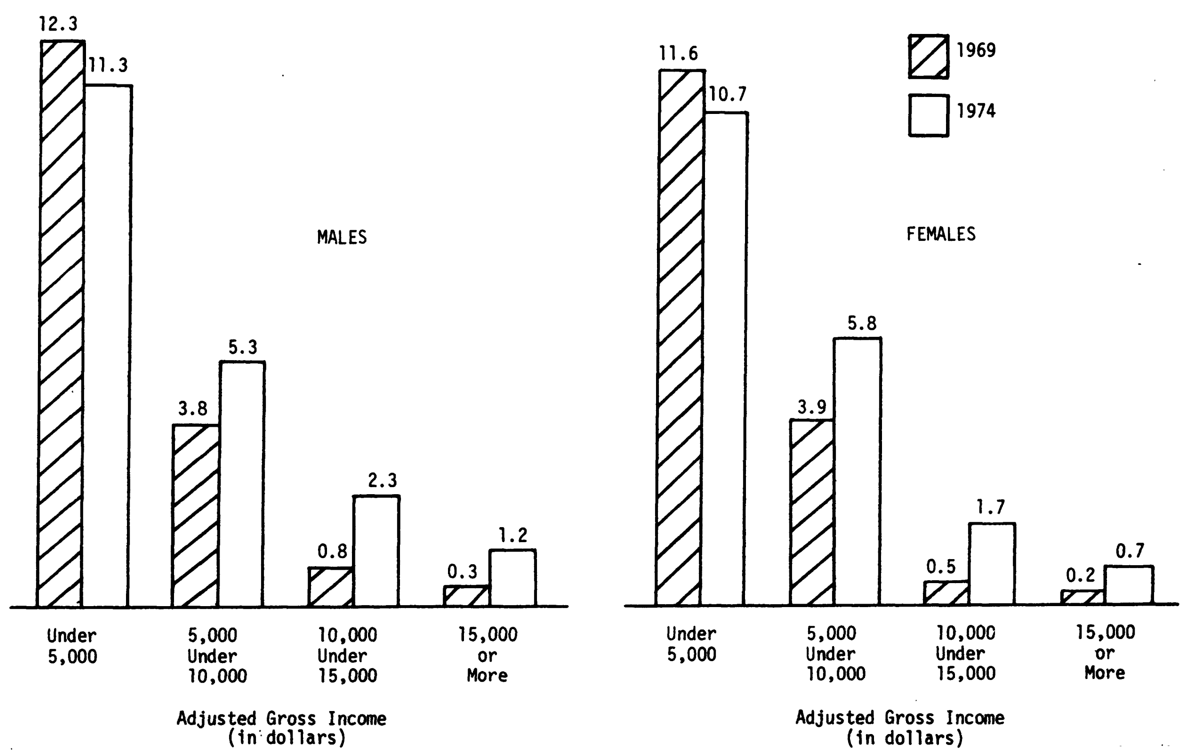
4. COMPARISON OF INCOME DIFFERENCES BY SEX, 1969 AND 1974

As mentioned earlier, 1974 was the second year in which IRS used Forms W-2 to produce statistics on men's and women's salaries. Originally, a major part of this paper was to be devoted to showing the changes in the relative economic positions of men and women since 1969. While the scope has changed slightly, a few observations should be made here. To begin with, the mean salary for women did rise from $3,456 to $4,790 in the five-year period. While this rise was sufficient to place the average female wage at 47 percent of the mean wage for males, it is interesting to note that that is exactly where it was in 1969.3/

Increases in Number of Female Taxpayers.--However, there were a few notable changes in the statistics. For instance, as is shown in table 1, a slightly higher proportion of wives had jobs in 1974: 55 percent of joint returns with salaries and wages included female wage-earners, as compared to 51 percent in 1969. Even more dramatic was the rise in the number of nonjoint returns filed by women--this category increased by 25 percent in five years, as compared to an increment of 15 percent in nonjoint returns filed by men.

Changes in Income Distributions.--Figure 2

Figure 2. --Nonjoint Returns by Sex and Size of Adjusted Gross Income, 1969 and 1974
(Number in Millions)



presents income distributions for nonjoint returns of men and women in 1969 and 1974. This chart shows that there was a considerable increase in the number of women with adjusted gross incomes in the $5,000 under $10,000 range during the five-year period. However, while there were also rises in the number of women with incomes of $10,000 or more, these increases were not nearly as great as those for men.

Table 3 compares data on these two years for joint returns with two wage earners. It shows almost no change in the percentage of total income which was contributed by wives of two-earner families. In both years, 42 percent of the wives contributed less than 25 percent of the joint income; 86 percent contributed less than 50 percent of the joint incomes.

Table 3.--Joint Returns with Two Wage Earners with Salaries and Wages from Form W-2, 1969 and 1974

(Number in thousands)

Type of Return	Number of Returns		Percent of Returns with Two Wage Earners	
	1969	1974	1969	1974
Returns with Male and Female Wage Earners..	17,283	19,076	100	100
Returns with Husband's Share of Wages:				
Under 25%.......	562	715	3	4
25 under 50%....	1,740	1,977	10	10
50 under 75%....	7,791	8,393	45	44
75% or more.....	7,190	7,992	42	42

5. THE EFFECT OF TAX RETURN LEGISLATION ON JOINT FILERS

Tax Reform Legislation.--The final topic to be dealt with in this paper is the effect that the tax reform legislation, passed between 1969 and 1974, had on married couples with two workers. In 1969 and earlier years, married people filing jointly had the advantage of a tax rate schedule with intervals equal to exactly twice the size of those in the single persons' schedule. In other words, the couple's taxable income, whether it was earned by one spouse or both, was taxed as if each was a single person and had earned exactly one-half. In 1974, married people filing jointly were still using the same schedule. However, single people with taxable incomes between $4,000 and $44,000 had a new tax rate schedule which had substantially lower rates than those which had been in effect for them in 1969. Furthermore, the standard deduction was liberalized considerably between 1969 and 1974. Yet, under the new rules, working couples received only half the benefit of this liberalization, since the two spouses together could claim only one standard deduction. In short, a working couple in which one partner earned considerably more than the other, and which had a fair amount of itemized deductions, could still expect to get a tax break out of marriage. However, a working couple in which both partners earned similar amounts of money, or which did not have much in the way of itemized deductions, could expect to pay more taxes than they would have had they been single.

Description of Comparison Study.--Without wanting to get involved in the controversy over the merits of the current system of taxing working couples, it was thought it would be interesting to get a few statistics which would quantify the effects of that system. (This was done by using the 1974 IRS tax model, [11] which will shortly be available for sale by the National Archives.) The remaining few comments attempt, with the help of the tax model, to determine how much tax working couples would have paid, had they been single.

In order to do this, of course, a number of assumptions had to be made. The only information available by sex was the amount of salaries that each spouse had earned (as obtained from the Forms W-2). Hence, the method used for distributing other sources of income (such as dividends and interest), itemized deductions (when applicable), and exemptions was simply to divide them up in the same proportions as salaries were distributed between the two spouses. However, if one spouse's share of any itemized deductions fell below the standard deduction, then, the standard deduction was used in computing the tax. If the couple used the standard deduction to begin with, then, each spouse was given the full standard deduction. For the sake of simplicity, no special tax computations (such as the alternative tax on capital gains) or tax credits (such as the foreign tax credit) were used in calculating either the separate or combined taxes of the two spouses.

Some Observations on the Comparisons.--Table 4 shows the results of these calculations. Please note that the table applies only to the 18 million couples with taxable incomes for which both spouses earned salaries. These returns accounted for over $315 billion in income. Had the incomes of the husbands and the wives been taxed separately under the assumptions made above, the total tax bill would have come to $43 billion. However, under the current system, the actual tax bill (before any reductions for special tax computations or credits) was nearly a billion dollars higher. This $1 billion was the net result of the fact that, due to marriage, 13 million couples had to pay an average of $149 more in taxes, and some five million couples paid an average of $177 less than would have been true had they been taxed as single individuals.

In view of what was stated earlier about income patterns on joint returns-- viz, that, in the vast majority of cases, wives in two-earner families

Table 4.--Federal Individual Income Taxes of Married Working Couples Who Filed Joint Returns for 1974

Item	Number of Returns	Amount (in thousands of dollars)
Total Adjusted Gross Income	18,451,158	315,487,042
Imputed Adjusted Gross Income for Husbands	18,451,158	226,929,830 1/
Imputed Adjusted Gross Income for Wives	18,451,158	88,557,212 1/
Tax on Husbands' Income, Taxed as Single Persons	17,865,346	32,628,295 2/
Tax on Wives' Income, Taxed as Single Persons	14,102,599	10,435,557 2/
Combined Tax of Husbands and Wives, Both Taxed as Single Persons	18,332,847	43,063,852 3/
Combined Tax of Husbands and Wives, Taxed as Married Couples	18,451,158	44,055,719 4/
Tax Savings Due to Joint Filing Status	5,319,043	934,125
Tax Losses Due to Joint Filing Status	12,924,304	1,934,991

Note: Only returns with taxable income were included in this analysis.

1/ Separate data on husbands' and wives' incomes were available only in the case of salaries and wages; all other sources of income received by the couple were imputed to the husband or his wife, based on the ratio of the husband's salaries and wages to those of his wife.

2/ Taxable income for each spouse was computed by subtracting from that spouse's imputed adjusted gross income his or her pro-rata share of exemptions and deductions (based on the ratio of the husband's salaries and wages to those of his wife). However, if the share thus arrived at was less than the standard deduction, the latter was used in computed taxable income. Tax was computed by applying the tax rate schedule for single persons to this taxable income. Additional tax savings which might be available to each taxpayer by using special tax computations, such as income averaging or the alternative computation for capital gains, were not considered; neither were any other possible offsets to tax, such as the foreign tax credit.

3/ Amount is the sum of the two previous items. The number of returns for this item is slightly less than total since some couples who were taxed under the 1974 system would have escaped taxation altogether if each could have filed as a single person.

4/ Tax computed in the ordinary method on each couple's taxable income. As in the case of the three previous items, neither tax savings due to special computations nor tax credits were considered. The actual tax paid by these taxpayers was, therefore, slightly less.

had substantially lower wages than their husbands--this result may come as something of a surprise. As it turned out, it is only for couples with combined adjusted gross incomes of $50,000 or more that the advantages of income-splitting outweighed the disadvantages of the lower standard deduction and less beneficial tax rate schedule. For every income class under $50,000, using the assumptions made for this tax model run, more couples lost than gained from having to file joint returns for tax year 1974.

This analysis is merely a small indication of the kinds of intriguing questions which might be examined with the data base described. Once the 1974 wage information by sex is released to the public, it is hoped that others will also find it of use and interest.

FOOTNOTES

* The author would like to express his gratitude to his colleague Shauna Anderson for going through hundreds of pages of computer printouts and boiling the data down to obtain the tables and charts included in this paper. Thanks are also due to Beth Kilss and H. Lock Oh, who assisted in preparing the charts and tables. Helpful comments were received from many persons, among them Wendy Alvey, Jack Blacksin, Beth Kilss, Fritz Scheuren, and Robert Wilson. Nonetheless, tradition dictates that all shortcomings of this paper be attributed to the author.

1/ The only exception to this exists in the case of married couples who, together, have more than $400 of dividends; such taxpayers are supposed to file a Schedule B, indicating which dividends accrued to the husband, to the wife, or to both jointly. However, even in such instances, only income from dividends is broken out by sex; not interest, wages, or other types of income.

2/ Adjusted gross income includes income from all sources subject to the Federal income tax, less allowable deductions for business-related expenses.

3/ Approximately two-thirds of this increase was attributable to inflation. The Consumer Price Index rose from 1.101, in July 1969, to 1.326, in July 1974. Using this index to deflate the 1974 figure, the rise from 1969 to 1974 in terms of 1969 dollars was from $3,456 to $3,977. An additional part of this increase in average salaries was due to changes in filing requirements between 1969 and 1974, which made it unnecessary for a number of wage-earners at the lower income levels to file returns for 1974.

REFERENCES

[1] Kilss, B. and Alvey, W. "Further Exploration of CPS-IRS-SSA Wage Reporting Differences for 1972," 1976 American Statistical Association Proceedings, Social Statistics Section.

[2] Vaughan, D. and Yuskavage, R. "Investigating Discrepancies between Social Security Administration and Current Population Survey Benefit Data for 1972," 1976 American Statistical Association Proceedings, Social Statistics Section.

[3] Alvey, W. and Cobleigh, C. "Exploration of Differences between Linked Social Security and Current Population Survey Earnings Data for 1972," 1975 American Statistical Association Proceedings, Social Statistics Section, 1976, pp. 121-128.

[4] Sailer, P. and Vogel, L. "Exploration of Differences between Linked Current Population Survey and Internal Revenue Service Data for 1972," 1975 American Statistical Association Proceedings, Social Statistical Section, 1976, pp. 129-137.

[5] Millea, M. and Kilss, B. "Exploration of Differences between Linked Social Security and Internal Revenue Service Wage Data for 1972," 1975 American Statistical Association Proceedings, Social Statistics Section, 1976, pp. 138-146.

[6] Scheuren, F. and Oh, H.L. "Fiddling Around with Nonmatches and Mismatches," 1975 American Statistical Association Proceedings Social Statistics Section, 1976, pp. 627-633.

[7] Internal Revenue Service. Statistics of Income--1969, Individual Income Tax Returns, U.S. Government Printing Office, Washington, D.C., 1971.

[8] Internal Revenue Service. Statistics of Income--1974, Individual Income Tax Returns, to be released in early 1977.

[9] Internal Revenue Service. Statistics of Income--1972, Personal Wealth, U.S. Government Printing Office, Washington, D.C., 1976.

[10] Bureau of Labor Statistics. Employment and Earnings, Washington, D.C., 1974

[11] Internal Revenue Service, 1974 Tax Model for National Estimates -- Individual Income Tax Returns, to be released in early 1977.

WHAT IS INTELLIGENCE?
A Re-examination

Douglas A. Samuelson

Intelligence is like charm or good looks: we all know what we mean by it, we all agree that it is good to have, but we cannot define just what it is or where it comes from. Most of us think of it as something like "general ability to solve complex or obscure mental problems," perhaps emphasizing abstract reasoning and pattern recognition.

Like "random" or "probability," though, the word holds pitfalls for the unwary. The more one knows about it, the more difficult it becomes to define. The Encyclopedia of Education (1971) states that no definition is universally accepted, then devotes several paragraphs to a discussion of what a good definition should include.[1]

The history of the concept is both enlightening and surprising. In 1905, Alfred Binet and his pupil T. Simon developed the first of what are now known as intelligence tests in response to a commission from the Paris public schools soliciting ways to distinguish "feeble-minded" from "normal" children so that the former could receive special instruction. The term "intelligence," derived from a Latin root meaning "to choose," was adopted by Binet and Simon to describe the mental attribute measured by their test: judgment, rather than sensory perception or memory. Indeed, the idea that "intelligence" is best thought of as one quantity, rather than a set of not-necessarily-related attributes, gained acceptance mainly because Binet's attempts using the single-valued tests succeeded after "cluster" and "profile" tests had failed.

What followed Binet's pioneering work is familiar to most students of psychology. Lewis Terman of Stanford adapted and improved the test for American children, and the test he devised, through its continuing evolution, has become the standard for the measurement of intelligence. The widespread use of the Stanford-Binet test to determine the mental abilities of World War I draftees popularized the test and the concepts behind it, and by the late 1920's its use in the schools was commonplace, together with many of the doubts, debates and abuses still common today.

However, the important constant in this history is that tests of intelligence have invariably been validated by their ability to predict success in the traditional Euro-American school system. As psychologist Arthur Jensen took care to note in his comprehensive and controversial article, "How Much Can We Boost IQ and Scholastic Achievement?"[2]:

> "The content and methods of instruction represented in this tradition, it should be remembered, are a rather narrow and select sample of all the various forms of human learning and of the ways of imparting knowledge and skills....We have accepted traditional instruction so completely that it is extremely difficult even to imagine, much less to put into practice, any radically different forms that the education of children could take. Our thinking almost always takes for granted such features as beginning formal instruction at the same age for all children (universally between the ages five and six), instruction of children in groups, keeping the same groups together in lock step fashion through the first several years of schooling, and an active-passive, showing-seeing, telling-listening relationship between teacher and pupils."

The consequences of this history were clear: "Satisfactory learning occurs under these conditions only when children come to school with certain prerequisite abilities and skills: an attention span long enough to encompass the teacher's utterances and demonstrations, the ability voluntarily to focus one's attention where it is called for, the ability to comprehend verbal utterances and to grasp relationships between things and their symbolic representations, the ability to inhibit large-muscle activity and engage in covert "mental" activity, to repeat instruction to oneself, to persist in a task until a self-determined standard is attained--in short, the ability to engage in what might be called self-instructional activities, without which group instruction alone remains ineffectual."

Therefore, the characteristics which Binet and Simon found most useful in distinguishing "bright" from "dull" children, and all the testing items and procedures which have descended from their efforts, reflect the value judgment that, as Jensen put it, "despite all the criticisms that can easily be leveled at the educational system, the traditional forms of instruction have actually worked quite well for the majority of children. And the tests that were specifically devised to distinguish those children least apt to succeed in this system have also proved to do their job quite well."[3] (Emphasis mine.)

Thus the relationship between tested intelligence and intelligence as socially defined depends upon the assumption that the skills most important to scholastic success are highly related to what we think of as "general mental ability." This is not an unreasonable assumption, since the educational systems have been directed toward teaching and improving the skills considered most important in later life; also, school success and standardized tests are major screening devices used in determining who will have access to the best opportunities later, which reinforces the apparent predictive value of intelligence tests.

However, recognizing this assumption explicitly has several important consequences. Let us examine the statistical consequences first.

IQ As a Statistic

As we have seen, IQ and other related single-valued measures of intelligence were developed from observation of behavior of "bright" and "dull" school children, and the tests were and are validated by prediction of scholastic success. Although Binet and Simon could not have known it, since the relevant statistical theory was not invented until many years later, their attempts to summarize the children's distinguishing traits in a single score can be viewed as a rather primitive, ad hoc sort of discriminant function analysis. They selected types of behavior, assigned weights to scores on items which measured different types of behavior (the weights presumably were selected so as to maximize the difference in average scores between "bright" and "dull" children), and then formed the weighted sum of these item scores.

Since this approach required the use of a wide variety of unrelated items varying over a continuum of difficulty, it should not surprise statisticians that the weighted sum had a roughly normal distribution. (In fact, Jensen's subsequent analysis suggested that the accumulated U.S. data can be fitted even better with a double-normal distribution, with one normal "hump" at mean of 55 to 60 for children with severe retardation caused by known genetic or nutritional defects, and a much larger normal "hump" with mean about 105 for all other children.[4] This would fit the proposed description of IQ as a discriminant function even better.)

Having achieved success with their method, Binet and Simon had little explanation for the underlying structures or attributes; but others were quick to supply ideas. Their results were published at the height of the "Social Darwinist" movement in American and European philosophy, as various thinkers -- notably Herbert Spencer -- asserted that man, like the animal species, is undergoing natural selection which manifests itself through the social class structure.

The approximately normal distrubution of IQ scores led nicely into the hypothesis that intelligence was a polygenic trait like height; the leading analysts who tackled the subject-- Francis Galton, Lewis Terman, Cyril Burt, Henry Goddard and Robert Yerkes, among others -- were all hereditarian and "Social Darwinist" in their thinking; several of these thinkers also belonged to one or more "eugenicist" societies, advocating breeding of humans as a means to improve the species and solve social problems.[5]

So soon after the idea was first developed, theory about intelligence was pushed into a hereditarian mold which has shaped all subsequent discussion. The discriminant function procedure, with its great robustness against violation of assumptions, would naturally tend to prevent subsequent data from providing a striking contradiction of theory -- or any other theory that had been adopted, for that matter.

When psychologists attempted to extend intelligence testing to adults, they soon found that the "mental age divided by chronological age" definition was meaningless, since the raw scores of adults of 50 were approximately the same as those of adults of 25. "Mental age" did not make sense for adults. Therefore, using the hereditarian theory which had since been developed, they simply assumed that IQ was normally distributed and standardized their test scores accordingly: that is, they assigned an IQ score of 100 to the mean raw score, 85 to the raw score attained or exceeded by 86% of the sample popluation, 115 to the raw score attained or exceeded by 14% of the population, and so forth.

All modern intelligence tests use this procedure, the last holdout -- the Stanford-Binet -- having converted to it in 1960.[6] Thus <u>the normal distribution of IQ scores is a direct consequence of the scoring procedure itself</u>; this distribution can no longer be said to prove anything about genetic influences.

In summary, then, it seems most reasonable to conclude that tested intelligence -- IQ and related measures -- can best be regarded as a statistical estimate of the probability of scholastic success, derived from a form of discriminant function analysis. This statistic is robust against the vagueness of the underlying theory and is insensitive to changes in critical assumptions, and may therefore understate or camouflage effects of theoretical errors. To illustrate the potential magnitude of this problem, let us consider a hypothetical example.

Viewing IQ From Afar: A Cautionary Tale

Imagine a school serving two neighborhoods, one primarily English-speaking, the other primarily Spanish-speaking. Suppose that instruction in the school is in English only, and that the two groups of children have equal means and distributions of IQ if tested in their own languages. Now suppose that a psychologist downtown at the school board office, ignorant of the situation, requests that a standard English IQ test be given to all the students in the school.

First of all, he would see that the scores for English-speaking students correlate well with teachers' opinions of them and with other measures of achievement and ability. The Spanish-speaking group would have depressed scores, with those who knew the most English doing best; this, too, would correlate highly with scholastic success. Finally, he might be tempted to conclude that the Spanish-speaking group was seriously inferior in general mental ability.

Naturally, the school counselor would soon propose a program to remedy the difficulty: tutoring the Spanish-speaking students in English. New IQ tests after several months of this program would show substantial gains for the Spanish-speaking students, and their school performance would also improve.

But what would happen after the tutoring ended? Retests a year or two later would show the Spanish-speaking students' IQ scores and school performance both declining again. The school board psychologist would say, sympathetically, "Well, it seemed like a good idea, but all you got was a 'hothouse effect.' As soon as you stopped the intensive tutoring, they started sinking back to their 'natural' level." And we can imagine him writing to his superiors, as Jensen did in his 1969 article,[7] "Compensatory education has been tried and it apparently has failed."

Now the frustrated counselor might try another approach: he translates the IQ test into Spanish and gives another school-wide test, with each child being tested in his own language. Would this change the situation? No, the school board psychologist, reviewing the new score distributions, would notice that the new test has a lower correlation with school success, and would therefore pronounce it "invalid."

This is why the validation of intelligence scores by prediction of school success is so important. If we have overlooked a major factor or group of factors which determine IQ, we may well be led to an error just a great, though not as obvious, as that of the hypothetical psychologist.

For example, for "English-speaking" and "Spanish-speaking" substitute "white" and "black"; for language, substitute tendency (perhaps largely or wholly genetic in origin) to react adversely under stress. Is it not possible that much of the observed difference in scores of black and white U.S. school children could be attributable to such factors? Even more important, given the way in which intelligence is evaluated and interpreted, how would we know whether such factors were involved? With these questions in mind, we turn to the core of the recent controversy: heritability.

Intelligence and Heritability

The concept of heritability, particularly as applied to race difference in intelligence, has been exhaustively reviewed and discussed; it has been the focus of the controversy over racial differences in intelligence. To summarize briefly, heritability has been estimated mainly from studies of twins reared apart, and of unrelated foster children reared together. Correlations between IQ scores of identical twins reared apart have averaged around .8; correlations between unrelated children reared together have averaged around .5 to .6. This evidence and corroborative estimates derived from the differences between correlations for monozygotic (identical) and dizygotic (fraternal) twins have led to a tentative conclusion, widely but not universally accepted, that the most reasonable estimate for heritability of IQ is 60 to 80 percent.[8]

Even if we concede, for the moment, that heritability is very important and high heritability is very discouraging to "environmentalists" -- a point to which we shall return later -- two important criticisms must be made. First, heritability estimates of IQ mis-attribute some environmental factors to heredity because of the structure of the experiments. Second, heritability estimates of IQ are artificially inflated because of the way the tests are constructed and updated.

The first point comes from the fact that heritability involves not two factors, but four: genotype; environment; covariance (the fact that genotypes may not be assigned to environments at random); and interaction (the fact that different genotypes may respond differently to a given environment). While some critics have raised questions concerning the possible neglect or understatement of the latter two factors,[9] most seem to have missed the fact that neither covariance nor interaction can be estimated from twin and foster-pair studies, and both effects will show up as "hereditary" in such studies.

To see why this is so, consider a hypothetical study in which two white twins, say Jim and John, and two black twins, say Bob and Bill, are separated at birth: Bob is reared with Jim and his white parents, John is reared with Bill and his black parents. Jim and Bob are considered by the methodology of twin studies to have the same environment, and accordingly all differences between their test scores would be attributed to heredity; in the same way, differences between John's and Bob's scores would be attributed to heredity.

But these differences include the different treatment Jim and Bob receive in school and elsewhere because one is white and the other is black. The same is true for John and Bill. Thus <u>the effects of direct racial discrimination</u> (interaction effects) <u>are attributed to heredity</u>.

Past studies and standardizations have suggested that interaction effects are small -- on predominantly white samples. A true "crossing experiment" such as the one just described has never been done, and probably never could be done within the code of ethics governing experimentation on human subjects.

In this example, covariance between heredity and environment would be zero. Similarly, though, it is clear that covariance, like interaction, cannot be estimated in a twin or foster-pair study, and the effects are most likely to be attributed to heredity. Actual adoptions are possible only for foster parents who are willing to adopt and who meet certain criteria set by the regulatory agencies. It cannot reasonably be argued that assignment of children to foster parents is anything resembling random. The tendency one would expect is, first, that the range of environments available would be substantially narrower than the range presented to natural children; and, second, that the white children would have a high probability of getting better environments. Again, similarities between white twins reared apart -- including the better-than-random assignment to environments -- would be attributed to heredity, and similarly for the effects of the non-random assignment of black children to environments. In addition, the limited range of environmental variation would further inflate heritability.

The second point, which no one to my knowlege has noted, is that the dependent variable (intelligence) changes over time: new test items are continually being developed, old items have been discarded as "no longer meaningful," and the rules of scoring and interpretation have changed accordingly. Intelligence, unlike, say, height, cannot be said with assurance to be the same thing in 1970 as in 1920. To understand the importance of this criticism, we must remember that items are dropped from IQ tests because they no longer predict well, and new items are added because they predict better. Thus <u>some of the most important effects of environmental changes are attributed in part to obsolescence of test items</u>.

For example, tests which included such "general knowledge" items as the identity of New York's baseball team were used in the 1920's to support the conclusion that recent immigrants from Eastern and Southern Europe were "inferior" to Americans of Western European ancestry.[18] Such items no longer predict success in school effectively, so they have been dropped and replaced; the conclusion drawn from the tests at the time, that Eastern and Southern European immigrants were "genetically inferior," has also been quietly discarded; but the heritability estimates derived from the tests remain in the scientific literature.

The IQ's of current descendants of those "inferior" immigrants are much higher than the high heritability estimates of that time would have led us to expect, but instead of being viewed as strong "environmentalist" evidence, these results are attributed to obsolescence or cultural bias of the tests items used then, if they are discussed at all.

These and other technical criticisms cast considerable doubt upon the value of heritability estimates in drawing scientific conclusions and in determining policy. Overshadowing all these arguments, however, there remains the question of what heritability of intelligence really means.

Is Heritability Important?

As we have seen, the usefulness of heritability in thinking about intelligence is diminished considerably by the technical difficulties in measuring heritability in a statistically sound manner. Even if heritability could be measured in a manner which would answer these objections, though, it is not clear what we would gain.

No matter how it is measured, heritability reflects only the effects of the environments and genotypes which were included in the sample from which heritability was estimated: for example, tuberculosis was highly heritable before the discovery of modern antibiotics, when incidence and severity of TB depended mainly on such largely genetic traits as respiratory allergies and general susceptibility to infection. In the same vein, it is possible to construct an experiment in which two groups of plants are selected from the same set of genotypes; one group is raised in light, the other in darkness. The heights of the plants will be highly heritable within groups, but the difference between groups will be environmental.

Similarly, the discovery of new environmental treatments which could alter IQ would also lower the heritability of IQ. High heritability does not "prove" that the trait cannot be modified; it suggests that modification <u>through treatments previously tried</u> is unlikely. It is unwarranted to concede, as many critics of the "Jensenist" position have implicitly done, that defensibly high heritability estimates for IQ would demolish the justification for trying to modify IQ through research and social programs. At best, heritability is useful primarily as an indication of which types of treatment seem unlikely to succeed. If one is tempted to agree with Jensen that methods of compensatory education hitherto tried have been proven unsuccessful, the conclusion should be to try altering the factors which have remained unaltered: starting age for school, home environment, prenatal nutrition, and so on. But perhaps the most important unchallenged factor is the assumption that the "intelligence tests" which predict success in school are actually measuring the general mental abilities which are needed for success in life.

Implications for Policy and Research

If the current controversy is seen, then, as much fuss over the wrong questions, where do we go from here? The first step, I think, is research on the right question: what do intelligence tests really measure? If tested intelligence is really a sort of discriminant function, why not see whether a properly designed discriminant function can't do better? Specifically, it should be possible to perform a discriminant function analysis on children classed as "bright" or "dull" according to school performance, using demographic factors and various test results as independent variables. If all the predictive information could be supplied by demographic variables, would that not change our ideas about the determinants of school success? Whether IQ provides independent information or not, would it not be useful and informative to learn what else is important in predicting school success?

Continuing research into motivational psychology and various observation-based studies of developing perception and reasoning -- the work of Piaget and his followers, for example -- is surely more promising than further explorations of the "nature-nurture" controversy. As was suggested earlier, experiments which could resolve the latter problem are almost certainly impossible under the code of ethics governing human experimentation; moreover, even if the experiments could be performed, their results might be rendered irrelevant by the discovery of a new treatment or by the acceptance of the idea that intelligence tests aren't really important anyway.

For policy, these conclusions are clear:

-- Remedial-intervention programs should concentrate directly on skills, not on raising IQ. If school performance changes, IQ scores will have to change in order for the tests to remain acceptable as predictors.

-- Research or policy based on "eugenic" considerations should be viewed with great suspicion until the meaning of tested intelligence is clearly understood.

-- Perhaps most important, before proceeding with massive programs to search for treatments to raise intelligence, we should consider not only whether raising intelligence is useful -- which, I suggest, is an unanswerable question until we understand clearly what intelligence is -- but also how the fair administration of the treatment could be controlled. If intelligence were largely hereditary, at least the resultant differences in education and occupational status could be rationalized sufficiently for society to function. But what if a simple environmental treatment -- say, a "smart pill" or hormone injection -- were found? Who would control who gets how much? And how would the "less fortunate" react?

Perhaps the most useful consequence of the race-IQ controversy is that social scientists and the public may be forced to realize that science cannot be divorced from its social context, as assumptions alter conclusions. The next time some data (such as the IQ scores) suggests to some analysts that the democratic ideal is unworkable, we should re-examine the analysis and the data before preparing to discard the ideal. Both science and democracy will be the better for it.

References

1. Encyclopedia of Education, (MacMillan, 1971), pp. 672-3.
2. Harvard Educational Review, Winter 1969, p. 7.
3. Ibid., p. 7.
4. Ibid., pp. 23-7.
5. This idea is developed, perhaps with more stridency than is warranted, in Leon Kamin's book, The Science and Politics of IQ (Wiley, 1974). Other useful references on the history of the concept include "Social Implications of IQ," by Sheldon H. White, in National Elementary Principal, March/April 1975 (there is also a short article by Kamin in the same issue), and Chapters 1 and 2 of Race and Intelligence, Ken Richardson and David Spears, editors, (Pelican, 1972).
6. Encyclopedia of Education, op. cit., pp. 131-2
7. Harvard Educational Review, op. cit., p. 2.
8. A well-written and fairly comprehensive review is Chapter 4 of Loehlin et. al., Race Differences in Intelligence (W. H. Freeman and Company, 1975).
9. Chapter 4 of Loehlin et. al. (op. cit.) briefly reviews the criticisms raised by Jencks (Inequality: Basic Books, 1972), Kamin (op. cit.), and Layzer (Science, 1974, 183, 1259-1266). The spring, 1969 issue of the Harvard Educational Review included an article by R. Light and P. Smith which was the first to introduce interaction into the discussion, although not very persuasively. W. Bodmer and L. Cavalli-Sforza were apparently the first to argue that the "nature-nurture" controversy can never be resolved under present circumstances because of confounding of interaction, covariance, and heredity. (Scientific American, October 1970, pp. 19-29.)
10. Kamin (op. cit.) discusses this at length.

AN IMPUTATION PROCEDURE FOR DETERMINING MISSING FACTOR LEVELS IN ANALYSIS OF VARIANCE

Ronny A. Schaul, Bureau of the Census
Jack C. Hayya, The Pennsylvania State University

A problem of missing independent variables arises in the processing of questionnaire data. Such a problem has occurred in the construction of a price index for new housing. The price of the house is the dependent variable, Y. *The independent variables,* X_i *are characteristics (square feet of floor, number of bathrooms, air-conditioning, etc.) consisting of several levels. The Construction Statistics Division of the Bureau of the Census obtains this information from a questionnaire survey. In some questionnaires only partial information is returned. A solution is to "guess" the missing levels of the* X_i *by classifying the dependent variable with the incomplete information on the* X_i *into one of several populations. The net effect in regression, for example, is to reduce the variance of the regression coefficients, thereby reducing the variance of the predicted* $\hat{Y}$.

I. INTRODUCTION

In a questionnaire survey, it may happen that some respondents do not complete all the items on the questionnaire. Thus some questionnaires with missing information are returned; and even though a follow-up may be initiated, there is no guarantee of obtaining the missing data when the respondent considers them confidential or proprietary (or he may simply be uninterested). But, in order for these questionnaires to be useful, all items in them must be completed. This happens, for example, when the questionnaire survey is used for regression purposes.

Given that certain necessary information is already available on an incomplete questionnaire, it seems desirable to impute the missing answers in order to reduce the loss of information. The imputation procedure considered in this paper requires that the items imputed can assume only a finite number of values; i.e., the imputation procedure is applicable in an ANOVA situation.

We give two examples from surveys conducted by the Construction Statistics Division of the Census Bureau. One example deals with a survey to determine a price index of new one family houses sold; the other deals with a survey to determine a cost index of residential buildings with two or more housing units.

A Price Index of New One Family Houses Sold

In order to establish a price index, respondents to a survey are asked to supply data concerning the following nine items:

1. The price of the house;
2. The size of the house (in terms of square feet of floor area grouped into 9 classes);
3. The number of stories;
4. The number of bathrooms;
5. The presence or absence of air-conditioning;
6. The type of parking facility (garage or no garage);
7. The type of foundation (basement or no basement);
8. The geographic location (in terms of 12 areas);
9. The metropolitan location (inside or outside the Standard Metropolitan Statistical Areas as defined by the Office of Management and Budget).

Items 2 to 9 above refer to eight characteristics of the house. Information regarding items 8 and 9 is never missing, since this information can be determined independently of the respondent. Consequently, the interest is in imputing a maximum of six characteristics.

In practice, about six percent of all questionnaires have at least one of the items 2 to 7 unanswered. Since the price of the house is practically a continuous variable, imputations for missing prices will not be considered here. Also since the price of the house is regressed against all eight characteristics, an imputation procedure for missing values of these characteristics seems desirable, especially when a questionnaire has only one or two entries missing.

In general, the nonresponse conforms to the following distribution:

Square feet only	32.0%
Basement only	6.1
Air-conditioning (AC) only	26.4
Garage only	6.1
Stories only	2.0
Bathrooms only	1.5
AC and garage	0.5
Garage and basement	0.5
AC, garage, and basement	1.5
All six characteristics missing	10.2
All other combinations	13.2
	100.0%

A Cost Index of Residential Buildings With Two or More Housing Units

Questionnaires similar to those for determining a price index of new one family houses are used here. About 40% of all questionnaires have at least one of the requested items on the characteristics of the buildings unanswered.

Review of Previous Work

Imputation procedures for similar problems have been used by the Bureau of the Census (See Chapman, [2, pp. 27-]). As described by Chapman, two main procedures are the Cold Deck procedure and the Hot Deck procedure. Both classify the data into cells according to the characteristics so that "responses will be relatively homogeneous within cells and heterogeneous between cells... For each missing item for a particular respondent to the...survey, the values of the appropriate completed items are noted to identify the relevant cell. The respondent is associated with the cell corresponding to the values of the items. A value is then selected from the responses in the cold deck included in the same cell. This value is usually selected at random or systematically" (Chapman, [2, p. 4]). Variations of this method consist in using "a moving average of values in a cell to substitute for a missing value," or "an imputed value...obtained from a regression of the particular item on several of the other items." (Chapman, [2, p. 6]). The difference between the two methods is that the Cold Deck procedure uses data from a previous survey, whereas the Hot Deck procedure uses data from the same survey. The interested reader may also refer to [4, 6, 7, 8, 9].

Currently, no imputation procedure is used by the Bureau of the Census in either of the two examples mentioned above. Incomplete questionnaires are not used when calculating the indices. It is planned to use the procedure proposed in this paper if the imputation improves the estimates of the regression coefficients which are used when computing the indices. Whether improvement is achieved or not is to be determined by simulation and by comparing the regression using the complete data set on the one hand, and the regression using the complete data set augmented by imputation on the other.

II. THE IMPUTATION PROCEDURE

The following is a modification of the Hot Deck procedure. It takes advantage of the fact that the variables for which imputations are made can assume only a finite number of values.

Determining Factors and Factor Levels

Suppose we have a factorial design with I factors, where factor K has L_K levels, $1 \leq K \leq I$. For instance, in the first example given in section I (the price index for new one family houses sold), the number of factors is I = 8. Also, the 8 factors and their number of levels are:

Factor 1. Size of the House, $L_1 = 9$ levels

2. Number of Stories, $L_2 = 3$ levels

3. Number of Bathrooms, $L_3 = 3$ levels

4. Presence or absence of Air-Conditioning, $L_4 = 2$ levels

5. Parking, $L_5 = 2$ levels

6. Type of Foundation, $L_6 = 2$ levels

7. Geographic Location, $L_7 = 12$ levels

8. Metropolitan Location, $L_8 = 2$ levels

The dependent variable is the price of the house and as said before is practically a continuous variable.

Determining Data Sets

Consider the total set of response data. Separate this set into 3 subsets:

- Subset D1. Complete information: the value of the dependent variable (here the price of a house) and the level of each one of the I factors are given; that is to say, an answer has been given for each question on the questionnaire.

- Subset D2. Incomplete information: the value of the dependent variable is given, but at least one of the levels of the factors was not reported.

- Subset D3. The dependent variable is not observed.

Since the dependent variable is continuous and the independent variables are discrete and since this imputation procedure requires the data to be classified in cells, imputation will be considered only for the independent variables, i.e., only for subset D2.

Classifying of the Data into Cells

Consider all $\prod_{K=1}^{I} L_K = C_1$ cells. (For the above factor levels, $C_1 = 15{,}552$). Classify the complete data of subset D_1 into these cells. Among these C_1 cells, consider the C_2 cells, $C_2 \leq C_1$, each containing "enough" observed values of the dependent variable so that a distribution can be assigned to these values. Only these C_2 cells will be considered in the imputation procedure. Therefore, the assumption is made that the incomplete questionnaires of the subset D2 can be classified appropriately using only these C_2 cells. If it happens that all C_2 cells contain enough observations so that the empirical distributions would be good enough approximations for the theoretical distributions, then the empirical distributions can be used instead of the theoretical ones.

Consider an incomplete questionnaire from subset D2. Determine from the C_2 cells those corresponding to the same level of non-missing factors which are given by the respondent. Suppose that C_3 cells are thus selected. (If the answers corresponding to factors 2 and 3 of our first example are missing, a maximum of 3 x 3 = 9 cells are isolated from the C_2 cells; these 9 cells are determined by the given levels of factors 1 and factors 4 through 8. There are nine such cells since factors 2 and 3 each have 3 levels. So here $C_3 \leq 9$.) The set of C_3 cells constitutes the possible non-empty cells from which the incomplete questionnaire may have originated.

Specifically, if all possible cells C_1 are made of boxes $B_1, B_2, B_3, \ldots, B_{15552}$, it may happen that the C_2 cells containing enough complete observations from the subset D1 are:

$$B_1, B_4, B_7, B_8, B_{11}, \ldots, B_{4000}.$$

Suppose that all possible cells, which an incomplete questionnaire of D2 could have come from, are $B_1, B_2, B_3, \ldots, B_9$. Then the C_3 cells selected from the C_2 cells will be the following 4 cells: B_1, B_4, B_7, B_8.

So here $C_3 = 4$.

The Classification of Incomplete Questionnaires into Cells

Assign costs of misclassification. Let $C(i|j)$ = cost of misclassifying an observation in cell i when in fact it is from class j; i, j = $1, 2, \ldots, C_2$, $i \neq j$. Suppose the distributions considered have pdf's. (Discrete probabilities can be treated in a similar fashion.) Let $p_i(x)$ be the pdf for the ith cell. Let R_i be the region of classification in the ith cell; i.e., if $y \in R_i$, then y (the price of the house) is classified in the ith cell. It follows that the probability of correctly classifying y into the ith cell is

$$P(i|i,R) = \int_{R_i} p_i(x)dx. \tag{1}$$

The probability of misclassifying y into the jth cell when in fact it comes from the ith cell is

$$P(j|i,R) = \int_{R_j} p_i(x)dx. \tag{2}$$

If a priori probabilities q_i are known, we can calculate the conditional probability of a given observation y coming from class i. It is

$$\frac{q_i p_i(x)}{\sum_{i=1}^{C_3} q_i p_i(x)}. \tag{3}$$

The expected cost of misclassification is

$$\sum_{j=1}^{C_3} q_j \left\{ \sum_{\substack{i=1 \\ i \neq j}}^{C_3} C(i|j)P(i|j,R) \right\}. \tag{4}$$

If we classify the observation in class j, the expected cost is

$$\sum_{\substack{i=1 \\ i \neq j}}^{C_3} \frac{q_i p_i(x) C(j|i)}{\sum_{k=1}^{C_3} q_k p_k(x)}. \tag{5}$$

The expected cost is minimized by choosing j so as to minimize (5). This is equivalent to considering

$$\sum_{\substack{i=1 \\ i \neq j}}^{C_3} q_i p_i(x) C(j|i), \tag{6}$$

and choosing that j which minimizes it.

If more than one j minimizes (6), any one of them could be chosen. If a priori probabilities are unknown, we can do the following:

The conditional cost if the observation comes from cell i is

$$r(i,R) = \sum_{\substack{j=1 \\ j \neq i}}^{C_3} C(j|i)P(j|i,R), \tag{7}$$

and the classification procedure chooses that i which minimizes (7). Suppose now that we have a priori probabilities for the population, say q_j for cell j, $1 \leq q_j \leq C_3$, where

$$q_j = \frac{\text{number of observations in cell j}}{\text{number of observations in the } C_3 \text{ cells}}. \tag{8}$$

This assumption simplifies the problem and is reasonable for the two examples given above. Suppose further that the costs of misclassification are equal. Let p_j be the pdf of the jth class. Then the observation considered is classified in class j if $q_j p_j > q_i p_i$ for all $i \neq j$, $1 \leq i \leq C_3$.

This procedure minimizes the expected cost of misclassification assuming that the costs of misclassification are equal. (Anderson [1, p. 148].)

III. OPTIMAL PROPERTIES

1. As shown in T. W. Anderson [1, pp. 142-147],

the procedure $R = \{R_1, R_2, \ldots, R_{C_3}\}$ is admissible.

2. In a regression situation, we have the following:

Using the set D1 of complete data:

$$Y^{(1)} = X^{(1)}\beta + E,$$
$$\hat{\beta} = (X^{(1)\prime}X^{(1)})^{-1}X^{(1)\prime}Y^{\prime(1)}. \tag{9}$$

Using the set $D1 \cup D'2$, where $D'2$ are those questionnaires of D2 which were completed by imputation, we have

$$Y = \begin{pmatrix} Y^{(1)} \\ Y^{(2)} \end{pmatrix} = X\beta + E = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} \beta + E. \tag{10}$$

Then given $X^{(1)}, X^{(2)}$,

$$\hat{\beta} = (X'X)^{-1}X'Y \text{ is unbiased} \tag{11}$$

=> it is unconditionally unbiased.

The generalized variance of $\hat{\beta}$ is

$$\frac{\sigma^2}{|X'X|} = \frac{\sigma^2}{|X^{(1)\prime}X^{(1)} + X^{(2)\prime}X^{(2)}|}. \tag{12}$$

Then the generalized variance of $\hat{\beta}$ is reduced if $X^{(1)\prime}X^{(1)}$, $X^{(2)\prime}X^{(2)}$ are positive definite matrices as shown below.

Theorem. If A^{nxn} and B^{nxn} are positive definite matrices, then

$$|A + B| > |A| + |B| > \max(|A|, |B|)$$

Proof. Use the following properties of positive definite matrices. (See Anderson, [pp. 333-341], for instance.) If A and B are positive definite matrices:

(i) their determinants are positive

(ii) there exists a non-singular matric C such that

$C'AC = I$ and $C'BC = D = \text{Diag.}\ (d_1, d_2, \ldots, d_n)$,

with $d_1, d_2, \ldots, d_n$ all positive.

Then $|A| + |B| > \max\ (|A|, |B|)$ holds. Next

$$|A + B| > |A| + |B| \iff |C'|\ |A + B|\ |C|$$
$$> |C'|\ |A|\ |C| + |C'|\ |B|\ |C|.$$
$$\iff |C'AC + C'BC| > |C'AC| + |C'BC|$$
$$\iff |I + D| > |I| + |D|, \text{using (ii) with}$$
$$D = C'BC$$
$$\iff \prod_{i=1}^{n} (1 + d_i) > 1 + \prod_{i=1}^{n} d_i,$$

which holds since $d_1, d_2, \ldots, d_n$ are all positive.

IV. SUMMARY AND CONCLUSIONS

An imputation procedure for missing factor levels in an ANOVA situation is described. This procedure uses discriminant analysis to classify incomplete questionnaires into cells, and the classification leads to completion of these questionnaires. The procedure is admissible, and under the usual assumptions of linear models the generalized variance of $\hat{\beta}$ in $Y = X\beta + E$ is reduced.

Work is presently undertaken on an actual problem at the Bureau of the Census, where 139 out of 197 incomplete questionnaires were rendered complete by this procedure. The question of assigning distributions has not been resolved; but it has been circumvented in this problem by assuming that the square root of the dependent variable is normally distributed. Also assumed are prior probabilities and equal costs of misclassification.

In conjunction with the above, the following research is being pursued:

1. The determination of distributions for the C_2 cells (non-empty cells containing completed questionnaires) in an automatic fashion;

2. The determination by simulation of the accuracy of the imputation procedure; and

3. The determination of its usefulness by comparison of the regressions on $D1 \cup D'2$, where D1 refers to the completed questionnaires and where $D'2$ is that subset of the originally incomplete questionnaires that were later completed by imputation.

ACKNOWLEDGMENT

The authors gratefully acknowledge the helpful suggestions concerning this problem given them by Dr. Roger C. Pfaffenberger of the University of Maryland and Dr. William E. Strawderman of Rutgers University.

REFERENCES

[1] Anderson, T.W., *An Introduction to Multivariate Statistical Analysis*. New York: John Wiley & Sons, 1958.

[2] Chapman, David W., "A Survey of Non-Response Imputation Procedures," paper presented at a meeting of the Methodology Section of the Washington Statistical Society, Washington, D.C., April 14, 1976.

[3] Ferguson, Thomas S., *Mathematical Statistics: A Decision Theoric Approach*. New York: Academic Press, 1967.

[4] Hansen, M. H., Hurwitz, W. N. and Madow, W. G., *Sample Survey Methods and Theory, Vol. I.* New York: John Wiley & Sons, Inc., 1953.

[5] Lilliefors, H. W., "On the Kolmogorov-Smirnov test for Normality with Mean and Variance Unknown," *Journal of the American Statistical Association*, Vol. 62, June 1967, pp. 399-402.

[6] Nordbotten, Svein, "The Efficiency of Automatic Detection and Corrections of Errors in Individual Observations as Compared With Other Means for Improving the Quality of Statistics." *Proceedings of the 35th Session, International Statistical Institute*, Volume XLI, Book I, Tome XLI, pp. 417-435, Beograd, 1965.

[7] Szameitat, K. and Zindler, H. J., "The Reductions of Errors in Statistics by Automatic Corrections." *Proceedings of the 35th Session, International Statistical Institute*, Volume XLI, Book I, Tome XLI, Beograd, 1965.

[8] U.S. Bureau of the Census, "1960 Censuses of Population and Housing - Processing the Data." Washington, D.C., 1962.

[9] U.S. Bureau of the Census, "The Current Population Survey - A Report on Methodology." Technical paper No. 7, U.S. Government Printing Office, Washington, D.C., 1963.

EVALUATION OF THE OCCUPATION ONE-YEAR-AGO ITEM IN THE JANUARY 1973 CPS

Paula J. Schneider, Bureau of the Census

I. INTRODUCTION

The January 1973 Current Population Survey (CPS) included supplementary questions on employment status, industry, occupation, and class of worker as of one year previous (January 1972). Since a portion of the persons who answered these retrospective employment questions had also been interviewed in the January 1972 CPS, it was possible to compare responses to the retrospective items with the actual reports given in January 1972. A computer match of the January 1972 and January 1973 CPS files on the basis of household and person identification information (age, race, sex, and line number) netted 29,491 completely matched persons.[1/] The CPS weights were applied to these cases, and tabulations were produced for persons 18 years old and over and employed in January 1973.

This evaluation of retrospectively reported data was undertaken because of the expressed interest in including similar questions in the 1980 decennial census and because an evaluation of the 1970 census retrospective occupation question (a 5-years-ago question) had raised serious doubts about validity.[2/] Specifically, the level of consistency between the 5-years-ago retrospective occupation response and the actual response given 5 years earlier was only about 22 percent. However, since the CPS items involved only a 1-year recall period and included an additional question on the starting date for the current job, it was felt that the CPS data might be better.

In the analysis which follows, the terms "response" and "report" are used to designate final occupational classification, not purely the person's response. In other words, both response and coding accuracy are reflected in the classification.

II. CONSISTENCY OF RETROSPECTIVE AND ACTUAL OCCUPATION RESPONSES

A. Overall Consistency.--The January 1972 occupation distributions of employed persons based on the retrospective responses and on the actual reports at that time are quite similar (table A). No occupation group differs by more than half a percentage point between the two distributions. However, when the actual and retrospective classifications are compared on a case-by-case basis, inconsistencies become evident (table B). It should be noted, though, that the inconsistencies cannot all be attributed to recall problems; coding errors can also affect the consistency level. Agreement between the two responses was measured at three levels of occupational classification--detail (3-digit occupation code), intermediate, and major group. Of all persons employed in January 1972, 59 percent retrospectively reported an occupation for that period which agreed at the detailed level with the occupation reported at that time. The consistency rate increases to 69 percent based on an intermediate group comparison and to 77 percent on a major group comparison. The levels of agreement vary somewhat among occupations. Basing the rate on a major group comparison, farmers and professionals show relatively high response consistency (91 percent and 84 percent, respectively), whereas the consistency rate for laborers is only about 56 percent. The remainder of the occupation groups have consistency rates ranging from 71 to 82 percent.

B. Consistency by Sex, Race, and Age.--Women have somewhat more consistent classifications at the major group level then men--82 percent versus 74 percent. Consistency does not vary much by race or age, except that the rate of agreement for the youngest group (18 and 19 years old) is relatively low (56 percent). This was caused primarily by the failure to report being employed in 1972. Only about 77 percent of the persons 18 and 19 years old who had been employed in 1972 retrospectively reported this employment. This is not surprising since younger workers normally have a more tenuous attachment to the labor force.

C. Consistency by Mobility Status.--Although the overall quality of the retrospective occupation data is relatively high, the primary use of such data is to identify the characteristics of occupationally mobile persons. Therefore, it is important to determine how well the retrospective questions define this group. For this study, occupational mobility is defined as a change in major occupation group between January 1972 and January 1973 on the basis of the reports at these dates (i.e., matched comparison). It should be noted, however, that this is not a perfect standard for comparison since coding error and erroneous response differences do cause some false occupational mobility.

Of the persons who were occupationally mobile according to the matched comparison, only about 12 percent reported a retrospective occupation which agreed at the major group level with the occupation reported in January 1972. This compares to a consistency rate of 96 percent for workers who were not mobile based on the matched comparison. As shown in tables C and D, the quality of the retrospective data for occupationally mobile persons, regardless of occupation group, sex, or age, is considerably lower than that for nonmobile workers. The consistency rate for mobile workers does not exceed 20 percent for any major occupation group and is less than 10 percent for three groups --managers and administrators, craft and kindred workers, and private household workers. In contrast, the consistency rate for nonmobile workers exceeds 95 percent for most groups.

D. Comparison With Other Data.--Overall, the accuracy or consistency of the CPS retrospective occupation data compares favorably with the evaluation of the 1970 census item on occupation 5 years ago and with evaluation studies on current occupation data from the 1960 and 1970 censuses (table E). The accuracy of the 1-year-ago occupation item is somewhat higher than that measured for the 5-years-ago question and is equal to or better than the quality measured for the current occupation data.3/ However, consistency between the actual and retrospective responses for occupationally mobile persons (based on the matched comparison) was lower for the 1-year-ago CPS occupation question than was true for the 5-years-ago question evaluated in 1968.4/ The overall consistency rate for the 5-years-ago item was 22 percent, as opposed to 12 percent for the 1-year-ago item in this study. One would expect that recall would be less accurate over 5 years than over 1 year. This perhaps indicates that the quality of occupation coding for the CPS items was less accurate than the coding in the 5-years-ago study, and that the "true" consistent response rate should be somewhat higher. The retrospective data in the 5-years-ago study were coded by experts in Washington, whereas the CPS responses were all coded by regular CPS coders. This fact lends more support to the assumption of poorer quality coding for the 1-year-ago data. However, even a consistency rate of 22 percent, as in the 5-years-ago study, cannot be considered encouraging.

III. COMPARISON OF LEVELS OF OCCUPATIONAL MOBILITY

The levels of occupational mobility between January 1972 and January 1973 measured by the retrospective comparison and the matched comparison of occupation groups are quite different. However, in interpreting these differences, the problems in comparing retrospective and matched data should be considered. Occupation data has historically been subject to variability in response and coding. Such errors cause an overstatement in the level of occupational mobility when the measure is based on a comparison of occupations reported in different surveys (i.e., matched data). On the other hand, retrospectively reported occupation data are often biased by a tendency to report one's past occupation to be the same as the current occupation, even when this is not the case. 5/ This results in an understatement in the level of occupational mobility. Therefore, the "true" mobility rate is most likely some value between the measures derived from matched and retrospective comparisons of occupations.

As shown in table F the level of occupational mobility derived from the retrospective data is only about one-fifth that based on the matched comparison (4.2 percent vs. 22.8 percent). The difference between the retrospective mobility rate and the matched mobility rate varies slightly among major occupation groups. When the rates are based on the occupation of origin (1972 occupation), the ratio of the retrospective to the matched mobility rate ranges from .11 for managers and administrators to .28 for sales workers.

Differences between mobility levels based on the retrospective data and the matched data increase by age (table G). For workers 18 and 19 years old, the retrospective approach measured about one-half the mobility level derived from the matched data; for persons 45 years old and over, the matched data yielded mobility rates which were more than 10 times as high as those based on the retrospective data comparison.

IV. CONCLUSION

The immediate impression one may get from this investigation is that neither matched data nor retrospective data provide adequate measures of occupational mobility. If this is the case, perhaps a better approach to the measurement of mobility is through a detailed survey instrument which can focus on job history, including promotions, changes in activities and levels of responsibility, as well as employer changes and the dates pertaining to all of these. Obviously, such detailed inquiries would not be possible in the decennial census proper. However, before ruling out the use of retrospective occupation inquiries in measuring mobility, the data in this study need to be investigated further.

FOOTNOTES

1/ Normally, more persons would be included in the matched universe, but the occurrence of CPS redesign within this time period decreased the size of the overlapping sample.

2/ Walsh, Thomas C. and Paula J. Buckholdt, Accuracy of Retrospectively Reporting Work Status and Occupation Five Years Ago, E15, No. 3, Bureau of the Census, 1970.

3/ See the series of evaluation studies from the 1970 and 1960 censuses for the methodologies of the CPS-Census Match studies and the Employer Record Check studies.

4/ See footnote 2.

5/ Walsh, Thomas C. and Paula J. Buckholdt, Effect on the "Same as Present Occupation" Circle on the Reporting of Occupation Five Years Ago, E-15, No. 2, Bureau of the Census, 1970.

Table A.--OCCUPATION DISTRIBUTION IN JANUARY 1972 BASED ON THE JANUARY 1972 CPS AND THE RETROSPECTIVE DATA IN JANUARY 1973 CPS

(Numbers in thousands. Persons 18 years old and over and employed in January 1973)

Occupation	January 1972 CPS		Retrospective data		Difference	
	Number	Percent	Number	Percent	Number	Percent
Total employed	20,110	100.0	19,988	100.0	-122	(X)
Professional, technical, and kindred workers	3,055	15.2	3,034	15.2	-21	0.0
Managers and administrators	2,367	11.8	2,396	12.0	+29	+0.2
Sales workers	1,331	6.6	1,360	6.8	+29	+0.2
Clerical and kindred workers	3,733	18.6	3,633	18.2	-100	-0.4
Craft and kindred workers	2,905	14.4	2,848	14.2	-57	-0.2
Operatives, except transport	2,387	11.9	2,313	11.6	-74	-0.3
Transport equipment operatives	830	4.1	793	4.0	-37	-0.1
Laborers, except farm	832	4.1	816	4.1	-16	0.0
Private household workers	232	1.2	227	1.1	-5	-0.1
Service workers, except private household	2,010	10.0	2,012	10.1	+2	+0.1
Farmers and farm managers	285	1.4	312	1.6	+27	+0.2
Farm laborers and supervisors	144	0.7	148	0.7	-4	0.0
Occupation not reported	(X)	(X)	97	0.5	(X)	(X)

(X) Not applicable.

Table B.--CONSISTENCY OF RETROSPECTIVE OCCUPATION RESPONSE 1/ AND JANUARY 1972 OCCUPATION RESPONSE BY MAJOR OCCUPATION GROUP, SEX, RACE, AND AGE

(Numbers in thousands. Persons 18 years old and over employed in January 1973)

Major occupation group, sex, race, and age	Total based on January 1972 CPS		Retrospective response--percent--			
				Occupation 1 year ago agreed with 1972 CPS at:		
	Number	Percent	Employed in Jan. 1972	Major group	Inter-mediate	Detail
Total employed in January 1972	20,110	100.0	95.8	76.8	69.1	59.1
Professional, technical, and kindred workers	3,055	100.0	96.2	84.0	78.5	67.8
Managers and administrators	2,367	100.0	97.5	71.4	65.6	61.4
Sales Workers	1,331	100.0	94.7	72.1	70.2	57.5
Clerical and kindred workers	3,733	100.0	95.7	80.0	66.9	53.8
Craft and kindred workers	2,905	100.0	96.4	75.6	67.5	61.1
Operatives, except transport	2,387	100.0	95.5	73.4	60.8	49.5
Transport equipment operatives	830	100.0	97.2	77.3	67.6	63.0
Laborers, except farm	832	100.0	93.1	55.9	51.2	40.3
Private household workers	232	100.0	84.5	73.7	(X)	56.0
Service workers, except private household	2,010	100.0	95.0	82.3	78.3	64.2
Farmers and farm managers	285	100.0	97.5	90.9	(X)	89.5
Farm laborers and supervisors	144	100.0	93.8	72.2	(X)	61.1
Male	12,995	100.0	96.2	73.9	67.4	58.7
Female	7,116	100.0	94.9	82.0	72.1	59.7
White	18,024	100.0	95.8	77.0	69.3	59.4
Black and other races	2,087	100.0	95.2	75.4	67.4	56.3
18 and 19 years old	585	100.0	76.6	56.4	51.3	31.1
20 to 24 years old	1,615	100.0	91.6	72.0	61.4	45.4
25 to 34 years old	4,122	100.0	96.7	78.6	71.0	59.4
35 to 44 years old	4,533	100.0	96.6	78.2	71.3	61.8
45 to 54 years old	5,011	100.0	97.0	76.5	68.3	60.9
55 to 64 years old	3,536	100.0	96.7	78.4	70.2	61.5
65 years old and over	709	100.0	96.2	80.0	75.7	67.8

1/ Occupation 1 year ago as reported in January 1973.

X Not applicable. No intermediate occupation groups.

Table C.—CONSISTENCY OF RETROSPECTIVE OCCUPATION RESPONSE 1/ AND JANUARY 1972 OCCUPATION RESPONSE BY OCCUPATIONAL MOBILITY STATUS 2/ AND MAJOR OCCUPATION GROUP

(Numbers in thousands. Persons 18 years old and over employed in January 1973)

Occupational mobility status and major occupation group in January 1972	Total based on January 1972 CPS		Retrospective response--percent--			
				Occupation 1 year ago agreed with 1972 CPS at:		
	Number	Percent	Employed in Jan. 1972	Major group	Intermediate	Detail
OCCUPATIONALLY MOBILE						
Total	4,577	100.0	92.0	11.8	10.8	8.5
Professional, technical, and kindred workers	442	100.0	91.6	10.0	8.8	7.2
Managers and administrators	654	100.0	96.3	6.0	5.8	5.8
Sales workers	403	100.0	91.1	17.9	17.1	2.5
Clerical and kindred workers	693	100.0	93.8	12.6	11.4	9.5
Craft and kindred workers	682	100.0	93.1	8.1	7.2	6.5
Operatives, except transport	623	100.0	92.1	12.2	10.1	9.1
Transport equipment operatives	204	100.0	96.1	15.7	14.2	14.2
Laborers, except farm	392	100.0	90.0	12.8	12.0	10.2
Private household workers	48	100.0	56.3	6.3	(X)	6.3
Service workers, except private household	365	100.0	86.3	19.2	18.9	15.6
Farmers and farm managers	24	100.0	91.7	13.6	(X)	13.6
Farm laborers and supervisors	47	100.0	80.9	19.1	(X)	19.1
NOT OCCUPATIONALLY MOBILE						
Total	15,533	100.0	96.9	96.0	86.2	77.5
Professional, technical, and kindred workers	2,613	100.0	97.0	96.5	90.3	79.6
Managers and administrators	1,713	100.0	98.0	96.4	88.3	85.2
Sales workers	928	100.0	96.1	95.6	93.3	81.5
Clerical and kindred workers	3,040	100.0	96.2	95.4	79.4	68.2
Craft and kindred workers	2,223	100.0	97.3	96.4	86.1	80.8
Operatives, except transport	1,764	100.0	96.7	95.1	78.6	69.0
Transport equipment operatives	625	100.0	97.8	97.6	85.1	84.3
Laborers, except farm	440	100.0	96.1	94.3	86.1	77.5
Private household workers	184	100.0	91.8	91.8	(X)	70.7
Service workers, except private household	1,646	100.0	96.8	96.3	91.4	79.6
Farmers and farm managers	261	100.0	98.5	98.1	(X)	97.7
Farm laborers and supervisors	97	100.0	100.0	97.9	(X)	90.7

X Not applicable. No intermediate occupation group.

1/ Occupation 1 year ago as reported in January 1973.

2/ Mobility status based on a comparison of major occupation groups as reported in January 1972 and in January 1973.

Table D.--CONSISTENCY OF RETROSPECTIVE OCCUPATION RESPONSE 1/ AND JANUARY 1972 OCCUPATION RESPONSE BY OCCUPATIONAL MOBILITY STATUS 2/, SEX, AND AGE

(Numbers in thousands. Persons 18 years old and over employed in January 1973)

Occupational mobility status, sex, and age	Total (based on January 1972 CPS)		Retrospective response--percent--			
			Employed in Jan. 1972	Occupation 1 year ago agreed with 1972 CPS at:		
	Number	Percent		Major group	Intermediate	Detail
Occupationally Mobile						
Male	3,410	100.0	93.2	10.8	10.1	8.4
Female	1,167	100.0	88.8	14.7	13.1	8.8
18 and 19 years old	267	100.0	69.7	27.3	25.1	19.1
20 to 24 years old	502	100.0	83.7	23.7	21.7	17.5
25 to 34 years old	913	100.0	93.4	16.2	14.6	11.9
35 to 44 years old	958	100.0	94.5	10.8	9.8	7.0
45 to 54 years old	1,095	100.0	96.0	4.3	4.0	3.3
55 to 64 years old	724	100.0	94.2	6.4	6.2	4.8
65 years old and over	118	100.0	98.3	2.5	2.5	1.7
Not Occupationally Mobile						
Male	9,584	100.0	97.3	96.4	87.8	80.4
Female	5,949	100.0	96.1	95.3	83.6	72.8
18 and 19 years old	319	100.0	82.1	80.9	72.7	57.4
20 to 24 years old	1,113	100.0	95.3	93.7	79.1	68.7
25 to 34 years old	3,209	100.0	97.7	96.3	87.1	78.1
35 to 44 years old	3,574	100.0	97.2	96.4	87.7	79.7
45 to 54 years old	3,915	100.0	97.4	96.7	86.2	78.4
55 to 64 years old	2,812	100.0	97.4	96.9	86.7	77.6
65 years old and over	591	100.0	95.8	95.4	90.4	81.4

1/ Occupation 1 year ago as reported in January 1973.

2/ Mobility status based on a comparison of major occupation groups as reported in January 1972 and in January 1973.

Table E.--COMPARISON OF THE ACCURACY OF THE RETROSPECTIVE 1-YEAR-AGO OCCUPATION DATA WITH THE ACCURACY OF OTHER RETROSPECTIVE AND CURRENT OCCUPATION DATA

Major occupation group	Accuracy rates of occupation data 1/					
	Retrospective occupation data		Current occupation data			
	CPS Occupation 1 year ago	Occupation 5 years ago 2/	1970 CPS-Census Match	1970 Employer Record Check 3/	1960 CPS-Census Match	1960 Employer Record Check
Total	80.2	74.0	76.8	80.6	80.6	83.2
Professional, technical, and kindred workers	87.5	85.8	84.9	83.6	89.4	89.7
Managers and administrators	73.6	68.8	57.2	60.0	65.0	64.1
Sales workers	76.3	67.5	78.5	79.5	83.3	91.7
Clerical and kindred workers	83.9	77.9	82.0	84.9	88.0	84.7
Craft and kindred workers	78.8	74.8	75.2	76.9	79.2	84.1
Operatives, except transport	77.5	75.1	81.1	81.8	82.9	86.5
Transport equipment operatives	79.9		79.6	95.6		
Laborers, except farm	60.0	47.5	50.4	51.7	59.3	54.0
Private household workers	87.7	80.0	80.6	(NA)	92.1	(NA)
Service workers, except private household	87.1	79.9	82.9	87.0	86.0	85.2
Farmers and farm managers	94.2	77.1	83.0	(NA)	84.6	(NA)
Farm laborers and supervisors	77.6	50.0	77.2	(NA)	75.0	(NA)

Note: Since the other studies measured only the accuracy of occupation responses and did not consider errors in reporting work status and nonresponse, data from this study were adjusted accordingly. The universe was restricted to persons who correctly reported (retrospectively) their work status in January 1972 and also reported an occupation. Thus, the rates given here do not agree with others shown in this report.

NA Not available.

1/ Accuracy rates are based on a comparison at the major occupation group level.

2/ Based on the 1968 Subject Response Study.

3/ Preliminary data.

Table F.--OCCUPATIONAL MOBILITY RATES[1/] FROM THE MATCHED AND RETROSPECTIVE COMPARISONS BY OCCUPATION OF ORIGIN AND OCCUPATION OF DESTINATION

Occupation	Mobility rates by occupation of origin			Mobility rates by occupation of destination		
	Matched comparison	Retro-spective comparison	Ratio: Retrospec-tive/ matched	Matched comparison	Retro-spective comparison	Ratio: Retrospec-tive/ matched
Total, 18 years and over..	22.8	4.2	0.18	22.8	4.2	0.18
Professional, technical, and kindred	14.5	2.2	0.15	15.9	2.7	0.17
Managers and administrators.........	27.6	2.9	0.11	31.2	4.6	0.15
Sales workers......................	30.3	8.6	0.28	30.2	5.4	0.18
Clerical and kindred workers..........	18.6	3.3	0.18	17.4	4.1	0.24
Craft and kindred workers...........	23.5	3.0	0.13	24.1	4.5	0.19
Operatives, except transport........	26.1	5.4	0.21	24.9	7.0	0.28
Transport equipment operatives......	24.6	5.6	0.23	22.6	5.7	0.25
Laborers, except farm...............	47.2	11.1	0.24	44.7	9.0	0.20
Private household workers...........	20.6	2.4	0.12	15.6	1.8	0.12
Service workers, except private household.........................	18.1	4.8	0.27	17.0	5.0	0.29
Farmers and farm managers...........	8.4	1.8	0.21	14.4	1.3	0.09
Farm laborers and supervisors.......	32.9	7.3	0.22	19.2	0.7	0.04

1/ Percent of all persons working in January 1972 and January 1973 who changed major occupation groups.

Table G.--OCCUPATIONAL MOBILITY RATES[1/] FROM THE MATCHED AND RETROSPECTIVE COMPARISONS ACCORDING TO OCCUPATION OF ORIGIN BY SEX AND AGE

Sex and age	Matched comparison	Retrospective comparison	Ratio: retrospective/ matched
Male................	26.2	4.7	0.18
Female..............	16.4	3.4	0.21
18 and 19 years..........	45.6	23.1	0.51
20 to 24 years...........	31.1	11.7	0.38
25 to 34 years...........	22.2	5.7	0.26
35 to 44 years...........	21.1	3.5	0.17
45 to 54 years...........	21.9	1.7	0.08
55 to 64 years...........	20.5	1.5	0.07
65 years and over........	16.6	0.6	0.04

1/ Percent of all persons working in January 1972 and January 1973 who changed major occupation groups.

ONE-STEP WEIGHTED LINEAR REGRESSION FOR DICHOTOMOUS DEPENDENT VARIABLE

Wen-Fu P. Shih, Florida Atlantic University

Introduction

A general linear approach to the analysis of qualitative data was developed by Gizzle, Stamer and Koch (GSK) in 1969.[3] Lehnen and Koch also applied this technique to the field of political science in 1974.[4]

Based on the measurement of dependent variables and the hypothesis of interest, both dummy variable regression and analysis of variance approaches have been applied to categroical data. The analysis of variance apporach requires the assumption of homogeneity of variance. However, when the dependent variable is nominal or ordinal this assumption needs some modification. Also the homogeneity of variance is not adjusted by the dummy variable regression technique. Thus, the GSK approach uses the analysis of variance type of application to nominal data without the homogeneity assumption. This method is based on the general weighted least squares to estimate appropriate functions of the cell proportions in the complex contingency table.

Most social science data are either from sample surveys with large sample sizes or from government or private institutions with mass data stored on the tapes. To apply the GSK method, researchers have to start from contingency table analysis. Then, based on this table output, a design matrix for regression models are set up. It requires tremendous man power and computer time to handle large samples. Especially, when setting up the design matrix for a saturated model for a large number of independent variables, the computation of interaction terms become very complicated and likely causes errors.

Therefore, the purpose of this study is to demonstrate a weighted linear regression analysis for a dichotomous dependent variable without going through contingency table analysis. It is called a one-step weighted linear regression (OWLR). Based on this OWLR technique, a user can directly apply regression analysis to the raw data and the unequal variance of the dependent variable can be simultaneously adjusted. This approach also can be generalized to the case of a dependent variable with more than two levels by adjusting the variance-covariance matrix.

Analysis of Technique

The social behavior relationship in some cases can be formulated as a dichotomous dependent variable. For example, a person can send his children to either public school or private school; or, he either protests school desegregation or does not protest. According to Godberger[2], this type of variable is naturally formulated into a regression equation. The dependent variable of this regression Y_i has only two values, which can be coded as 0 and 1 without losing generality.

I.e.: $$Y_i = \begin{cases} 0 \text{ if negative} \\ 1 \text{ if positive} \end{cases}$$

The general linear regression model can be written as

$$Y = X\beta + \varepsilon \qquad (1)$$

where Y is an n x 1 vector of observations on a dependent variable, n is a dimension of the vector, X is an n x p matrix of nonstochastic regressors with rank p, β is a p x 1 vector of unknown regression coefficients, and ε is an n x 1 vector of unknown disturbances.

To illustrate, let behavior toward school desegregation be a dependent variable Y, and let income and education be independent variables X_1 and X_2, respectively. The model can be written as

$$Y_i = \mu + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i \qquad (2)$$

where

$$Y_i = \begin{cases} 1 \text{ if protest} \\ 0 \text{ if not protest} \end{cases}$$

$$X_{1i} = \begin{cases} 1 \text{ if high income} \\ -1 \text{ if low income} \end{cases}$$

$$X_{2i} = \begin{cases} 1 \text{ if education high} \\ -1 \text{ if education low} \end{cases}$$

i = 1, 2, ..., n and n is the number of observations. μ, β_1, and β_2 are regression coefficients, and ε_i are unknown disturbances.

This model can also be considered as a conditional probability function P_r (Protest given $X_1, X_2) = P(Y_i = 1 \mid X_1, X_2)$ (3)

Denote $P(Y_i = 1 \mid X_1, X_2) = p$ for all i (4)

then

$$E(Y_i) = 0 \cdot (1-p) + 1 \cdot p = p \qquad (5)$$

$$E(Y_i^2) = 0^2 \cdot (1-p) + 1^2 \cdot p = p \qquad (6)$$

$$Var(Y) = E(Y_i^2) - (E(Y))^2 = p(1-p) \qquad (7)$$

$$Cov(Y_j, Y_j') = E(Y_j Y_j') - E(Y_j)E(Y_j') = 0 \qquad (8)$$

The dependent variable Y_i is a Bernouli trial under the condition of $X_1 = x_{1i}$, $X_2 = x_{2i}$. According to Feller[1], the probability function is

$$P(Y=y) = p^y (1-p)^{1-y}, \text{ y=0 and 1 for each trial.} \qquad (9)$$

The likelihood function of p is

$$L(p) = P(Y = y_1, Y = y_2, \ldots, Y = y_n) = p^{\sum_i y_i} (1-p)^{n - \sum_i y_i} \qquad (10)$$

To maximize likelihood, take log of L

$$\ln L = \sum_i y_i \ln p + (n - \sum_i y_i)\ln(1-p) \qquad (11)$$

and take derivative of L with respect to p,

$$\frac{\partial \ln L}{\partial p} = \frac{\sum_i y_i}{p} - \frac{n - \sum_i y_i}{1-p} = 0 \qquad (12)$$

The maximum likelihood estimator of p is

$$\hat{p} = \frac{\sum_i y_i}{n} = \bar{y} \qquad (13)$$

From (7) and (13) the estimator of Var(Y), s^2, is obtained as

$$s^2 = \hat{p}(1 - \hat{p}) \qquad (14)$$

Given $X_1 = x_{1i}$ and $X_2 = x_{2i}$, where x_{1i}, x_{2i} = -1 or 1, the maximum likelihood estimators of p and Var (Y) are noted as

$$\hat{p}_j = \frac{\sum_{i=1}^{n_j} y_{ij}}{n_j} \qquad (15)$$

and $$s_j^2 = \hat{p}_j(1 - \hat{p}_j) \qquad (16)$$

where n_j equals the total frequencies in the subgroup j, with (X_1 = -1 or 1 and X_2 = -1 or 1).

Σy_{ij} equals the total number of 1's in the jth subgroup.

In this particular case there are four subgroups, for j = 1, Group 1 = (X_1 = -1, X_2 = -1) with n_1 observations, for j = 2, Group 2 = (X_1 = -1, X_2 = 1) with n_2 observations; for j = 3, Group 3 = (X_1 = 1, X_2 = -1) with n_3 observations; for j = 4, Group 4 = (X_1 = 1, X_2 = 1) with n_4 observations. The total number of subgroups equal $\prod_k \lambda_k$, where λ_k equals the levels of X_k.

Therefore, the variance-covariance matrix Σ becomes

$$\left[\begin{array}{c|c|c|c} s_1^2 & & & \\ \ddots & & & \\ s_1^2 & & & \\ \hline & s_2^2 & & \\ & \ddots & & \\ & s_2^2 & & \\ \hline & & s_3^2 & \\ & & \ddots & \\ & & s_3^2 & \\ \hline & & & s_4^2 \\ & & & \ddots \\ & & & s_4^2 \end{array}\right] = {}_n\Sigma_n \qquad (17)$$

where $n = n_1 + n_2 + n_3 + n_4$.

As can be seen from equation 17, the disturbance belongs to heteroskedastic, and the weighted linear regression coefficient β can be estimated by

$$\hat{\beta} = (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1}Y \qquad (18)$$

the equation 18 is equivalent to a function used to solve the linear regression coefficient β, i.e., a linear regression equation is expressed as

$$Y^* = X^*\beta + \Sigma \qquad (19)$$

in which

$$\hat{\beta} = (X^{*\prime}X^*)^{-1}X^{*\prime}Y^*, \qquad (20)$$

$$X^* = \Sigma^{-\frac{1}{2}}X, \qquad (21)$$

and

$$Y^* = \Sigma^{-\frac{1}{2}}Y. \qquad (22)$$

In practical application, the X^* and Y^* are replaced as x_{ij}^* and y_{ij}^*, respectively; and the x_{ij}^* and y_{ij}^* are expressed as

$$x_{ij}^* = \frac{x_{ij}}{s_j} \qquad (23)$$

$$y_{ij}^* = \frac{y_{ij}}{s_j} \qquad (24)$$

where

$$s_j = (\hat{p}_j(1 - \hat{p}_j))^{\frac{1}{2}} \qquad (25)$$

Based on the equations 19, 20, 23, 24, and 25, a systematic flow chart as shown in Figure 1 is developed for this OWLR technique. This flow chart has been also converted to a computer program with Fortran IV language. The input data of this program are only raw dependent variables and dichotomous independent variables. All interaction terms will be generated by the program itself. The results of the program output are the estimators of weighted linear regression coefficients, the χ^2 test for each regression, and the goodness of fit of the model.

Another computer program with the same language is also developed to generate the independent variables with more than two levels.

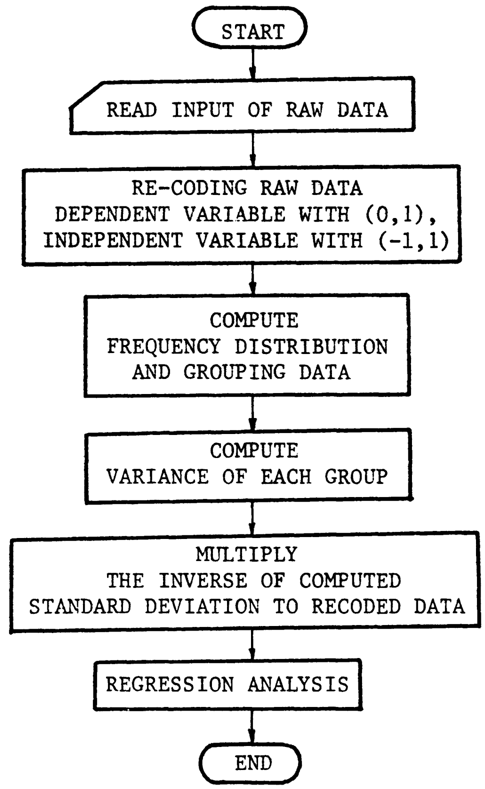


Figure 1: Systematic flow chart of OWLR technique.

Example of Application

Data used in this example are obtained from a survey of parents in Duval County, Florida.

On the basis of responses to questionnaire items by parents whose children attend desegregated public schools, two groups are classified:

(1) Those who did not protest against desgregation: Y = 0

(2) Those who did protest: Y = 1.

The goal is to study the impact on protest of income, education, the percent of black change in assigned school from 1971-72 to 1972-73, and racial prejudice. The regression model for the ith observation can be written as:

$$\text{(A)} \quad Y_i = \mu + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \varepsilon_i \qquad (26)$$

$$\text{or} \quad P(Y_i/X_1,X_2,X_3,X_4) = \mu + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \varepsilon_i \qquad (27)$$

i = 1, 2, ..., n

where $X_{1i} = \begin{cases} 1 & \text{for high income} \\ -1 & \text{for low income} \end{cases}$

$X_{2i}=\{ \begin{array}{l} 1 \text{ for high education} \\ -1 \text{ for low education} \end{array}$

$X_{3i}=\{ \begin{array}{l} 1 \text{ for high racial prejudice} \\ -1 \text{ for low racial prejudice} \end{array}$

$X_{4i}=\{ \begin{array}{l} 1 \text{ for the school with black ratio increase} \\ -1 \text{ for the school that did not change or decreased the black ratio} \end{array}$

There are sixteen possible combinations of X's. Hence, sixteen subgroups are used in the sample. The frequency in each subgroup of data are shown in Table 1. For convenience, the equation 27 can be expressed as

$$s_j = (\hat{p}_j (1 - \hat{p}_j))^{\frac{1}{2}}$$

$$= \left(\frac{\sum_i y_{ij}}{n_j} \left(1 - \frac{\sum_i y_{ij}}{n_j}\right)\right)^{\frac{1}{2}}$$

$$= \frac{1}{n_j} \left(\sum_i y_{ij} \left(n_j - \sum_i y_{ij}\right)\right)^{\frac{1}{2}} \qquad (28)$$

As can be seen from Table 1, the GSK method requires all these Y's and X's values for the analysis of the weighted linear regression coefficient. However, on this study, the values of the weighted linear regression can be calculated directly from raw data by using the OWLR program, and the Y is a dichotomous dependent variable instead of using all frequencies as shown in Table 1. The result of regression obtained by both GSK method and OWLR techniques is the same and is expressed as

$$Y = 0.2841 + 0.0332X_1 + 0.0700X_2{**} + 0.0724X_3{**} + 0.0339X_4 \qquad (29)$$

**significant at the α = .01 level.

As can be seen from equation 29, the education X_1 and changing black ratio X_4 are nonsignificant, but income X_2 and racial prejudice X_3 are highly significant. These imply that the proportion of protest of school desegregation is highly affected by the parent's income and racial prejudice, but is not affected by the parent's education and black ratio change. The code system (-1, 1) is used in the study. Based on the previous study reported by Shih, et. al.[5] (1975), the regression coefficients are twice larger than the percentage wise effect on the dependent variable which is based on the (0, 1) code system.

TABLE 1
Protest Cross-Classified by
Education, Income, Racism and Percent of Black Change

Subgroup	Non-protesters 0 Y	Protesters 1	Education X_1	Income X_2	Racism X_3	% Black Change X_4	$(P(1-P))^{\frac{1}{2}}$
1	1	1	-1	-1	-1	-1	.5000
2	32	6	1	-1	-1	-1	.3646
3	2	1	-1	1	-1	-1	.4714
4	30	8	1	1	-1	-1	.4077
5	6	2	-1	-1	1	-1	.4330
6	33	14	1	-1	1	-1	.4573
7	6	2	-1	1	1	-1	.4330
8	27	23	1	1	1	-1	.4984
9	2	1	-1	-1	-1	1	.4714
10	23	6	1	-1	-1	1	.4051
11	1	1	-1	1	-1	1	.5000
12	12	8	1	1	-1	1	.4099
13	7	2	-1	-1	1	1	.4157
14	19	7	1	-1	1	1	.4436
15	3	2	-1	1	1	1	.4899
16	11	15	1	1	1	1	.4940

Thus, the result b_2 = .07 can be interpreted to mean that the group with the higher income is 3.5% more likely to protest school desegregation, or that the proportion of protestors is increased 3.5% from lower income to higher income. The value of b_3 = .0724 implies that the proportion of protestors is 3.62% higher due to the higher racial prejudice.

The model with two factor interactions of this data can be written as

$$y_j = \mu + \sum_{i=1}^{4} \beta_i X_{ij} + \sum_{\substack{i=1 \\ i' \neq i}}^{4} \beta_{ii'} X_{ij} X_{i'j} + \varepsilon_j$$

The estimated regression coefficients and their χ^2 tests are shown in Table 2. As Table 2 shows, further analysis is probably meaningless because all b's are nonsignificant.

Summary and Conclusions

The Grizzle, Starmer and Koch (GSK) approach has been widely applied to qualitative (ordinal) data in the social sciences to perform general weighted linear regression. The application of the GSK method to most social science data requires two steps: the first is to find the cell frequencies for each subgroup; the second is to construct a design matrix and apply the method. When the sample size is very large, as in most social science data, this approach is too tedious and cumbersome to use because there are too many interaction terms and much computing time is involved. Therefore, a one-step procedure without going through the procedures of finding cell frequencies and constructing the design matrix has been mathematically modified in this study to provide the weighted linear regression. This

TABLE 2

Independent Variable	b	χ^2
Education	-.0039	.0062
Income	.0510	1.4700
Racism	.0130	.0707
% Black Change	.0332	.6329
Education x Income	.0385	.8594
Education x Racism	.0668	1.9232
Education x % Black	.0081	.0383
Income x Racism	.0278	1.1462
Income x % Black	.0363	1.8352
Racism x % Black	-.0147	.2971

All χ^2's are nonsignificant at $\alpha = .05$.

one-step weighted linear regression (OWLR) for dichotomous dependent variables has been computerized. The imput data of this program require only a raw dependent variable and dichotomous independent variables. All interaction terms are generated by the computer itself. The result of the program output are the estimators of weighted linear regression coefficients, χ^2 test for each regressor, and the goodness of fit of the model.

Responses to questionnaire items by parents whose children attend desegregated public school in Duval county, Florida, are used to exemplify the application of OWLR technique. Comparisons of these OWLR results with the GSK approach indicate that OWLR not only is applicable but also that the computing time can be reduced significantly.

References

1. Feller, W. (1964). "An Introduction to Probability Theory and Its Application," John Wiley & Sons, Inc., New York.

2. Goldberger, A.S. (1964). "Econometric Theory." John Wiley & Sons, Inc., New York.

3. Grizzle, J.E., Stamer, C.F., and Koch, G.G. (1969). "Analysis of Categorical Data by Linear Model," Biometrics, Vol. 25: pp 489-504.

4. Lehnen, R.G. and Koch, G.G. (1974), "A "A General Linear Approach to the Analysis of Nonmetric Data: Application for Political Science." American Journal of Political Science, May 1974, pp 283-313.

5. Shih, W. F. P., Giles, M. W., Cataldo, E. F. and D. S. Gatlin (1975). "A Contrast of Two Systems of Contrast Coding," Social Statistics Section, Proceeding of the American Statistical Association, pp 640-645.

RANDOMIZED RESPONSE TECHNIQUE IN A NATIONAL SURVEY

Iris M. Shimizu and Gordon Scott Bonham
National Center for Health Statistics

1. INTRODUCTION

Up until the Supreme Court ruling in 1973, abortions were illegal in many States and greatly restricted in most other States. Asking women to report abortions was often asking them to report illegal behavior and to potentially open themselves to criminal prosecution. Even where abortions were legal, the sensitivity of the subject led to underreporting. The importance to fertility and health research of knowing about the amount of abortion, however, was still great.

Surveys designed to measure abortions by direct questionning had been able to elicit little reporting of abortion [10]. Registration or reports of legal abortions in the United States [2, 9] did not include illegal abortions nor abortions from non-reporting jurisdictions or facilities. In addition, little information is available from registration on the characteristics of women whose abortions are reported.

Randomized response is a technique introduced by Warner [8] to obtain estimates of behavior that is normally underreported, most often because it is sensitive or may have social stigma attached to it. The randomized response technique allows the respondent to answer truthfully without the interviewer being able to know whether or not the respondent has the sensitive characteristic. Horvitz and his associates [4] found that the randomized response technique produced an estimate of illegitimacy almost as high as the known illegitimacy in the selected sample. Another North Carolina study yielded an estimate of abortions for the United States that was in line with previously hypothesized numbers, if the total United States experience was similar to that of North Carolina [1].

The randomized response tehcnique of interviewing on sensitive topics was used by the National Survey of Family Growth (NSFG) to ascertain the incidence of abortion within a twelve month period, and yet preserve the individual respondent's complete privacy. The model chosen for the survey was the two unrelated questions model investigated by Folsom, et al [3]. The results indicate that a substantial amount of abortion was reported, even though the survey yielded divergent estimates for the amount of abortion.

2. IMPLEMENTATION OF RANDOMIZED RESPONSE MODEL

2.1 The National Survey of Family Growth

The NSFG is based on a multi-stage probability sample of women 15 to 44 years of age, living in households in the conterminous United States. To be eligible for the NSFG, a woman must be either ever married, or single (i.e., never married) with her own children living with her. The data are weighted to provide national estimates. The survey is conducted biennially by the National Center for Health Statistics. Data for the present discussion comes from the first cycle of the survey, which began in 1973.

The field work for that cycle was conducted by the National Opinion Research Center of the University of Chicago between July 1973 and February 1974. The median date of the interviews was about September 15, 1973. The NSFG measure of abortion was in reference to the 12 months preceding the date of interview. Although this measurement reflects no exact calendar time period, discussion in this paper will be made as if the twelve month incidence was congruent with the period September 16, 1972, through September 15, 1973. All figures shown are provisional.

2.2 Application in the Survey

The randomized response model employed two random half-samples. All sample households were numbered sequentially at the time the sample was drawn. If the last digit of the household number was odd, any eligible respondent interviewed in the household fell into the first half-sample. This included 4,926 respondents. During the interview, they were asked the innocuous question:

> This time last year, did you live in a different county or State than this one?
>
> (IF THE RESPONDENT NOW LIVES IN LOUISIANA, SAY: "different parish or State."
>
> IF R. NOW LIVES IN THE NEW ENGLAND STATES, SAY: "different township or State.")

At the end of the interview, respondents in half-sample 1 were handed a card with the questions:

> [Heads] In the past 12 months, I had an abortion done to end a pregnancy.
>
> [Tails] Was your mother born in April?

They were instructed to toss a penny and if a head showed, then the respondent was to answer the question after the head on the card. If a tail showed, then the respondent was to answer the question after the tail on the card. They were to answer only "yes" or "no".

If the last digit of the household number was even, the interviewed respondent was in half-sample 2. This half-sample included 4,871 women who were directly asked:

> In what month and year was your mother born?

At the end of the interview, half-sample 2 respondents were given the same instructions as the half-sample 1 respondents and then given a card with the questions:

> [Heads] In the past 12 months, I had an abortion done to end a pregnancy.
>
> [Tails] This time last year, I lived in a different county or State than this one.

The responses to the above questions were used to obtain the following estimates:

$\hat{\lambda}_1^r$ = estimated probability of a "yes" response to the randomized question asked in the first half-sample.

$\hat{\lambda}_1^d$ = estimated probability of a "yes" to the direct question in the first half-sample and is an estimate of the proportion of the population who lived in a different county or State last year.

$\hat{\lambda}_2^r$ = estimated probability of a "yes" response to the randomized question in the second half-sample.

$\hat{\lambda}_2^d$ = estimated proportion of the population whose mothers were born in April, derived from the second half-sample.

The probability of selecting the sensitive question on abortion is assumed to be P = 1/2, since a penny was the randomizing device. This assumption, together with the above estimates, leads to the following unbiased estimates of π_A, the proportion of the population having abortions. That is

$$\hat{\pi}_A(1) = \frac{1}{P}\left\{\hat{\lambda}_1^r - (1-P)\,\hat{\lambda}_2^d\right\} = 2\hat{\lambda}_1^r - \hat{\lambda}_2^d \quad (2.1)$$

$$\hat{\pi}_A(2) = \frac{1}{P}\left\{\hat{\lambda}_2^r - (1-P)\,\hat{\lambda}_1^d\right\} = 2\hat{\lambda}_2^r - \hat{\lambda}_1^d. \quad (2.2)$$

The final estimate is then the weighted average

$$\hat{\pi}_A = W\,\hat{\pi}_A(1) + (1-W)\,\hat{\pi}_A(2) \quad (2.3)$$

where W was chosen to minimize the variance of $\hat{\pi}_A$. If

$$\Sigma_1^2 = \text{Var}\,\hat{\pi}_A(1) \quad (2.4)$$

$$\Sigma_2^2 = \text{Var}\,\hat{\pi}_A(2) \quad (2.5)$$

$$\Sigma_{12} = \text{Cov}[\pi_A(1),\ \pi_A(2)] \quad (2.6)$$

then the value of W that minimizes the variance of $\hat{\pi}_A$ is

$$W_{opt.} = (\Sigma_2^2 - \Sigma_{12})/(\Sigma_1^2 + \Sigma_2^2 - 2\Sigma_{12}) \quad (2.7)$$

and the variance of $\hat{\pi}_A$ becomes

$$\text{Var}(\hat{\pi}_A) = (\Sigma_1^2\,\Sigma_2^2 - \Sigma_{12}^2)/(\Sigma_1^2 + \Sigma_2^2 - 2\Sigma_{12}). \quad (2.8)$$

Estimates of the variances and covariances of $\hat{\pi}_A(1)$ and $\hat{\pi}_A(2)$ were substituted in (2.7) and (2.8). A balanced half-sample replication procedure [5] was used to compute the variances and covariances of $\hat{\lambda}_1^r$, $\hat{\lambda}_1^d$, $\hat{\lambda}_2^r$, and $\hat{\lambda}_2^d$, which values were then used in the straightforward formulae for estimating the variances and covariance of $\hat{\pi}_A(1)$ and $\hat{\pi}_A(2)$.

2.3 Effectiveness of the Technique

Of the 9,797 respondents, 98.5 percent accepted the randomized response "game" and gave an answer of "yes" or "no". Seven respondents did not answer because they could not read and twenty refused to give an answer. Another 124 were not asked the question or did not give a codable answer. Of these 151, about 55% and 45% were in the odd and even samples respectively. Interviewers reported that some respondents felt flipping a coin was foolish, and immediately answered "no" to both, but no count was made of these respondents. There was no other report of adverse reactions, and the small rate of non-response on this question compares favorably with other items on the questionnaire. Non-response was actually higher for the innocuous questions when asked directly during the course of the interview than for the randomized response question itself. Of the half-sample asked when their mother was born, 8.2 percent did not know the month and an additional 3.0 percent were erroneously not asked the question. For the question on whether they lived in the same county and State last year, there were no "don't know" responses but 3.0 percent of the women were erroneously not asked the question. Most of the not ascertained cases were due to interviewer confusion--asking the wrong question for a woman in a specified half-sample, or completely missing the question. In the NSFG, missing data were imputed from similar respondents where "similar" respondents are defined to be of the same age, race, and, for migration, education.

3. FINDINGS

3.1 NSFG Estimates

The provisional estimates of women with abortions produced through the randomized response technique are shown in the table, along with all the components of the estimator. The overall estimated proportion of the ever married plus single women with their own children in the household who had abortions is 3.0 percent with a standard error of 0.8 percentage points. Based on an estimated 31,018,000 women ever married or single with their own children, an estimated 930,000 women in these categories had an abortion within the 12 month reference period with a standard error of the 248,000 assuming the randomized response procedure achieved accurate reporting.

However, there is wide variation in the half-sample estimates of abortion. According to the table, 5.3 and 0.6 percent of the women ever married or single with their own children, had abortions in half-samples 1 and 2, respectively. The difference in the two half-sample estimates is 3.0 times the standard error of the difference. This great a difference would not be expected by chance. Potential sources of error or bias are: (1) selection of the two half-samples and differential weighting; (2) the randomizing device or its application; (3) interviewers asking inappropriate questions, failing to ask appropriate questions, or incorrectly wording questions; (4) respondents failing to answer questions;

(5) respondents misunderstanding the questions; (6) respondents falsifying their answers; (7) interviewer recording, coding, or keypunch error; or (8) data editing and imputation procedures. Although all these errors probably exist in every survey, the present concern is a bias in one half-sample versus the other half-sample. The only differences between the two half-samples (other than the respondents) were the questions that appeared differently for the two sets of respondents. Hence, the difference in estimates appears to be respondent misunderstanding on these questions.

It was observed during interviewer training for the second cycle of the survey that "county" was often misread as "country". In half-sample 1, the interviewer verbally asked the innocuous question on migration. In half-sample 2, the innocuous question on migration was printed on a card for the respondent to read. The Cycle 1 trained interviewers could not have made the wording mistake often as the NSFG estimate of migration from half-sample 1 is more similar to the Census estimate of inter-county migration than of inter-state migration. However, the respondents were not trained, and might have misread the word as the untrained interviewers tended to do at first.

In view of these possible problems of respondent misunderstanding in half-sample 2, it has been suggested that the result from half-sample 2 be discarded in favor of that from half-sample 1, by itself. Since the NSFG sample was randomly divided, the first half-sample is still a probability sample.

Whether one accepts as best the weighted average of the two half-sample estimates or just the estimate based on the first half-sample, it is still likely that abortions are undercounted in the NSFG since respondents may be hesitant to report having an abortion even if they are convinced that no one will know their answer.

3.2 Comparison with Other Data

How do the findings on abortion using the randomized response compare to other measures of abortion? It was possible for women to volunteer abortions in the NSFG during the time they were asked about birth control methods. Some encouragement was given by abortion being included in a list of family planning methods handed to the respondent during this part of the interview. Relying on this volunteered information alone would have produced an estimate of only 28,000 women having abortions in the previous twelve months, or 3.0 percent of the estimate produced using randomized response techniques. In other surveys, directly asking women about abortions during interview has achieved very little reporting of abortion. The National Fertility Study, conducted in late 1970 and early 1971, found 1.5 percent of the ever married women 15-44 years old had ever had an abortion, and only 0.3 percent reported an abortion within the 12 months prior to interview.1/ This latter estimate is much smaller than the corresponding estimate of 2.8 percent of ever married women produced in the NSFG.

The major sources of abortion data have been the Center for Disease Control and the Alan Guttmacher Institute. In 1973, the Center for Disease Control collected reports on 615,831 legal abortions from the central health agencies in 24 States and the District of Columbia and from hospitals and other facilities in the remaining States [2]. These reports did not cover all facilities performing legal abortions, nor did all States report for the full 12 months. Also in 1973, the Alan Guttmacher Institute surveyed health service providers and arrived at an estimate of 745,400 legal abortions with an estimated underreporting of 5-10 percent [9]. Direct comparison cannot be made between the NSFG estimates of abortion and those from the Center for Disease Control or the Alan Guttmacher Institute because the data relate to different populations of women--most notably the NSFG did not interview single women without their own children in the household.

While the Center for Disease Control did not tabulate data on abortions separately for widowed, divorced, and separated women or for never married women, it does report that 27.4 percent of the legal abortions in the 24 States reporting marital status were to currently married women.2/ If this percentage applied to all conterminous States, 167,000 of the legal abortions reported to the Center for Disease Control and 203,000 of the legal abortions reported to the Alan Guttmacher Institute would be to currently married women.

Additional adjustments would be needed to restrict the data from the Center for Disease Control and the Alan Guttmacher Institute to produce estimates that are comparable to the estimates from the NSFG. However, this would require data on how many of the reported abortions occurred to currently married women below age 15 or above age 44, and how many of the reported abortions were the second, third, etc., to an individual woman during 1973. In addition, abortions occurring in the last quarter of 1973, well after the Supreme Court ruling on abortions, would need to be deleted and abortions occurring in the last quarter of 1972, before the ruling, would need to be added. All of these adjustments would further reduce the number of reported abortions in the Center for Disease Control and the Alan Guttmacher Institute data. Therefore, failure to make these adjustments should result in an underestimate of the difference between the magnitude of abortion measured in the NSFG and the magnitude of legal abortion reported to the Center for Disease Control and the Alan Guttmacher Institute.

7. CONCLUSION

The use of the randomized response method of obtaining information on sensitive, and therefore underreported, behavior has produced higher estimates of abortion than have previously been achieved. Its use in the National Survey of Family Growth was therefore valuable. Great confidence cannot be placed in any single estimate of abortion, however, because of apparent problems in questionnaire design and the additional variance introduced by the randomized response technique. The survey produces an estimate of

3.0 percent of the surveyed women, who include both ever married women and single women if their own children live with them, having had an abortion within the preceeding 12 months. It is possible, however, that an estimate slightly greater than 5.0 percent is more appropriate. This indicates that between 3 and 8 times as many abortions were performed in 1973 as were obtained through the present abortion reporting systems.

The second cycle of the National Survey of Family Growth, with fieldwork scheduled during the first half of 1976, should provide further estimates on abortion and further testing of this application of the randomized response technique. Three changes have been made for the second cycle: (1) the way a pregnancy ended is asked directly as well as questions about abortions by the randomized response method, (2) a question on being the only child replaces the question on whether the mother was born in April, the latter having had high rates of "don't know" in Cycle I, and (3) the card handed to the respondent in half-sample 2 reads:

> This time last year, I lived in a different county in this State--or a different State--than this one.

Although the randomized response models have been used for at least ten years, there is still need for work on the Field Administration and subsequent analysis of these models. This analysis is but a first step in the investigation of the technique, especially as it relates to measuring the incidence of abortion. Additional analysis can and will be made using the NSFG data from Cycle I, with further clarifications being made once the Cycle II data are available.

FOOTNOTES

1/ Tabulations made by the authors from the 1970 National Fertility Study data file obtained from the Data and Program Library Service and Larry Bumpass, Department of Sociology, at the University of Wisconsin. The survey was conducted by Norman B. Ryder and Charles F. Westoff, Office of Population Research, Princeton University. See [10] for a discussion on the abortion findings.

2/ This is not too different from results found in 1970 where 29.9 percent of the abortions were to currently married women, 14.2 percent to widowed, divorced, or separated women, and 55.9 percent to single women [6].

REFERENCES

[1] Abernathy, James R., Bernard G. Greenberg and Daniel G. Horvitz, "Estimates of Induced Abortion in Urban North Carolina," Demography 7 (February 1970), 19-29.

[2] Center for Disease Control, Abortion Surveillance: 1973, (May 1975).

[3] Folsom, Ralph E., Bernard G. Greenberg, Daniel G. Horvitz, and James R. Abernathy, "The Two Alternate Questions Randomized Response Model for Human Surveys," Journal of the American Statistical Association, 68, (September 1973), 525-30.

[4] Horvitz, Daniel G., B. V. Shah, and Walt R. Simmons, "The Unrelated Question Randomized Response Model," Proceedings of the Social Statistics Section, American Statistical Association, 1967, 65-72.

[5] National Center for Health Statistics, "Replication, An Approach to the Analysis of Data from Complex Surveys," Vital and Health Statistics, PHS Pub. No. 1000 - Series 2 - No. 14. Public Health Service, Washington, D.C., United States Government Printing Office, April 1966.

[6] Tietze, Christopher and Sarah Lewit, "Joint Program for the Study of Abortion (JPSA): Early Medical Complications of Legal Abortion," Studies in Family Planning, 3, (June 1972), 97-122.

[7] United States Bureau of the Census, "Mobility of the Population of the United States: March 1970 to March 1971," Current Population Report, Series P-20, No. 235, (April 1972), Table 7.

[8] Warner, Stanley L., "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," Journal of the American Statistical Association, 60, (March 1965), 63-9.

[9] Weinstock, Edward, Christopher Tietze, Frederick S. Jaffe, and Joy G. Dryfoos, "Legal Abortions in the United States Since the 1973 Supreme Court Decisions," Family Planning Perspectives, 7, (January/February 1975), 23-31.

[10] Westoff, Charles F. "The Yield of the Imperfect: The 1970 National Fertility Study," Demography, 12, (November 1975), 573-80.

Provisional Estimates for Number and Percent of Women 15 to 44 Years Old Who had Abortion in the 12 Months Prior to Interview, by Marital Status: United States 1973. (Standard Error of Estimates Shown in Parentheses)

		Total	Single with Children	Ever Married	Marital Status: Currently married	Marital Status: Widowed divorced, separated
Total number of women (thousands)		31,018 (395)	771 (59)	30,247 (390)	26,646 (364)	3,601 (128)
Number of women with an abortion in 12 months (thousands)		930* (248)	77* (29)	847* (242)	693* (213)	194* (61)
Percent						
$\hat{\pi}_A$	- Women with an abortion in 12 months, combined	3.0* (0.8)	10.0* (3.7)	2.8* (0.8)	2.6* (0.8)	5.4* (1.7)
$\hat{\pi}_A(1)$	- Women with an abortion in 12 months, half-sample 1	5.3 (1.1)	9.0* (4.8)	5.3 (1.1)	5.1 (1.1)	6.4* (2.3)
$\hat{\pi}_A(2)$	- Women with an abortion in 12 months, half-sample 2	0.6* (1.1)	11.4* (5.6)	0.3* (1.1)	-0.2* (1.1)	4.0* (2.7)
$\hat{\lambda}_1^r$	- Yes to randomized question, half-sample 1	7.0 (0.5)	7.6 (2.3)	7.0 (0.5)	6.8 (0.5)	8.0 (1.0)
$\hat{\lambda}_1^d$	- Migrated during 12 months, half-sample 1	12.0 (0.6)	5.4 (1.8)	12.2 (0.6)	12.7 (0.7)	8.4 (1.2)
$\hat{\lambda}_2^r$	- Yes to randomized question, half-sample 2	6.3 (0.4)	8.4 (2.7)	6.2 (0.4)	6.3 (0.5)	6.2 (1.2)
$\hat{\lambda}_2^d$	- Mother was born in April, half-sample 2	8.6 (0.5)	6.3 (1.4)	8.7 (0.5)	8.6 (0.5)	9.6 (1.1)
$Cov[\hat{\pi}_A(1), \hat{\pi}_A(2)]$		0.2	0.4	0.1	0.2	0.1

*Relative standard error greater than 25.0 percent

THE EPIDEMIOLOGY OF DRUG ABUSE: A NEW DATA BASE, NEW TECHNIQUES AND NEW FINDINGS

Eduardo N. Siguel and William H. Spillane
National Institute on Drug Abuse

INTRODUCTION

The Client Oriented Data Acquisition Process (CODAP) is a data collection system developed and operated by the National Institute on Drug Abuse (NIDA) in treatment facilities (clinics) that receive Federal funds. Its purpose is to provide current information which describes clients and the treatment provided to them in order to aid in planning, management and evaluation activities. This report presents a concise summary of CODAP and describes some uses of CODAP in regard to the epidemiology of drug abuse.

HISTORY

CODAP has been evolving since 1972. During this time there has been a substantial increase in the number of clients admitted to drug abuse treatment programs and reported on the CODAP system. The system has been adapted in order to be more responsive to user needs.

CODAP was initially designed by the Special Action Office for Drug Abuse Prevention (SAODAP) to satisfy the requirements outlined in P.L. 92-255. When the system design was completed, the responsibility for its implementation was transferred to NIDA's predecessor, the Division of Narcotic Addiction and Drug Abuse of the National Institute of Mental Health, U.S. Department of Health, Education and Welfare. CODAP, which has been operational since 1973, requires reporting by all recipients of Federal funds designated for the provision of drug treatment and rehabilitation services.

During CODAP's first year of operation, over 130,000 Admission Reports for clients entering treatment were processed from nearly 900 reporting clinics. Because drug abuse activities were expanding so rapidly at that time, an evaluation of CODAP was performed by the staff members of participating Federal and State agencies. The system was found to be lacking flexibility in the collection of significant information for Federal, state and program management requirements. It was determined, however, that with some revision CODAP would not only meet more sophisticated management needs but also would fulfill, to a greater degree, basic state information requirements. From April to August of 1974, a pretest of the revised CODAP system was performed in three states and several local programs. The pretest effort and a series of conferences with 36 participating Single State Agencies (SSA's) and other local treatment agencies resulted in the current design of CODAP. This revised version became operational November 1, 1974.

This article was written by Dr. Siguel and Dr. Spillane in their private capacity. No official support or endorsement by NIDA is intended or should be inferred.

CURRENT STATUS

Approximately 1,600 clinics currently report to CODAP on a regular basis. These clinics account for more than 30,000 client admissions and discharges each month. Agencies participating with NIDA in the CODAP data collection effort include: The Veterans Administration (VA), the Bureau of Prisons (BOP), the Law Enforcement Assistance Administration (LEAA) and the Department of Housing and Urban Development (HUD). While clinics may receive funds from several Federal or non-Federal sources, most of the reporting clinics receive NIDA funding. The SSA's responsible for administering drug abuse services also participate in the CODAP effort. In several states the SSA's implement the standard CODAP data collection instruments on a state-wide basis, while in others, SSA's obtain the necessary data from existing systems modified to comply with Federal standards for the collection of CODAP data.

The current CODAP system offers not only greater flexibility than the previous version, but also provides additional analytical potential and increased protection of client confidentiality. Its emphasis is on collecting client-related data at the points of admission to and discharge from treatment. CODAP forms are completed by trained clinic staff and do not reflect either the opinions or value judgments of the client. The revised CODAP provides descriptions of:

> drug abuse phenomena such as:
>
> demographic characteristics of drug abusers seeking treatment; geographic location (county, Standard Metropolitan Statistical Area (SMSA), state) of acute drug problems; indicators of the incidence and prevalence of drug abuse as well as patterns of abuse including the number and type of drugs being abused, the frequency of use of these drugs and the time interval from onset of drug use to continuing use and treatment.
>
> the treatment process such as:
>
> the number of clients being treated in different modality/environment treatment regimens, including the number of individuals waiting to be treated; the kinds of treatment services being provided; the impact of various treatment endeavors on specific patterns of drug abuse; the characteristics of clients likely to complete treatment and the reasons for discharge from and the length of time spent in treatment.

Data described in both of these categories are necessary for the development and execution of an effective strategy to counter drug abuse. Specific illustrations of potential managerial applications of these data are presented in the last section.

FUTURE OUTLOOK

As with any new national system of this magnitude, the development of CODAP has not been without problems. The training of a few thousand clinic staff involved in the data collection process was a major undertaking. In addition, extensive education was required to demonstrate to the clinics and the public that client confidentiality safeguards were adequate. After one year of operation, new considerations are appearing that were not evident at the time of the pilot test. For example, discharge reports are more difficult to match to admission records than was anticipated. Also, certain data items are proving to be less useful than was expected while others require more clarification than was originally foreseen.

In the near future, as the system becomes further refined and the information becomes more available to the states, increased utilization and many new applications of the data are anticipated. For example, an SSA may request copies of its state CODAP data tape files from NIDA. NIDA will provide the tapes as well as any technical assistance required to interpret them via standard statistical packages so that SSA's may produce tabulations to be used in management decision-making at the state level. Procedures have been developed to provide CODAP data to the scientific community for research and evaluation.

DATA COLLECTED

CODAP provides drug abuse program management with not only client-related data as to the type and pattern of drug usage, educational and employment status and demographic characteristics, but also with clinic activity data indicating treatment approaches, services provided and the number of clients treated.

In contrast to the original CODAP system which collected aggregate data on a quarterly basis, the revised CODAP collects individual client data on a monthly basis:

> Admission Report - This form is completed for each client as he is admitted to a clinic for treatment after being screened and accepted. It provides data regarding the date of admission to the clinic, admission type (first admission, readmission, transfer), treatment modality (detoxification, maintenance, drug free, other), treatment environment (prison, hospital, residential, day care, outpatient), medication(s) prescribed, legal status (voluntary or involuntary), demographic characteristics (sex, year of birth, race or ethnic group), employment status (employed or unemployed), educational status (last formal school year completed and educational or skill development program enrollment at admission), number of prior treatment experiences (including months since last treatment experience), and pattern of drug abuse for the primary, secondary and tertiary drug types. The pattern of drug abuse is specified in terms of the drug type used, frequency of use at admission, year of first use, year of last continuing use and whether or not the client has a problem with more than three drugs.
>
> Discharge Report - This form is completed for each client as he is discharged from a clinic. It provides data concerning the date of discharge from and admission to the clinic, the reason for discharge (completed treatment-no drug use, completed treatment-drug use at discharge, treatment no completed, transfer, referral, noncompliance to program rules, incarceration, or death), modality and environment at time of discharge (see description of Admission Report for categories), medication(s) prescribed, length of time in treatment (in any and all clinics in the program), employment and educational status and drug(s) used at time of discharge.
>
> Client Flow Summary - This report provides a client census by modality and environment as of the last day of the report month, a summary of screening and referral activity during the month, the number of individuals currently on the waiting list and the number of admission and discharge reports submitted for the month.
>
> Client Progress Report, Bureau of Prison (BOP) - This report is submitted quarterly only for BOP clients and provides information regarding their individual progress.

Submission of the Activity Report, which provided data concerning clinic activities during the month and characteristics of clients in treatment at the end of the report month, was made optional to the SSA's as of November 1, 1975.

In addition, NIDA maintains a separate Control File for each state which contains clinic administrative data including identification information, funding sources and program (organizational) linkages. The Control File is updated quarterly with about one-third of the states being processed each month during the quarter.

PROCEDURES FOR DATA COLLECTION AND PROCESSING

NIDA provides training in the completion of the above forms to all programs and clinics receiving Federal funds. In addition, the Institute trains SSA staff in the CODAP data collection requirements, definitions, and procedures so that they can assist in the data collection process.

Trained clinic personnel are responsible for form completion and submission either directly to NIDA, or to the SSA's which in turn send the completed material to NIDA. Completed forms for each report month are due by the 7th of the following month. After editing and keypunching, data files for a given month are available approximately one and one-half months after the end of the report month. For example, data regarding activities during the month of October are ready for analyses around the 15th of December. The actual forms are generally available to NIDA after the middle of November.

By using the CODAP reports and descriptive information for each clinic obtained from the Control File, the following files are prepared: the Admission and Discharge Files which contain the Admission and Discharge Reports, respectively; the Clients-in-Treatment File which contains records for all clients admitted but not yet discharged; the Client Flow Summary File which contains all data submitted on the Client Flow Summary Reports; Historical Clients File which contains all data on clients who have been discharged - Admission and Discharge Reports are matched and stored together on this file to allow for analyses of discharge data as a function of admission data. For example, the reasons for discharge and length of time in treatment as a function of the pattern of drug abuse at admission can be evaluated through the use of this file.

A client identifier number assigned by the treatment clinic appears on each Admission and Discharge Report form. The primary function of this client identifier is to serve as a means of matching admission and discharge data for the client. Data files with the original data including the client identification number are available to selected NIDA staff in order that they may either provide information to the clinics in case some or all of their records are lost or to correct any erroneous or missing information.

QUALITY CONTROL AND CONFIDENTIALITY

The accuracy of the system is being monitored continuously. Clinics that are late in reporting are contacted to determine the reasons for the delay. Continuous delays are not allowed. If necessary, additional training and technical assistance are provided. Tests for internal validity and report consistency have been computerized. Internal validity tests are performed on the incoming forms to make certain that the responses are within acceptable ranges as defined in the CODAP National Management Handbook. Data items found to be invalid become part of an error report produced by the data processing system to facilitate manual resolution. Internal consistency tests are performed to insure that the reported data items are consistent with one another. For example, each discharge date is tested to determine if it is after the admission date and the year of birth is tested to see if it is before the year of admission. Even though the tabulations for January-March 1975 indicated that after editing and keypunching, less than five percent of the forms contained invalid codes a comprehensive strategy has been developed to test the internal consistency of CODAP data. Also, a separate study of the external validity of the data which compares the items reported with other clinic records is currently under way.

The CODAP system fully adheres to the requirements for confidentiality established through the amendments (section 303 (a) of PL-93-282) to section 408 of the Drug Abuse Office and Treatment Act of 1972. Safeguards have been implemented to prevent the possible identification of clients from data. Only the clinic maintains files of the previously mentioned client numbers that allow cross-referencing between client numbers and individuals. Furthermore, CODAP data cannot be linked to other Federal data (such as U.S. Bureau of the Census data) because the data describing the client is not sufficient to match records. Thus, no Federal agency has a cross-reference index to the client's identification.

DATA LIMITATIONS

There are several factors which should be considered when interpreting CODAP data. While the universe of federally-funded drug abuse treatment clinics should be reporting all of the requested CODAP information consistently and accurately, there are limitations to the collection of data which can lead to some inconsistencies and incomplete reporting. The universe of reporting clinics varies from time to time because new clinics are being created, old ones are closing and not all of the existing clinics are reporting consistently each month. Although the characteristics of the population of nonreporting federally-funded clinics are unknown, most of the nonreporting clinics are new and have relatively few clients. Thus, the percentage of clients actually being reported is higher than the percentage of clinics reporting.

This large core of clinics consistently reporting to CODAP provides a broad data base with which to perform useful analyses. While the absolute numbers reflect only those clinics which reported, the percentage relationships derived from these numbers adequately reflect activity in all federally-funded clinics, and provide indicators of drug abuse phenomena outside the clinics. Thus, profiles of drug abuse phenomena related to treatment activity can be developed through cross-tabulations of CODAP data. The large number of clients for whom data are collected allows analyses to be performed which are not feasible when only a few thousand individuals are surveyed. Any one variable against any other(s) can be tabulated for a defined subset of the CODAP population to provide insight into specific abuser characteristics. For example, the following can be tabulated: 1) the percentage of women under 21 with a high school education who use heroin; 2) the probability that the successful completion rate increases or decreases as the number of prior treatment experiences; 3) the age, race and sex distribution of clients admitted to treatment and their patterns of drug abuse; 4) length of time in treatment can be analyzed for purposes of evaluating use of resources; 5) selected indicators of clinic's characteristics, clients in treatment and treatment approaches which can be compared across clinics or programs.

ILLUSTRATIVE APPLICATIONS OF CODAP

This section illustrates some potential applications of CODAP data to the epidemiology of drug abuse. It is not the purpose of this presentation to analyze drug abuse or drug abuse treatment. Therefore, specific data and tables are presented merely for purposes of illustration. The total number of clients reflected in each table are not always the same. This is due to the fact that the number of missing values depends on the client variables presented. Therefore, the number of observations excluded from a table because of missing data was dependent on the nature of the particular variables under consideration.

The more acutely managers are aware of the characteristics of drug abusers and their drug problems, the better they will be able to direct the activities of intervention programs to respond to abusers' needs for treatment. Data describing both demographic characteristics of drug abusers and drug abuse patterns are available from the CODAP Admissions File.

Pattern of Drug Abuse

Age, race, sex, educational background, and employment status are among the data collected for each client at the time of admission to treatment. Table 1 depicts a tabulation of selected demographic and drug use variables of clients admitted to treatment during 1975. Knowledge of selected demographic and social characteristics of clients in treatment will result in greater utilization of more appropriate treatment approaches. For example, it is important that members of the treatment staff be able to communicate with clients. This may require the matching, to some extent, of such characteristics as age, race, and sex for certain staff positions with those characteristics for the client population. Educational background and job status can be used in conjunction with age, race, and sex to plan counseling approaches. A large number of young clients with weak educational backgrounds would indicate that program emphasis be placed on continuing education or vocational training. Older unemployed clients with a high school education, however, may require more emphasis on job training and job placement. Current estimates indicate that the employment rate increases when age increases and about 25% of the clients admitted to treatment are employed.

Characteristics of clients' drug abuse problems are also particularly valuable in developing treatment strategies. CODAP not only provides data identifying the drugs being abused by clients at time of admission but also their patters of previous drug abuse. This includes data pertaining to the combinations of drugs used, the number of prior treatment experiences, and the time intervals between the various stages of drug abuse for each client. This data can naturally be cross-tabulated with other client characteristics.

It can be seen from Table 1 that race (or ethnicity) and age are related to opiate utilization. While the use of opiates increases with age, the use of marihuana decreases. The relationship between age, race and primary drug of abuse can be used to anticipate and plan for the kinds of drug treatment appropriate for a clinic based on the age/race distribution of its potential clients. This relationship suggests the types of treatment approaches that may be appropriate for particular age groups. For instance, if one were to establish a drug counseling program in a youth center that caters principally to individuals younger than 18, the emphasis could be placed on marihuana as the primary drug of abuse. On the other hand, a community center that attracts individuals over 30 may want to emphasize treatment for alcohol and heroin use.

Tabulations of various client data pertaining to the severity of drug problems can be produced for individual treatment programs in addition to all programs within specific geographical areas. The number of prior treatment experiences for clients can be cross-tabulated with the number and types of drugs abused and other drug abuse pattern variables such as the length of time between first use and first continuing use. Tables of this nature will provide the managers with an indication of the trends in drug abuse activity in particular areas and will facilitate comparisons among programs, states, and other geographical regions.

Often preconceptions exist pertaining to the types and combinations of drugs being abused. Since such notions influence the general approach taken to combat drug abuse, it is worthwhile to examine data from CODAP in order to acquire more factual evidence.

Trends in Drug Abuse

Clients' drug histories can be viewed as a series of three critical points in time for each particular drug used. The first of these is the year when a drug is first used (referred to as onset of drug abuse). The second is the year of first continuing use, and the third is the year of admission to a treatment program. CODAP includes this drug history data and thus allows for analysis of the time intervals among these three points in time. These data may help managers to better direct prevention and treatment activities at target populations before continuing drug usage occurs by identifying the age at onset. The age at first use of a given drug may be used to develop programs aimed to those individuals with greatest risk of beginning to use drugs (peak of curve describing year of first use). CODAP data may be used to monitor trends in the age at first use. Population groups with specific demographic and drug abuse pattern characteristics can be compared in terms of the distributions of the above time intervals to determine if there are basic differences between the groups. Such information can be used to estimate future demands for treatment based on current problems and expected time lags between first use, continuing use, and need for treatment. Changes over time (such as trends) in patterns of drug abuse can be described using the above time intervals. Figure 1 shows the percent of individuals who began to use heroin during 1975. The peaks around 1969 are interpreted as evidence of a 1969 epidemic while the peak around 1972 is related to the peak of waiting time to enter treatment. Mathematical models are being developed to study waiting times to first treatment, waiting times between consecutive treatment experiences and other related times. For example, preliminary analysis suggests that a Weibull distribution fits data on waiting times between first and continuous use.

The number of times clients have previously been treated for drug abuse also indicates the severity of their problem. Preliminary tabulations suggest that as the number of prior treatment experiences increase, the probability of a client being admitted to a detoxification modality increases and his probability of [redacted] being placed in a drug free modality decreases. Such insights could be valuable in planning strategies for admission and treatment. For example, the number of opiate users with prior treatment experiences could be used to estimate the demand for detoxification programs.

SUMMARY

CODAP can be an extremely valuable tool for managers at the national, state and local level. CODAP helps to answer a myriad of questions regarding the problem of drug abuse and drug abusers, such as the determination of target groups for prevention efforts based on patterns and history of drug abuse; the allocation of resources; and the planning for the demand for treatment approaches (such as detoxification and maintenance); the estimation of incidence and prevalence; and the evaluation of the effectiveness of treatment programs. Trends in selected indicators can be used to monitor the performance of clinics. Applications of CODAP will increase using 1975 Admissions and Discharges (without client, clinic or program identification) which the National Institute on Drug Abuse made available to the scientific community during 1976.

Acknowledgements

Several people provided valuable comments on the preparation of this paper. In particular, we would like to thank the staff of the Division of Scientific and Program Information for their suggestions and editing.

References

1. CODAP National Management Report. Quarterly reports prepared by Division of Scientific and Program Information, National Institute on Drug Abuse, based on CODAP data.

2. CODAP National Management Handbook, NIDA, November 1974. This handbook defines each data item in the CODAP forms.

3. Statistical Series, Quarterly Report. Quarterly reports on demographic and social characteristics of clients in treatment prepared by Division of Scientific and Program Information, National Institute on Drug Abuse, based on CODAP data.

4. The Client Oriented Data Acquisition Process (CODAP) by Siguel, E.N. and Spillane, W.H. National Institute on Drug Abuse, 1976.

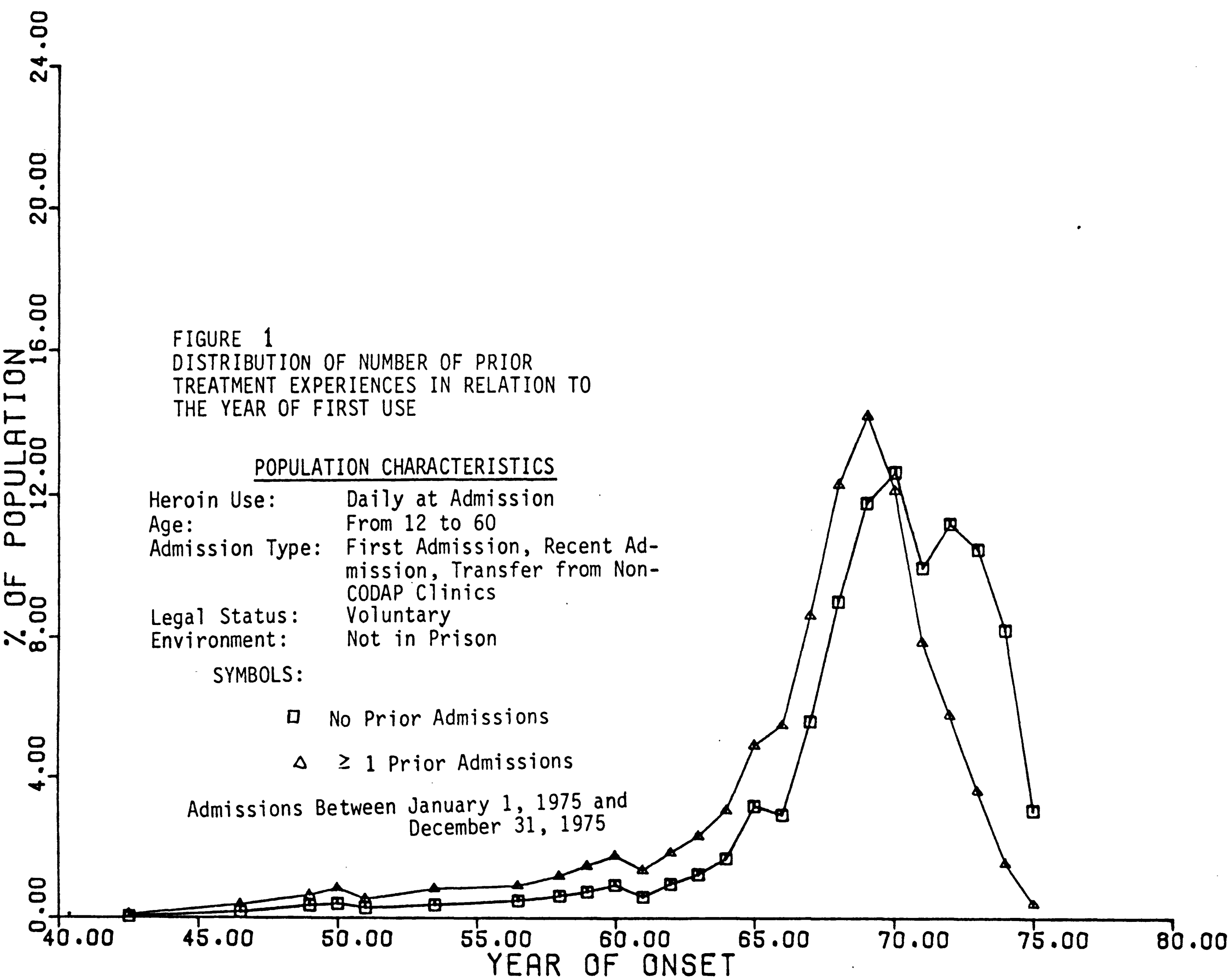


TABLE 1 -- TYPES OF PRIMARY DRUG USED BY SEX-AGE-RACE CHARACTERISTICS OF CLIENTS (IN PERCENTAGES)

	SEX-AGE-RACE CHARACTERISTICS OF CLIENTS										
	SEX		AGE					RACE			
PRIMARY DRUG	Male	Female	18	18 - 20	21 - 25	26 - 30	30	White	Black	Spanish	Other
None	2.7	4.2	6.2	3.0	2.3	2.5	3.0	2.4	4.2	2.8	2.4
Opiates	59.6	53.7	3.9	33.3	69.7	77.7	67.2	42.6	77.0	73.6	36.2
Alcohol	8.2	5.4	7.7	5.5	3.4	4.6	18.4	10.1	4.9	2.8	9.0
Barbiturates	4.3	6.4	8.4	8.8	4.7	2.8	2.5	7.4	1.9	1.9	4.3
Amphetamines	4.2	5.1	5.5	7.2	4.7	3.7	2.2	7.3	1.2	1.1	4.3
Marihuana	14.1	16.4	50.4	29.7	9.7	4.6	2.1	19.5	8.3	11.0	26.9
Other	6.9	8.9	17.9	12.5	5.5	4.1	4.5	10.6	2.6	6.7	16.8
Total Percentage	100.0	100.1	100.0	100.0	100.0	99.9	99.9	99.9	100.1	99.9	99.9
Total N	35460	12298	5784	6357	15643	10586	9188	24946	16864	5191	668
Row Percent	74.2	25.8	12.2	13.4	32.9	22.3	19.3	52.3	35.4	10.9	1.4

A NOTE ON THE RANDOMIZED RESPONSE TECHNIQUE

Jagbir Singh, Temple University

1. Introduction

In surveys related to delicate questions, Warner (1965) introduced a randomized response technique for eliciting information, and thus estimating the proportion π of the population in the sensitive category A. The technique consists in providing a spinner, with two outcomes A or not A with associated probabilities p and $\bar{p} = 1-p$ to each respondent. The respondent spins the spinner unobserved by the interviewer and answers yes if he possesses the characteristic indicated by the pointer and no otherwise. It is obvious that the probability of a yes response, say θ, is

$$\theta = p\pi + \bar{p}(1-\pi)$$
$$= \bar{p} + (2p-1)\pi.$$

If n_1 is the number of yes responses from a total of n interviews, and if all responses are truthful, Warner claimed that the maximum likelihood estimator (m.l.e.) of θ is $\hat{\theta} = n_1/n$ and consequently that of π is

$$\hat{\pi} = [\frac{n_1}{n} - (1-p)]/(2p-1)$$

provided $p \neq 1/2$. The purpose of this note is to point out that $\hat{\theta}$, as well as $\hat{\pi}$, are not the m.l.e.'s of the respective parameters. Indeed the m.l.e. of θ, say $\tilde{\theta}$, presented in this note is uniformly better than $\hat{\theta}$ with respect to squared error loss. In other words, $\hat{\theta}$ is not even admissible with respect to squared error loss. The same, of course, is true with $\hat{\pi}$.

2. The Maximum Likelihood Estimator

Let us assume without any loss of generality that $p > 1/2$. It is of crucial significance to note that θ is restricted to be in the interval $(\bar{p},p)$, since π is in (0,1). Thus the m.l.e. $\tilde{\theta}$ of θ is

$$\tilde{\theta} = \begin{cases} n_1/n & \text{if } \bar{p} < n_1/n < p \\ \bar{p} & \text{if } \bar{p} \geq n_1/n \\ p & \text{if } n_1/n \geq p \end{cases}$$

Let $b(x: n,\theta)$ be the binomial probability density function of n_1. We can write

$$E(\tilde{\theta}) = \theta + B$$

where the bias B is

$$B = \frac{1}{n}[\sum_{x<n\bar{p}} (n\bar{p}-x)b(x: n,\theta) + \sum_{x>np} (np-x)b(x: n,\theta)].$$

The values of the first and the second summations are respectively positive and negative. However, overall magnitude and the sign of B will depend upon the values of the parameters n, p, and π. For fixed n, and p, it can be shown that the bias evaluated at π is equal in magnitude to that at $(1-\pi)$ but opposite in sign. We computed the bias magnitude for a limited choice of the parameters, and found it to be very small. It is easy to verify that $\tilde{\theta}$ is consistent.

To see that $\tilde{\theta}$ is uniformly better than $\hat{\theta}$ with respect to squared error loss, we notice that:

$$E(\tilde{\theta}-\theta)^2 - E(\hat{\theta}-\theta)^2 = \sum_{x\leq n\bar{p}} \{(\bar{p}-\theta)^2 - (\frac{x}{n} - \theta)^2\}$$

$$b(x: n,\theta) + \sum_{x\geq np} \{(p-\theta)^2 - (\frac{x}{n} - \theta)^2\}b(x: n,\theta).$$

Both the summations are negative. In the first summation $x/n \leq \bar{p}$. This means $x/n - \theta \leq \bar{p}-\theta < 0$. Hence $(\bar{p}-\theta)^2 - (x/n - \theta)^2 < 0$. This guarantees the value of the first summation to be negative. By the similar argument the second summation is also negative. Thus $\hat{\theta}$ is not even admissible.

The m.l.e. $\tilde{\pi}$ of π is obtained from $\tilde{\theta}$, since π is linearly related to θ.

$$\tilde{\pi} = \begin{cases} [\frac{n_1}{n} - (1-p)]/(2p-1) & \text{if } \bar{p} < n_1/n < p,\ p \neq 1/2 \\ 0 & \text{if } \bar{p} \geq n_1/n \\ 1 & \text{if } n_1/n \geq p \end{cases}$$

Of course, statements which are true for $\tilde{\theta}$ relative to $\hat{\theta}$ continue to be true for $\tilde{\pi}$ relative to $\hat{\pi}$. Asymptotically, both $\tilde{\pi}$ and $\hat{\pi}$ are equally efficient, but for small surveys $\tilde{\pi}$ is highly efficient. For example, if n = 10, p = 0.8, $\pi = 1/12$, then the relative efficiency of $\tilde{\pi}$ is 1.62.

In conclusion, it should be said that similar observations can be made for the follow-up research on this topic.

3. Summary

In surveys related to delicate questions, Warner (Journal of the American Statistical Association, 60 (1965), 63-69) introduced a randomized response technique for eliciting information, and thus estimating the proportion π of the population in the sensitive category. The estimator of π obtained was called, not only by Warner but also subsequently in the literature, as the maximum likelihood estimator. It is shown here that Warner's estimator is indeed neither the maximum likelihood estimator nor even admissible. The maximum likelihood estimator given in this note is uniformly better than that of Warner's with respect to squared error loss.

References

[1] Warner, S. L., "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," Journal of the American Statistical Association, 60 (March 1965), 63-69.

DESIGN EFFECTS IN RETROSPECTIVE MORTALITY SURVEYS

Monroe G. Sirken and Patricia N. Royston
National Center for Health Statistics

1. INTRODUCTION

Recently, we conducted an experiment to measure some error effects of selected design factors in retrospective household sample surveys on dual system estimators of mortality. These estimators have been used to estimate the number of deaths [6] and the completeness of death registration [4]. The dual system mortality estimator implies two operational stages:

- Conducting a survey in which household respondents report retrospectively deaths that occurred during a prior calendar period.
- Matching the deaths enumerated in the survey against a file of registered deaths.

Our experiment was designed to measure the error effects of survey design factors on the number of deaths enumerated in the survey that were matched with their death certificates.

The experiment was based on a sample of deaths registered in North Carolina. First, an address frame of the households that would be eligible to report these deaths in a retrospective survey was compiled. Second, retrospective surveys were conducted on subsamples of addresses. Finally, the reported survey deaths were matched against the file of registered deaths in North Carolina. The objectives and procedures of each stage of the experiment are summarized in Exhibit 1.

2. DESIGN FACTORS AND OPTIONS

Design factors represent the manipulatable features of the survey design. There are many design factors, including the sample design, estimator, data collection method, questionnaire design, etc. Each factor usually presents several design options, including the null option, and each option has its cost and error effects. In this frame of reference, the survey design problem may be stated as follows: to select an option for every design factor such that the set of options selected is best in the sense that it produces a smaller mean square error for fixed costs than would be produced by any other option set.

The first and second columns of Exhibit 2 list and define respectively the five design factors that were investigated in the survey experiment. The particular options that were investigated for each factor are listed in the third column. In view of our particular interests in the counting rule strategy [5], the counting rule and the counting rule weight were the principal design factors that we investigated in the experiment. In this connection, a recent paper by Nathan [3] is noteworthy. These factors are probably less familiar to the reader than the other listed design factors. Hence, the information presented for counting rules and counting rule weight in Exhibit 2 is amplified below.

Counting Rules

In retrospective mortality surveys, counting rules specify the conditions that make decedents eligible to be enumerated at households. Five different counting rule options are listed in Exhibit 2. For instance, option 1.1 links decedents to their last places of residence. In compliance with this rule, the household respondent would be asked: "Did anyone die during the [reference period] while he was living here?" Option 1.1 is a conventional rule since it has the property of making a decedent eligible to be enumerated at one and only one household. The remaining options 1.2-1.5 listed in Exhibit 2 are multiplicity rules since they do not uniquely link decedents to one and only one household. In compliance with each of these options, the appropriate question becomes: "Does anyone living here have a [specified relative] who died during the [reference period]?"

It is noteworthy that the counting rule could be based on any subset of the five options listed in Exhibit 2 including options taken 1, 2, 3, 4 and 5 at a time. Summing these subsets, we obtain a total of 31 different counting rule possibilities.

Options 1.2-1.5 were used to link middle aged decedents to the residences of surviving relatives. Somewhat different options were tested for decedents in the youngest and oldest age groups. Children under 17 years were linked to the residences of their mother, and their maternal grandparents, aunts, and uncles. Decedents 85 years and older were linked to residences of their spouse, children and siblings.

The experiment also investigated counting rules that specify the proximity of the decedents' last residence to the residences of their surviving relatives. There were three options: the decedent and his surviving relative lived in the same County, in the same State, or in the U.S. To apply these geographic rules requires collecting information on the location of last residence for every decedent enumerated in the survey.

Counting Rule Weights

Survey estimators based on multiplicity counting rules adjust for the multiple chances of a decedent being enumerated by appropriately weighting each household that reports him in the survey. These weights are called counting rule weights. The survey estimator is unbiased if the sum of weights assigned to the households eligible to report the same decedent is equal to one.

Exhibit 1.

Stages in the Design of the Survey Experiment: Objectives, Procedures and Products

Stage	Objective	Procedures	Types of Errors Measured
1	To compile an address frame of households eligible to report deaths	Selected a sample of death records from the files of registered deaths; queried death record informants to obtain the addresses where the deaths would be enumerable in a survey.	Sampling errors Counting rule bias
2	To conduct retrospective surveys based on households selected from the address frame	Conducted mail and personal interview surveys; conducted reinterviews with adults who did not originally respond for themselves.	Nonresponse bias Response bias
3	To match the deaths enumerated in the survey to their registered death certificates	Matched the deaths enumerated in the survey against the complete file of registered deaths using the Health Department's standard matching procedures.	Matching bias

Exhibit 2.

Survey Design Factors: Definitions and Options

Design Factors	Definitions	Options
1. Counting Rule	Defines the households where the deaths are eligible to be enumerated in the survey	1.1 Last residence of decedent 1.2 Residence of surviving spouse 1.3 Residence of surviving parents 1.4 Residences of surviving children 1.5 Residences of surviving siblings
2. Counting Rule Weight	A weight assigned to every household for every death it is eligible to report	2.1 Inverse of the number of households eligible to report the death 2.2 Fraction of the eligible relatives residing in the household
3. Data Collection Method	The method of querying the households in the survey	3.1 Mail survey 3.2 Personal interview survey
4. Respondent Rule	Defines the persons that are eligible to respond in the survey	4.1 Related adults are eligible to respond for one other 4.2 Adults are eligible to respond for themselves
5. Length of Reference Period	The elapsed time between the date of the person's death and date the household eligible to report the death is surveyed	5.1 Within 6 months 5.2 Within 9 months 5.3 Within 12 months

(See the appendix for the formulation of the estimator and the derivation of the unbiasedness conditions.) The information needed to calculate the counting rule weights is obtained in the survey from the household that reports the decedent.

We refer to the counting rule options 2.1 and 2.2 that are listed in Exhibit 2 as the unit and the element counting rule weights respectively. Although both options are shown in the appendix to satisfy the unbiasedness conditions, the questions asked in the survey to obtain the information needed to calculate them is somewhat different. They have one question in common, namely

"How many [specified relatives] does the decedent have?"

The unit weight requires two additional questions:

"What are their names?"

"Which of them are living together?"

The element weight requires only one question in addition to the common question:

"How many of the [specified relatives] are living in this household?"

Our experience has been that it takes less time and effort to collect the information for the element than for the unit weight.

3. ERROR EFFECTS OF DESIGN FACTORS

For every design option listed in Exhibit 2, the experiment investigated their separate and combined effects on the five types of errors listed and defined in Exhibit 3. In addition to sampling errors, four types of bias errors were investigated in the experiment. (The stage of the survey experiment that measured these errors is shown in the right hand column of Exhibit 1.) It is noteworthy that many types of errors were not measured at all by the experiment. Excluded, for example, were nonsampling variance and bias errors due to erroneously enumerating or matching deaths.

Design factors are selective in their error effects. Exhibit 4 identifies the types of errors that are affected by each of the design factors. Thus, the counting rule is the only factor that affects all five types of errors. Counting rule weights affects four types of errors. Each of the other design factors affect three types of errors, but only two of them affect the same types of errors.

The error effects of the design factors are not independent. For instance, the effect of a self respondent rule may be quite different when combined with a conventional counting rule than when combined with a counting rule that links deaths to the households of surviving relatives. The findings of the experiment will make it possible to compare the error effects of about 500 different option sets for the five design factors.

4. CONCLUDING REMARKS

We have formulated a strategy for designing efficient data systems, which involves selecting the set of design options that minimizes the sampling and measurement errors. Somewhat similar strategies have been proposed by Dalenius [1] and Nathan [2]. To implement the strategy proposed here requires a matrix of information on the cost and sampling and nonsampling error effects of design factors and their options. We have proposed a partial structure of the design matrix for retrospective mortality surveys and have described a survey experiment that was conducted to measure some of the sampling and nonsampling error effects of selected options for a few design factors. We have implied some of the cost effects in terms of the supplementary information required by some of the options. The remaining structure of this design matrix needs to be defined, and more experiments need to be conducted to compile information for the additional option sets.

Although the design matrix for retrospective mortality surveys may be applicable to some other types of retrospective surveys, it is not likely to be applicable to most data systems. It is timely to begin to formulate the design matrices for different types of data systems, and to construct these matrices from information that is already available or by designing the necessary experiments. If nature is kind, we will discover some generalities in the cost and error structures so that by means of a relatively small number of design matrices of reasonable size we will be able to handle many different types of data systems.

ACKNOWLEDGEMENTS

The survey experiment is being conducted for the National Center for Health Statistics by the Research Triangle Institute.

References

[1] Dalenius, Tore, "Designing Descriptive Sample Surveys" in New Developments in Survey Sampling, Wiley-Interscience 1969, pp. 390-416.

[2] Nathan, Gad, "Utilization of Information on Sampling and Nonsampling Errors for Survey Design - Experiences in Israel. Bulletin of the International Statistical Institute, Vienna, 1973.

[3] Nathan, Gad. "Evaluation of Different Counting Rules and Weighting Procedures for Surveys with Multiplicity." Paper presented at the 1976 Meetings.

[4] Sirken, Monroe G., "Design of Household Sample Surveys to Test Death Registration Completeness," Demography, August 1973, Vol. 10, No. 3, pp. 469-478.

Exhibit 3.

Types of Errors and Their Definitions

Type of Error	Definitions
A. Sampling errors	Errors resulting from the sample selection procedure
B. Counting rule bias	Deaths not enumerated because they are not linked to any household by the counting rule
C. Nonresponse bias	Deaths not enumerated due to household nonresponse in the survey
D. Response bias	Deaths not enumerated because the responding households fail to report eligible deaths
E. Matching bias	Survey deaths that fail to match with their death certificates

Exhibit 4.

Error Effects of Design Factors

Design Factor	Type of Error				
	A Sampling error	B Counting rule bias	C Nonresponse bias	D Response bias	E Matching bias
1. Counting rule					
2. Counting rule weight					
3. Data collection method					
4. Respondent rule					
5. Length of the reference period					

[5] Sirken, Monroe G., "The Counting Rule Strategy in Sample Surveys," American Statistical Association Proceedings of the Social Statistics Section, 1974, pp. 119-123.

[6] Sirken, Monroe G., "Dual System Estimators Based on Multiplicity Surveys," Developments in Dual System Estimation of Population Size and Growth, The University of Alberta Press, 1976, in print.

APPENDIX. DERIVATION OF THE ELEMENT AND THE UNIT COUNTING RULE WEIGHTS

Let

D_α ($\alpha = 1, \ldots, N$) represent the N deaths in the population Π

H_i ($i = 1, \ldots, M$) represent the M households in the sampling frame

A survey is conducted based on a sample of H_{i_j} ($j = 1, \ldots, m$) housing units to estimate N. Deaths are enumerated at the m sample households in compliance with a counting rule adopted in the survey.

The counting rule link between a decedent and the household eligible to report him is represented by the indicator variable

$$\delta_{a,i} = \begin{cases} 1 \text{ if a relative of } D_\alpha \text{ resides at } H_i \\ 0 \text{ otherwise.} \end{cases}$$

Thus,

$$S_\alpha = \sum_{i=1}^{M} \delta_{ai} = \text{number of households containing relatives of } D_\alpha.$$

The linear estimate of N,

$$\hat{N} = \frac{M}{m} \sum_{\alpha=1}^{N} \sum_{j=1}^{m} W_{\alpha,i_j} \delta_{\alpha i_j}$$

is unbiased if and only if

$$\sum_{i=1}^{M} \delta_{\alpha i} W_{\alpha i} = 1 \ (\alpha = 1, \ldots, N).$$

The $W_{\alpha i}$'s are the counting rule weights.

If the weights are assigned such that $W_{\alpha i} = W_\alpha$, the unbiasedness condition becomes

$$W_\alpha = \frac{1}{\sum_{i=1}^{M} \delta_{\alpha i}} = \frac{1}{S_\alpha}.$$

The unbiasedness condition is also satisfied if

$$W_{\alpha,i} = \frac{R_{\alpha i}}{R_\alpha}$$

where

$R_{\alpha i}$ = number of D_α's relatives residing in H_i

$R_\alpha = \sum_{i=1}^{M} R_{\alpha i}$ = number of D_α's relatives.

The W_α's and $W_{\alpha i}$'s are referred to as the unit and the element counting rule weights respectively.

THE CONDITION OF THE DETERMINANT SURFACE IN THE SPACE OF THE EXPERIMENTAL DESIGN LEVELS AS A FUNCTION OF THE INITIAL PARAMETER "GUESTIMATES" IN NONLINEAR PARAMETER ESTIMATION

Dwight E. Spence, University of Pennsylvania

Introduction: Professors G.E.P. Box and H.L. Lucas (1959) considered the matter of selecting those factor levels (ξ) which would result in the estimation of a given nonlinear (w.r.t. parameters) model whose parameter estimates (θ) would have the property of smallest variance. For least squares estimates of these parameters the criterium would entail minimizing the absolute value of the determinant Cov(θ), which is approximately the generalized variance. The Variance-Covariance matrix cov(θ) is proportional to $(F'F)^{-1}$ where F is the pseudo-design matrix (Gallant, 1975) and is a function of both the parameters (θ) and the factor levels (ξ). Used to explain this criterium are the relationships between the parameter space, the sample space and the solution locus, which constitutes a subset of sample space, as follows. For P dimensional points θ in parameter space, P(θ) defines a transformation to N dimensional sample space constituting a subset of sample space, i.e., the solution locus. Using an orthogonal transformation of sample space coordinates, solution locus points in the near neighborhood of a point P(θ) may be reduced to P dimensionality. Whe P(θ)=P(θ_0), θ_0 being the true and unknown values, the determinant $|(F'F)^{-1}|$ equals the square of the Jacobian of the transformation of these reduced-dimension points (near P(θ)) to image points in parameter space. Thus, minimization of the absolute value of $|(F'F)^{-1}|$ implies minimization of the volume covered by these points mapped to parameter space.

Determinant Surface:
The primary illustration of the above involves a two stage chemical reaction: A→B→C. Box and Lucas concern themselves with the model $\eta = \theta_1/(\theta_1-\theta_2)\{e^{-\theta_2\xi} - e^{-\theta_1\xi}\}$ where η is the percentage yield of substance B. Given the restriction that the number of data points to be selected equal the number of parameters to be estimated, note that the problem of minimizing the absolute value of $|(F'F)^{-1}|$ reduces to maximizing the absolute value of $|F|$ which is of course still a function of both parameters (θ) and the factor levels (ξ). To proceed toward the selection of the desired factor levels initial estimates of the parameters must be chosen. This brings us to the focal question on which the present paper is based: What are the consequences of choosing one set of parameter estimates (θ^*) over one of the many other possible sets? In this particular example the parameters are interpreted as indicating the rates of chemical reactions. Often there may not be even this much information on which to base the "questimates" (θ^*). The initial estimates used by Box and Lucas are θ_1^* =.7 and θ_2^*=.2. Utilizing a computer and an extensive grid search technique, we can plot the contours of the response surface of $|F^*|$ in the space of the factor levels, i.e., ξ_1 and ξ_2. Figure A4 shows, as did Box and Lucas, that the result of grid search is a multi-extremum surface with stationary points at $|F^*| \doteq .039$ and $|F^*| \doteq -.81$. If instead we consider θ_1^*=.2 and θ_2^*=.7, we get through grid search the contour surface indicated in Figure B4. For this case and in the region considered here, the determinant response surface is simpler having only one stationary point, a valley at $|F^*| \doteq -.1176$. In these two cases the consequent factor levels, the indicated numerical values of ξ_1 and ξ_2, differed fairly negligibly as we will speak more about later. If we go on to consider initial parameter estimates of θ_1^*=1.0 and θ_2^*= 0.0, figure C4 shows that the result is a rising ridge type surface indicating that, within the region under consideration, the absolute value of $|F^*|$ is maximized at $|F^*|$ = -12.0. Finally, figure D4 displays the results of using θ_1^* =.5 and θ_2^* =.4 as the starting estimates; and again we find a multi-extremum surface with a peak at $|F^*|$ =.000093 and a valley at $|F^*|$ = -.3770. Evidently, reasonably small changes in the initial estimates of the parameters result in quite a variety of determinant surfaces.

Experimental Design:
Should circumstances require that one attempt to locate these experimental design points, i.e., ξ_1 and ξ_2 which optimize $|F^*|$, by the use of numerical methods (Box & Wilson, 1951) such as steepest descent (ascent), different results can be expected depending on the nature of the particular determinant surface. Figures A3, B3, C3, and D3, reveal the approximated determinant contour surfaces for each of the four sets of initial estimates and based on quadratic equations whose data points are presented in Table 1. (Note how rapidly the approximation deteriorates as we move away from the center of the selected data points.) Table 1 also indicates the four sets of factor level coordinates, ξ_1 and ξ_2, obtained as a result of calculating the stationary points of the corresponding quadratic equations. With respect to θ_1^*=1.0 and θ_2^*= 0.0 the values of ξ_1 and ξ_2 are dictated in

part by the limits of the experimental region since the quadratic equation in this case indicates a stationary point far outside the experimental region considered here. The graphs shown in figures B_1, B_2, B_3, and B_4 (as well as those in figures A_1, A_2, A_3, and A_4) when available provide important information concerning the selection of the data points on which the quadratic equations are generated. In each case of graphs of type B, graphs in the space of the partial derivatives ($D_2\eta$), the objective is to determine the triangular simplex of greatest volume such that one vertex is at the origion; the remaining two vertices indicate values of ξ_1 and ξ_2 which define a point very near (or, ideally, exactly equal to) the desired stationary point. This information would thus eliminate the need for most all of the steepest descent (ascent) methodolgy.

Finally, Table 1 (Bottom) gives the results of translating and orthogonally rotating the x_1 and x_2 axes by way of canonical reduction of selected quadratic equations (Bradley, 1958). The x_1 and x_2 coordinate values are simple functions of ξ_1 and ξ_2 as also indicated in Table 1. The graph of figure A_3 illustrates these results with canonical axes X_1 and X_2 centered at the stationary point.

REFERENCES

Box, G.E.P. and Lucas, H.L. (1959): "Design of Experiments in Nonlinear Situations," Biometrika, 46, 77-90.

Box, G.E.P. and Wilson, K.B. (1951): "On the Experimental Attainment of Optimum Conditions," Journal of the Royal Statistical Society, Series B, Vol. XIII, No. 1.

Bradley, Ralph A. (1958): "Determination of Optimum Operating Conditions by Experimental Design," Industrial Quality Control, 15, 16-20.

Gallant, A.R. (1975): "Nonlinear Regression," American Statistican, Vol. 29, 73-81.

TABLE 1

The MODEL, under the assumption of normality:

$$\eta = \theta_1/(\theta_1-\theta_2)\{\exp(-\theta_2\xi)-\exp(-\theta_1\xi)\}$$

$$\eta_i^* = \eta_i + \epsilon_i,\ E(\eta_i^*) = \eta_i \text{ and } E(\epsilon_i\epsilon_j) = \begin{cases}\sigma^2 & i=j\\ 0 & i\neq j\end{cases}$$

$$\phi_1 = \partial\eta/\partial\theta_1 = [1/\theta_1 - 1/(\theta_1-\theta_2)][\theta_1/(\theta_1-\theta_2)]\exp(-\theta_2\xi) - [1/\theta_1 - 1/(\theta_1-\theta_2) - \xi][\theta_1/(\theta_1-\theta_2)]\exp(-\theta_1\xi)$$

$$\phi_2 = \partial\eta/\partial\theta_2 = [1/(\theta_1-\theta_2) - \xi][\theta_1/(\theta_1-\theta_2)]\exp(-\theta_2\xi) - [1/(\theta_1-\theta_2)][\theta_1/(\theta_1-\theta_2)]\exp(-\theta_1\xi)$$

	A	B	C	D
INITIAL PARAM. ESTIMATES:	$\theta_1^*=.7, \theta_2^*=.2$	$\theta_1^*=.2, \theta_2^*=.7$	$\theta_1^*=1.0, \theta_2^*=0.0$	$\theta_1^*=.5, \theta_2^*=.4$
OPTIMAL LEVELS INDICATED.	$\xi_1=1.23, \xi_2=6.86$	$\xi_1=1.61, \xi_2=5.93$	$\xi_1=1.0, \xi_2=36.0$	$\xi_1=1.37, \xi_2=5.32$

A. The following data points

ξ_1	ξ_2	Δ
1.10	6.5	-.803 336
1.20	6.5	-.807 992
1.30	6.5	-.806 709
1.10	7.0	-.804 758
1.10	7.5	-.798 920
1.20	7.0	-.809 632
1.23*	6.87*	-.810 413*

$x_1 = 5(\xi_1 - 1.1)$
$x_2 = (\xi_2 - 6.5)$

resulted in a Quadratic equation of the form

$$\Delta = -.8033 - .01531x_1 - .01013x_2 + .01195x_1^2 + .01454x_2^2 - .00067x_1x_2$$

B. The following data points

ξ_1	ξ_2	Δ
1.6	4.5	-.108 211
1.7	4.5	-.106 370
1.8	4.5	-.104 087
1.6	5.5	-.117 023
1.7	5.5	-.116 342
1.6	6.5	-.116 487
1.65*	6.0*	-.117 618*

$x_1 = 5(\xi_1 - 1.6)$
$x_2 = (\xi_2 - 5.5)$

resulted in a Quadratic equation of the form

$$\Delta = -.11715 - .00033x_1 - .00415x_2 + .00519x_1^2 + .00491x_2^2 - .0009x_1x_2$$

C.

The following data points

ξ_1	ξ_2	Δ
.8	25.0	-8.62712
.9	25.0	-8.78190
1.0	25.0	-8.82911
.8	30.0	-10.42443
.9	30.0	-10.61147
.8	35.0	-12.22175
1.0*	35.0*	-12.50790*

$x_1=5(\xi_1-.8)$
$x_2=(\xi_2-25.0)/10$

resulted in a Quadratic equation of the form

$$\Delta=-8.62487-.43559x_1-3.61406x_2+.2331x_1^2+.01793x_2^2-.08416x_1x_2$$

D

The following data points

ξ_1	ξ_2	Δ
1.3	4.5	-.363 893
1.4	4.5	-.362 252
1.5	4.5	-.358 417
1.3	5.0	-.375 072
1.4	5.0	-.374 863
1.3	5.5	-.375 811
1.39*	5.34*	-.377 209*

$x_1=5(\xi_1-1.3)$
$x_2=(\xi_2-4.5)$

resulted in a Quadratic equation of the form

$$\Delta=-.36387+.00089x_1-.03315x_2+.00455x_1^2+.02124x_2^2-.0049x_1x_2\ .$$

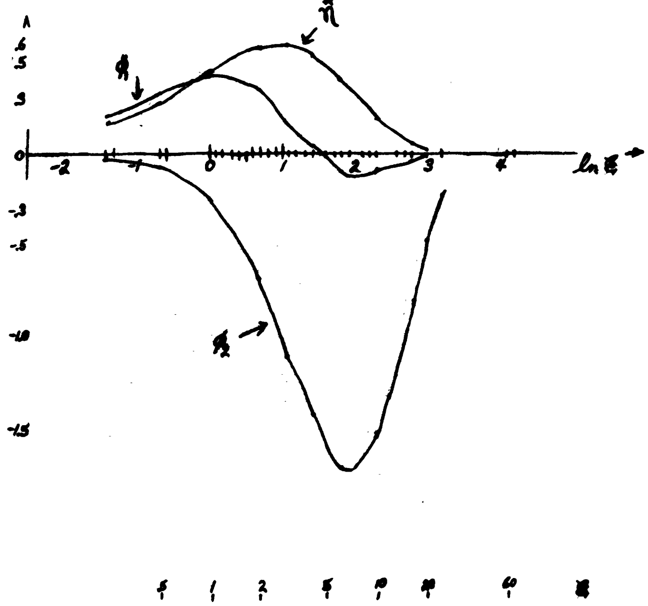
Reduction of the above to canonical form gives the following:

	A	B	C	D
Eigenvalues:	$B_1=.0146$	$B_1=.0055$?	$B_1=2.410$	$B_1=.0216$
	$B_2=.0119$	$B_2=.0046$	$B_2=.0099$	$B_2=.0042$
Eigenvectors:	$\begin{bmatrix} .1262 \\ -.9920 \end{bmatrix}_1$		$\begin{bmatrix} .9827 \\ -.1853 \end{bmatrix}$	
	$\begin{bmatrix} .9920 \\ .1262 \end{bmatrix}_2$		$\begin{bmatrix} .1853 \\ .9827 \end{bmatrix}$	

*Asterik values are the coordinates of the stationary point based on the quadratic equation generated from the initial six data points.

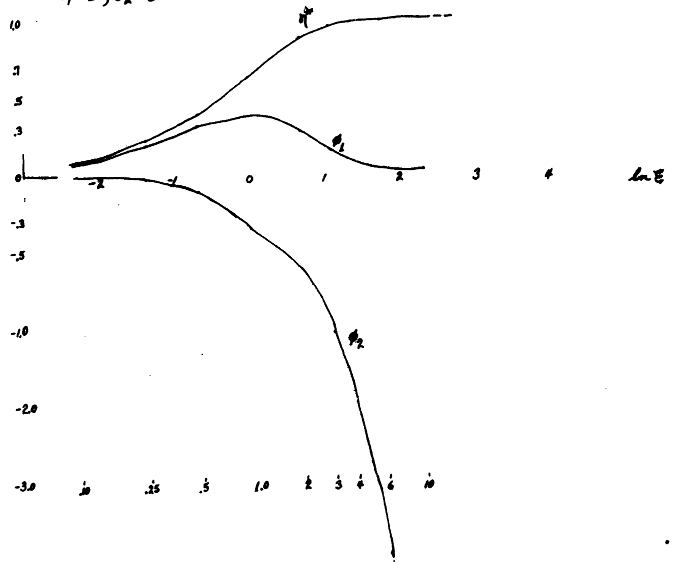
A_1

$\theta_1^* = .7,\ \theta_2^* = .2$



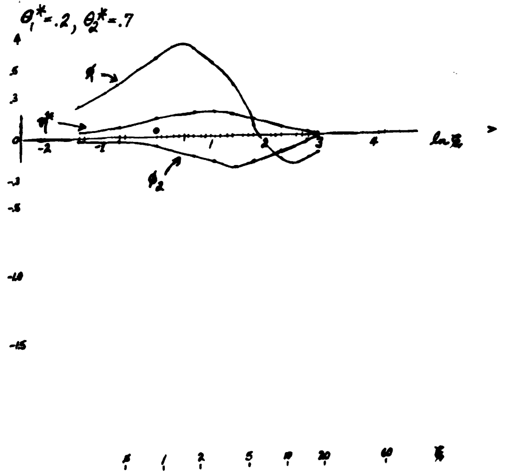
C_1

$\theta_1^* = 1,\ \theta_2^* = 0$



B_1

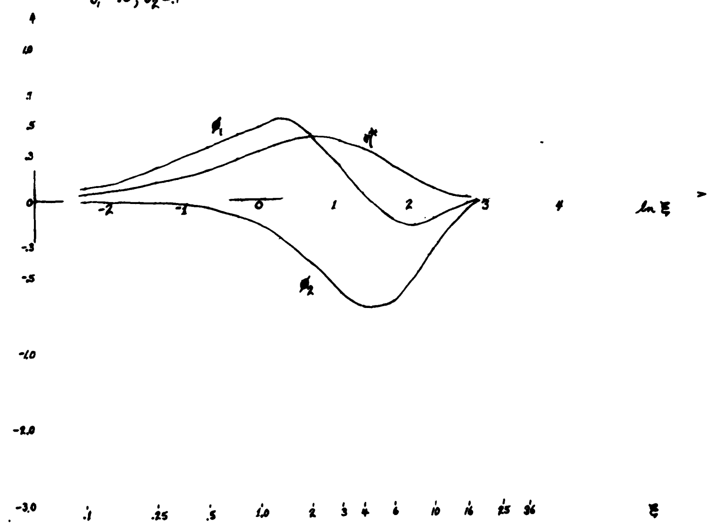
$\theta_1^* = .2,\ \theta_2^* = .7$



$$\hat{\eta} = \frac{\theta^*}{\theta_1^* - \theta_2^*}\left\{e^{-\theta_2^*\xi} - e^{-\theta_1^*\xi}\right\}$$

D_1

$\theta_1^* = .5,\ \theta_2^* = .4$



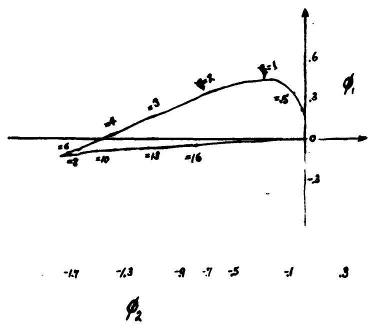
A₂

7, $\theta_2^* = .2$

Design Locus:

$\phi_1(\xi) = [.8 + 1.4\xi]e^{-.7\xi} - .8e^{-.2\xi}$

$\phi_2(\xi) = [2.8 - 1.4\xi]e^{-.2\xi} - 2.8e^{-.7\xi}$

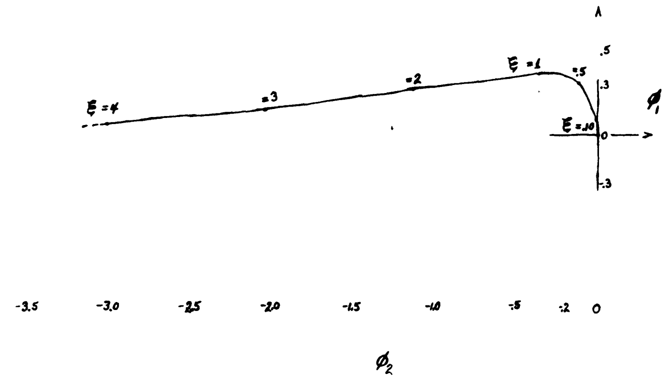


C₂

$\theta_1^* = 1$, $\theta_2^* = 0$

$\phi_1(\xi) = \xi e^{-\xi}$

$\phi_2(\xi) = 1 - \xi - e^{-\xi}$



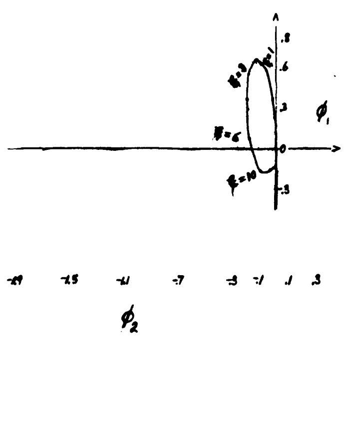
B 2

$\theta_1^* = .2$, $\theta_2^* = .7$

Design Locus:

$\phi_1(\xi) = [2.8 - .4\xi]e^{-.2\xi} - 2.8e^{-.7\xi}$

$\phi_2(\xi) = [.8 + .4\xi]e^{-.7\xi} - .8e^{-.2\xi}$

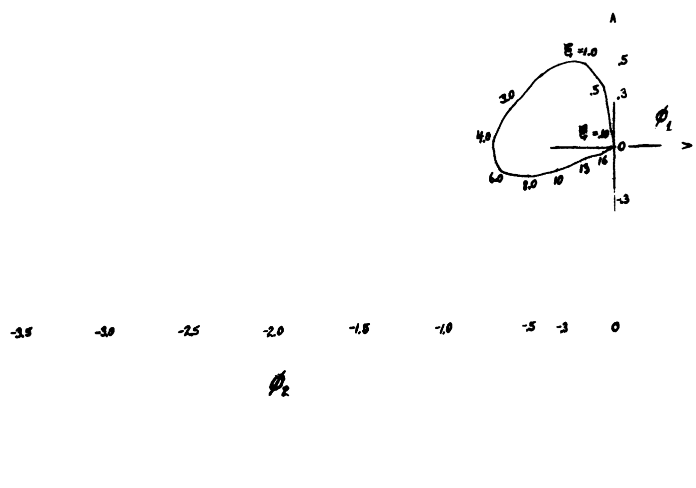


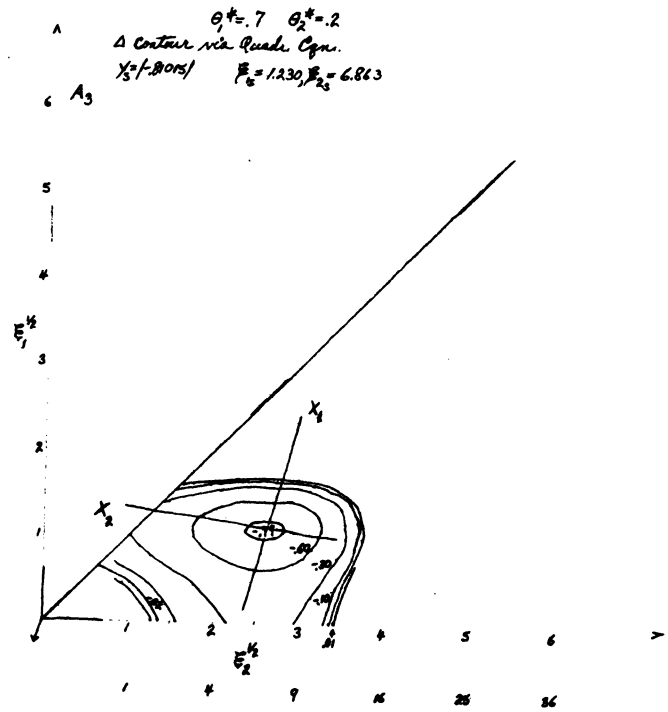
D₂

$\theta_1^* = .5$, $\theta_2^* = .4$

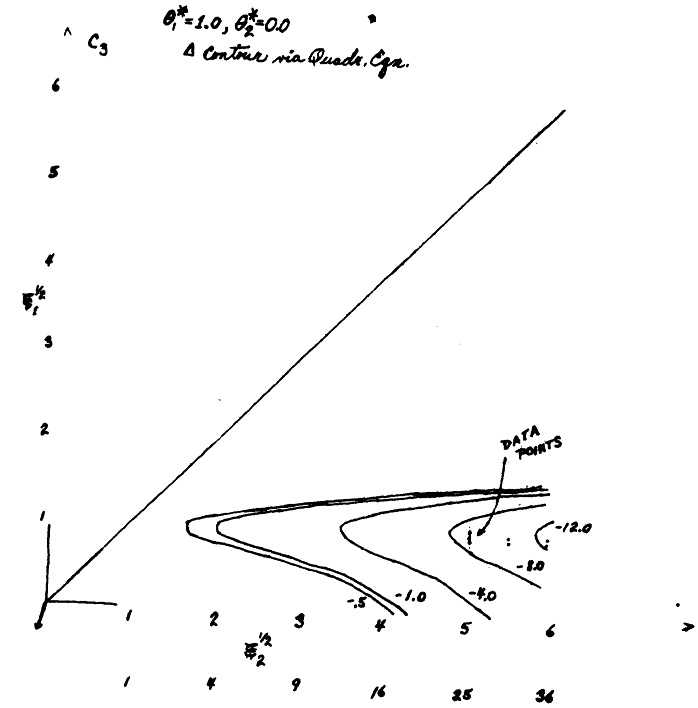
$\phi_1 = (40 + 5\xi)e^{-.5\xi} - 40e^{-.4\xi}$

$\phi_2 = (50 - 5\xi)e^{-.4\xi} - 50e^{-.5\xi}$

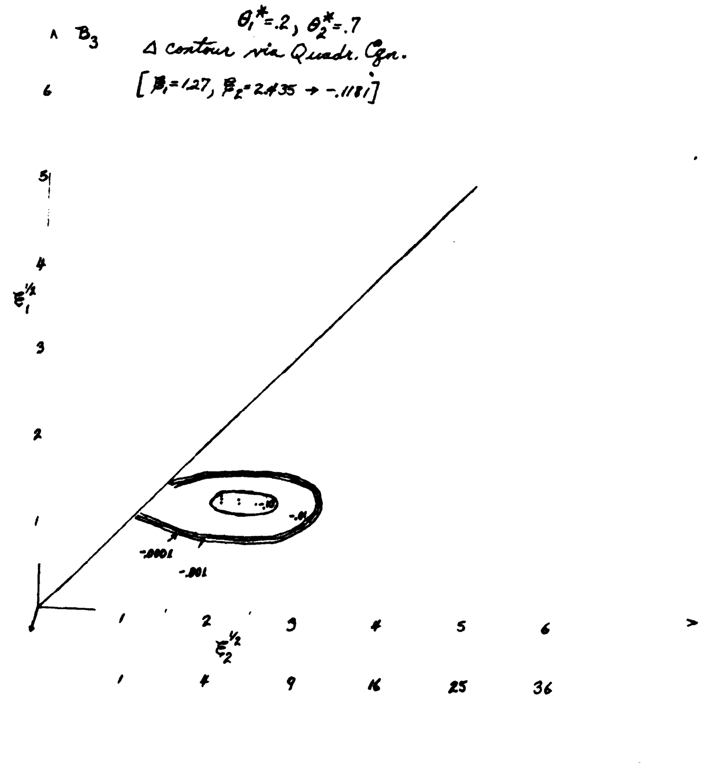




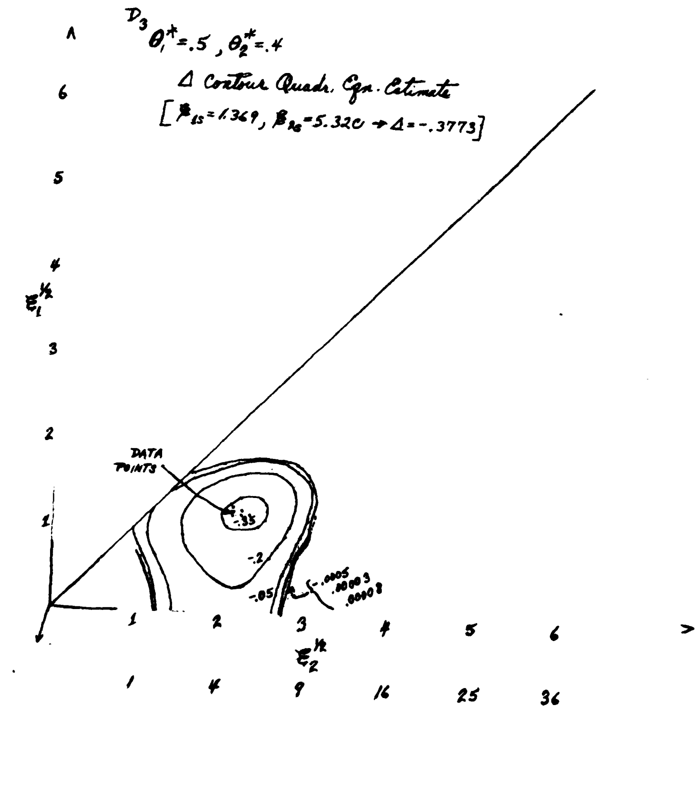
A3
$\theta_1^* = .7 \quad \theta_2^* = .2$
Δ contour via Quadr. Eqn.
$\xi_1^{1/2}$
$\xi_2^{1/2}$
X_1
X_2



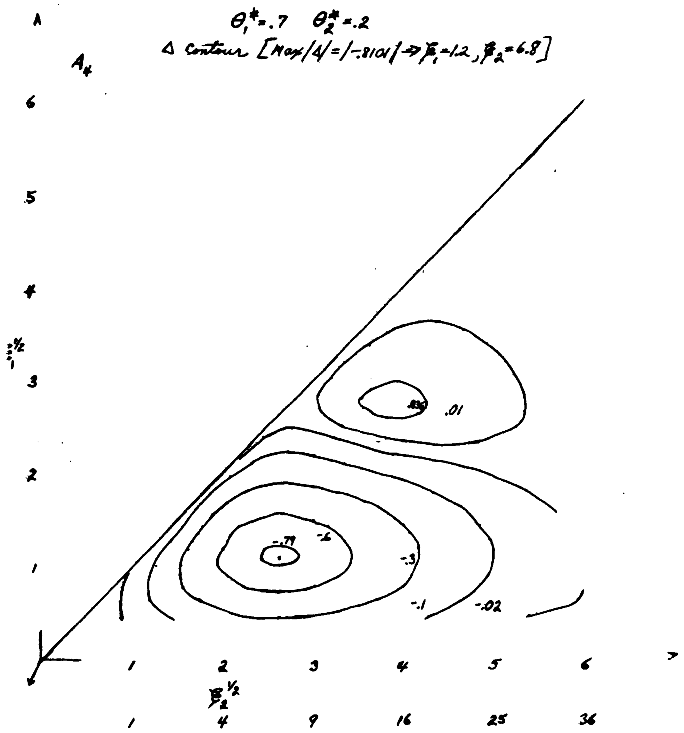
C3
$\theta_1^* = 1.0,\ \theta_2^* = 0.0$
Δ Contour via Quadr. Eqn.
DATA POINTS
-12.0
-8.0
-4.0
-1.0
-.5
$\xi_1^{1/2}$
$\xi_2^{1/2}$



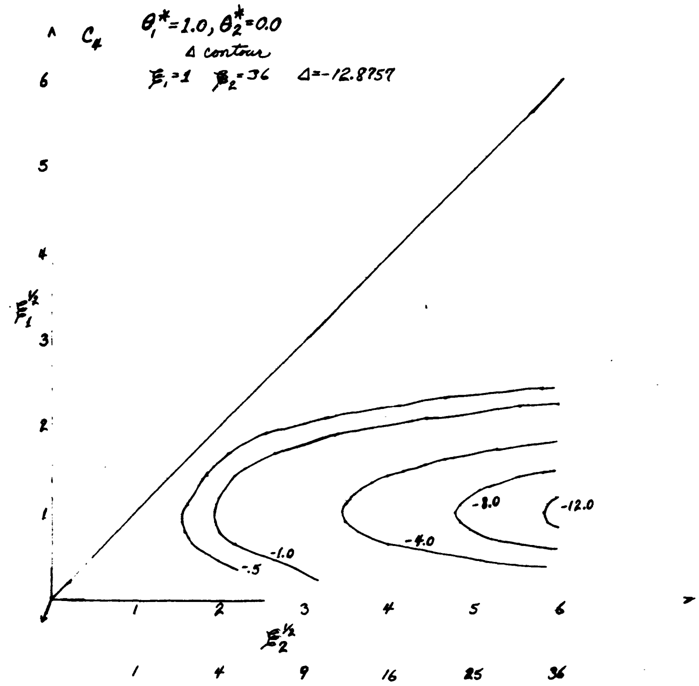
B3
$\theta_1^* = .2,\ \theta_2^* = .7$
Δ contour via Quadr. Eqn.
-.0001
-.001
$\xi_1^{1/2}$
$\xi_2^{1/2}$



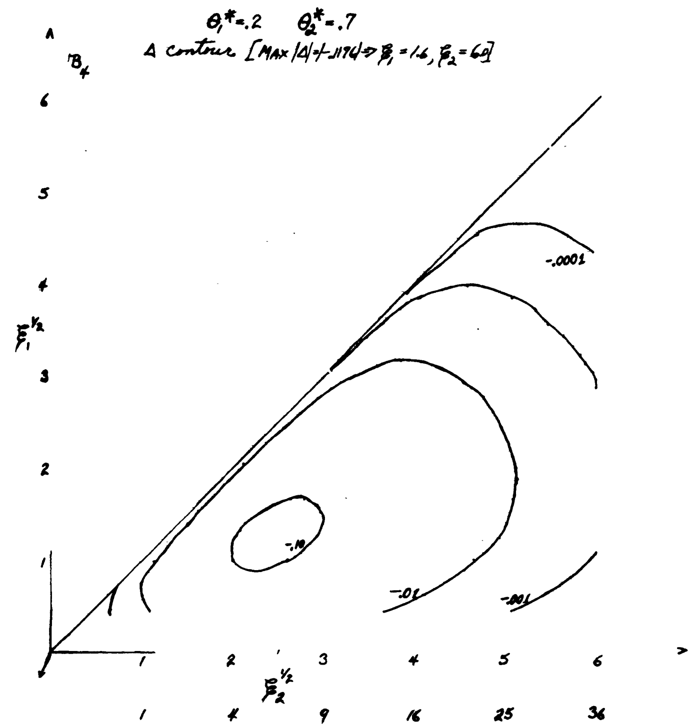
D3
$\theta_1^* = .5,\ \theta_2^* = .4$
Δ Contour Quadr. Eqn. Estimate
$[\beta_{13} = 1.369,\ \beta_{23} = 5.320 \rightarrow \Delta = -.3773]$
DATA POINTS
-.35
-.2
-.05
$\xi_1^{1/2}$
$\xi_2^{1/2}$



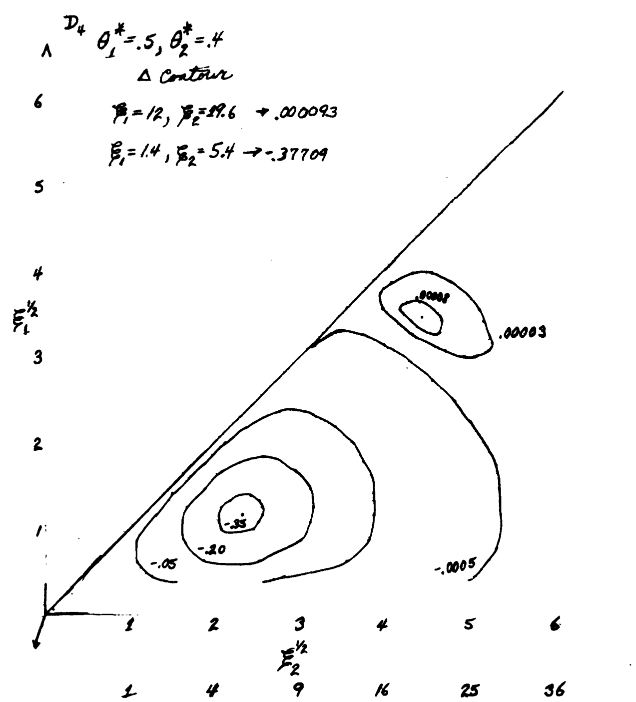
A_4
$\theta_1^* = .7 \quad \theta_2^* = .2$
Δ contour [Max |Δ| = |-.8101| → ξ_1 = 1.2, ξ_2 = 6.8]
.01
-.79
-.6
-.3
-.1
-.02
$\xi_1^{1/2}$
$\xi_2^{1/2}$



C_4
$\theta_1^* = 1.0, \theta_2^* = 0.0$
Δ contour
ξ_1 = 1 ξ_2 = 36 Δ = -12.8757
-8.0
-12.0
-4.0
-1.0
-.5
$\xi_1^{1/2}$
$\xi_2^{1/2}$



B_4
$\theta_1^* = .2 \quad \theta_2^* = .7$
Δ contour [MAX |Δ| = -.1194 → ξ_1 = 1.6, ξ_2 = 6.0]
-.0001
-.10
-.01
-.001
$\xi_1^{1/2}$
$\xi_2^{1/2}$



D_4
$\theta_1^* = .5, \theta_2^* = .4$
Δ contour
ξ_1 = 12, ξ_2 = 19.6 → .000093
ξ_1 = 1.4, ξ_2 = 5.4 → -.37709
.00008
.00003
-.35
-.20
-.05
-.0005
$\xi_1^{1/2}$
$\xi_2^{1/2}$

THE RELATIONSHIP BETWEEN VARIOUS INTERVIEWER CHARACTERISTICS AND THE COLLECTION OF INCOME DATA

Joyce A. Stevens and Barbara A. Bailar, U.S. Bureau of the Census*

Introduction

In a supplement to the Current Population Survey (CPS), income data are collected each March. Some interviewers are more successful than others in eliciting responses to income questions. This study is a preliminary attempt to determine if specific interviewer characteristics or attitudes are related to the successful collection of income data in the CPS March Supplement.

Methodology

The best method to have used to develop relationships between interviewer characteristics and the successful collection of income data would have been by means of a controlled experiment in which households would have been randomly assigned to interviewers of different types. It was not possible to do this kind of experimentation in the CPS; therefore, a different kind of project was carried out.

Every CPS interviewer who had worked on the March 1975 CPS Supplement was sent a questionnaire. The questions asked included the interviewer's age, educational level, and family income; whether the interviewer was more at ease about asking certain types of income questions; whether the respondents were more cooperative about answering certain types of income questions; whether the respondents understood the different types of income questions; whether the respondents had the information available to answer the income questions; and whether the interviewers felt it was appropriate for the Census Bureau to ask about certain subjects. The results from this questionnaire were linked to selected items from the March 1975 CPS Income Supplement. These items included nonresponse rates for different income questions, measures of nonprobing for income data, and some characteristics of the interviewer's assignment area. Thus, the new record created for each interviewer contained the answers to the questions about the interviewer's characteristics, attitudes, and opinions, and the nonresponse rates and nonprobing rates that the interviewer achieved on the March 1975 Income Supplement.

Two measures of the interviewer's "success" in collecting income data are used in this analysis. They are the nonresponse rates by type of income, for households and for persons, and the nonprobing rates by type of income for persons. Nonprobing for Social Security and interest income is said to have occurred when the annual dollar amount reported is exactly divisible by four. Should an interviewer accept a quarterly (or monthly) dollar amount and multiply by four (or 12) to arrive at an annual figure instead of probing to determine if the same amount of money were received in each quarter (or month), an incorrect annual total would result. For example, there was a change in the Social Security benefit payment in July, 1975. If the respondent multiplied his most recent payment by 12 to arrive at an annual figure, the annual amount would be incorrect. Since a generalized technique for determining nonprobing for all kinds of income items was desirable, divisibility by four was used as the criterion for nonprobing.

There are certain limitations to this analysis. First the nonresponse and nonprobing rates do not relate directly to the quality of the income data collected. For instance, an interviewer could have no nonresponses but the reported amounts could be inaccurate. Second, since a randomized controlled study was not the method used, no causal relationships can be inferred. All that can be asserted is that relationships between variables do or do not exist. Third, when an interviewer was asked her opinion or attitude about an event pertaining to all respondents, a particularly unusual occurrence could influence her answer. For example, the memory of a particularly unpleasant refusal may have influenced the interviewer to report that more people were uncooperative in 1975 than in the past when that may not have been the case.

The universe for this study consists of all of the CPS interviewers who worked on the March 1975 Income Supplement and who contacted at least 10 households. However, some interviewers did not return a questionnaire and some who returned one did not sign their names. About 85 percent of the CPS interviewers who worked on the March Supplement completed the attitude questionnaire. About 81 percent signed their names. Because it was necessary to have their names for matching the interviewer data from the two sources, 81 percent was the resultant success rate. Thus, 802 interviewers had matched data available. This group is regarded as the population and no sampling errors are calculated since all differences are regarded as differences occurring in the population.

Since it is possible that those interviewers who refused to fill out or sign the questionnaire were different from those who did, a comparison of the two groups was made for selected items of information from the March Supplement. Table 1 shows this comparison.

Very few differences are apparent. The nonrespondents may have had somewhat more black households, more Spanish-speaking households, and more elderly households. They had somewhat more households not reporting dividends or interest. However, the assignment sizes were about the same, or somewhat smaller and the percentage of self-respondents was about the same. Of course, a similarity between those who cooperated and those who did not on this set of items does not mean that there was a similarity for other types of items that are presented in this paper.

Attitude Survey - Capsule of Findings

The interviewer survey indicated that the "average" CPS interviewer was a 48-year old married woman who was a high school graduate and had a family income of about $17,000. About 80 percent of the interviewers had worked on the March Supplement at least twice. Those interviewers who had previous experience with the March Supplement were asked if they noticed any change in respondent attitudes towards asking income questions in the past year or two. Just under a half said they found respondents less cooperative, nearly two-fifths found respondents about the same, and approximately one-tenth found respondents more cooperative. (Fewer than one out of every 20 interviewers mentioned a difference but did not say in which direction.)

March 1975 CPS - Capsule of Findings

Item by item nonresponse rates were tabulated for two different universes, for persons and for households, and for three categories of income: interest income, Social Security income, and income other than wages and salary. Among the major items included in income other than wages and salary are Social Security, Railroad Retirement, Supplemental Security Income, interest, dividends, royalties, estates or trusts, rental income, self-employment income, farm income, welfare, unemployment, and alimony. For the sake of brevity, income other than wages and salary will be referred to as "other income" and "Social Security income" will be used to mean Social Security, Supplemental Security, and Railroad Retirement income, unless otherwise indicated. Similarly, "interest" will refer to interest, dividends, royalties, estates or trusts, and rental income.

Person nonresponse rates varied only from 7.8 percent for those who had Social Security income but gave no amount, to 9.9 percent for persons with interest but with one or more interest items not reported. The percentage of households with one or more "other income" items not reported was 16.2 percent. (These nonresponse rates do not include noninterview cases for which there was no response to the entire questionnaire.) Although there are other causes or factors contributing to nonresponse, for this analysis, it is assumed that the level of the nonresponse rate is a function of the interviewer.

Another set of items which is assumed to be a function of the interviewer is the nonprobing rate. The nonprobing rate varied from 9.6 percent for Social Security items to 12.3 percent for interest items.

Matched File (Attitudes & March Supplement) - Findings

Table 2 shows the average nonresponse and nonprobing rates for selected income items by interviewer characteristics. Except for the group of interviewers 55 years of age or older, no apparent relationship between nonresponse or nonprobing and age exists. For this one age group, there seems to be a general tendency for lower nonresponse and nonprobing rates.

When the nonresponse and nonprobing rates were compared for interviewers with different levels of educational attainment, it was found that those interviewers who had completed 4 or more years of college had higher rates than interviewers with less than a college degree.

Several interesting relationships appear when the interviewer's family income level is crossed separately by measures of nonresponse and nonprobing, and by the characteristics of the area in which the interviewer worked. High income interviewers (interviewers whose family income was $25,000 or more) had the highest nonresponse and nonprobing rates of all interviewers. This finding may be correlated with the high nonresponse and nonprobing rates associated with highly educated interviewers. The higher nonresponse rates for high income interviewers cut across all kinds of income and, presumably, across all respondent income levels. For instance, the high income interviewers had the highest nonresponse rates for household other income, for persons with pensions (private, military, State and local government pensions; alimony, child support, or other regular contributions) and for persons on welfare (AFDC). The nonresponse rates appear to be a function of the interviewer, and not the area. While the high income interviewers had high nonresponse rates for welfare income, their total household assignments had low percentages of households in poverty and households with Negro heads, as can be seen in Table 2.

Multiple sources of income, such as interest, dividends, and royalties, are associated with high income areas. An interviewer who works in an area with a high percentage of multiple sources of income has more chances to make errors than the interviewer who works in an area where one source of income predominates. As shown in Table 2, the high income interviewers worked in areas with the highest mean income and with the highest percentage of respondents who had multiple sources of income. It is possible, then, that the high income interviewers had high nonresponse rates because of the many chances to make errors.

An experienced interviewer is generally believed to collect better data. To see if such a relationship exists for income data, the number of times an interviewer worked on the March Supplement was crossed by her nonresponse and nonprobing rates from the 1975 March CPS. The interviewers who worked on the March CPS for the first time in 1975 had the lowest nonresponse rates. The interviewers who had interviewed on three March Supplements had the highest nonresponse rates. Those who had worked on the March CPS the longest (four or more times) had rates similar to those who had interviewed twice on the March Supplement. These latter nonresponse rates were slightly lower than those for interviewers who had interviewed three times. Length of experience did not seem to be associated with nonprobing.

It is possible that the first time interviewers being freshly trained were full of zeal and paid particular attention to detail. Those who had worked on four or more CPS Supplements may represent a mixed group: persons who were very dedicated and precise, and those who were indifferent. Further work is needed to break out the differences in the more experienced group. It appears, though, that the high nonresponse rates may again be more a function of the interviewer than of the area. The interviewers who had worked four or more times had the lowest percentage of persons with multiple sources of income. Thus, it may be that they did not ask about some sources of income.

If, in fact, a high nonresponse rate was more a function of the interviewer than of the areas in which she worked, correlations between the interviewer's attitudes and her nonresponse rates should be very informative. Because of space limitations it is not possible to analyze all of the available relationships between attitudes and nonresponse and nonprobing rates. A small subset illustrates the relationships.

Interviewers were asked if they thought it were appropriate for the Census Bureau to ask about a large number of items. Included were sources and amounts of income. Almost all of the interviewers who felt the source was appropriate also felt that the amount was appropriate to ask. The same was true for those who felt it was inappropriate, had no opinion or did not answer the question.

As is shown in Table 2, those who found asking amount or source of income inappropriate were more likely to have higher nonresponse rates. However, there was no relationship between their attitude about the appropriateness of the question and the nonprobing rates. One difficulty with this study is that there is the tendency to infer a causal relationship between the interviewers' attitude that it is not appropriate to ask income questions and the "resultant" higher nonresponse rates. However, it may have been that several refusals to those questions caused the interviewers to feel that the questions were inappropriate.

Interviewers were also asked whether they noticed if there were differences in the cooperation of respondents depending on respondent income level, and, if so, which group they found least cooperative. About 35 percent of the interviewers found the high income respondents to be least cooperative while only 8 percent found the low income group to be the least cooperative and 21 percent thought the average income group was least cooperative. About 36 percent noticed no difference or did not answer the question. Those interviewers who perceived the average income group to be the least cooperative had slightly higher nonreponse rates.

Another question asked of the interviewers was whether they noticed a change in the respondents' attitudes according to whether they were asked about certain types of income. They could answer that the respondents' attitudes were the same as towards other income questions, that respondents were more cooperative, less cooperative, or that the interviewers could not tell. Of those interviewers who could tell about a respondent's attitude, the interviewers who said respondents were less cooperative toward interest questions had the highest nonresponse rates. The same was true for questions about dividends. There appeared to be no difference in the nonprobing rates. It should be noted that the percentage of persons with one or more interest items not reported is directly related to the nonresponse rates for the separate attitude questions on interest and on dividends; and the percentage of households with one or more other income items not reported includes households with one or more interest and dividend income items not reported.

Perception of respondent attitudes was also related to the collection of Social Security income. However, the differences between groups of interviewers was not as great as those for dividends and interest income. The percentage of persons not reporting Social Security and Supplemental Security Income corresponds directly to the questions on the respondents' attitudes toward Social Security and Supplemental Security Income. The interviewers who said that respondents seemed to be less cooperative toward Social Security income questions had slightly higher nonresponse rates, but the interviewers who said that respondents seemed to be less cooperative toward Supplemental Security Income questions had slightly lower nonresponse rates. Again, there appeared to be no differences in the nonprobing rates.

This paper represents a small fraction of the data available from this project. Further exploration of the association of interviewer attitudes and characteristics, as well as area characteristics, with the quality of income data will take place over the coming year.

TABLE 1.--A COMPARISON OF SELECTED ITEMS FROM THE MARCH 1975 CPS SUPPLEMENT FOR INTERVIEWERS WHO COMPLETED A SURVEY FORM AND THOSE WHO DID NOT

Interviewer assignment characteristics	Percent of CPS interviewers	
	Who completed and signed a questionnaire (N = 802)	Who did not complete or did not sign a questionnaire (N = 189)
	100.0	100.0
Number of households in assignment:		
One to 19	3.9	3.2
20 to 39	22.8	27.5
40 to 59	59.2	54.5
60 or more	14.1	14.8
Percentage of black household heads:		
0.0 to 4.9	59.3	55.6
5.0 to 19.9	24.9	27.0
20.0 or over	15.8	17.5
Percentage of households with Spanish speaking head:		
0.0 to 4.9	72.6	69.8
5.0 to 19.9	17.9	18.5
20.0 or over	9.5	11.6
Percentage of household heads 65 or over:		
0.0 to 9.9	13.8	9.6
10.0 to 24.9	59.5	61.7
25.0 or over	26.7	28.7
Percentage of telephone interviews:		
0.0 to 9.9	9.6	7.4
10.0 to 29.9	16.4	19.6
30.0 to 49.9	22.4	24.3
50.0 to 69.9	35.0	36.5
70.0 or more	16.6	12.2
Percentage of poverty households:		
0.0 to 9.9	45.7	41.3
10.0 to 19.9	35.7	41.3
20.0 or over	18.6	17.5
Percentage of households not reporting interest or dividends:		
0.0 to 4.9	30.5	28.0
5.0 to 14.9	48.2	46.0
15.0 to 24.9	15.6	18.5
25.0 and over	5.7	7.4
Percentage of self-respondents:		
25.0 to 49.9	86.8	85.7
50.0 or more	13.2	14.3

TABLE 2.--AVERAGE NONRESPONSE AND NONPROBING RATES FOR SELECTED INCOME ITEMS, OR AVERAGE AREA CHARACTERISTICS BY INTERVIEWER CHARACTERISTICS

Percentage of persons or households of specific type	Interviewer characteristic				
	AGE				
	Total[1/] (802)	25-34 (67)	35-44 (221)	45-54 (345)	55 or over (166)
Households, 1+ other income item NA	16.2	16.8	16.2	16.5	15.2
Persons, interest and/or dividends NA	9.9	10.5	10.2	10.2	8.8
Persons, interest and/or dividends NP	12.3	12.2	12.2	12.5	12.0
Persons, Social Security items NA	7.8	7.5	8.0	8.3	6.8
Persons, Social Security items NP	9.6	9.4	9.4	9.8	9.6
	EDUCATION				
	Total[1/] (802)	1 to 11 years (26)	12 years (410)	College 1 to 3 years (235)	College 4 years or more (128)
Households, 1+ other income items NA	16.2	17.1	15.9	16.1	17.3
Persons, interest and/or dividends NA	9.9	8.6	9.6	10.2	10.2
Persons, interest and/or dividends NP	12.3	12.5	11.9	12.0	13.8

Percentage of persons or households of specific type	Interviewer characteristic					
	FAMILY INCOME					
	Total[1/] (802)	Less than $10,000 (139)	$10,000 to $14,999 (161)	$15,000 to $19,999 (202)	$20,000 to $24,999 (176)	$25,000 or more (115)
Households, 1+ other income items NA	16.2	14.8	16.1	16.6	15.1	19.0
Persons, interest and/or dividends NA	9.9	8.8	9.7	10.4	9.4	11.3
Persons, interest and/or dividends NP	12.3	10.9	12.3	11.9	12.7	14.1
Persons, Social Security items NA	7.8	6.8	7.6	8.2	7.7	8.9
Persons, Social Security items NP	9.6	9.5	9.9	9.9	9.4	9.0
Persons, pensions, alimony items NA	7.1	6.7	6.6	7.4	6.6	8.4
Persons, welfare items NA	6.8	6.2	6.3	7.3	6.3	8.2
Persons with multiple income sources	36.7	35.6	36.5	36.0	37.6	38.2
Households in poverty	12.5	14.7	13.2	12.4	12.3	9.7
Households, black head	8.1	8.2	9.0	8.5	8.0	6.3
Average household income	12,470	11,815	11,485	12,425	12,657	13,862

Percentage of persons or households of specific type	Interviewer characteristic				
	YEARS OF EXPERIENCE WITH MARCH SUPPLEMENT				
	Total[1/] (802)	1975 first time (165)	Second time (125)	Third time (133)	Fourth time or more (376)
Households, 1+ other income items NA	16.2	13.8	16.6	17.9	16.4
Persons, interest and/or dividends NA	9.9	8.5	10.0	11.0	10.0
Persons, interest and/or dividends NP	12.3	12.6	12.0	12.9	12.2
Persons, Social Security items NA	7.8	6.7	7.6	9.0	8.1
Persons, Social Security items NP	9.6	9.6	9.7	9.4	9.6
Persons with multiple income sources	36.7	37.7	37.6	38.4	35.6
Households in poverty	12.5	12.3	13.0	11.5	12.8
Average household income	12,470	12,367	12,140	12,690	12,550

Percentage of persons or households of specific type	Interviewer characteristic			
	APPROPRIATENESS OF ASKING SOURCES OF INCOME			
	Total (802)	Appropriate (437)	Inappropriate (176)	No opinion or NA (189)
Households, 1+ other income items NA	16.2	15.5	17.8	16.3
Persons, interest and/or dividends NA	9.9	9.2	11.5	9.9
Persons, interest and/or dividends NP	12.3	12.3	12.3	12.1

TABLE 2.--AVERAGE NONRESPONSE AND NONPROBING RATES FOR SELECTED INCOME ITEMS, OR AVERAGE AREA CHARACTERISTICS BY INTERVIEWER CHARACTERISTICS - continued

Percentage of persons or households of specific types	Interviewer characteristics				
	APPROPRIATENESS OF ASKING INCOME AMOUNTS				
	Total	Appropriate	Inappropriate	No opinion or NA	
	(802)	(436)	(176)	(190)	
Households, 1+ other income items NA	16.2	15.5	17.9	16.2	
Persons, interest and/or dividends NA	9.9	8.5	11.6	9.9	
Persons, interest and/or dividends NP	12.3	12.3	12.4	12.2	
	LEAST COOPERATIVE INCOME GROUP				
	Total	High	Average	Low	Can't tell or NA
	(802)	(283)	(169)	(63)	(287)
Households, 1+ other income items NA	16.2	16.3	17.2	15.6	15.6
Persons, interest and/or dividends NA	9.9	10.0	10.2	9.4	9.6
Persons, interest and/or dividends NP	12.3	12.6	12.2	12.3	12.0
	ATTITUDE OF RESPONDENT TOWARD INTEREST QUESTIONS COMPARED WITH OTHER INCOME QUESTIONS				
	Total	More cooperative or same	Less Cooperative	Can't tell or NA	
	(802)	(118)	(650)	(34)	
Households, 1+ other income items NA	16.2	14.4	16.5	16.1	
Persons, interest and/or dividends NA	9.9	8.6	10.1	10.6	
Persons, interest and/or dividends NP	12.3	12.1	12.3	12.7	
	ATTITUDE OF RESPONDENT TOWARD DIVIDEND QUESTIONS COMPARED WITH OTHER INCOME QUESTIONS				
	Total	More cooperative or same	Less Cooperative	Can't tell or NA	
	(802)	(163)	(582)	(57)	
Households, 1+ other income items NA	16.2	13.5	16.7	18.2	
Persons, interest and/or dividends NA	9.9	8.2	10.3	11.2	
Persons, interest and/or dividends NP	12.3	11.8	12.4	12.1	
	ATTITUDE OF RESPONDENTS TOWARD SOCIAL SECURITY QUESTION COMPARED WITH OTHER INCOME QUESTIONS				
	Total	More cooperative or same	Less Cooperative	Can't tell or NA	
	(802)	(730)	(37)	(35)	
Persons, Social Security items NA	7.8	7.8	9.9	7.2	
Persons, Social Security items NP	9.6	9.5	9.9	10.8	
	ATTITUDE OF RESPONDENTS TOWARD SUPPLEMENTAL SECURITY INCOME QUESTION COMPARED WITH OTHER INCOME QUESTIONS				
	Total	More cooperative or same	Less Cooperative	Can't tell or NA	
	(802)	(639)	(36)	(127)	
Persons, Social Security items NA	7.8	7.9	6.9	7.8	
Persons, Social Security items NP	9.6	9.6	10.0	9.2	

NA - Not answered
NP - Not probed

1/ Numbers may not sum to total because a few interviewers did not answer certain questions.

FORECASTERS SHOULD BE SOCIAL SCIENTISTS

John Stimson, William Paterson College
Ardyth Stimson, Kean College

The techniques of forecasting are thought of as econometric. Because of this historical development we tend to use economic variables in our forecasting models. The recently interrupted period of sustained growth has led us to mistakenly believe that societal forecasting could be accomplished without leaving the closed system of the economy. Forecasting has been defined as predicting the appropriate growth rate and economic growth has been seen as the driving wheel.

Recently, the end of population growth and the post-60's reassessment have redirected attention toward social values and goals and the collective behaviors that change them. There is a growing suspicion that many causative variables may lie outside the economic system. We hope to reinforce this mood by discussing the relevance of a few sociological variables in determining the economy. We will maintain that social definitions of value priorities are always at the base of economic movements. (That economy is after all a sub-system of society).

The authors are sociologists who recently returned to the campus after a lengthy association with economic and security market forecasting.[1] Our employment was not the beginning of a trend. We were originally included in the forecasting team as computer managers, not sociologists. If our sociological training was mentioned it was as a curiosity. As our stay lengthened the econometricians found ways to use our peculiar abilities to help them understand social change. That is the trend, economists are recognizing that social facts determine economic facts, but their reaction is not to consult sociologists, but to become "sociologists" themselves. This trend has been greatly accelerated by the recent failure of forecasting.

This note should not be interpreted as a criticism of quantification. Econometric techniques have made remarkable advances. The goal is to force attention back to the variables that these techniques are manipulating.[2]

Economists have not yet, and will never be able to, prepare adequate forecasts alone, because the basic causal variables they seek are decisions based on estimates of the expectations of consumers, managers, financers and potential parents. Economists think they can handle such decisions inside a rational-utility mode, but, these decisions are not rational, they are arbitrary. They are based on changing priorities and values. It is difficult to rationally model a social process when the values of goals keep changing. It is impossible to model such a process effectively unless one recognizes the arbitrary nature of social definitions of priorities.

We will return to "irrational" considerations. Let us now review some variables recognized as powerful influences on possible economic futures.

Demography is the area in which it is expected that there be an overlap of sociological and economic usage. We doubt, however, that sociologists know just how wide that overlap is. The Conference Board has published A Guide to Consumer Markets, (Axel, 1973) since 1960. The statistics collected there are at least as sociological as economic; (earnings, labor force participation, and life style preferences). The first 73 pages, the largest portion of the book, describe population time series. At least by this measure, then, demography is the most studied area of consumer behavior.

Demographics are so extensively studied because they provide the only long term, structured, indicator of future growth of the society. This potential growth of the consumer pool and labor force provides the context for economic work. This is the first entry point through which most forecasters become acquainted with society. They discover that population growth is not inevitable, nor does it behave totally rationally in the timing of its lapses and growth surges.

Unless you have directly experienced it you cannot imagine the theoretical dedication to growth that exists in most business or government planning. The recent end of the baby boom is important not just because it presages possible downturns; it is more important because it took away the only context of planning. Any interpretation of, or alternative to, zero population growth is listened to with interest. Sociologists now have an opportunity to provide other variables or contexts that could become basic forecasting tools.

A second illustration is the accelerating feedback relationship between Woman's Liberation and the economic structure. Let one semi-plausible scenario represent many other possibilities. Some sociologists have discussed the possible instability of the Black family in terms of the shifting relative importance of male and female roles when women find it easier to find work. Admittedly, it is an area that is little understood, but could the Black family be the precursor of the post-liberation family? There would be numerous attendant economic and societal changes if a substantial proportion of families found the woman to be the most reliable breadwinner.

A concept that sociologists should have developed in supernumerary income (Axel, 1973: 144). This variable measures the long term secular growth of affluence, by aggregating the amount of income in excess of $15,000 flowing to each family unit. For example, in 1955, 5.6 per cent of U.S. families had supernumerary income, by 1965, 14.6 per cent, and in 1975 it is estimated that 33.5 per cent of families will have "affluent" income. Where this income will be directed is a sociological, life-style, question. Whether or not supernumerary income will continue to grow, and which segment of the population will receive the extra income, and what they are likely to use it for, are also sociological questions. Economists will not be able to understand or predict these changes without assistance.

It has recently been forcibly demonstrated that consumers' decisions to spend or save create business cycle turning points and determine the difference between recession or depression. Economists, however, generally consider consumers to be followers of the economy. That is, they feel consumers tend to change their opinions after other indicators have turned. Their concentration on prediction makes economists occasionally look at the indications instead of the real underlying process. Because the present indicators are unstable, and appear non-rational, economists tend to downgrade the causal weight of consumers' attitudes because they appear to lag behind other predictors. There is no doubt, however, that consumer demand determines the system. The reason other parts of the economy turn first and react more rapidly is because they are trying to anticipate demand. The answer is real social indicators.

This unrecognized centrality of consumer attitudes provides at least two major opportunities for social scientists. First, and most obvious, the measurement of consumer moods is still primitive. The surveys that are regularly carried out are simple, and oriented toward prediction of specific purchases. Application of the societal perspective, especially through disaggregation of "consumers" into a meaningful typology, that might straighten out overlapping irrationalities could be productive.

Secondly, attempts to work out the social-psychological interdependencies between the various levels of consumer decisions and their anticipations by producer decisions, must be understood. We live in a competitively and semi-informally planned economy. The producer of consumer, or government, goods must correctly guess demand--but so must all of his suppliers, and their suppliers, for a dislocation of the economy to be avoided. The complexity and interconnectedness of the system of overlapping decisions is clearly demonstrated by use of input-output economics. An input-output table can show the interdependence of the economy's industries. It is a remarkable advance in systematic analysis. One is able to follow the consumer dollar paid for an automobile through the whole economy and see which industry used up which portion, steel, iron, coal, railroads, or electronics, etc.

Looked at with a sociologist's bias the input-output table demonstrates the difficulty of producers consistently guessing what their various consumers will need in one, two, or three years. Each consumer, whether he is a professional purchasing agent or not, is subject to the existing moods of the country, as they a are filtered through his life-style. The total process is confused and volatile. Any study that sociologists could contribute to regularize it would make for a more stable planning system.

A more complicated example of the interconnection of societal and economic forces can be seen in recent studies of the effects of housing programs.[3] Very little support has been found for the proposition that government should ever again try to improve social conditions by building new housing. It also seems that the decisions of where and when to start new construction have really been made for economic reasons, (i.e., to stimulate employment or finance). These new studies uncover a picture of massive economic programs that are supported only by the societal notion that a man's home is his castle and that a man or his family can't be effective without a respectable castle. How are changing definitions going to effect housing programs in the future? Is this powerful economic stimulant going to be withdrawn? Will it only be defined as valuable outside of cities? Will the fact that programs seem to have no ameliorative effect have any impact at all? The potential results require a mix of sociological and economic forecasting skills.

We have been listing justifications for social scientists other than economists claiming that the basic processes that are dealt with in economic predictions are really in their data-territory. One final reminder that consumers decisions are dependent on major societal value shifts. One economist was quoted recently in the N.Y. Times as wondering if, in ten years, todays' dog food producers would be canning food for pets or pets for food. This specific remark may have been partly inspired by a temporary feeling of insecurity, but it also emphasizes that specific attitudes support a major industry that feeds certain animals to other animals. This set of attitudes is no less idiosyncratic than the values that determine the style of America's economy. Americans spend very high proportions of their dollar on autos and the welfare of others. There are reactions underway to both these preferences. The strength of the reaction will determine future life styles, and therefore, economic structures.

FOOTNOTES

[1]We have recently formulated a program of ways in which sociologists can help in forecasting, and the outline of a course that could prepare them for this job. "Sociologists should be put to work as forecasters" American Sociologist 1976:49-56.

[2]We apologize because the style of what follows deliberately approaches economists in a combative fashion. It is only because we have found that other social scientists feel some unwarranted insecurity when they face the edifice of modern econometrics. One should not be concerned. Most economists do not understand it either.

[3]Mitchell, Robert E. "Sociological Research on the Myths of Housing" Social Problems 1974: 259-79.

THE VARIANCE ESTIMATOR AND SAMPLE SIZE IN VARIABLES SURVEY SAMPLING

Vernon W. Stone and Russell W. Irvine, Georgia State University

PROBLEM

While attributes sampling focuses on the question "How many?," variables sampling focuses on the question "How much?." In the objective determination of sample size, $\underline{pq}$ is the variability factor in the former, and $\underline{\sigma}$ is the variability factor in the latter. In variables sampling, depending upon the various and sundry estimators employed to estimate variance and/or standard deviation, a wide range of sample sizes can result. The formula application requires the direct use of the standard deviation. The tabular reading requires the use of the standard deviation as the denominator in the ratio of the sampling error to the standard deviation. The foregoing, then, suggests various estimates of variance and/or standard deviation, pursuant to a variety of estimators, and the subsequent objective determination of sample sizes employing such estimates.

HYPOTHESIS

The basic hypothesis may be stated as follows: Depending upon the estimator(s) employed, marked variations can be observed among the variance estimates and, consequently, the objectively determined sample sizes.

PROCEDURES

Method

The descriptive methodology, making use of the survey technique was employed.

Subjects

College professors, graduate students, and research practitioners, totaling 115 volunteers, in attendance at a 1976 sampling workshop, constituted the sample. The volunteers represented a variety of behavioral-science disciplines. Moreover, each participant possessed recognized preparation and expertise in the general sampling area.

Instrumentation

A questionnaire, detailing the following contrived situation, was distributed to the 115 prospective participants:

1,000 college freshmen are annually enrolled in Mathematics 105, Mathematical Analysis, at City Junior College. During the last five academic years (1970-71 thru 1974-75), an annual mathematics achievement test has been administered to all of the 1,000 students. For the 1975-76 academic year, it is necessary, because of budgetary problems, to use a sample from the 1,000 students.

The following information was attached to the questionnaire: (1) Miscellaneous experimental data, including numerical indices of the achievement-test variances for the last five academic years; and (2) Population frame of the 1976 mathematics achievement scores for the 1,000 freshmen.

Also, the following sampling specifications were indicated: (1) Population is 1,000 ($N=1,000$); (2) Confidence level is to be 95% (z (standard variate)$=1.96$; and (3) Sampling error is to be 2 ($E=2$).

The sampling task was to estimate the population variance--by any estimator desired--and report a numerical index for the variance estimate.

In addition to the foregoing, the respondent was requested to indicate, on the reverse side of the questionnaire, the estimator category, or categories, and the estimator sub-category, or sub-categories, employed in arriving at the reported numerical index of the variance. The checklist presented both the estimator categories and estimator sub-categories as follows:

1.__Empirical Bases
   a.__Past experience
   b.__Conventional practice
   c.__(Other)____________

2.__Ad Hoc Samples
   a.__One-sample statistics
   b.__Two-or-more-sample statistics
   c.__(Other)____________

3.__Central Limit Theorem Applications
   a.__Derived-formula basis
   b.__Normal distribution basis
   c.__(Other)____________

4.__Systematic Determinations
   a.__Computed values
   b.__Tabular values
   c.__(Other)____________

5.__Subjective Approaches
   a.__Intuition
   b.__Whim
   c.__(Other)____________

__(Employed Estimator Category Not Listed)

Treatment

On the basis of the reported variances, sample sizes were determined by both formula and table.

(1) By formula. The starting point was the reliability, or precision, formula, based upon the standard error of the mean. Then, the sample-size (n) formula was derived, and the finite population correction was applied thereto, thus resulting in the following composite formula which is applicable to variables sampling:

$$n = \frac{N\,(z\,\sigma)^2}{(N\,E^2) + (z\,\sigma)^2}$$

(2) By table. The ratios of the sampling error to the standard deviation were computed. Then, each ratio was related to population size and confidence level. For purposes of this study, the appropriate sample-size table by Arkin was employed.

Except for minor variations, the formula and the table yield comparable sample sizes.

RESULTS

Table 1, "The Frequency and Proportion of Subjects Employing Each of the Estimator Categories," indicates that the number of usable responses was 36, or approximately 31 per cent of the number of questionnaires distributed. The frequency of use of the five estimator categories ranged from 4 to 12, showing Systematic Determinations as the modal estimator category.

Table 1

THE FREQUENCY AND PROPORTION OF SUBJECTS EMPLOYING EACH OF THE ESTIMATOR CATEGORIES

Estimator Category	Frequency	Proportion of Respondents ($n=36$)
Empirical Bases	4	.11
Ad Hoc Samples	7	.20
Central Limit Theorem Applications	5	.14
Systematic Determinations	12	.33
Subjective Approaches	8	.22
Totals	36	1.00

Table 2, "The Estimated Ranges of Variance and Standard Deviation Reported for Each of the Estimator Categories," shows that the Subjective Approaches are the most variable. Variance ranges from 36 to 361, and standard deviation ranges from 6 to 19. The Empirical Bases are observed to be the least variable. Hence, the lowest variance is 36, and the highest variance is 361, representing standard deviations of 6 and 19, respectively.

Table 2

THE ESTIMATED RANGES OF VARIANCE AND STANDARD DEVIATION REPORTED FOR EACH OF THE ESTIMATOR CATEGORIES

Estimator Category	Estimated Range for Variance (σ^2)	S.D. (σ)
Empirical Bases	81-121	9-11
Ad Hoc Samples	64-225	8-15
Central Limit Theorem Applications	100-169	10-13
Systematic Determinations	81-196	9-14
Subjective Approaches	36-361	6-19

Pursuant to the five specified estimator categories, it was expected that Subjective Approaches would be the most variable since there is a discernible relationship between variability and subjectivity. It is unclear, however, as to why Empirical Bases are the least variable, insofar as a degree of subjectivity is involved therein. It was anticipated that Systematic Determinations would be the least variable because of the objective nature of this estimator category.

A comparison of Tables 1 and 2 shows that there is a relationship between the highest frequency (Table 1) and the most variable estimator category (Table 2). A Spearman calculation results in a coefficient of .70. Notwithstanding, the small number of estimator categories renders the index insignificant.

Table 3, "Sample Sizes on the Formula Basis (Standard Deviation) and the Tabular Basis (Ratio of Sampling Error to Standard Deviation) Determined for the Reported Standard Deviation Range (S.D. 6 thru S.D. 19)," makes use of the computational formula by using the standard deviation values and the table by using the ratio of the sampling error to standard deviation.

It will be noted that minor differences are to be observed in most instances with respect to the values of the two sample-size determinations. Observing that the standard deviations range from 6 thru 19 (variances 36 thru 361), the sample sizes, by formula, range from 34 thru 258 and, by table, from 41 thru 278.

Table 3

SAMPLE SIZES ON THE FORMULA BASIS (Standard Deviation) AND THE TABULAR BASIS (Ratio of Sampling Error to Standard Deviation) DETERMINED FOR THE REPORTED STANDARD DEVIATION RANGE (S.D. 6 thru S.D. 19)

Standard Deviation (σ)	Sample Size By Formula (n)	Ratio of Sampling Error to Standard Deviation (E/σ)	Sample Size By Table (n)
6	34	.33	41
7	45	.29	42
8	58	.25	58
9	73	.22	88
10	88	.20	88
11	105	.18	89
12	122	.17	146
13	140	.15	146
14	159	.14	164
15	179	.13	186
16	198	.13	187
17	218	.12	211
18	238	.11	241
19	258	.10	278

Table 4, "The Cost of Incremental Sample Units When Sample Size Is Based Upon Formula Determined for the Reported Standard Deviation Range (S.D. 6 thru S.D. 19)," underscores the direct, positive relationship between incremental sample units and cost, using the sample sizes, by formula, for the illustration. Pursuant to a request, a testing service representative suggested that the subject mathematics achievement test, under the conditions indicated, would cost approximately $1,180 for the first 75 students when group tested. Thereafter, a pro rata average cost of $30.00 for each incremental sample unit above 75 students would be reasonable on the basis of current educational costs. Even a degree of conservatism was discerned with respect to these cost figures. From a S.D. of 6 thru a S.D. of 9, a maximum sample of 73 is involved. Hence, $1,180 would represent the basic cost. From a S.D. of 10 thru a S.D. of 19, however, the incremental sample units at $30.00 each bring the total sample cost to $1,570 thru $6,670, representing sample sizes of 88 and 258, respectively.

Table 4

THE COST OF INCREMENTAL SAMPLE UNITS
WHEN SAMPLE SIZE IS BASED UPON FORMULA
DETERMINED FOR THE REPORTED STANDARD DEVIATION RANGE (S.D. 6 thru S.D. 19)

Standard Deviation (6)	Sample Size (n)	Incremental Sample Unit Cost at $30.00 Each	Total Sample Cost Given Minimum Cost for 75 Units ($1,180)
6	34	$ ---	$ 1,180
7	45	---	1,180
8	58	---	1,180
9	73	---	1,180
10	88	390	1,570
11	105	510	2,080
12	122	510	2,590
13	140	540	3,130
14	159	570	3,700
15	179	600	4,300
16	198	570	4,870
17	218	600	5,470
18	238	600	6,070
19	258	600	6,670

CONCLUSIONS

The subject hypothesis was confirmed, clearly demonstrating that there are, depending upon the estimator(s), marked variations among variance estimates and, consequently, sample sizes.

The study constituted a pilot, designed to suggest the confirmation or disconfirmation of the hypothesis. To be sure, a tighter, more telescopic study may reveal different numerical results. Notwithstanding, there is ample reason to anticipate that such results will underscore the general direction of the findings reported herein.

Recognizing that samples which are too small and samples which are too large can adversely affect findings, it behooves the researcher to attend such with great care. At the threshold of the Third Century, researchers have an obligation to avoid both "overkill" and "underkill" via sample size.

REFERENCES

Arkin, Herbert. Statistical Sampling for Auditors. N.Y.: Price Waterhouse & Co., 1960.

Arkin, Herbert. "Statistical Sampling in Auditing," The New York Certified Public Accountant, July, 1957.

Arkin, Herbert and Henry P. Hill. Sampling in Auditing. N.Y.: Price Waterhouse & Co., 1959.

Cochran, William G. Sampling Techniques, 2/e. N.Y.: John Wiley & Sons, Inc., 1963.

Deming, W. Edwards. Sample Design in Business Research. N.Y.: John Wiley & Sons, Inc., 1960.

Hansen, Morris H., William N. Hurwitz, and William G. Madow. Sample Survey Methods and Theory, Vol. I. N.Y.: John Wiley & Sons, Inc., 1953.

Slonim, Morris J. Sampling in a Nutshell. N.Y.: Simon and Schuster, 1960.

Stephan, Frederick F. "History of the Uses of Modern Sampling Procedures," Journal of the American Statistical Association, 43 (241), 1948.

Trueblood, Robert M. and R.M. Cyert. Sampling Techniques in Auditing. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1957.

Vance, Lawrence L. and John Neter. Statistical Sampling for Auditors and Accountants. N.Y.: John Wiley & Sons, Inc., 1956.

SENSITIVITY OF ALTERNATIVE FERTILITY INDICES

C.M. Suchindran, J.W. Lingner, A.N. Sinha, E.J. Clark, University of North Carolina at Chapel Hill

I. Introduction

The purpose of this paper is to examine the sensitivity of selected fertility indices to true differences in fertility levels over time and space. Such examination requires an understanding of the statistical variations in these indices. Variability is to be expected even in the case of complete and accurate measurement of fertility, since the behavioral and biological mechanisms which produce births are subject to random variations. When data from sample surveys are used to construct indices, another type of variation is produced by the sampling mechanism itself. Better understanding of the effect of these variations on alternative fertility indices is essential if efficient and accurate indicators of fertility change are to be identified.

This investigation focuses on the sensitivity of certain conventional fertility indices as well as indices based on the birth intervals. The data used in this study were generated through a Monte Carlo simulation model, POPREP. POPREP is a computer simulation model developed by the Department of Biostatistics at the University of North Carolina. It generates the reproductive histories of a hypothetical female population under a set of assumed conditions regarding probabilities of marriage, marital dissolutions, deaths of women and their spouses, conception, spontaneous and induced abortion, use of contraception, and other factors related to fertility performance. The structure of the model is similar to that of POPSIM, but unlike POPSIM, POPREP explicitly considers biological factors underlying fertility. This biological approach to population simulation is closely related to the REPSIM model developed by Ridley and Sheps (1966). Further details of the model are given in the User's Manual for POPREP (1975).

II. Design of Experiments

The first step in the present investigation was to establish an initial population with characteristics approximating that of a developing society. For the sake of internal coherence, and because empirical data were more readily available than for other countries, most of the parameters used in establishing the initial population were derived from data on India.

The initial population was assumed to consist of 10,000 women between age 0 and age 50 who had survived to the initial simulation year. The age distribution was based on census data for India for 1961 and resembles the broad-based age pyramid characteristic of societies with relatively high growth rates.

Although marriage (which is here defined as representing any form of sexual union) was assumed not to occur before age 15, the population was characterized by an early age at marriage. Slightly over half of the women were assumed to have married by age 16. The age of husbands was determined as a function of the age of wives at marriage, but in general, husbands were about seven years older than their spouses. Divorce was assumed not to occur. The probability of widowhood was based on the probability of male mortality from Regional Model Life Table South, Level (e_o^o = 51.0). Remarriage from widowhood was assumed not to occur. Table 1 shows the distribution of women by marital status at the start of simulation.

The assumptions made in regard to factors affecting reproduction within marriage were as follows: Fecundability, the monthly chance of conception in the absence of contraception, varied among women and, for a given woman, with her age. The parameters used to determine fecundability implied a mean fecundability of approximately .20 at the age when fecundability was highest. The probability of a spontaneous fetal loss was determined as a function of age, reaching a minimum for women in their twenties at a level of about .20 and rising to a maximum of over .50 for women in the oldest age groups. The distribution of age at sterility postulated for these simulations implied a mean age of 42.14 years, with a standard deviation of 4 years. The distribution of length of pregnancies ending in fetal losses had a mean of 1.9 months and a variance of 3.1 months, while all pregnancies ending in live births were assumed to last exactly nine months. The distribution of length of postpartum anovulation following a live birth was assumed to be relatively long, with a mean of 12 months; such a distribution would be consistent with a society where breastfeeding was practiced widely and for substantial periods of time.

These parameters were assumed to operate during the first five years of simulation to establish a baseline against which the three experimental patterns of contraceptive use could be compared. During the five year baseline period, mortality among women was also assumed to occur. Probabilities of female deaths were derived from Regional Model Life Tables, South, Level 13 (e_o^o = 50.0).

Three patterns of contraceptive use were postulated to begin to operate at the end of the fifth year of simulation. In all three, the chance of becoming a contraceptor was assumed to vary negatively with age and positively with parity. The rate of acceptance for the first contraceptive use pattern implied moderate acceptance of contraception and the second pattern assumed somewhat higher acceptance rates.

In both the first and the second contraceptive pattern, contraception was assumed to be 100% effective, and the women who adopted contraception were assumed to practice until the end of their reproductive lives. Thus, these two patterns can be viewed as simulating the results of a sterilization program. In the third pattern, however, although women were assumed to accept contraception according to the higher hazard function of pattern 2, effectiveness of contraception was assumed to be 90%. Moreover, in this pattern, if a woman practicing contraception became pregnant, she was assumed to return to the noncontracepting state and her chance of reaccepting contraception was no different from that of a woman of the same age and parity who had never previously used contraceptive methods. Thus pattern 3 can be thought of as representing the use of a fairly effective method of contraception for a period of time determined by the advent of the next conception. Each of these patterns of acceptance was assumed to operate over a ten year period.

Table 2 shows the number of acceptors and the acceptance rates for each of the patterns of acceptance. Somewhat surprisingly, the number of acceptors declines over the 10 year period of program operation for each acceptance pattern. That this is in part due to reduced numbers of women eligible for acceptance is indicated by the data on person-years exposure to the risk of acceptance. These also decline from year 5 to about year 12, and then become approximately stable. The annual acceptance rates decline for the two high acceptance rate patterns. The sharpest decline is observed for the high acceptance rate, 100% effectiveness pattern. Under this pattern, the women who are non-acceptors at the end of the simulation period are likely to be young women of low parity and a correspondingly reduced risk of becoming a contraceptor. In the high acceptance rate, 90% effectiveness pattern, acceptance rates are influenced by the number of women who have accidental pregnancies and, thus, drop out of contraceptive practice. When these women complete their accidental pregnancies, they are again exposed to the risk of acceptance and may again become acceptors. This process of dropping out and reentering operates to keep acceptance rates from declining as much as in the high acceptance rates, 100% effectiveness pattern. In the moderate acceptance rates, 100% effectiveness pattern, annual acceptance rates are consistently lower than for the other two patterns, but the pattern of decline over time is not nearly so marked as in the high acceptance rate pattern.

III. Sensitivity of Fertility Indices: An Examination of Patterns

Conventional Fertility Indices

The results of these simulations, expressed in terms of specific fertility rates and indices summarizing these rates, are given in Tables 3 and 4. Table 3 shows age specific fertility rates for each of the five years of the baseline period when no contraception is used. The most noteworthy feature of this table is the substantial fluctuation in rates from one year to the next. The largest fluctuations, with a range of 8.3% of the highest value, appear for the gross reproduction rate, which for these simulations, has been calculated on the basis of the number of female births actually occuring in each simulation, rather than as a multiple of the total fertility rate. The smallest variation, about 2.4% of the highest observed value, appears for the general fertility rate. For the total fertility rate, the range of variation is approximately 3%.

Table 4 shows age specific fertility rates and associated summary indices for each pattern of contraceptive use over the ten year period. As expected, all of these fertility rates show a tendency to drop. The decline is sharpest between simulation years 6 to 10. The first and second contraceptive patterns, those with 100% effectiveness, continue to show a moderate change in years 11 through 15. These changes in fertility rates are quite consistent with the data on contraceptive acceptance shown in Table 2. However, fertility indices for the third contraceptive pattern, with 90% effectiveness, reach a low point in the tenth year of program operation and remain virtually constant thereafter. This latter result is perhaps to be expected in view of the consequences of discontinuation of contraception which occur only under the 90% effectiveness pattern. Under this pattern of acceptance, the reduction in fertility attributable to new acceptors appears to be counterbalanced by the increase in fertility resulting from "accidental" pregnancies and discontinuance of use.

It should be noted that whereas the trend of fertility is continuously downward for each successive year of the first five years of program operation, irregularities occur in all three patterns during the eleventh to the fifteenth year.

Birth Interval Indices

In recent years, considerable interest has been expressed in using data on the intervals between successive births to develop sensitive indicators of fertility change. The commonly used indices are the mean of all closed birth intervals, the mean of the last closed birth intervals and the mean of the interval since the last birth (the open interval). The latter intervals are often calculated on a parity specific basis. Despite the intuitive appeal of these indices, they are subjected to large biases due to effect of truncation (for details, see Sheps et. al. (1969)). A number of attempts have been made to construct refined birth interval indices which control for the effects of truncation as well as for age and parity composition. These indices include life table estimates (Sheps (1965)) and the estimates developed by Poole (1973).

Parity specific interval indices have been computed for each of the first nine birth intervals. These indices are summarized by weighting the results for each of the first nine intervals by the number of women attaining each parity. Table 5 presents these weighted index for four survey points, years 5, 7, 10 and 15. As can be seen, only the unadjusted mean open interval shows a consistent tendency to lengthen at each successive time point. Moreover, the differences between year 5 and year 7 are quite small, with the largest difference appearing for the 90% effectiveness contraceptive use pattern. By year 10, differences are much more distinct. The longest interval is observed for the high acceptance rate, 100% effectiveness pattern, but the 90% effectiveness pattern continues to produce a longer interval than the moderate acceptance rates. Not until year 15 are the differences among the three patterns distinct.

The summary adjusted indices fail to show any change in birth interval length at year 7. Although longer intervals are consistently indicated by Poole's Index in year 15, the observed differences are slight.

IV. Power and Sensitivity of Fertility Indices Under Alternative Sample Sizes: Some Estimates from Replication

The results presented thus far pertain to the total simulated populations. As will be examined in more detail in the next section, decisions made on the basis of samples are subject to additional sources of error, even when the period of observation is relatively long. The major question to be answered is whether, when changes in fertility are to be inferred on the basis of sample values rather than population values, the values observed in a sample of a given size are consistent with what is occurring in the total population.

To investigate the variability of fertility indices for these experimental populations, a series of replicated samples was created. For each series, samples of women who were alive at a given time point and were between the ages of 15 and 50 were selected. A particular woman could appear only once in each sample (sampling without replacement) but could be selected into more than one sample in a given series.

Table 6 shows the results from the various sets of replicated samples. The relative variation of the sampling distributions of these indices, as measured by their coefficients of variation (the standard deviation of the values of the indices from each set of replicated runs divided by the mean value of the index), exhibits a quite consistent pattern for all sets of replicates. The coefficients of variation are highest for the age specific birth rates, at an intermediate level for the total and general fertility rates and the open interval, and lowest for the other interval measures. This pattern of relative variability sheds some light on the total population results presented earlier. The very small amounts of change seen in most of the indices based on closed intervals is likely to be in part due to the extreme stability of these indices. The somewhat erratic behavior of the age specific indices is, in contrast, partly due to their relatively high standard errors.

To investigate the significance of observed differences in fertility, the value of each fertility index was determined at years 5, 10, and 15 for each replicated sample. From these data, it was possible to calculate sets of replicated differences between year five and year fifteen, a ten year span, and over the two five year intervals from year five to ten and from year ten to fifteen. The standard errors of the differences were estimated from the distribution of the differences between time points in each replicate. Standard normal tests were then applied to test the hypothesis of no change. The results are shown in Table 7. For the high acceptance rate, 100% effectiveness pattern, all birth rate indices show significant changes over the ten year period between year five and year fifteen, except for the age specific birth rates for the youngest and oldest age groups. Among the interval indices, however, only the open interval shows a significant change. The power of these indices to detect changes declines when comparisons are made over shorter spans of time. When year five results are compared with those of year ten, only the total fertility rate and the open interval indicate a significant change. When year ten is compared with year fifteen, the open interval is the only index which would lead to rejection of the "no change" hypothesis. Quite possibly this measure is influenced by changes in the relatively distant past. To the extent that this is true, the open interval may be an inadequate index for studying current fertility changes.

When the high acceptance rate, 90% effectiveness pattern is considered, fewer indices differ significantly. Only the general fertility rate and the open interval detect the change occurring over the ten year period from year five to year fifteen. The total fertility rate shows a statistically significant difference for the period from year five to year ten, while the open interval indicates a difference from year ten to year fifteen, again raising a question about the short term sensitivity of interval indices.

These results indicate that the general fertility rate, the total fertility rate and the open interval are the measures that are likely to detect changes in fertility. Table 9 shows the power of these indices in detecting an assumed amount of change for various periods. The power is determined as the proportion of the samples showing a significant change when the amount of change observed for the total population is assumed as the true change. For all contraceptive use patterns, the general fertility rate and the total fertility rate appear to be more powerful in detecting changes over the first five years of program operation than the open interval. Over the ten year period and for the period between year ten and year fifteen, the open interval appears to be more powerful than the other two indices, except in the case of the last five years of the 100% effective pattern. The table also shows a 10% increase in the power of the total and general fertility rate with doubling the sample size from 500 to 1000 in the comparison of year five with year fifteen, 100% effectiveness.

Table 9

Power of Selected Fertility Indices in Detecting an Assumed Amount of Change for Various Periods

	High Acceptance Rates, 100% Effectiveness			
	Year 5 Vs. 15		Year 5 Vs. 10	Year 10 Vs. 15
	N=1000	N=500	N=500	N=500
General Fertility Rate	0.9984	0.9049	0.8159	0.4168
Total Fertility Rate	0.9990	0.9162	0.8315	0.6950
Open Interval	1.0000	1.0000	0.6443	0.4013
	High Acceptance Rates, 90% Effectiveness			
		N=500	N=500	N=500
General Fertility Rate	- -	0.6736	0.6443	0.0359
Total Fertility Rate	- -	0.4960	0.6950	0.0455
Open Interval	- -	0.8289	0.4013	0.5160

V. Conclusions

It should be noted that the conclusions drawn about the statistical properties of these indices pertain to the particular set of conditions governing this investigation. Their generality to other circumstances may be limited. It is especially important to note that the simulated data studied here are free of response and other types of nonsmapling errors which are often the major problems of real-life data. Extension of this research by incorporating these factors might be very useful.

Nevertheless, these results do suggest that conventional fertility measures such as the total and general fertility rates may be more sensitive to short term changes than other types of measures. They also indicate that, under the conditions postulated for those simulations, long term effects of alternative patterns of contraceptive practice may not be easily estimated from short term results.

References

Poole, K. (1973). Fertility Measures Based on Interval Data. Theoretical Population Biology 4: 357-387.

Sheps, M.C. (1965). An Analysis of Reproductive Patterns in an American Isolate. Population Studies 19: 68-80.

Sheps, M.C., Menken, J.A., Ridley, J.C., and Lingner, J. W. (1970). The Truncation Effect in Birth Interval Data. Journal of American Statistical Association 65: 678-694.

Ridley, J.C., and Sheps, M.C. (1966). An Analytical Simulation Model of Human Reproduction with Demographic and Biological Components. Population Studies 19: 297-310.

User's Manual for POPREP and INTRVL. (1975). Department of Biostatistics, University of North Carolina at Chapel Hill.

Table 1

Distribution of Women by Age and Marital Status at the Beginning of Simulation

Age	Single	Currently Married	Widowed	Total
<1	405	0	0	405
1-4	1470	0	0	1470
5-9	1523	0	0	1523
10-14	1295	0	0	1295
15-19	441	674	3	1118
20-24	105	862	22	989
25-29	16	812	45	873
30-34	7	652	37	696
35-39	5	565	66	636
40-44	6	443	93	542
45-49	7	350	97	454
Total	5280	4358	363	10001

Table 2

Number of Acceptors Per Year, Estimated Number of Person Years Exposed to the Risk of Acceptance, Annual Acceptance Rates Per Currently Married Woman, and Proportion of Currently Married Women Using Contraception at the End of Each Simulation Year by Pattern of Contraceptive Use

	Year									
	5	6	7	8	9	10	11	12	13	14
	Moderate Acceptance Rates, 100% Effectiveness									
Number of Acceptors	298	323	323	283	285	257	268	262	273	221
Estimated Number of Person Years Exposed to Risk of Acceptance	3737	3597	3446	3363	3267	3190	3113	3070	3055	3065
Annual Acceptance Rates	7.97	8.98	9.37	8.42	8.72	8.06	8.61	8.73	8.94	7.21
Proportion Using At End of Year	.0581	.1146	.1697	.2057	.2431	.2690	.2960	.3183	.3405	.3506
	High Acceptance Rates, 100% Effectiveness									
Number of Acceptors	473	468	435	396	361	336	319	267	280	235
Estimated Number of Person Years Exposed to Risk of Acceptance	3671	3420	3185	3004	2845	2720	2653	2604	2594	2619
Annual Acceptance Rates	12.88	13.68	13.66	13.18	12.69	12.35	12.02	10.25	10.79	8.97
Proportion Using At End of Year	.0903	.1717	.2411	.2949	.3424	.3748	.4016	.4170	.4346	.4407
	High Acceptance Rates, 90% Effectiveness									
Number of Acceptors	497	495	404	425	365	389	381	352	350	313
Estimated Number of Person Years Exposed to Risk of Acceptance	3665	3388	3185	3072	2993	2919	2890	2910	2949	2933
Annual Acceptance Rates	13.56	14.61	12.69	13.84	12.20	13.33	13.18	12.10	11.87	10.66
Proportion Using At End of Year	.0938	.1681	.2172	.2599	.2848	.3070	.3239	.3286	.3359	.3347

Table 3

Age Specific Fertility Rates and Marital Fertility Rates by Year Since Start of Simulation, Year One to Five, No Contraception

	Year				
	1	2	3	4	5
Age Group	Age Specific Fertility Rate				
15-19	189.9	203.1	196.0	205.1	205.3
20-24	369.5	352.1	366.8	364.5	369.3
25-29	384.9	395.8	397.6	359.8	384.3
30-34	312.1	344.7	291.1	341.1	289.6
35-39	212.2	215.2	220.3	219.3	207.8
40-44	87.0	80.8	74.3	81.9	91.8
45-49	2.2	2.1	2.0	6.0	7.8
General Fertility Rate	247.8	252.6	246.6	249.9	247.6
Total Fertility Rate	7789.0	7969.0	7740.0	7889.0	7780.0
Gross Reproduction Rate	3678.0	3982.0	3653.0	3895.0	3693.0

Age Group	Age Specific Marital Fertility Rate				
15-19	318.5	344.6	323.6	337.8	342.7
20-24	418.5	397.8	412.9	411.4	414.9
25-29	410.1	422.0	425.2	385.4	412.6
30-34	334.8	365.6	312.8	368.8	313.1
35-39	237.7	240.0	246.5	243.6	225.7
40-44	103.3	97.1	88.1	94.9	107.9
45-49	2.9	2.8	2.6	5.2	10.3
General Marital Fertility Rate	300.7	306.1	297.9	301.4	299.0
Total Marital Fertility Rate	9128.9	9349.1	9058.8	9235.7	9135.9

Table 4

Age Specific Fertility Rates by Pattern of Contraceptive Use and Year of Simulation

Age Group	6	7	8	9	10	11	12	13	14	15
	Moderate Acceptance Rates, 100% Effectiveness									
15-19	191.2	183.2	193.9	190.6	181.1	197.2	199.6	200.1	213.0	190.8
20-24	383.1	334.2	330.2	333.1	323.1	296.2	285.5	317.6	306.9	298.5
25-29	397.7	358.8	345.9	334.6	317.4	304.8	306.1	286.7	263.6	281.2
30-34	318.6	307.8	278.3	263.0	248.4	239.1	222.1	233.6	201.0	216.8
35-39	224.0	218.1	162.0	149.7	164.8	133.5	133.8	142.4	129.3	98.2
40-44	80.5	78.4	82.4	44.9	49.9	33.1	36.1	41.6	33.1	35.9
45-49	5.6	9.2	8.9	0.0	5.1	3.4	5.0	1.7	0.0	3.3
General Fertility Rate	253.4	234.7	224.0	214.0	208.2	197.8	194.8	201.0	191.5	187.0
Total Fertility Rate	8004.0	7448.0	7008.0	6580.0	6449.0	6036.0	5941.0	6118.0	5733.0	5624.0
Gross Reproduction Rate	3929.0	3579.0	3538.0	3116.0	3255.0	2958.0	2867.0	2811.0	2778.0	2834.0
Age Group	High Acceptance Rates, 100% Effectiveness									
15-19	189.6	193.5	185.4	185.6	183.5	186.2	203.5	204.8	204.4	188.4
20-24	370.8	337.2	327.9	323.0	304.2	268.0	289.2	280.3	300.0	269.9
25-29	391.3	357.2	318.8	302.9	282.6	259.9	246.1	257.1	243.5	263.3
30-34	305.2	305.3	246.9	230.2	225.5	171.1	169.4	169.7	150.5	156.5
35-39	230.1	178.9	142.1	150.3	129.0	114.6	122.2	85.7	104.1	82.78
40-44	80.4	70.2	55.8	42.7	41.9	26.8	41.7	34.2	13.04	39.2
45-49	5.6	5.5	1.8	5.2	1.7	3.3	3.3	0.0	1.6	0.0
General Fertility Rate	248.5	230.8	207.1	201.5	190.9	170.5	178.3	173.4	172.4	167.2
Total Fertility Rate	7865.0	7239.0	6394.0	6204.0	5842.0	5150.0	5377.0	5159.0	5087.0	5000.0
Gross Reproduction Rate	3899.0	3466.0	3057.0	3081.0	3001.0	2531.0	2458.0	2499.0	2660.0	2425.0
Age Group	High Acceptance Rates, 90% Effectiveness									
15-19	189.6	192.0	193.3	184.0	181.3	191.9	187.9	210.1	197.3	197.8
20-24	395.3	320.7	332.8	329.4	307.6	326.0	325.4	308.2	329.7	300.2
25-29	384.0	345.8	339.6	333.0	302.1	334.3	320.8	336.8	311.2	334.2
30-34	309.8	309.4	288.7	255.8	256.7	235.2	239.4	245.6	254.7	238.1
35-39	214.2	170.8	210.0	180.6	156.2	136.7	147.4	133.3	119.7	145.3
40-44	77.2	62.4	57.5	62.3	39.1	52.8	44.2	42.4	39.3	31.9
45-49	5.6	3.7	5.3	1.7	0.0	0.0	3.4	1.7	3.3	4.9
General Fertility Rate	250.5	224.4	227.5	215.6	201.3	208.2	206.6	209.7	206.5	205.0
Total Fertility Rate	7883.0	7024.0	7137.0	6734.0	6215.0	6384.0	6342.0	6391.0	6276.0	6261.0
Gross Reproduction Rate	3882.0	3316.0	3546.0	3350.0	3130.0	3021.0	3204.0	3033.0	2980.0	3121.0

Table 5

Weighted Mean Interval Indices,
Currently Married Women by Year of Simulation
(in months)

	Unadjusted Means			Median Poole's Index	Life Table Median
	Open	Last Closed	All Closed		
Year	Moderate Acceptance Rates, 100% Effectiveness				
5	34.00	31.44	29.05	26.47	26.67
7	34.44	31.60	29.02	26.49	26.59
10	38.06	30.96	28.69	26.47	26.81
15	48.39	29.79	27.78	26.56	27.22
Year	High Acceptance Rates, 100% Effectiveness				
5	34.00	31.44	29.05	26.47	26.67
7	34.84	31.70	29.01	26.50	26.64
10	40.22	31.12	28.57	26.75	27.24
15	54.32	30.07	28.13	27.11	27.93
Year	High Acceptance Rates, 90% Effectiveness				
5	34.00	31.44	29.05	26.47	26.67
7	35.00	31.60	29.04	26.48	26.57
10	38.31	31.35	28.68	26.65	27.00
15	43.25	32.12	29.01	26.87	27.48

Table 6

Comparison of the Effect of Varying Sample Size,
High Acceptance, Rate, 100% Effectiveness Pattern of Contraceptive Use

Year	5						15					
Type of Index	(50 Replicates of 500)			(50 Replicates of 1000)			(50 Replicates of 500)			(50 Replicates of 1000)		
	$\bar{x}$	S.D.	C.V.	$\bar{x}$	S.D.	C.V.	$\bar{x}$	S.D.	C.V.	$\bar{x}$	S.D.	C.V.
<u>Age Specific Rates</u>												
15-19	175.6	35.6	20.27	178.1	24.4	13.70	163.9	33.3	20.32	154.6	24.8	16.04
20-24	365.3	49.8	13.63	361.2	37.4	10.35	275.3	43.2	15.69	289.8	33.8	11.66
25-29	395.4	52.7	13.33	400.6	35.2	8.79	251.6	37.7	14.98	247.4	34.2	13.82
30-34	308.3	47.9	15.54	301.4	34.9	11.58	162.9	51.7	31.74	176.4	34.7	19.67
35-39	230.0	66.2	28.78	223.4	36.8	16.47	102.1	46.7	45.74	97.9	22.2	22.68
40-44	100.5	45.2	44.98	95.6	25.1	26.26	38.4	25.7	66.93	36.9	13.7	37.13
45-49	12.9	19.4	150.40	10.9	10.6	97.25	1.4	5.7	407.14	2.7	5.5	203.70
General Fertility Rate	245.0	19.5	7.96	240.5	11.6	4.82	162.7	16.8	10.35	163.7	9.3	5.68
Total Fertility Rate	7938.3	653.0	8.23	7856.1	401.4	5.11	4978.3	556.9	11.19	5028.4	333.6	6.63
<u>Interval Measures</u>												
Open	33.66	2.40	7.13	34.40	1.51	4.39	54.23	2.28	4.20	54.03	1.68	3.11
Last Closed	31.19	0.95	3.05	31.23	0.69	2.21	29.99	0.90	3.00	30.04	0.63	2.10
All Closed	28.82	0.57	1.98	28.81	0.39	1.35	28.07	0.60	2.14	28.11	0.42	1.49
Median Poole's Index	26.35	0.49	1.86	26.36	0.35	1.33	27.19	0.57	2.10	27.11	0.37	1.36
Life Table Median	26.30	0.68	2.59	26.41	0.51	1.93	27.80	0.70	2.52	27.85	0.45	1.62

Table 7

Comparison of Sampling Distribution Characteristics for Selected Time Points,
High Acceptance Rates, 100% Effectiveness Pattern of Contraception Use
(50 Replicates of 500)

Year	5			7			10			15		
Type of Index	$\bar{x}$	S.D.	C.V.	$\bar{x}$	S.D.	C.V.	$\bar{x}$	S.D.	C.V.	$\bar{x}$	S.D.	C.V.
<u>Age Specific Rates</u>												
15-19	175.6	35.6	20.27	162.4	36.1	22.23	161.0	29.0	18.01	163.9	33.3	20.32
20-24	365.3	49.8	13.63	318.2	48.8	15.34	315.5	46.0	14.58	275.3	43.2	15.69
25-29	395.4	52.7	13.33	366.1	46.7	12.76	281.2	42.8	15.22	251.6	37.7	14.98
30-34	308.3	47.9	15.54	320.1	52.8	16.49	229.9	52.3	22.75	162.9	51.7	31.74
35-39	230.0	66.2	28.78	200.0	59.3	29.65	142.9	44.4	30.83	102.1	46.7	45.74
40-44	100.5	45.2	44.98	70.0	35.0	50.00	52.2	26.8	51.34	38.4	25.7	66.93
45-49	12.9	19.4	150.40	11.2	15.3	136.61	1.1	5.2	472.73	1.4	5.7	407.14
General Fertility Rate	245.0	19.5	7.96	225.3	19.2	8.50	188.4	12.3	6.51	162.7	16.8	10.35
Total Fertility Rate	7938.0	653.0	8.23	7240.0	638.8	8.82	5920.0	398.7	6.73	4978.3	556.9	11.21
<u>Interval Measures</u>												
Open	33.66	2.40	7.13	34.47	2.56	7.43	39.92	2.34	5.86	54.23	2.28	4.20
Last Closed	31.19	0.95	3.05	31.88	1.10	3.45	30.67	1.08	3.52	29.99	0.90	3.00
All Closed	28.82	0.57	1.98	29.05	0.67	2.31	28.43	0.57	2.00	28.07	0.60	2.14
Median Poole's Index	26.35	0.49	1.86	26.55	0.52	1.96	26.74	0.55	1.93	27.19	0.57	2.10
Life Table Median	26.30	0.68	2.59	26.50	0.86	3.25	27.03	0.73	2.70	27.80	0.70	2.52

Table 8

Observed Differences in Fertility Indices, Estimated Standard Errors and Z Values for Various Periods

High Acceptance Rate, 100% Effectiveness

Type of Index	Year 5 Vs. Year 15			Year 5 Vs. Year 10			Year 10 Vs. Year 15		
	Observed Differences	S.E.[1]	Z	Observed Differences	S.E.[2]	Z	Observed Differences	S.E.[2]	Z
Age Specific Rates									
15-19	-16.9	34.42	-0.49	-21.8	46.56	-0.47	4.9	40.27	0.12
20-24	-99.4	56.50	-1.76*	-65.1	76.81	-0.85	-34.3	62.66	-0.55
25-29	-121.0	50.59	-2.39***	-101.7	70.49	-1.44	-19.3	58.10	-0.33
30-34	-133.1	48.09	-2.77***	-64.1	56.67	-1.13	-69.0	70.74	-0.98
35-39	-125.0	46.43	-2.69***	-78.8	81.42	-0.97	-46.2	65.17	-0.71
40-44	-52.6	28.24	-1.86*	-49.9	46.28	-1.08	-2.7	37.78	-0.07
45-49	-7.8	11.26	-0.69	-6.1	20.85	-0.29	-1.7	7.94	-0.21
General Fertility Rate	-80.40	17.54	-4.58***	-56.7	22.31	-2.54***	-23.7	19.75	-1.20
Total Fertility Rate	-2.78	0.59	-4.71***	-1.94	0.75	-2.59***	-0.84	0.68	-1.24
Interval Measures									
Open	20.32	2.33	8.72***	6.22	3.09	2.01**	14.10	3.47	4.06***
Last Closed	-1.37	1.09	-1.26	-0.32	1.51	-0.21	-1.05	1.38	-0.76
All Closed	-0.92	0.62	-1.48	-0.48	0.76	-0.63	-0.44	0.91	-0.48
Median Poole's Index	0.64	0.56	1.14	0.28	0.74	0.38	0.36	0.70	0.51
Life Table Median	1.26	0.79	1.59	0.57	0.80	0.71	0.69	0.90	0.77

High Acceptance Rates, 90% Effectiveness[2]

Type of Index	Year 5 Vs. Year 15			Year 5 Vs. Year 10			Year 10 Vs. Year 15		
	Observed Differences	S.E.[1]	Z	Observed Differences	S.E.[2]	Z	Observed Differences	S.E.[2]	Z
Age Specific Rates									
15-19	-7.5	50.55	-0.15	-24.0	44.30	-0.54	16.5	52.24	0.32
20-24	-69.1	62.76	-1.10	-61.7	71.44	-0.86	-7.4	63.63	-0.12
25-29	-50.1	77.16	-0.65	-82.2	72.56	-1.13	32.1	82.60	0.39
30-34	-51.5	81.16	-0.63	-32.9	67.53	-0.49	-18.6	72.71	-0.26
35-39	-62.5	73.03	-0.86	-51.6	80.16	-0.64	-10.9	71.38	-0.15
40-44	-59.9	51.40	-1.17	-52.7	59.12	-0.89	-7.2	51.65	-0.14
45-49	-2.9	26.28	-0.11	-7.8	19.47	-0.40	4.9	13.54	0.36
General Fertility Rate	-42.60	20.35	-2.09**	-46.30	22.99	-2.01**	3.70	24.43	0.15
Total Fertility Rate	-1.52	0.93	-1.63	-1.57	0.73	-2.15**	0.05	0.85	0.05
Interval Measures									
Open	9.25	3.57	2.59***	4.31	3.10	1.39	4.94	2.92	1.69*
Last Closed	0.68	1.44	0.47	-0.09	1.23	-0.07	0.77	1.28	0.60
All Closed	-0.04	0.88	-0.05	-0.37	0.86	-0.43	0.33	0.91	0.36
Median Poole's Index	0.40	0.73	0.55	0.18	0.65	0.28	0.22	0.78	0.28
Life Table Median	0.81	1.00	0.81	0.33	0.88	0.38	0.48	1.07	0.45

* significant at 5% level
** significant at 1% level
*** significant at 0.1% level

[1] Standard Errors estimated on the basis of 50 replications of 1,000 women aged 15-49.
[2] Standard Errors estimated on the basis of 50 replications of 500 women aged 15-49.

The Assessment of Stability of Factor Analytic Scales

Phyllis R. Sweet, University of Lowell

This paper relates one of a series of studies intended to develop and refine a paper- and pencil measure of perception of pain. The ultimate aim of these studies is to develop a battery of measures which can be administered in absentia to persons not in pain. This battery is intended to predict the subject's health orientation and care.

This study is a replicative study intended to assess the stability of factors found in a smaller (N=68) sample of the same population. A sample of 163 undergraduates enrolled in psychology courses for non-majors at state universities voluntarily completed the inventory. In the original study, twelve dimensions were isolated. With the larger sample size, 10 dimensions appeared to underlie the data. The Pain Perception Inventory (PPI) is the instrument used in both studies. Three options for response were allowed for all thirty items that made up the inventory. Subjects indicated their agreement, disagreement, or lack of opinion to items such as "I have experienced a great deal of pain".

Naming the factors (or scales) that result from a factor analytic study is a first step to determine stability. On a broad level, the ability to retain scale names and constructs is the sin qua non of factorial stability. However, the assignment of names is always tentative and, at least in the initial instrument stages, largely intuitive.

The process of assigning meaningful names to the factor scales, used for both the original study (factor analytic study I) and the replicative study (factor analytic study II), is given here in detail for whatever insight into this intuitive procedure such details may offer.

I. Factor analytic study I - Twelve scales.

a. Using SPSS, a principle components factor analysis with oblique rotation was undertaken, resulting in twelve factors.

b. The factor pattern matrix was analyzed to determine which items presented a factor loading of ± 0.500. These items were intended to define each scale. (Four items did not have a high enough factor loading to be assigned according to the criterion. These were assigned to that factor upon which the item had the largest weight, factor loadings = 0.37;0.40;0.42;0.43)..

c. A seperate table for each factor was compiled. Each set of items and their factor loadings, together with an explanation of direction of subjects response to them, made up the tables.

d. Each table was content analyzed in the attempt to assign a name to the scale. More often than not content analysis amounted to a sophisticated term to indicate that one key word within an item was selected as reasonably represenitive of the theme of the items taken together. Occasionally, no single theme appeared. In this case, either the highest loading item or the most prevelant theme was selected.

e. Scale names and sample items and their loadings were then organized. Descriptions of high scores were also prepared.

II. Factor analytic study II - Ten scales.

a. Steps a through c were repeated with the data collected for the second study. These resulted in ten factors.

b. In an attempt to preserve, to the extent possible, the names of the scales, specific items were traced from the original study to the replication. Five scale names were retainable. The majority of weight, judged by the magnatude of the comparitive factor loadings, on these scales resulted from the same items. For example, scale 5 from the original study was called "Delay". It was named from the items, "I would seek medical treatment if pain was bothering me for more than a week." This item had a factor loading of -0.853 in study I; it had the same loading in study II, although secondary items varied between the two studies. In order to ditinguish between this scale and another scale from study II also emphasizing delay, the name of scale 5 was qualified to add "short-term". The other scales retaining the same name and overall theme were:Tolerance, Hindrance, Empathy, and Cognizance.

c. The remaining 4 factors were named as previously. Names of the resulting scales and description of the meaning of scores of all ten scales are given below. (High scores are described except for bi-polar factors which are presented in terms of both sides of the factor.)

1. Medical utilization (includes original Inconsistency scale) - One should prompty seek medical care for pain although cost may be a facto considered. There is acknowledgement that some social pain-causing conditions exist. Both experience and defiinitional are included.

2. Endurance - a high score indicates that pain has some ultimate meaning. If still necessary after attempting to discover the meaning of pain, one should seek medical help. A low score indicates one should endure pain.

3. Long-term delay (inclddes original Definition scale) - A low score indicated recognition of the existence of pain but no seeking medical advice unless forced. A high score does not associate pain with misery. Presumably, therefore, the high scorer would not have to be force to seek medical care.

4. Tolerance - Once one is in pain there is no escape, no toleration. Therefore, one attempts to avoid pain.

5. Short-term delay - One should give pain a chance to go away by itself; then, if necessary, seek medical treatment.

6. Stoicism - One does not complain when one is in pain.

7. Hindrance - (includes the original Helplessness scale) - A low score indicates that pain is viewed as restrictive. High scorers believe that medical help is not necessarily sought for phisical pain nor is such pain found enjoyable. People are not the cause of pain. High scorers on this scale may present a sport-connected ideology whereby pain is viewed as a warning (which may not be heeded) to stay outof the game or practice. Trainer care (for male athletes at least) rather than medical care would be sought and the cause of pain would be physical rather than social.

(This scale will be expanded and hypotheses related to sports activity will be tested in a future study.)

8. Empathy - Pain is the physical sensation which presents absence on comfort. One can learn what pain is like by observation of others in pain.

9. Intellectualization - Low scores indicate pain does not disrupt consentration. This may be related to pain tolerance itself, that is, a low score on the intellectualization factor may be related to high pain tolerance. "Your attitude toward pain can control the sensation to a large extent" is the meaning of high scores.

10. Cognizance - Although pain is an emotional sensation, one tries to ignore it's presence.

The naming procedure is essentially a straightforward process. In attempting to determine stability, the process may be compicated by the imposed condition to retain as many constructs and associated names as possible. There was an obvious lack of stability of factors with increasing sample size when stability is viewed through the naming precedure.

Another possibility in assessing stability is to look at what percentage of variance is accounted for by stable factors as compared to non-stable factors. The stable factors from study I accounted for 33.0% out of a total of 73.2% for the twelve factors extracted. The stable factors for study II accounted for 22.5% out of 64.3%. Thirty-five percent of the variance accounted for by the ten factors extracted is attributable to stable factors, whereas the stable factors represent 45.1% of the original 12 factors accounted for in study I. The factors extracted first in these factor analysis procedures were less stable. Since they accounted for a greater percent of the variance, to the extent they were not replicated, the stable factors ration reported above will decrease.

Obviously factor stability is a function of sample size. How markedly will factors change if the study is repeated with a new population? With the number of factor analytic studies being done in educatinnal psychology today, to what extent can we count on the conclusions drawn?

Would these factors change again if once more data was collected on the same population? How stable would these factors be across populations? How many more items and subjects are needed for the joint conditions of stable constructs and reliable scales? These questions are still unanswered, leaving only a partial approach to the assessment of stability. Further research is necessary.

CLUSTER ANALYSIS ON AGE PATTERNS OF FERTILITY

Prem P. Talwar and Peter A. Lachenbruch, International Fertility Research Program

1. Introduction

The age-specific fertility rate curve, in general, is a bell-shaped, unimodal curve which first rises slowly and then sharply in the age group 15-19, attains its modal value somewhere between ages 20-29 and then declines, first slowly and then steeply, until it approaches zero around the age of 45 years. Its two dimensions are the area and the shape. The demographic term for the former dimension, area, is the total fertility rate or the completed family size, according to whether period data or cohort data are utilized for computation of such rates. The latter dimension, namely, the shape of the curve, may be termed the fertility pattern or the pattern of reproduction or the age pattern of fertility.

The shape of the age-specific fertility rate curve (subsequently referred to as ASFR) has several useful applications in demographic analysis (1,2,3,8). In a study of relationship between different fertility indices, total fertility rate has a one-to-one relationship with either crude birth rate or general fertility rate for a given age pattern of fertility (shape of the age-specific fertility rates). In comparing the levels of fertility at two time periods or between two or more populations, it may be necessary to control for differences in age patterns of fertility. Coale and Tye (2), in an attempt to explain differences in the levels of fertility in two ethnic groups in Singapore, had to consider fertility patterns of births. In population projections, where total number of births are needed for the future time periods, it may be essential to consider changes in the age pattern of fertility, especially in populations where fertility has shown decline.

Despite the need for considering age pattern of fertility in demographic analysis, the only detailed study devoted to this topic is one by the United Nations (10). They noted the variations in terms of two basic characteristics of the curves: the peak-fertility age group and the degree of concentration around this peak and determined some basic fertility patterns in which most of the populations belong. This study attempts to determine such basic fertility patterns by using more information from the individual curves. The goals of this study are: 1/

(1) to investigate the extent to which fertility patterns vary among populations, and

(2) to investigate whether basic age patterns of fertility exist, and if so, what they are.

2. Sources of Data

The United Nations has compiled data on fertility rates specific for quinquennial age groups for 73 populations in and around the year 1960. This study uses their data (10). The UN examined the data for various inaccuracies and adjusted for most of them. For 36 low-fertility populations (TFR < 4.12) 2/, the numerator of the ASFR was derived from registration statistics on births by age of mother. It was divided by the official estimates of age-specific females population to get the fertility rates. The registration statistics were reported to be complete in these countries except for Greece where under-registration was not extensive. The 37 high-fertility populations have TFR above 4.12 and their data are of satisfactory quality for the purposes of this study. Data of doubtful accuracy were omitted.

For the present study the data on period ASFR's given in the UN publication were converted to age-specific relative fertility rates (ASRFR) by dividing them by the sum of all ASFRs (equalizing the area for each fertility curve). In other words, if $g_x\,dx$ is the area of ASFR curve between x and $x + dx$, f_x is defined as the proportion of the total fertility achieved in age x and x + dx and is obtained by making total area under the curve equal to unity. In mathematical notations

$$f_a = \frac{\int_a^{a+5} g_x\,dx}{\int_{w_0}^{w_1} g_x\,dx}$$

w_0 and w_1 are the beginning and end ages of reproductive life

which, in case the fertility rates are specific for five age groups, becomes

$$f_a = \frac{5g_a}{5\sum_1^7 g_a}$$

This transformation was done in order to investigate primarily variations in fertility patterns or age patterns of fertility. In the further description f_x will be referred to as age-specific relative fertility rates (ASRFR).

3. Variations in Individual Fertility Patterns

A large variation was observed in the individual age patterns of fertility. No two populations had exactly the same pattern; though there are some similarities in the patterns among different populations. The mean relative fertility rates for the 73 populations under study are 8.4, 26.7, 27.4, 19.5, 21.1, 4.8 and 1.0 and are different from the relative fertility rates observed in individual populations. For instance, the mean relative fertility rate for the age group 15-19, is 8.4 compared to the individual population rates which range from 1.0 for Ireland to 21.4 for Gabon. The coefficients of variantion calculated for each age group reveal that groups 15-19, 40-44 and 45-49 have relatively larger variations. The larger variation in the earlier part of the reproductive life can be attributed to variations in the patterns of marriage and in patterns and incidence of illegitimacy among countries. That in thè latter part is due to several biological and socio-cultural factors which affect termination of the reproductive life.

The fertility pattern of a high-fertility country is likely to be different from that of a low-fertility country because both these groups of countries differ not only in their family size norms but also in the effects of biological and socio-cultural factors. Hence the first part of the study considered whether all 73 populations should be considered together or in sub-groups. For such study the populations were grouped into two sub-groups--those with TFR greater than 4.12 and those with less than that. This grouping was adopted in view of the U.N. observation (10) that no other social or economic indicator separates countries so well into two groups as does the TFR level of 4.12. Step-wise discriminant analysis was adopted to obtain an equation which best discriminated fertility patterns of these two groups. A posterior classification was obtained on the basis of this equation. It was found that fertility patterns of high-fertility countries are different from those for low-fertility countries. The fertility pattern of a low-fertility population typically has lower relative fertility rates at ages 45-49 and 15-19 and higher relative fertility rates at ages 25-29 and 20-24. The typical fertility pattern for a low-fertility country is 7, 29, 30, 19, 11, 3.5 and 0.5 for seven quinquennial age groups as compared to 10, 24, 24, 20, 14, 6 and 2 for a high-fertility. Since the fertility pattern for a high-fertility country is different than for a low-fertility country, the two groups were considered separately to determine basic age patterns for each.

4. Approach to the Problem

The United Nations attempted to determine basic age patterns on the basis of two characteristics of the curve: (1) modal fertility age group and (2) spread around the modal age group. We will use cluster analysis to apportion the total number of age patterns of fertility into the "best" number of homogeneous disjoint subsets(4). The average pattern of each subset is the basic age pattern for that group of countries.

Cluster analysis identifies homogeneous classes or groups in a larger set of data (7). The first step in its use is to choose a similarity index which provides a measure of closeness between two units. Such an index is calculated for all pairs of units. In the second step clustering algorithm is used to form a few clusters so that the units in one cluster are more similar than those in different clusters. A "stopping rule" is ultimately needed to terminate the clustering process. The third step in the cluster analysis applicable in our case, is to choose the "best" set of clusters if there are more than one set of clusters formed by taking more than one similarity indexes.

5. Cluster Analysis on Age Patterns of Fertility

a. Similarity Index

The indices of similarity were defined by subtracting distances between the fertility patterns of two populations from the maximum possible distance. Two types of distances between populations i and j considered in this study are:

$$d_{ij}^2 = \frac{1}{7}\left[(f_i - f_j)^T(f_i - f_j)\right]$$

$$D_{ij}^2 = \frac{1}{7}\left[(F_i - F_j)^T(F_i - F_j)\right]$$

where T = Transpose of the matrix

f_i = Column vector of relative fertility rates for seven quinquennial age groups for country i

These distances were converted to measures of similarity by subtracting them from

$$D^2 = \frac{100^2+100^2}{7} = \text{The maximum distance between two fertility curves.}$$

Thus the similarity indices, for our analysis, were defined as

$$s_{ij} = D - d_{ij}$$

$$S_{ij} = D - D_{ij}$$

The first index considers the discrepancy in fertility pattern curves in each age group, while the second considers discrepancy in cumulative reproduction up to different age groups. In our analysis, we identified these similarity indices as those based on simple distance and on cumulated distance respectively. The clusters formed by these two similarity indices for low-fertility and high-fertility countries are discussed separately.

b.1 Process of Cluster Formation

The cluster formation process consisted of two parts. The first part calculated the (n x n) symmetric matrix S of similarity coefficients between every pair of units. The second part formed clusters by adopting agglomerative algorithm (4). Specifically, the cluster formation program proceeded in the following fashion:

(1) Initially, the n units were considered to constitute n clusters. Each unit was designated by the subscript i (i = 1,2 ......n)

(2) In the first step, the (n x n) symmetric similarity matrix was scanned for the maximum value of S_{ij}, i<j. We denote $\max_{\text{all } i<j} (S_{ij}) = S_{ij}^*$ Then S_{ij}^* became the level of similarity at which clusters were formed at this step.

(3) In case the value of the similarity coefficient between units i and k (k = 1, 2.....n; k ≠ j) or between units j and k were tied with S_{ij}^*, then random selection decided

whether a cluster was formed by ($\underline{i},\underline{j}$), ($\underline{i},\underline{k}$) or ($\underline{j},\underline{k}$) units. For this example, let us assume that the units $\underline{i}$ and $\underline{j}$ clustered together.

(4) The group which clustered $(\underline{i},\underline{j})^{th}$ units together is represented by the i^{th} unit (note: $\underline{i} < \underline{j}$)whose row and column elements were replaced by those of

Min (S_{ik}, S_{jk}); $(k = 1,2,....n; \neq i \text{ or } j)$
Elements corresponding to unit $\underline{j}$ were replaced by those of the $\underline{n}^{th}$ unit. Thus each cluster was represented by a unit in the next step of the similarity matrix whose dimension was reduced by one.

(5) The value of S^*_{ij} and various clusters and their constituent units were recorded

(6) Steps (2) to (5) were repeated until all the units were included in one cluster.

This algorithm was used to form two sets of clusters from two indexes of similarity considered here. The two indexes are the simple distance and the cumulated distance between two units. It was noted that the process of clusters formation was slower for the low-fertility counties for both these sets. This might be an indicator of more heterogeneity in the fertility patterns among the low-fertility countries because of greater control on their reproductive lives.

b. 'Stopping rules' for decision on the optimum number of clusters

The clustering process started from the stage when each unit was an individual cluster to the last stage when all units formed one cluster. Some rule was needed to decide on the optimum number of 'homogeneous' clusters. One possible approach is to use distribution theory, although fully aware that it is not applicable in the usual statistical sense. This approach looks at some statistics of within-cluster and between-cluster variations at each step of the clustering process.[3/] Though these statistics are generally used to test the statistical significance, they were used here to study their changed pattern when the clustering process passes from one stage to the next. An abrupt change in the expected changing pattern can be suggestive of a stopping stage.

The statistics based on between clusters and within clusters variations considered but not pursued further were (1) those generally used in multivariate testing--largest root, trace or the likelihood criterion, and (2) multivariate outlier test statistics, because of smaller number of units per cluster compared to the number of variates (six corresponding to seven quinquennial age groups). The other statistics computed at different stages of the clustering process and whose results were used to decide on the optimum number of clusters will be described here. The purpose of looking at several statistics was to check the consistency in decision about the stopping stage since no one statistic appeared to be better than the other. Two types of measures were used: (1) statistics based on unweighted distance, and (2) statistics based on weighted distance. The word distance has been used here though analogy with univariate analysis of variance is very close. Like analysis of variance, total variance is being partitioned into within-cluster and between-clusters. The first type of measure is F-ratio test in the analysis of variance and the second measure is similar to the generalized $\underline{D}^2$ given by Rao(6).

The measure based on unweighted distance at a step when there were $\underline{g}$ clusters is defined as

$$UD_1 = \frac{(B/(g-1))}{W/(n-g)}$$

where

W = overall within-cluster variance = $\sum_{k=1}^{g} W_k$

and

W_k = within-cluster variance for k^{th} cluster and is calculated as

$$\sum_{k=1}^{n_k} (\underset{\sim}{f}_i^{(k)} - \underset{\sim}{\bar{f}}^{(k)})^T (\underset{\sim}{f}_i^{(k)} - \underset{\sim}{\bar{f}}^{(k)})$$

where $f_i^{(k)}$ = column vector for relatively fertility rates for seven age groups for country i which belongs to cluster k, and

$\bar{f}^{(k)}$ = column vector for mean relative fertility rates for seven age group in cluster k.

T = transpose of the column vector, and

n_k = number of countries in cluster k

B = overall between-cluster variance calculated as

$$\sum_{k=1}^{n_g} n_k (\underset{\sim}{\bar{f}}^{(k)} - \underset{\sim}{\bar{f}}^{(\cdot)})^T (\underset{\sim}{\bar{f}}^{(k)} - \underset{\sim}{\bar{f}}^{(\cdot)})$$

where $\bar{f}^{(\cdot)}$ = column vector for overall mean relative fertility rates for seven age groups.

g = number of clusters in which n countries have been grouped with $n_1, n_2 \ldots n_g$ countries in each cluster

n = total number of countries

The measure based on weighted distance at a step when there are $\underline{g}$ clusters is defined as

$$WD = \frac{WD_1}{(g-1)\ (p-1)}$$

where

$$WD_1 = \sum_{k=1}^{g} n_k\ (\bar{\underset{\sim}{f}}^{(k)} - \bar{\underset{\sim}{f}}^{(.)})^T\ (S_w^{-1})(\bar{\underset{\sim}{f}}^{(k)} - \bar{\underset{\sim}{f}}^{(.)})$$

$$\text{and } S_w = \sum_{k=1}^{g} \frac{\sum_{i=1}^{n_k} (\underset{\sim i}{f}^{(k)} - \bar{\underset{\sim}{f}}^{(k)})(\bar{\underset{\sim i}{f}}^{(k)} - \bar{\underset{\sim}{f}}^{(k)})^T}{n-g}$$

P = number of independent variates = 6 in our case

Obviously, the larger values of $\underline{UD_1}$ or $\underline{WD}$ would mean larger distance between clusters and smaller distances within clusters. Thus larger values of these indices are preferred for the choice of stopping stage.

These two measures were calculated at each stage of the clustering process. It may be noticed that both these measures consider the number of clusters at each stage which is essential since these measures are calculated at different stages of the clustering process when there are different numbers of clusters. Tables 1 and 2 give values of $\underline{UD_1}$ and $\underline{WD}$ at different stages of the clustering process.

Table 1 gives results for the case when the simple distance was taken as the similarity index and Table 2 when the cumulated distance was used. In general the values of UD_1 and WD in Tables 1 and 2 decline with the decline in the number of clusters. But our choice of the "best" number of clusters was based on two desirable properties. First, the number of clusters should not be too large and secondly, the within clusters distance should not be too large (or between clusters distance should be large)i.e. larger value of the indices were to be preferred. These two considerations suggest five clusters for the low-fertility countries from Table 1. Both UD_1 and WD suggest a great increase in the within clusters variance compared to between clusters if one considers a choice of less than five clusters (sharp decrease in UD_1 and WD). The choice has to be on eight clusters if the within clusters variance has to be further reduced. The same considerations suggest three clusters for the high-fertility countries in Table 1.

In the case of Table 2, $\underline{UD_1}$ suggests seven clusters for the developed countries but WD suggests six, although it seems that the value of $\underline{WD}$ remains stable from step 3 to 6 except for random fluctuations. Hence, the choice in this case was seven clusters for the low-fertility countries. For the high-fertility countries, the value of $\underline{UD_1}$ remains stable from steps 4 to 7, while that of $\underline{WD}$ is stable from steps 3 to 6 except for the random fluctuations. Hence, the choice was three clusters for the high-fertility countries.

In reality, there are only five clusters for the low-fertility countries based on Table 2, since Japan and Ireland have unique fertility patterns and form clusters of their own. The same argument suggests three clusters for the high-fertility countries in Table 1. The basic pattern for this cluster is obtained by averaging the fertility patterns of all populations forming that particular cluster. Although some countries belong to the same cluster on the basis of the two similarity indices considered here, there are other countries which belong to different clusters if different similarity indices are used.

c. The Choice of "best set of age patterns of fertility"

The choice of the two similarity indices lead to the two sets of age patterns of fertility both for the low-fertility and the high-fertility countries. The next decision was to make a choice of the best set out of the two available. The first set based on simple distance as a measure of similarity had five fertility patterns for the low-fertility and three for the high-fertility countries. The second set (based on cumulated distance as a matter of similarity) had seven fertility patterns for the low-fertility and three for the high fertility countries.

The choice of the "best" set was based on statistical and demographic considerations. It was desired that the basic fertility patterns should be distinct. In other words, countries within a pattern (cluster) should be more homogeneous than those in different patterns. Hence the statistical considerations were based on smaller within-clusters distance or larger between-clusters distance. The demographic considerations required that the "best" set should cover extreme shapes of the age-specific relative fertility curves. The following paragraphs discuss the results of various statistical and demographic techniques utilized for investigations into the choice of the "best" set. A method based on ranks of the pair-wise distances was also used for this purpose.

Step-wise discriminant analysis was used to study whether different fertility patterns within a set were distinct and non-overlapping. In Set I all low-fertility countries were correctly classified in the three clusters; only one of the 37 high-fertility countries was misclassified; it had 0.45 posterior probability of belonging to the correct group. For Set II, all high-fertility countries were correctly classified, but one of the low-fertility countries was misclassified with 0.37 probability of correct classification. If the number of misclassified cases and their posterior probabilities are taken as an index of clear-cut and distinct fertility patterns in a set, then Set I seems to be slightly better by this analysis.

A statistical test for outliers was used to determine whether all countries forming a cluster have come from some basic pattern. It provided a test for homogeneity of all fertility patterns in a cluster. In its application, we have assumed that the basic distributional assumptions are fulfilled by our data. Wilks' (11) test for multivariate outliers was applied to individual clusters within the two sets[4/]. In Set I, no outliers were found. In Set II, there was one outlier. If the number of outliers is taken as an index of heterogeneity of clusters, then this test, too, tends to favor Set I.

Demographically, the fertility patterns should cover the extreme shapes since they have effect on the population birth rates and the growth rates. Stable population parameters were estimated for each fertility pattern within the different sets under consideration. It is found that the fertility patterns in Set I covered the larger range of the stable population parameters.

The rank order of distances method considers all the $\binom{n}{2}$ pair-wise distances between countries. The principle underlying this technique is that two countries with shorter distances between them are more likely to belong to the same cluster. That is, pair-wise distances between units in the same cluster should be smaller than between those in different clusters. Let there be g clusters C_1, $C_2 \ldots C_g$ with $n_1, n_2 \ldots n_g$ number of units such that $n = \sum_{k=1}^{g} n_k$. The total number of pair-wise, within-cluster distances are given by

$$N = \binom{n_1}{2} + \binom{n_2}{2} + \ldots\ldots\ldots + \binom{n_g}{2}$$

where N is less than $\binom{n}{2}$ unless g = 1.

Now if all $\binom{n}{2}$ distances are ranked in an array of descending order then ideally, the last N distances should be from countries which belong to same clusters. Operationally, all the $\binom{n}{2}$ distances were ranked in descending array; 'S' was marked for distance between those countries both of which were in the same cluster and 'D' for those in the different, then the per cent of S's in the last N distances were taken as an index of distinct and clear-cut grouping in clusters within a set. The ideal value of this index is 100. This index was calculated for both the sets; Table 3 shows the results for the developed and the developing countries. Set I shows the largest percentage of S's though they are far from the ideal value of 100 per cent. Thus this method also suggests Set I as the "best" choice.

Different techniques adopted in the investigation on the "best" set have generally shown results in favor of Set I.

e. Results:

The analysis discussed above shows that the basic age patterns of fertility, i.e. percent of the total fertility rates accounted for by different quinquennial age groups, for the developing countries are:[5/]

Pattern I= UP_1: 19.7, 24.1, 21.2, 16.6, 11.5, 4.8, 2.2

Pattern II= UP_2: 12.0, 25.9, 24.1, 18.4, 12.6, 5.4, 1.6

Pattern III = UP_3: 5.9, 21.7, 25.8, 22.2, 15.7, 7.0, 1.6

and those [6/] for the developed countries are given as follows:

Pattern I = DP_1: 13.1, 37.9, 26.6, 13.6, 6.6, 2.0, 0.2

Pattern II = DP_2: 6.5, 30.9, 30.4, 18.8, 9.7, 3.3, 0.3

Pattern III = DP_3: 3.4, 23.4, 31.5, 22.6, 13.6, 5.0, 0.6

The subscripts 1,2 and 3 have been assigned with a view of suggesting typology to the basic fertility patterns. The patterns with subscripts 1's (UP_1 and DP_1) had lower mean, median and modal ages at reproduction compared to the other subscripts in the same group(U or D).

The results obtained by cluster analysis technique were confirmed when individual fertility patterns were graphically represented by

(a) Parameters of a mathematical curve fitted to the fertility patterns(5), and

(b) First two factors in scores in factor analysis on six-variate fertility patterns

Different basic age patterns of fertility occupied different and non-overlapping sub-spaces and thus reinforced the fertility patterns obtained in our earlier analysis.

6. Summary

Age-specific fertility rate data for the period around 1960 from 36 low-fertility and 37 high-fertility countries were converted to age-specific relative fertility rates by equalizing the area for each fertility curve. This was done primarily to investigate variations in fertility patterns.

It was found that the fertility patterns for the high-fertility countries were different from those for the low-fertility countries. Thus these two groups of countries were considered separately in the investigation of the basic age patterns. A cluster analysis technique was utilized to determine a "reasonable" number of patterns. It was found that both the low-fertility countries, excluding Japan and Ireland, and the high-fertility countries could be grouped into three basic patterns each. Their basic patterns, i.e. percent of the total fertility rates accounted for by different quinquennial age groups, are given in the paper.

TABLE 1

Some Descriptive Statistics Calculated at Different Steps, to Determine the Stopping Stage, of the Clustering Process with Simple Distance as the Similarity Index, for the Low-Fertility and the High-Fertility Countries

Low-Fertility Countries				High-Fertility Countries			
Step	Number of Clusters	Unweighted Distance[a] (UD_1)	Weighted Distance[a] (WD)	Step	Number of Clusters	Unweighted Distance[a] (UD_1)	Weighted Distance[a] (WD)
1	19	46.2	35.0	1	21	25.2	25.9
2	11	45.2	24.1	2	11	17.4	13.4
3	8	41.5	20.9	3	6	22.2	15.5
4	6	36.7	19.5	4	3	28.1	23.1
5	5	39.2	19.4	5	2	26.2	17.2
6	3	20.6	9.1				
7	2	30.0	10.4				

[a]See text for the definitions of these notations.

TABLE 2

Some Descriptive Statistics Calculated at Different Steps, to Determine the Stopping Stage, of the Clustering Process with Cumulated Distance As the Similarity Index, for the Low-Fertility and the High-Fertility Countries

Low-Fertility Countries				High Fertility Countries			
Step	Number of Clusters	Unweighted Distance[a] (UD_1)	Weighted Distance[a] (WD)	Step	Number of Clusters	Unweighted Distance[a] (UD_1)	Weighted Distance[a] (WD)
1	21	38.8	59.9	1	25	21.5	55.2
2	13	36.3	54.6	2	12	18.9	24.4
3	10	38.7	22.2	3	9	16.0	27.3
4	8	38.8	27.5	4	6	24.8	34.5
5	7	38.8	22.3	5	4	24.2	23.5
6	6	26.8	22.7	6	3	21.4	27.5
7	5	27.4	17.1	7	2	24.5	10.6
8	4	26.7	11.9				
9	3	29.7	10.7				
10	2	27.8	9.5				

[a]See text for the definitions of these notations.

TABLE 3

Number of S's And D's in the Last N of the $\binom{n}{2}$ Pair-Wise Ranked (Descending) Distances by Sets of Clusters, for the Low-Fertility and High Fertility Countries.

Country Type		Set No. I	Set No. II
Low-fertility	N	213	136
Percent S of total N		81.2	72.1
High-fertility	N	284	252
Percent S of total N		75.7	65.9

FOOTNOTES

1/ This is a part of the first author's Ph.D. dissertation, which goes beyond what has been reported here. Various factors responsible for differential fertility patterns have been studied. Many demographic applications have also been shown.

2/ The U.N. divides all 73 populations into two groups, those with GRR>2.0 and those with GRR < 2.0. We translated these into total fertility rates (TFR).

3/ The cluster analysis program groups n units into a few clusters and the total variance into two components: between-clusters variation and within-clusters variation. Each successive step of the program tends to increase the within-clusters component (pooled within variance) and thus reduces the one corresponding to the between-clusters variation.

4/ A cluster must have seven or more countries for this test to be applicable.

5/ "U" stands for underdeveloped populations. Similarly "D: will be for the developed ones.

6/ Japan and Ireland are omitted, since both of them have unique fertility patterns and form clusters of their own. Japan's unique fertility pattern is explained by a very high incidence of abortion. A very high age at marriage and a very high incidence of spinsterhood in Ireland is responsible for its unique fertility pattern.

REFERENCES

1. Coale, A.J.(1965)"Birth rates, death rates and rates of growth in human population". Public Health and Population Change. Eds. Mindel C. Sheps and Jeanne C. Ridley. University of Pittsburgh Press, Pittsburgh,242-65.

2. Coale, A.J. and Tye, C.Y.(1961). The Signifiof Afe Patterns of Fertility in High Fertility Population. Milbank Memorial Fund Quarterly 39(4), 631-46.

3. Coale, A.J. and Zelnik, M. (1963). New Estimates of Fertility and Population in the United States. Princeton University Press. Princeton.

4. Gower, J.C.(1967)"A Comparison of Some Methods a Cluster Analysis", Biometrics 23(4).

5. Johnson, N.L. (1949). Systems of Frequency Curves Generated by Methods of Translation. Biometrika 36(1&2), 149-76.

6. Rao, C.R. (1952). Advanced Statistical Methods in Biometric Research. John Wiley & Sons, New York.

7. Sokal, R.R., and Sneath P.H. (1963). Principles of Numerical Taxonomy. W.H. Freeman and Company, San Francisco and London.

8. Talwar, P.O.(1970). Age Patterns of Fertility. Institute of Statistics Mimeo Series No. 656, University of North Carolina, Chapel Hill, North Carolina.

9. Talwar, P.P. (1971). "Variations in Fertility Patterns Among Populations". Presented in All-India Seminar on Demography and Statistics, Demographic Research Center, Banaras Hindu University, Varanasi, April 23-25, 1971.

10. United Nations (1965). Population Bulletin of the United Nations No. 7, 1963: Conditions and Trends of Fertility in the World. New York

11. Wilks, S.S. (1963). "Multivariate Statistical Outliers". Sankhya Series A, 25(1).

QUALITY CONTROL OF COMPUTERIZED GEOGRAPHIC INFORMATION

Kathryn Thomas, William C. Davie, John S. Linebarger
Bureau of the Census

A street map is a graphic representation of the physical features which can be found in our environment. Computerized information files have been developed which contain data and allow the user to associate this data to its specific geographic location in much the same way the use of a map enables us to relate to our physical environment. Combined into one package, these planning tools, maps and computer files become an invaluable data reference source. A geographic base file, then, may be defined as a reference file designed to relate various types of factual data to their geographical location.

In conjunction with the 1970 Census of Population and Housing, the Census Bureau began developing a system whereby a computer could pinpoint the location of a specific address in relation to other addresses, its street location, and so forth. This system -- called the Geographic Base (DIME) File, or "GBF/DIME" -- provided a fast and accurate means of locating addresses to their geography.

The Bureau's Geographic Base File System, currently undergoing expansion to include most major metropolitan areas, was developed jointly by the Bureau of the Census and local area governments and/or planning organizations. The technique for creating the Geographic Base File is a method of translating geographic information such as street names and address ranges from maps and other source materials into a form which can be read by a computer.

The GBF/DIME File is composed of segment records. A segment is defined as a length of a street or other feature between two distinct vertices, or nodes. Other features include imaginary lines defining political or other boundaries; topological features such as rivers and shorelines; other physical map features such as railroad tracks; and any other feature defining a boundary. Nodes are points where features begin, end, intersect, or curve sharply. The Geographic Base File incorporates the DIME (Dual Independent Map Encoding) feature, which gives an identification number at each end of a segment and includes both sides of the street in a single record. Using the node identifications, it is possible to link records together to form a single block. This feature helps in identifying inconsistencies in the information, such as when the computer cannot identify all segments bounding a block. Edits have been formulated to detect such errors. The Bureau provides these edits to local agencies so that the inconsistencies in a file can be eliminated to help improve the file's accuracy.

In order for the file to remain useful over time, the Bureau recommends that local agencies have a regular program for updating and maintenance. The Bureau of the Census has established the CUE program -- a nationwide, standardized approach to the Correction, Update, and Extension of the Geographic Base File/Dual Independent Map Encoding (GBF/DIME) System. The Bureau provides local agencies with clerical procedures, processing methodology, and the computer programs necessary to carry out the CUE operations.

Currently, no formal quality control program exists for the CUE operation, although such a program presently is being tested as a result of recent research discussed in this paper. The Bureau has relied heavily on the local areas to do a high-quality job and on various computer edits to uncover inconsistencies in the file. A formal QC program to evaluate available local reference materials does exist for areas initially setting up a file, but no quality checks are done after the initial reference material evaluation to insure continued updating and high-quality output by every local area.

The purpose of this paper is to discuss the evaluation studies that were undertaken in a few selected areas to determine the adequacy of the creation and correction plans currently in use and to determine whether there was a need for quality control plans for the CUE operation. One site, City A, was chosen for a study to identify phases of the operation where quality control procedures are needed in the "create" portion of the GBF/DIME System program. In order to identify particular phases of the operation where quality control procedures were needed in the CUE program, a separate evaluation was conducted in a different city, City B. Based on these studies, recommendations on quality control requirements were made.

City A Evaluation 1/

The evaluation of the "create" portion of the GBF/DIME System file was designed with two distinct objectives in mind. One of these was to evaluate the quality of current local reference materials and to verify the blockside error rate estimated from the quality control procedures used in "creating" the file in April of 1974. The procedures employed and analysis of resultant data from this portion of the evaluation study are described below. Further objectives of this geographic evaluation were to compute a segment error rate and an address error rate for the currently existing City A GBF/DIME System file. These results and a description of the methods employed in obtaining them are also described in this document.

1/ It should be noted that evaluation in only one area is inadequate to conclude about the quality of reference materials in all areas. However, this evaluation did produce information that was helpful in planning the QC program. Later evaluations from applications in different areas will give more adequate readings on the quality of reference materials.

Evaluation of Local Reference Materials in City A

The study consisted of systematically selecting (following a random start) a 1-in-40 sample of blocks from a listing of all blocks in City A. This sample was the identical sample of blocks used in evaluating the quality of local reference materials during the creation of that particular GBF/DIME System File in 1974.

The reference materials were assigned one blockside error whenever one or more address units existing on the blockside could not be located on that blockside via the 1975 reference materials. The results of checking 42 sample blocks containing 175 blocksides yielded a blockside error rate of 2.9 percent with 95 percent confidence limits of 0.3-5.5 percent. No comparison was made to the GBF/DIME System file.

Five blocksides were found to differ when comparing the findings on the ground with the set of 1975 reference materials. Four of the errors were lower and/or higher addresses found on the ground than shown in the reference material. The fifth error was a blockside having odd-even street numbers reversed as to the side of the street on which they were located.

Table A1 below shows the distribution of the sample and error rates for each of the 10 map sheets for City A. Confidence limits in the table were based on the binomial probabilities assuming simple random sampling. When consideration was given to the cluster effect, similar estimates were obtained. The intraclass correlation2/ was calculated at -0.023, indicating that there was little clustering of errors within a block.

Based on this evaluation, the overall quality of the reference materials was assessed as adequate for use as a source for creating a GBF/DIME System file. However, as indicated in Table A1, the variation on the quality of the materials by map sheet could be large. Therefore, it was concluded that, while the 1-in-40 block sample may be a good means of evaluating overall reference materials, the sample is much too small and spread too thinly to produce good estimates of the quality of the reference materials for a particular map sheet.

Computation of Segment and Addressable Unit Error Rates

A sample was selected from the City A GBF/DIME System file (file after edit dated September 1974); resulting from the CREATE phase of the GBF/DIME System program. The Metropolitan Map Series were used to locate the sample segments in the field. A team of two persons located each segment on the ground verifying the name with street signs, along with the names of intersecting streets. High and low addresses were recorded for both sides of the segment; the total number of addresses contained in the segment (both sides of the street) were noted; odd and even address placement was verified; "out of sequence" addresses and addresses falling outside the file range were documented. Persons in the field had access to the file data, so reconciliation of any differences was conducted at the same time as the original field check. Therefore, all results in this report are after field reconciliation.

After removal of non-street features and inaccessible segments from the sample, there were 280 sample segments. From these it is estimated that 18.6 percent of the segments in the GBF/DIME System file were in error. The 95 percent confidence limits of this estimate are 14.0 and 23.2 percent. Sixty percent of the segment errors (projected at 11.1 percent of the total file) were critical errors; the remaining 40 percent of the segment errors (projected at 7.5 percent of the total file) were non-critical errors. A "critical" error is defined as one that would cause an incoming address to be coded to the wrong geographic area, i.e., block or Census tract. A "non-critical" error segment is defined as one which has an error in the file but which will not result in miscoding of addresses to an incorrect geographic area.

TABLE A1

Sample and Error Distribution by Map

Map No.	No. of Blocks	No. of Blocksides	Blocksides in Error	Blockside Error Rate (%)	95 % Confidence Limits
TOTALS	42	175	5	2.9	0.3, 5.5
1	2	9	1	11.1	0.0,35.5*
2	-	-	-	-	-
3	1	4	-	-	0.0,60.3*
3NW	10	42	2	4.8	1.7, 7.9
3SW	8	32	1	3.1	0.2, 6.0
4	3	12	-	-	0.0,26.5*
4NE	14	60	-	-	0.0, 6.0*
4SE	3	12	-	-	0.0,26.5*
5	1	4	1	25.0	0.0,68.4*
6	-	-	-	-	-

*Constructed by using the binomial distribution for computation of confidence limits. See Hansen, Hurwitz and Madow, "Sample Survey Methods and Theory," Vol. 1, pp. 135-136.

About 23 percent of the critical errors fell in the category "lower and/or higher addresses in segment on the ground than allowed for in the file." Another 26 percent of the critical errors were an "odd-even anomaly;" that means that at least one but not necessarily all of the addresses in the segment were on the wrong side of the segment (i.e., an odd address on the even side of a street or vice versa). It should be noted that the quality control plans that were in existence

2/ Deming, William Edward, "Some Theory of Sampling," p. 203.

for the CREATE phase would not necessarily detect an error of the latter type. A breakdown of all critical errors is contained in Table A2.

TABLE A2

Critical Errors and Non-Critical Errors

In addition to naming the errors, a code for the possible sources of the errors are shown in this table. The sources listed are the best explanation for probable causes of the errors; however, no evidence is available to substantiate or refute these reasons.

Error Type	Number of Segments in Error	Percent of All Critical/ Non-Critical Errors	Percent of File	Possible Source of Error*
Total Errors	52		18.6	
Total Critical Errors	31	100.0	11.1	
1. Odd-Even Sides Reversed	2	6.5	0.7	1, 2, 3
2. Incorrect Block Number	2	6.5	0.7	2, 3
3. Different Address Range on Ground than in File	3	9.7	1.1	1
4. Incorrect or Illegal Place Code Number	3	9.7	1.1	2, 3
5. Lower and/or Higher Address in Segment Than File Allows	7	22.6	2.5	1
6. Incorrect Feature Name	2	6.5	0.7	1, 2
7. Two or More Segments on Ground; One in File	3	9.7	1.1	1
8. Odd-Even Anomaly	8	25.8	2.9	1, 2, 3
9. Improper Placement of Corporate Boundary	1	3.2	0.4	1, 2, 3
Total Non-Critical Errors	21	100.0	7.5	
1. Paper Streets (exist on Map and in File, but not on the Ground)	18	85.7	6.4	1
2. Two or More Segments in File, One on Ground	1	4.8	0.4	1
3. Zero Addresses in File, Numbered Addresses on Ground	1	4.8	0.4	1
4. Duplicate Node Numbers	1	4.8	0.4	1, 2

* (1) Source Materials or Improper Use of Them
(2) Clerical or Transcription Error
(3) Keying/Card Punching Error

Of the non-critical errors, 86 percent were those where the GBF/DIME file indicated that a segment existed, but the segment did not exist on the ground. These are referred to as paper streets and are usually a utility right-of-way, footpath, a street not yet built, or a yard (with no room to put a street). Other non-critical errors are detailed in Table A2. Non-critical errors would not cause an incoming address to be coded to the wrong geographic area.

Speculating on probable causes of the errors, faulty reference materials might first come to mind. However, the first part of this report indicates that the reference materials are of acceptable quality; therefore, wrong interpretation of good source materials is a probable source of error. Also, clerical and transcription errors may have added to the problem.

In each segment the number of individual units were counted and the addressable unit error rate for City A was calculated to be 5.3 percent. The 95 percent confidence limits of this estimate are 2.0 and 8.7 percent. These limits are based on cluster sampling techniques since the selected sampling unit was a segment and addressable units are clustered within the segments.3/ The percentage of addressable units in the sample which would be miscoded to the wrong geographic area was 3.1 percent with 95 percent confidence limits 0.6, 5.6. An additional 2.2 percent (95 percent confidence limits of 0.2, 4.3) of the addresses would not be coded by the file. In an actual Census operation, these would be identified, field-checked and a code would be assigned manually.

The addressable unit error rates by map sheet range from zero (95 percent confidence limits 0.0, 60.2) to 23.5 percent (95 percent confidence limits 0.0, 55.8) -- the confidence intervals are too wide to distinguish among the map sheet error rates. An error does not necessarily cause all addressable units within a segment to be miscoded or no-coded. Some of the addressable units may still be coded to the correct geographic area, even though an error has been made in the segment.

The intraclass correlation coefficient for clustered addressable units within segments was calculated to be 0.69. This is a strong indication that errors are clustered within segments; i.e., if an error affects one addressable unit within a segment, it is likely to affect more of the addressable units within that segment.

The intraclass correlation was calculated (from Hansen, Hurwitz, and Madow4/) using the formula for an estimate of the intraclass correlation coefficient from a sample, made in terms of the variances between and within ultimate clusters. Segments which contained no addressable units were eliminated from the calculations; also, those segments with only one addressable unit were eliminated from this calculation since by definition these cases do not apply.

The City A file was the CREATE file and had not undergone the CUE program. Several cycles through correction and update would be expected to improve the quality of the existing file.

City B Evaluation

Before a QC plan could be devised a decision was required as to what acceptable quality was in terms of a local GBF/DIME file. Cities that used such a file locally would probably need less stringent verification than those that do not use

3/ Cochran, William G., "Sampling Techniques," John Wiley & Sons, Inc., Formula (3.26), p. 65.

4/ Hansen, Morris H., William N. Hurwitz, and William G. Madow, Sample Survey Methods and Theory, Vol. I: Methods and Applications, p. 266.

their files, simply because they would be more likely to discover and correct any errors themselves. City B was selected for evaluation solely because it was further into the CUE program than most other places. City B should not be considered as representative of all cities in the program, but merely a city that was selected as a means of obtaining some insight into the type of problems a city faces in keeping a GBF/DIME file updated and to evaluate the findings with the objective of obtaining a realistic picture of the quality control needs of a CUE program.

In November 1974, a group of 12 field representatives were sent into City B to obtain geographic information for a sample of about 2,000 street segments that were contained in the GBF/DIME System file for City B. These sample segments were each checked on the ground by field personnel and selected address information was obtained independent of the files. These field data were then compared to the current City B GBF/DIME System file as it existed at the time of the study. Subsequently, a field reconciliation was made for all segments where a difference existed between the GBF/DIME System files and the ground.

The evaluation study entailed the matching of two Geographic Base (DIME) System files -- (1) the one prior to the initiation of the CUE program in City B, and (2) the current GBF/DIME System file updated to January 1, 1974. The match was based on a unique six-digit serial number for each segment record. The matched records were then compared for exactness.

The current GBF/DIME System file contained approximately 32,000 street and 6,000 5/ non-street segments. Each street and non-street record was sorted into one of the following classes, based on their status in the two files cited above: (1) Changes, (2) No Changes, (3) Additions, (4) Deletions. The first two classes were comprised of records which matched on serial number and contained discrepant geographic information (1) or identical geographic information (2). Unmatched records in the January 1974 file comprised class (3), while class (4) consisted of unmatched records in the original file. The first output was a count of the number of records for each sorted class with a subsequent systematic random sample of segments selected from each of the four strata.

A systematic sample of 1,900 segments was selected in order to analyze the data with an adequate degree of confidence. A variable sampling rate was used for each of the four classes to assure adequate representation in each class. The first three classes were selected from the January 1974 file and the fourth class, Deletions, was selected from the first file prior to the initiation of the CUE program.

The total sample was comprised of 1,974 coded segments spread over the four classes as follows:

Class	# of Cases	Sampling Fraction
Changes (same serial number, but some difference in record)	849	1/27
No Changes (no difference in the two files)	357	1/27
Additions (not in the file prior to CUE) and	572	1/10
Deletions (not in the file following Update)	196	1/4
	1,974	

The following table shows the distributions of the updated City B file over the above four classes (first column) and the segment error rate for each portion of the file over the same four classes (second column):

Class	Distribution of Updated City B File (%)	Segment Error Rate (%)
Changes	58.2	20.1
No Changes	25.0	14.0
Additions	14.8	18.9
Deletions (from pre-CUE file)	2.0	7.1

The weighted overall segment error rate for City B was derived by weighting each of the above four classes by the inverse of their sampling fraction. These calculations resulted in a weighted segment error rate for City B of 18.1 percent (see attached Table B1). The 2-sigma limits on this estimate are 16.3 and 20.1 percent. 6/

One portion of the CUE process is the application of computer edits for checking parity of address ranges, bounding a block, and address-range completeness. An evaluation of the effectiveness of these edits showed that they had caused a reduction in the overall segment error rate from 21 percent to the 18.1 percent level.

More than three-fourths of the file errors were classified as "critical" errors. The remaining errors were classified as "non-critical" errors (See Table B1 attached.) The weighted "critical" error rate is 15.1 percent. The weighted "non-critical" error rate is 3.0 percent. A "critical" error is defined as one that would cause an incoming address to be coded to the wrong geographic area, i.e., block, tract, place or MCD. A "non-critical" error is defined as one that will not

5/ All error rates shown in this report are calculated with a base of 39,288 street and non-street segments. Non-street segments including railroad tracks, etc., were not included in the sample as they would not contribute to the error rate. However, corporate boundaries were field-verified and would contribute to the error rate. Overall, the error rates shown may be underestimates of error rates which are based on a check of all segments.

6/ Cochran, William G., "Sampling Techniques," John Wiley & Sons, Inc., Formula (5.43), p. 106.

cause miscoding of addresses to a geographic area. The major category accounting for about 50 percent of the "critical" errors was defined as a smaller and/or larger address found existing on the ground than allowed for in the files. The major category of "non-critical" errors was defined as a segment with no addresses in the file and numbered addresses on the ground, accounting for about one-third of the total "non-critical" errors. This type of error was classified as non-critical since it would not cause an incoming address to be coded to the wrong geographic area. This error type would initially result in no coding. However, in a census, the item would be identified by the computer as a problem, followed up in the field, and a code eventually assigned.

The range of error rates by map sheet for City B areas is a low of 9.7 percent in the northeast to a high of 30.2 percent in the southwest. The error rates for the central part of City B (Maps 7, 7NE, 7NW) are somewhat lower than the areas surrounding the central city. The highest error rates appear in the west and southwest sections of the city (Maps 5, 6, 10 and 11). This distribution pattern of error rates was expected since the central city areas are usually more stable and more systematically arranged than the suburban areas.

The corrective process for CUE has an error associated with it. The "Change" category makes up the largest part of the City B file (about 3/5 of the file) and had the largest error rate (both number of errors and percent of total errors) among the four strata, Changes, No Changes, Additions and Deletions. Because the "Change" category is a significant part of the total City B file, further analysis was done to determine the extent to which a change occurred and either corrected the error or did not correct the error.

Table B3 below shows the source of errors (combined critical and non-critical) for the "change" category as an aid to determining whether an existing error was corrected (77.0 percent of 849) or was left uncorrected by the change action (11.3 percent of 849), whether the change replaced one error with another (9.3 percent of 849) or whether the change caused an error to a previously correct record (2.4 percent of 849).

In column 4, a count is given by map of those segments for which a change was made to a portion of the segment but did not affect the error; that is, the error was present in the original file (dated 1971) and remains after the change (49.2 percent of the 195 errors).

In column 5, a count is given of those segments for which changes were made to the error portion of the record. However, an error remains following the change (40.5 percent of 195 errors).

Column 6 gives the count of those segments for which the change created an error. The portion of the segment which is now in error was correct in the original file (10.3 percent of 195 errors).

TABLE B3

Status of Error Changes Made to the "Changes"

Map No.	Total Change*	Change Corrected Error	Error Exists After Change		
			Original Error Not Corrected by Change	Change Replaced One Error with Another	Change Introduced Error
(1)	(2)	(3)	(4)	(5)	(6)
Total	849	654	96	79	20
1	28	25	2	0	1
2	113	81	23	8	1
3	31	25	3	1	2
4	150	119	9	18	4
5	52	33	7	12	0
6	81	49	14	11	7
7	44	36	3	5	0
7NW	137	116	12	8	1
7NE	80	71	4	4	1
8	66	50	8	6	2
9	13	9	2	2	0
10	31	24	4	2	1
11	23	16	5	2	0

*There was a total of 849 segment records in the "Change" category, of which 195 were in error prior to the review of Topological and Address Range Edits.

An attempt was made to estimate the address error rate for the City B GBF/DIME System file. All segments that contained "critical" errors were reviewed and an estimate of the number of error addresses was derived. For those segments where a larger and/or smaller address existed on the ground in the sample segment than allowed for by the file, one address was assumed as the minimum number of address errors and a maximum was derived by taking the difference between the actual and the allowable (file) addresses and dividing by two. (Example -- if the first address of one sample segment on the ground was 93, and the low address in the file for the segment was 99, the difference of six was divided by two and the resulting three addresses were considered to be in error -- 93, 95, and 97).

When an entire segment would have been miscoded -- a wrong block number or wrong MCD/Place was in the file -- eight addresses were assumed to be in error (eight addresses as the average per segment was derived from the estimated total number of addresses in City B divided by the number of street segments in the City B file). Using these methods, the range for the address error rate for the City B GBF/DIME System is estimated to be 8 percent to 21 percent.

The GBF/DIME System file for the City B SMSA had an overall segment error rate of 18 percent at the completion of the first update cycle. Further updates, using reference material of current quality and with continued acceptable quality on manual coding operations, should result in a segment critical error rate of about 10 percent. This estimate is derived, based on the assumption that the computer edits would effectively locate and identify errors for clerical correction. Reference material quality and the quality of the resultant GBF/DIME System is not inconsistent with quality levels achieved in the 1970 Census of Population for similar processes.

On-site investigation and review of sample information indicate that manual coding operations are of acceptable quality and that the designed edit operations of the CUE program are well-conceived and adequate for purposes of detecting file inconsistencies. An area which could be improved through careful quality control processes is the reference or source material used in the CUE operations. Overall, based on a small sample, the reference material for the City B area is estimated to have an 8.5 percent segment error rate, and it is felt that a feasible goal would be to achieve an average outgoing quality limit of 5 percent for references and sources. A quality control plan for this purpose, and incorporating further checks on manual processes, is now under development.

CONCLUSIONS

1. The reference material used in the creation of the City A file is of acceptable quality based on the quality control evaluation and verified by this study using a similar set of reference materials updated to 1975.

2. The 1-in-40 block quality control sample is sufficient for an overall estimate of the quality of reference material for a large area. However, it is too small a sample and spread too thinly to provide error rates on a map sheet or smaller area basis.

3. Based on the estimated 18.1 percent segment error rate in City B, a quality control program is needed for the GBF/DIME System CUE operation.

RECOMMENDATIONS

Based on the findings of the evaluation studies, the following recommendations were made:

1. While the 1-in-40 block sample is adequate for overall error rate estimates, it is recommended that a larger sample, such as that used in the procedures recommended for CUE, be selected to provide detailed error rates by map sheet. Based on the City B experience, the variability of error rates between map sheets may be great. Thus, an area might be acceptable from an overall determination of the error rate, but one or more map sheets or small areas might not be acceptable when individual map sheet or small area error rates are known.

2. When possible, use the same type of quality control plans for CREATE as for CUE.

3. More attention must be given to the clerical and keying operations, since even with good reference materials, a high ultimate error rate can occur in GBF/DIME. It is suspected that the 1-in-40 block sample is too small to quality check on individual coders -- similar to "by map" problem.

As a result of the recommendations based on these evaluation studies, a set of quality control plans for the creation, correction, update and extension of the GBF/DIME System file has been proposed and is currently being tested in several areas. Results of these tests will be documented, and based on these results, quality control plans will be designed for use in all areas in the CUE program.

REFERENCES

"Geographic Base Files - Plans, Progress, and Prospects," Conference Proceedings from Jacksonville, Florida, April 1-2, 1971.

"CUE - Correction-Update-Extension," Procedural Manual for the Review and Correction of the Address Range Edit (ADDEDIT) Listing, January 1973.

"CREATE Coder's Manual," CUE, Correction-Update-Extension, January 1, 1974, Revised July 1, 1974.

CUE Evaluation Study--Field Procedures, October 1974.

"Final Quality Control Specifications for CUE Operation" from C. Jones, SMD to J. Silver, Geography Division, February 23, 1976.

"An Outline on the Geographic Base File and CUE Maintenance Program to the Ohio Department of Transportation," by Kenneth Graska, May 31, 1975.

Table B1

City B CUE Evaluation

Weighted Breakdown of Source of Errors by Type of Error

Error Type (1)	Estimated Number of Segments with Errors (2)	Percent of Total Errors (3)	Percent of Total File (4)
All Errors	7103	87.3	18.1
Non-Critical Errors	1188	14.6	3.0
Critical Errors	5915	72.7	15.1
Source Materials	3978	48.8	10.2
1. Smaller and/or larger addresses in segment than file allows	3658	44.9	9.3
2. Wrong Address Range	101	1.2	0.3
3. 2 segments on ground, 1 segment in file	219	2.7	0.6
Clerical			
(Transcription, Keying, Card Punching)	1475	18.1	3.8
1. Wrong block No.	1058	13.0	2.7
2. Wrong MCD/Place	336	4.1	0.9
3. Wrong Tract No.	81	1.0	0.2
Combination of Source Materials and Clerical/Computer			
1. Odd-even sides reversed	462	5.7	1.2

STATISTICAL ANALYSIS OF VARIABLES WITH SHARED TERMS

Havens C. Tipps U.S. Commission on Civil Rights

A potential problem with Pearsonian correlation coefficients has been recognized almost since the development of the coefficient itself. This paper describes the results of experimental tests of this statistical problem which has been causing concern and distress among a sizeable number of researchers. The problem arises most often when variables are created through making ratios, proportions and various kinds of indexes. It has been maintained that when certain types of pairs of numbers with common terms are correlated, statistics are produced which are spurious, misleading, or in other ways inappropriate. If the problem is as severe as some have asserted, a significant part of our empirical findings are suspect. This paper examines the nature of the problem and provides evidence that the problem is indeed worth stressing under certain limited conditions.

The pairs of variables below illustrate some of the types of combinations which have been described as problematic Pearsonian correlation coefficients:

1. Organizational size with administrative intensity:
(administrators plus non-administrators)
correlated with
(administrators/admin plus non-admin.) or:
(organization size)
correlated with
(administrators/organizational size)
2. Population size with suicide rate:
(population)
correlated with
(number of suicides/population) or:
(suicides plus non-suicides)
correlated with
(suicides/suicides plus non-suicides)
3. Percent black with population density:
(blacks/population)
correlated with
(population/area)
4. Population size with growth:
(population at time 1)
correlated with
(population at time 2) or:
(population at time 1)
correlated with
(population at time 1 plus change)
5. Proportion nonwhite with education index:
(nonwhites/population)
correlated with
(those with some college/population)
6. Social origin with social mobility:
(father's status)
correlated with
(son's status - father's status)

Needless to say, we are talking about types of variables and types of relationships which are very important to social scientists. In each of those pairs of variables numerical terms from one variable are duplicated in the other. In some situations this factor has been described as having the variables "definitionally dependent" on each other so that it is virtually assured that a non-zero correlation will be found between the two variables. The issues for this study are the extent to which the shared terms influence the value of the correlation coefficients and the meaning of this influence (if any) for actual researchers correlating variables with shared terms.

In the past few years interest in this problem has spread as with the publication of several important and often misinterpreted articles. The problem has also been extended to path coefficients by Karl Schuessler.[1] This project actually started with an article by Freeman and Kronenfeld (2) that bothered me and renewed a debate on the issue between myself and some colleagues. The article was titled "Problems of Definitional Dependency: The Case of Administrative Intensity." Freeman and Kronenfeld focused on a particular research problem in the study of organizations which involved the problematic condition of correlating variables with shared terms.

The interesting and theoretically important variable at the center of the Freeman and Kronenfeld analysis is the composition of organizations in terms of the administrative and productive components. Some organizations have a relatively large administrative component and some have a relatively small administrative component. A key theoretical issue involving this variable is the relationship between the organization's size and the administrative component. There are reasons to believe that as organizations get larger the proportion of the workers who hold administrative positions systematically changes. With the advancement of work in this area different patterns will probably be discovered for different types of organizations.

Freeman and Kronenfeld noted that many researchers have found a negative relationship indicating that the larger organizations have a smaller proportion of the work force holding administrative positions. The statistical concern of the authors is over the fact that the correlation of the size of an organization with the proportion of those workers who are in administrative positions involves the issue of "definitional dependency" as the authors call it. The correlation is between (x+y), with "x" being the administrative and "y" being the non-administrative workers, and either (x/y) or (x/x+y) depending on the method of measuring the administrative variable as a ratio or a proportion.

Through a mathematical demonstration they conclude that the reason why almost all data from research on this question exhibit an "exponentially shaped decreasing pattern," is that "this pattern is primarily the result of the coordinate transformation, and not because of any inherent relationship between x and y." (p. 112)

I had a difficult time accepting the implication of a correlation such as this being due to definitional dependence. There are just too many studies in the literature and especially in unpublished work where the correlations are zero and near zero under similar conditions to be easily convinced that there is a definitional dependence when a unit's size is correlated with a proportion or rate based on the unit's size. The field of urban sociology seems filled with near zero correlations of that type. My two primary concerns were to provide experimental evidence concerning the extent and nature of the problem and to emphasize the limited applicability of the problem.

To create experimental conditions approximating the problematic situation data sets were randomly generated so variables could be created fitting the required specifications. The IBM Fortran Scientific Subroutine Package, using the GAUSS subroutine, was used to generate 100 data sets of 50 observations for each of three variables (X,Y, and Z). The generation process produces a normally distributed random number for each request with a given mean and standard deviation. In this case the mean used was 500 and the standard deviation was 125. The ratios, sums, and differences were then calculated and correlated for each of the 100 data sets.

Table 1 provides some of the statistics calculated from the 100 correlations produced for each type of relationship. Although 100 statistics (correlation coefficients in this case) is clearly not enough to generate a sampling distribution it seems to be a large enough sample of samples to reach some important conclusions about some of the problematic correlations. It is ample, it seems, to demonstrate that there is indeed a "problem" with some of the coefficients and the problem is numerically substantial. Equally important is that it seems that some of the coefficients which Freeman and Kronenfeld (1973) have described as spurious from the analysis of the formulas seems not to be "definitially dependent" in the samples generated for this study.

The table contains twenty-six types of relationships which involve shared terms. Freeman and Kronenfeld refer mainly to relationship number 4 and number 8.

The mean of both of these sets of correlations is close to zero and the number of significant correlations (12) and 9 respectively) is not especially far from the expected five at the .05 level. But most of the other sets of correlations are clearly far above the expected five significant correlations and seven have all 100 correlations being significant. In additional experiments performed by some colleagues at the University oc Cincinnati the means and the standard deviations of X,Y, and Z were modified in various ways for some of the types of relationships.3 Under conditions of varying means and standard deviations of the components, even relationships like number 4 and number 8 exhibit correlations that are vastly different from zero. Clearly there is room for concern over this issue.

Table 1

Statistics from the distribution of 100 correlation coefficients (r) with X,Y, and Z being random numbers*

	Form of the Relationship		Mean of r's	Std. Dev.	Range: Min.	Max.	Number of r's P .05
1.	X	Y	-.003	.13	-.33	.32	6
2.	X	X/Y	.623	.12	.13	.83	98
3.	X/Y	Y	-.694	.07	-.84	-.49	100
4.	X+Y	X/Y	-.040	.16	-.38	.32	12
5.	X	X+Y	.709	.06	.51	.84	100
6.	X	X-Y	.711	.07	.42	.84	100
7.	X	X/(X+Y)	.702	.08	.40	.86	100
8.	X+Y	X/(X+Y)	.022	.15	.36	.41	9
9.	X/(X+Y)	X-Y	.983	.01	.96	.99	100
10.	X+Y	X-Y	.017	.14	-.39	.30	4
11.	X/Y	X-Y	.935	.06	.51	.98	100
12.	X	Y/Z	-.018	.13	-.43	.27	2
13.	X+Y	Y/Z	.419	.11	.15	.66	87
14.	X+Y	Z/Y	-.484	.10	-.77	-.18	96
15.	X/Y	Z/Y	.569	.13	.20	.99	98
16.	X/Y	Y/Z	-.433	.10	-.63	-.13	92
17.	X/(X+Y)	Z/Y	.521	.11	.17	.77	98
18.	X/(X+Y)	Y/Z	-.433	.11	.69	-.10	91
19.	X-Y	Z/Y	.499	.09	.21	.74	98
20.	X-Y	Y/Z	-.444	.11	-.68	-.12	92
21.	X+Y	Y-Z	.486	.10	.20	.74	95
22.	X-Y	Y-Z	-.501	.10	-.71	-.19	98
23.	X/(X+Y)	Y-Z	-.489	.10	-.71	-.14	99
24.	Z/(X+Y+Z)	X/(X+Y+Z)	-.490	.10	-.73	-.19	96
25.	Z/(X+Y)	X/(X+Y)	-.007	.15	-.38	.38	7
26.	Z/(X+Y)	X+Y	-.578	.10	-.81	-.34	100

+Each of the 100 data sets has 50 observations for X,Y and Z, with each value selected as a normally distributed random number with a mean of 500 and a standard deviation of 125.

Practical Implications for Researchers

Given that there is ample evidence that correlations between composite variables with shared terms often exhibit high correlations when the components are not correlated, we must ask what the impact of this should be for researchers. What seems to need stressing most is the limited applicability of the problem for the theoretical and empirical issues which researchers typically face. The problem of "definitional dependence" is only "a problem" when the researcher is really interested in the variables X,Y and/or Z rather than the ratios, proportions or differences, which the researchers constructs. This is very rarely the case. Generally, it seems, when we use percentages, ratios or differences, we are actually interested in the percentages, ratios and differences as meaningful varioables in themselves. As Fuguitt and Lieberson reported:

> Discussions of this problem have centered on the purpose and assumptions of the analysis to be undertaken. Several writers have regretted Pearson's choice of the word spurious to refer to this phenomenon. A number have pointed out that there is nothing intrinsically spurious about the correlation, through interpretations may indeed be spurious, as in inferring from a ration correlation the size or direction of a component correlation or vice versa. A basic distinction here is whether the ratio or difference score is taken to be the basic variable describing the population under study or whether one's major interest really focuses on the component measures. If the former is the case, some authors argue that spurious correlation is not a problem (Yule, 1910;Kuh and Meyer, 1955; Rangarajan and Chatterjee, 1969). Logan (1972,p.67) gives as an example the association of speed (miles per hour) with gasoline consumption (miles per gallon). The basic interest here is in whether cars that go faster burn gasoline at a greater rate, and not in the associations between component variables _ miles traveled, time elapsed, or gasoline consumed. Just as this example utilizes a common numerator in the two ratios (miles), sociologists may likewise try to claim an inherent interest in ratios with a common denominator; for example, the correlation between per capita energy consumption of nations and their per capita gross national product.4

Freeman and Kronenfeld, among many other statistical analysists, fail to realized or appreciate the importance of the composite variables. Administrative intensity is an important variable itself (as a ratio or a proportion) and as such the relationship between the components of the composite variables are irrelevant when the variable is correlated with organizational size. In fact, I can not think of a single major research finding that should be discredited because of the definitional dependence issue. The isolated case where researchers construct composite measures, correlate them with variables sharing some terms, and then try to infer to the components of the composite variables is clearly inappropriate. In such cases the construction of composite variables is usually to control for an additional variable. Although the construction of composite measures is not a very appropriate method for multivariate controls, the standard methods of control still apply and should be used. Thus standardization, partial correlations and multiple regression can offer solutions to the problem of "definitional dependency" if the primary interest is in the component variables.

In the normal situation rates, proportions, ratios and differences are used as variables with theoretical importance. Correlations between proportions, rates, ratios, or differences and other variables which may or may not share terms are subject to the same sources of misinterpretation as other correlations. We must always be aware of possible effects of other variables, sampling error, etc. But is has not been established that special problems exist for this type of correlation. In fact, since Yule clarified the issue in 1910 by advising that one simply state in advance whether one was interested in the components of the ratios or the ratios themselves,5 there has not been a dispute over the statistical issue. The experimental evidence provided in this paper further documents the existence of cause for concern. The key issue remaining is the validity of studying variables which are in the form of ratios, proportions and differences under the types of conditions discussed above. This question of the validity of the variables is a theoretical question that can only be answered on the grounds of the particular substantive areas. In the more common areas of investigation such variables are definitely established as legitimate and often they are the most important variables. Often they seem to be the purest form of measurement we have since they are based on counting, although they are not an artifact of the size of the populations subject to the count.

References

1. Karl Schuessler, "Ratio Variables and Path Models," in Arthur S. Goldberger and O.D. Duncan (eds), Structural Equation Models in Social Science, New York: Seminar Press, 1973.

2. J.H. Freeman and J.E. Kronenfeld, "Problems of Definitional Dependency: The Case of Administrative Intensity," Social Forces, Vol. 52, Sept. 1973.

3. W.E. Feinberg, J. Trotta and H.C. Tipps, "Correlations of Ratio Variables Involving two components: Investigation of an Alternative Significance Test," unpublished paper, 1976.

4. Glenn V. Fuguitt and Stanley Lieberson, "Correlation of Ratios or Difference Scores Having Common Terms," in H.L. Costner (ed.), Sociological Methodology 1973-1974, San Francisco Jossey-Bass, 1974. Yule, G.U., "On the interpretation of correlations between indicies or ratios," Journal of the Royal Statisical Society 73 (June):644-647, 1910. Kuh, E. and Meyer, J.R., "Correlation and regression estimates when data are ratios," Econometrica 23:400-416, 1955. Rangarajan, C. and Chatterjee, S., "A note on comparison between correlation coefficients of original and transformed variables," American Statistician 23 (October): 28-28, 1969.

5. G.U. Yule: see above.

INVESTIGATING DISCREPANCIES BETWEEN SOCIAL SECURITY ADMINISTRATION AND CURRENT POPULATION SURVEY BENEFIT DATA FOR 1972

Denton R. Vaughan
Social Security Administration
and
Robert E. Yuskavage*
Bureau of Economic Analysis

This paper is the first in a series which will examine errors in the reporting of 1972 social security benefit income in the March 1973 Current Population Survey (CPS). Earlier studies conducted at Social Security by Bixby, Haber, and Finegar in connection with the 1963 and 1968 surveys of the aged [1,3,4] and by Projector - Bretz [9] for the March 1971 CPS have documented the quality of social security benefit reporting in the CPS and similar surveys. By-and-large, this previous work has emphasized the demographic and programmatic characteristics associated with misreporting. Although some mention has been made of the actual sources of reporting error [1,3], to our knowledge, presentation of a detailed analysis of this sort has not been undertaken. We intend to explore the actual sources of reporting error in some depth.

DELIMITING THE PROBLEM OF ERROR

While the causes of misreporting in the CPS are, no doubt, many and varied, we consider the following to be among the more important:

1. failing or refusing to report benefit income or attributing it to another source;
2. reporting benefit income when it was not received;
3. reporting income from another source together with benefit income;
4. combining the benefits for two or more persons in the record of a single recipient;
5. misspecifying the elements needed to derive an annual benefit figure for the survey, such as, the dollar value of the monthly benefit, the number of months benefits were received, deductions, etc.;
6. estimating the annual amount, including more or less haphazard guesses, because of ignorance or uncooperativeness on the part of the respondent; and,
7. processing errors, from those made in recording information in the original interview through creation of the machine readable file.

The present analysis will focus primarily on the fifth type of error which results when the elements needed to derive the annual social security benefit amount have been misspecified. The third and sixth sources will also receive some attention. In all, six possibilities are considered:

1. mistiming benefit increases,
2. misspecifying the number of months of benefits,
3. ignoring benefit increases,
4. mishandling Supplemental Medical Insurance (SMI) deductions,
5. Social Security-Railroad Retirement dual recipiency,1/ and
6. rounding errors.

THE DATA BASE AND STUDY UNIVERSE

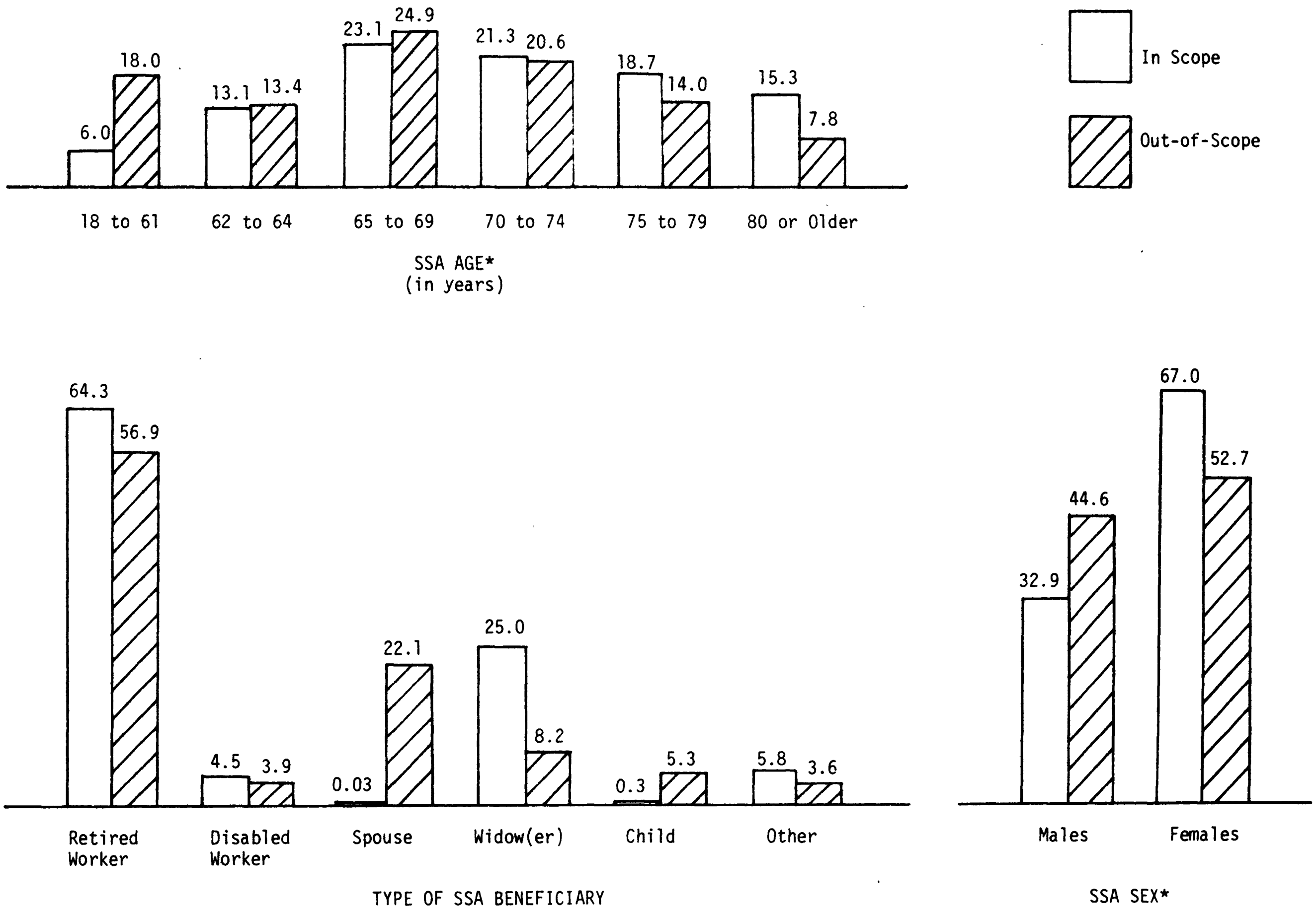
The analysis draws on a data file which was obtained by matching the March 1973 CPS to Social Security earnings and benefit records.2/ Because of problems associated with the reporting of social security income for spouses and minor children in the CPS, we have focused our attention on persons 18 years or older who were the sole SSA beneficiary in their recipiency unit.3/ Among this group, we will further limit ourselves to recipients with social security income amounts in both the SSA and CPS record. Furthermore, individuals were also excluded if their CPS benefit income had to be imputed due to refusal or nonresponse.4/

Estimates for the number of beneficiaries and total benefits in the overall SSA-CPS universe age 18 and older and in the study universe were obtained using sample weights which incorporate some adjustments for nonmatches and mismatches. Population controls used in the weighting include adjustments for institutionalized, overseas, and decedent beneficiaries not eligible for interview. The derivation of the sample (initial raking) weights is described in detail elsewhere [15].

The study universe is significant in both its size and diversity. It consists of 9.9 million beneficiaries who received $15.1 billion in benefits, each about 40 percent of the respective 1972 totals for the age 18 and older CPS-eligible beneficiary universe. As indicated in figure 1, the beneficiaries included in the study tend to be somewhat older and have a considerably greater representation of females than the out-of-scope segment. While retired worker recipients constitute the dominant beneficiary type in both groups, the proportion of widowed recipients was considerably higher, and that of spouse beneficiaries, considerably lower, in the study universe.

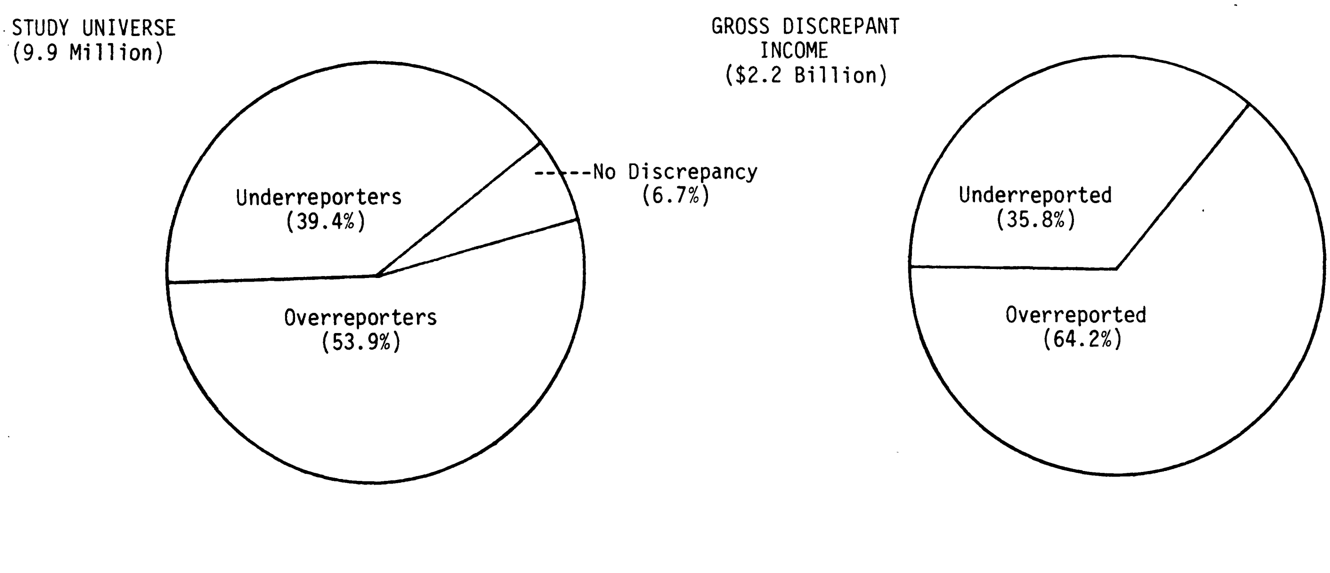
Figure 2 classifies study universe recipients and gross discrepant income by the direction of the observed SSA-CPS discrepancy. It shows that more than half of the recipients (54 percent) were "CPS overreporters," that is, the amount reported in the CPS was greater than in the SSA record. About two-fifths of the recipients were "CPS underreporters" (39 percent), with the amount in the CPS record understating the SSA benefit amount. For the remaining 7 percent, the survey and administrative amounts agreed exactly (1 percent) or fell within a narrow tolerance ($10 or less). The gross dollar discrepancy associated with the misreporting in the CPS amounted to $2.2

Figure 1. --Percent Distribution of Study Universe and Exclusions by SSA Age, Sex, and Type of Beneficiary



* Totals do not add to 100 percent because unknown age and sex have not been included.

Figure 2. --Distribution of Cases in the Study Universe and Discrepant Income by Direction of Discrepancy



billion, or 15 percent of all benefit income paid to members of the study universe. Nearly two-thirds of this discrepancy was overreported.

It should be noted that this preponderance of overreporting in the study population is not characteristic of the overall benefit reporting error found in the CPS [2, 9] or in other surveys of the beneficiary population carried out in a similar setting [1, 3, 4]. At the present time, all that can be said is that the net underreporting known to exist in the 1973 March Supplement data [2] does not stem from that part of the beneficiary population under consideration here.

DETECTING SOURCES OF REPORTING ERROR

While the sources of disagreement between survey and administrative information may lie with either the survey or the administrative record system (and, in some instances, neither may be correct), in a great majority of cases, the administrative data are likely to yield a superior estimate of the "true" amount. Consequently, for the purposes of this paper, the information in the administrative record will be used as the criteria for evaluating the benefit data reported in the CPS.

Methods of Detection.-- Two general approaches were used to detect the source of SSA-CPS benefit discrepancies.5/ For errors stemming from misspecifying months of benefits, mistiming or ignoring benefit increases, and mishandling SMI deductions, we attempted to reproduce the observed discrepancies by holding constant one or more of the known elements from the administrative system. Primarily because we considered it likely that the monthly benefit amount would be rounded in deriving the annual benefit income figure for the CPS (at a minimum, to the nearest dollar), we did not expect to be able to exactly replicate the observed discrepancies. Consequently, in testing for a match between the observed and simulated differences, a tolerance of $6 to $10 was permitted.

Discrepancies due to Social Security-Railroad Retirement dual recipiency and rounding errors were treated differently. It is not possible to distinguish between Social Security and Railroad Retirement income amounts in the CPS when the beneficiary receives payments from both sources. However, both the SSA and CPS records indicate the occurrence of dual recipiency. Consequently, discrepancies of CPS overreporters whose SSA record disclosed dual recipiency were attributed to the inclusion of Railroad Retirement income in the CPS. The mean value of the discrepancies detected in this fashion, when disaggregated by type of annuitant, closely approximates the corresponding average Railroad Retirement benefit for 1972.6/

CPS benefit amounts were tested for possible rounding errors if their ending digits were '000', '500', '00', or '50'. It was assumed that amounts ending in '000' might have been rounded to the nearest thousand dollars, those ending in '500', to the nearest 500 dollars, and so forth. Observed discrepancies were attributed to rounding if differences of $1 to $500 were associated with ending digits of '000', if errors of $1 to $250 were associated with ending digits of '500', and so on for amounts ending in '00', or '50'.

Handling Multiple Explanations.-- In a considerable number of cases it proved possible to attribute an individual discrepancy to more than one source of error. In other words, a particular discrepancy was subject to alternative explanations. In order to simplify the analysis, a single source of error was assigned such cases based on priorities established by the overall prevalence of each category of error. Thus, timing errors were given first priority, ignoring the benefit increase was given second priority, months benefits received, third, and dual recipiency, fourth priority. Primary SMI errors were exempted from the criteria and assigned unconditionally, since they often seemed to be confused with small discrepancies also attributable to timing problems. As a matter of judgment, we felt that this resulted in a more plausible final distribution of the different types of error. Rounding errors were also exempted from the frequency criteria. Since the tests associated with detecting rounding errors were much less precise than those employed for the other possible sources of error, competing explanations were always given priority over rounding.

FINDINGS

Table 1 displays the percentage distribution of all explained discrepant cases and respective discrepant benefit income for overreporters and underreporters by type of error, as assigned by the priority procedure discussed above. Sixty-nine percent of all discrepant cases were explained by one or more of six hypothesized discrepancy models. The error models proved to be slightly better at detecting errors of overreporting (70 percent) than of underreporting (66 percent). The explained discrepancies accounted for proportionately less of the gross dollar misreporting, just over half the total. A somewhat greater share of the dollar discrepancy was explained among underreporters than overreporters.

Overreporters.-- Turning now to the sources of explained discrepancy for overreporters, we note the importance of errors in specifying the timing of benefit increases. Beneficiaries were granted a 20 percent across-the-board benefit increase in September 1972, effective in their October checks [11, 16]. This error model assumes that the respondent remembered the benefit increase but mistimed its occurrence. CPS overreporters, in committing this type of mistake, proceeded as though the higher benefit amount had been in effect four months or more. Misspecifying the timing of the benefit increase was, by far, the most frequent source of discrepancy detected among overreporters. Seventy-six percent of explained overreporter cases and 47 percent of the

Table 1.--Percent Distribution of Explained Discrepant Cases and Associated Discrepant Income by Type of Primary Discrepancy

Type of Primary Discrepancy	Discrepant Cases (in thousands)			Gross Dollar Discrepancy (in millions)		
	Total	Over-reporters	Under-reporters	Total	Over-reporters	Under-reporters
Total	9,227	5,328	3,899	2,233	1,456	777
Unexplained:						
Number or Amount	2,896	1,586	1,311	1,075	717	358
Percent	31.4	29.8	33.6	48.1	49.2	46.0
Explained:						
Number or Amount	6,331	3,742	2,588	1,158	739	419
Percent	68.6	70.2	66.4	51.9	50.8	54.0
PERCENT OF EXPLAINED						
Total	100.0	100.0	100.0	100.0	100.0	100.0
Mistiming Benefit Increase	55.7	76.2	26.0	34.7	46.7	13.6
Misspecifying Months of Benefits	19.0	3.1	42.0	31.1	7.6	72.6
Ignoring Benefit Increase	13.0	14.4	11.0	13.0	17.5	5.3
Mishandling SMI* Deductions	9.3	2.1	19.6	3.4	0.7	8.1
Railroad Recipiency	2.1	3.6	-	17.2	26.9	-
Rounding Error	0.9	0.7	1.3	0.6	0.7	0.5

Note: Percents may not add to 100 due to rounding.
*Supplemental Medical Insurance

associated explained discrepant income was attributed to this error source.

A related type of discrepancy was produced by completely ignoring the effect of the benefit increase. In this instance, the respondent apparently failed to recall the benefit increase at all and proceeded as though the post-increase amount had been in effect for each month benefits were received. This kind of error accounts for an additional 14 percent of the explained overreporter discrepancies and 13 percent of the explained overreporter income discrepancy. More than 90 percent of the explained overreporting errors and almost 65 percent of the associated discrepant income were attributable to these two sources.

Underreporters.-- The relative importance of the six error types was somewhat different for underreporters. Since nearly 90 percent of the recipients in the study universe received benefits for a full 12 months, errors in specifying the number of months of benefits were more likely to result in an underreport than an overreport. In fact, this kind of mistake was the most frequent source of explained discrepancy among underreporters. It was responsible for forty-two percent of the explained underreporter cases and nearly 73 percent of the associated underreported income.

Twenty-six percent of the explained underreports were associated with possible errors in specifying timing of the benefit increases. Another 11 percent apparently resulted from ignoring the benefit increase. Together these two error sources accounted for about 20 percent of explained underreported SSA benefit income.

Supplemental Medical Insurance, a voluntary complement to the Medicare Health Insurance program, covers virtually all Old Age, Survivor, and Disability Insurance (OASDI) recipients, age 65 and older [13]. In 1972, monthly premiums of slightly less than $6 were deducted from their benefit checks [12]. The CPS interviewer was instructed to determine if the respondent was an SMI enrollee and, if so, to add the total premium to the net annual benefit amount. This adjustment provided the opportunity to make one of two kinds of errors: an overstatement of the CPS amount, for those who were not SMI enrollees but had an amount corresponding to the premium added to their actual benefits, or an understatement, for those who were SMI enrollees but did not have the premium amount added to their net benefit. Since nearly four-fifths of the study population was age 65 or older and, therefore, generally exposed only to the second kind of mistake, one would expect SMI errors to be associated primarily with underreporting in the CPS. In fact, mistakes in handling SMI premium deductions were attributed to nearly 20 percent of the explained underreports. However, since the errors were relatively small (about $68 on the average), they accounted for only 8 percent of explained underreported benefits. Furthermore, mishandling SMI deductions was not an important source of reporting error among overreporters.

Dual Recipiency and Rounding.-- The two remaining sources of discrepancy, dissimilar in both nature and method of detection from the previous four, are Social Security - Railroad Retirement dual recipiency and rounding in the CPS reported amount. Dual recipiency was associated with only 4 percent of explained overreporter cases, but, since the discrepancies tended to be quite large, it accounted for 27 percent of explained overreported income. Rounding errors, on the other hand, proved to be a relatively unimportant source of discrepancy for both over- and underreporters. They were associated with only about 1 percent of the explained discrepant cases and less than 1 percent of the explained discrepant income.

Dollar Value of Explained and Unexplained Discrepancies.-- Reflecting the finding noted earlier that the proportion of explained misreported benefit income was less than the proportion of explained discrepant cases, table 2 shows that the dollar size of explained reporting errors was considerably smaller than that of the residual or unexplained discrepancies. Three-quarters of the explained discrepancies were less than $200, while this was so for only 39 percent of the unexplained differences. The average explained dollar difference ($183) was slightly less than half that for the unexplained cases ($371). Furthermore, the ratio of mean explained to unexplained error was considerably smaller for overreporters than underreporters (44 percent vs. 59 percent).

SUMMARY AND FUTURE PLANS

We have shown that six relatively simple error models, when applied to the March 1973 CPS, explain somewhat more than two-thirds of the cases involving misreported benefit amounts and about half of the associated gross discrepant benefit income. However, the results also indicate that the error models were at a comparative disadvantage in detecting relatively large-sized discrepancies, particularly among overreporters.

In the future, we will develop and report on new models designed to detect the sources of these larger discrepancies. Another major task will involve the extension of the present analysis to multi-beneficiary units. In pursuing this facet of the research, we should also uncover the sources of CPS net underreporting. With the expectation of gaining useful insights into the background variables related to misreporting, we also intend to undertake analysis of the demographic and programmatic characteristics of the beneficiaries with explained and unexplained errors. In particular, size of discrepancy will be taken into account.

FOOTNOTES

* The authors would like to express their thanks to Wendy Alvey, Ben Bridges, Dan Radner, and Fritz Scheuren for their helpful comments on earlier versions of this paper and to Beth Kilss and Helen Kearney for their assistance with the charts and tables.

Table 2.--Percentage Distribution of Explained and Residual Discrepancies of CPS Overreporters and Underreporters by Size of Discrepancy

(Numbers in Thousands)

Size of Discrepancy (in dollars)	Total		Overreporters		Underreporters	
	Explained	Residual	Explained	Residual	Explained	Residual
Total Number ...	6,331	2,896	3,742	1,586	2,588	1,310
Percent of Total.......	100.0	100.0	100.0	100.0	100.0	100.0
1 to 99............	45.1	20.4	38.1	7.6	55.2	35.8
100 to 199...........	29.9	18.9	33.2	15.4	25.2	23.2
200 to 299...........	14.2	17.0	18.5	21.3	8.0	11.8
300 to 499...........	5.3	21.2	5.3	29.2	5.3	11.6
500 to 999...........	3.2	15.5	2.0	17.6	4.9	12.9
1000 to 1499..........	1.5	4.7	1.6	5.7	1.3	3.6
1500 or More..........	0.8	2.1	1.2	3.1	0.3	0.9
Mean Discrepancy (in dollars)..........	183	371	197	452	162	273

Note: Detail may not add to total due to rounding.

1/ While the reporting of Railroad Retirement in the CPS is, in fact, a source of discrepancy between the survey amount and the social security administrative figure, it does not actually constitute misreporting in the CPS. The social security record simply does not include Railroad Retirement benefits.

2/ At the 1975 ASA meetings in Atlanta, several papers were presented which focused on the conceptual and reporting differences among the linked CPS, IRS, and SSA data sets for calendar year 1972. Some of the preliminary analyses presented included comparisons between matched CPS and IRS income information [5] and similar comparisons for SSA and IRS wage data [8]. For more information on the basic study, see [7], which appears elsewhere in these 1976 Proceedings. See also[14].

3/ The particular type of recipiency unit we will be using is known as the dependency unit, a classification routinely used at Social Security to arrange CPS household members in kinship groups consisting of persons generally considered interdependent under social insurance programs. Administratively, it is analogous to the subgroup of CPS household or family members that would be considered when determining eligibility and benefits for a single disabled or retired worker and his or her dependents or survivors [6].

4/ A small number of recipients with an SSA age of 18 or older (less than 30 sample cases) are also excluded because their age in the CPS was reported to be under 14, and, hence, income information was not obtained for them during the interview.

5/ A detailed description of the methods used to detect the individual sources of error is available from the authors on request. Their mailing address is: Division of Economic and Long-Range Studies, Office of Research and Statistics, Social Security Administration, 1875 Connecticut Avenue, N.W., Washington, D.C. 20009.

6/ For a description of the dual recipient population and benefit levels for 1972, see [10].

REFERENCES

[1] Bixby, L. E., Finegar, W., Grad, S., Kolodrubetz, W., Lauriat, P., and Murray, J. Demographic and Economic Characteristics of the Aged: 1968 Social Security Survey, Social Security Administration, Washington, D.C., 1975.

[2] Bureau of the Census. Current Population Reports, "Money Income of Families and Persons in the United States," Series P-60, no. 90, Government Printing Office, Washington, D.C., 1973.

[3] Epstein, L. E and Murray, J. The Aged Population of the United States: the 1963 Social Security Survey of the Aged, Social Security Administration, Washington, D.C., 1967.

[4] Haber, L. "Evaluating Response Error in the Reporting of the Income of the Aged: Benefit Income," 1966 American Statistical Association Proceedings, Social Statistics Section, 1967.

[5] Herriot, R. and Spiers, E. "Measuring the Impact on Income Statistics of Reporting Differences between the Current Population Survey and Administrative Sources," 1975 American Statistical Association Proceedings, Social Statistics Section, 1976, pp. 147-158.

[6] Johnston, M. P. "Derivation of STATS Units," Studies in Income Distribution, Report no. 1, by D. Projector, M. T. Millea and K. Dymond, Social Security Administration, 1975, pp. 59-76.

[7] Kilss, B., and Alvey, W. "Further Exploration of CPS-IRS-SSA Wage Reporting Differences for 1972," 1976 American Statistical Association Proceedings, Social Statistics Section.

[8] Millea, M. T. and Kilss, B. "Exploration of Differences between Linked Social Security and Internal Revenue Service Wage Data for 1972," 1975 American Statistical Association Proceedings, Social Statistics Section, 1976, pp. 138-146.

[9] Projector, D. and Bretz, J. Measurement of Transfer Income, Social Security Administration, 1973.

[10] Railroad Retirement Board. "Dual (RRA-SSA) and Special Guarantee Benefits as of December 31, 1972," Research and Statistics Notes, no. 7-73, April 1973.

[11] Shulman, H. "OASDI Benefit Amounts under 1972 Amendments," Social Security Bulletin, 1973, vol. 36, no. 1, p. 30.

[12] Social Security Administration. Social Security Bulletin, Annual Statistical Supplement, 1972.

[13] Social Security Administration. Medicare: Health Insurance for the Aged, Section 2, Enrollment, 1972, Washington, D.C., 1975.

[14] Scheuren, F., et. al. Studies from Interagency Data Linkages, Report No. 4, Social Security Administration, Washington, D.C., 1975.

[15] Social Security Administration, Studies from Interagency Data Linkages, Report No. 8, Washington, D.C., (in preparation).

[16] U.S. Senate Committee on Finance. Social Security Amendments of 1972, Government Printing Office, Washington, D.C., 1972.

PROBABILITY PLOTTING OF MULTIPLE DISCRIMINANT RATIOS

Kenneth W. Wachter, Harvard University

This report is about some spinoff from the plotting methods for principal components described at the Ninth Interface Symposium (Wachter (1976a)). The results go back to joint work of Colin Mallows and the present author in 1969 at Bell Telephone Laboratories, Mallows & Wachter (1970). What is new is a rigorous proof of the theorem, stated below, on which the probability plotting methods for multiple discriminant ratios and related quantities from large multivariate data sets can be based. The proof is given in a Harvard Research Memo, Wachter (1976c), which is now being expanded for publication. The present report is a brief sketch of the content, uses, and limitations of the results.

A typical problem which our methods address arises when we have groups of observations on, say, a dozen variables and we want to judge whether discriminant analysis is worth pursuing. We offer a quick graphical method for comparing the data against a standard null hypothesis under which the variables contain no information relevant for discriminating among groups.

In all cases our methods rely on calculations of the asymptotic empirical measure of a set of eigenvalues or singular values or other roots. Figure 1 displays the distribution function of an empirical measure, a step function with a step of size 1/p at each of p values. We deal with asymptotic situations in which, in one way or another, the number of steps, p, is going to infinity, yielding an increasingly gentle staircase like Figure 1 in contrast to the few big jumps of Figure 2.

Figure 1.

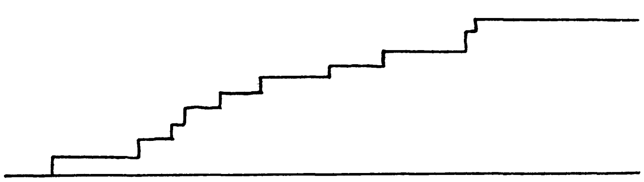
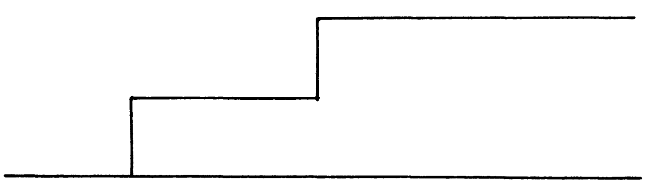


Figure 2.



* * *

The particular p values of interest to us are random variables $L_1 \ldots L_p$ whose joint density is proportional to

$$\Pi_{i=1}^{p} L_i^{(m-p-1)/2}(1-L_i)^{(n-p-1)/2}\Pi_{j<i}|L_i-L_j|$$

We call them the Fisher-Hsu-Roy-Girshick-Mood roots. They are the solutions to a familiar determinental equation in random matrices; their joint distribution was discovered simultaneously in 1939 by these five men. They play a role in multivariate analysis of variance and in canonical correlations as well as in discriminant analysis and go under many names. The equivalence of null hypotheses in these areas to the set of assumptions in Theorem 1 follows easily from the expositions in standard textbooks like Anderson (1958) or Dempster (1969).

Theorem 1 brings us the happy news that the empirical measure of the Fisher-Hsu-Roy-Girshick-Mood roots converges in distribution to a fixed limit with a density in simple closed form. The convergence takes place as the dimension and degree-of-freedom parameters p,m and n in the expression for the joint density go to infinity while their ratios approach fixed parameters.

Theorem 1: Suppose

1. Z or Z(n) is a p(n) by n dimensional matrix whose columns are mean-centered independent exchangeable multivariate normal random vectors and whose transpose is Z^*.
2. J is an n by n projection matrix of rank m.
3. K_n is the empirical measure of the p(n) positive solutions x to det $|ZJZ^* - xZZ^*| = 0$.
4. Prob $\mathbb{R}$ is the space of probability measures on the real line with the topology of weak convergence.
5. As $n\to\infty$, $p(n)/n\to\beta$ and $m(n)/n\to\mu$ with $\beta\leq\mu<1$. Then as $n\to\infty$, the random element K_n in Prob $\mathbb{R}$ converges in distribution to the fixed element of Prob $\mathbb{R}$ concentrated on $[A^2,B^2]$ with density

$$\sqrt{(y-A^2)(B^2-y)}\ /\ (2\pi\beta(1-y)y)$$

where

$A = \sqrt{\mu-\mu\beta} - \sqrt{\beta-\mu\beta}$ and $B = \sqrt{\mu-\mu\beta} + \sqrt{\beta-\mu\beta}$.

The difficulty of this theorem, responsible for the seven year gap between the discovery by Dr. Mallows and the author of the limit formula and the present proof, is stochastic degeneracy of the limit. For each triplet of finite values of p,m, and n, the empirical distribution function is a random step function. It is easy to imagine a limit which is still random rather than fixed. That the limit be fixed, that is, stochastically degenerate, is equivalent by Wachter (1974), 3.2, to the assertion that the roots be asymptotically independent. In effect we need this independence to simplify expressions for moments. The proof in Wachter (1974c) insures against random limits by a conditioning argument and throws the onus onto keeping convergence uniform over the conditioning events. It thus depends upon the new uniform convergence theorems for random matrix spectra proved in Wachter (1976b).

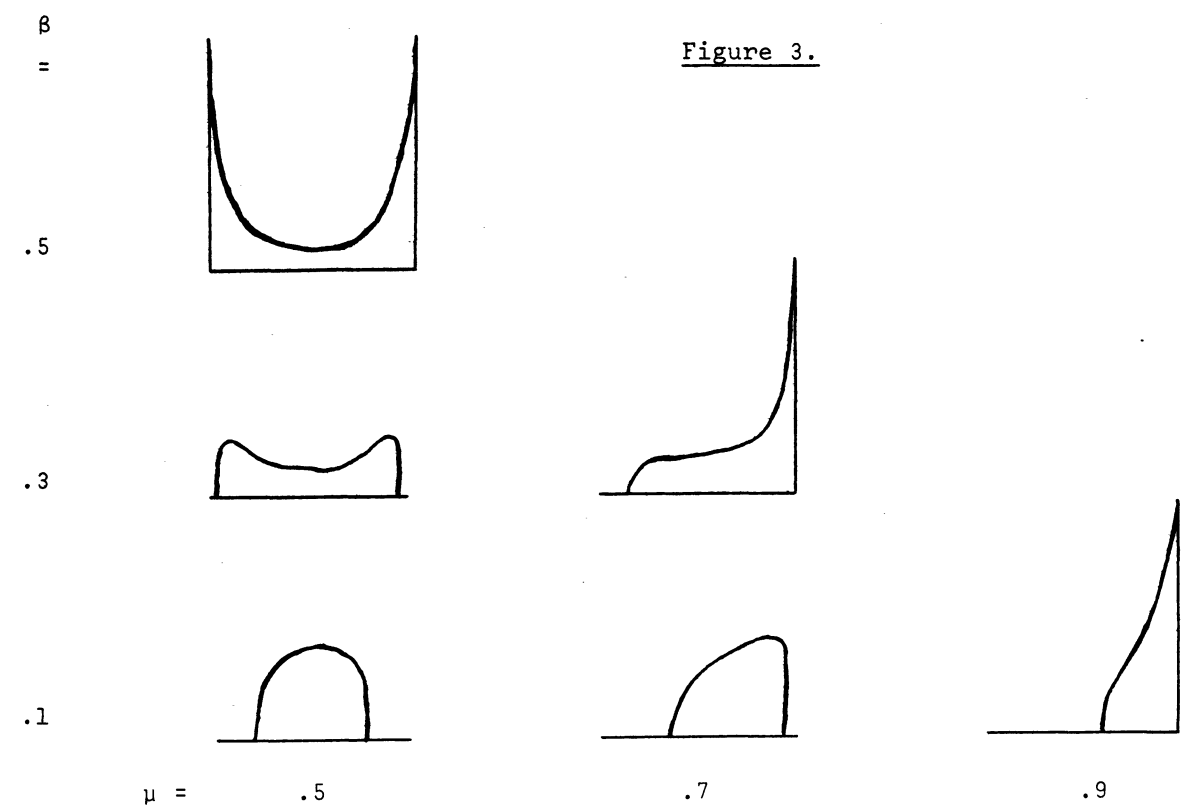
The assumptions in Theorem 1 are the standard null assumptions of the areas of application, but they are highly restrictive. In particular, they require multivariate normality. It is possible that uninteresting departures from normality, instead of interesting departures from the exchangeability of groups, could be responsible for bad fit of data to null hypothesis. Unfortunately, normality, via rotational invariance, is crucial to the proof of Theorem 1. On the other hand, the parallel theorems for principal components in Wachter (1976a and b) hold without any distributional assumptions beyond weak moment bounds, tempting us to conjecture the appropriateness of the limit formula even in the absence of normality.

No simulations have yet been done to test the speed of convergence in Theorem 1 itself, but simulations have been completed in the principal component case which suggest convergence rapid enough to make plots for p=12 or more informative.

Figure 3 shows densities for the limiting empirical measures of Theorem 1 for a variety of values of μ and β. Each graph is positioned in the figure according to its value of μ (for the x-axis) and β (for the y-axis). The operation of replacing μ by $1-\mu$ and reflecting each graph about the center would produce further cases.

The densities in Figure 3 give a more intuitive sense of how the Fisher-Hsu-Roy-Girschick-Mood roots tend to spread out when there are many of them than does, for instance, the expression for their joint density. Some surprises lurk in the graphs. Notice, for instance, that for small enough β and large enough μ most of the roots (which have an interpretation as squared correlations) clump above .99. Without theoretical guidance one might wrongly sieze on correlations this large as grounds for rejecting the null hypothesis. Notice also that when μ and β are near 1/2 the roots tend to separate into two bunches. Procedures which include linear discriminators in an analysis until a large gap between roots is reached would operate merrily in this situation, even though in fact the null hypothesis is true.

Figure 3.



Besides instructing our intuition, Theorem 1 provides plotting points for quantile-quantile plots of sets of roots from data against theoretical values under the null hypothesis. We illustrate this method with an example of a discriminant analysis, performed during research into talker identification at Bell Telephone Laboratories described in Bricker et al. (1971). We begin with measurements of p variables replicated k times on each of m groups. In this case the p=19 variables are spectral measurements on utterances. There are m=172 groups of utterances consisting of k=4 utterances each. Each group corresponds to a single talker repeating the same digit four times. The aim of the analysis is to discriminate between groups, with the hope of finding variables which will efficiently classify a new utterance into one of the groups, that is, assign an utterance to the person who spoke it.

Like much multivariate research, the data set in this problem is large. Many alternative analyses need to be tried and the cost of computing new variables in any one analysis can mount up. Thus there is a premium on methods which help us recognize hopeless from promising analyses at an early stage. It is precisely under such circumstances that our methods have something to offer.

Let W be the sample within-group covariance matrix pooled across groups and let B be a "between-group" covariance matrix, the sample covariance matrix of the group centroids or means. Our roots $L_1 \ldots L_p$, sometimes called discriminant ratios, are the p eigenvalues of $(W+B)^{-1}B$, though we can calculate them via singular value decompositions without actually forming W and B. The ordered roots for this data set appear in the second column of the table in Figure 4. Without a standard of comparison these eigenvalues may not seem especially illuminating. A standard of comparison is just what Theorem 1 provides.

Suppose R is the distribution function for the density in Theorem 1. In this problem (trading one degree of freedom for the grand mean) $n=172*4-1$ and we set $\beta=19/n=.028$ and $\mu=172/n=.249$. For each i define X_i to be the value such that $R(X_i)=i/(p+1)$. The computer routines listed in Wachter (1976a) can be used to find X_i if the subroutines DENSIT and RANGE are altered in an obvious fashion, but for pencil-and-paper accuracy a hand-calculator suffices. The values $X_1 \ldots X_{19}$ for this μ and β appear in Column C. We plot L_i on the y-axis against X_i on the x-axis for $i=1,2\ldots p=19$. Agreement between the data and the null hypothesis would manifest itself by points lying along a straight line through the origin at 45°.

Figure 4

i	L_i	X_i
1	.949	.369
2	.907	.349
3	.901	.331
4	.867	.317
5	.843	.301
6	.825	.289
7	.814	.278
8	.803	.265
9	.782	.252
10	.754	.241
11	.722	.229
12	.710	.219
13	.676	.207
14	.659	.198
15	.620	.185
16	.550	.175
17	.473	.162
18	.428	.151
19	.403	.138

Table of Eigenvalues L_i from Talker Identification Data and Quantiles X_i from Null Hypothesis with $\beta=.028$; $\mu=.249$.

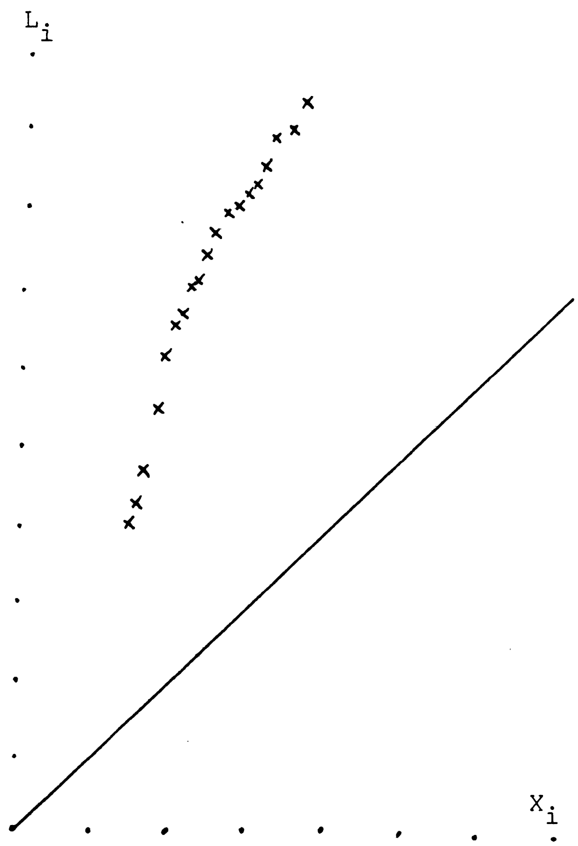
* * * * *

It is evident that the resulting plot, in Figure 5, shows little agreement with the null hypothesis. The points depart radically from the 45° line and their locus curves distinctly. In this problem this is welcome information, for it suggests that talkers are by no means exchangeable, so that additional labour invested in the calculation of discriminant coordinates ought to prove justified. In fact, as described in Bricker et al. (1971), identification strategies using five linear discriminators turned out to be highly successful at identifying new utterances by talkers in this population.

It would be desirable to go further and explore the departure from the null hypothesis indicated by Figure 5 by constructing plots of the L_i against quantiles derived for various alternative hypotheses of structure in the data useful for discrimination. For the parallel methods for principal components described in Wachter (1976a) quantiles under alternative hypotheses are available, but for discriminant ratios Theorem 1 does not furnish any information except under the null hypothesis. It offers no guidance as to how many discriminant coordinates we should retain in an analysis or how we should employ them. In this regard the present methods are severely limited in scope.

On the other hand, the quick graphical check of the data against the null hypothesis based on Theorem 1 becomes feasible for just those high-dimensional

Figure 5.



Q-Q Plot: Data versus Null Hypothesis

* * *

cases where tables of standard test statistics become sparse and intuition requires instruction. A by-product of more extensive results for principal components and general random matrix spectra, these methods fill a small but nagging gap in our collection of statistical tools for approaching multiple discriminant analysis and related fields.

REFERENCES

T.W. ANDERSON (1958). An Introduction to Multivariate Statistical Analysis, Wiley, N.Y.

BRICKER, GNANADESIKAN, MATHEWS, PRUZANSKY, TUKEY, WACHTER & WARNER (1971). "Statistical Techniques for Talker Identification," Bell System Technical Journal, 50, 1427-1454.

A.P. DEMPSTER (1969). Elements of Continuous Multivariate Analysis, Addison-Wesley, Reading, Mass.

C. MALLOWS & WACHTER (1970). "The Asymptotic Configuration of Wishart Eigenvalues," An Abstract," Annals Math. Statist. 41, 1384.

K.W. WACHTER (1974). "Exchangeability and Random Matrix Spectra," in Progress in Statistics, edited by Gani, Sarkadi, & Vince, North Holland Press.

K.W. WACHTER (1976a). "Probability Plotting Points for Principal Components," in Computing and Statistics: Proceedings of the Ninth Symposium on the Interface, edited by Hoaglin & Welsch.

K.W. WACHTER (1976b). "The Strong Limits of Random Matrix Spectra for Sample Matrices of Independent Elements, Parts I and II," typescript awaiting publication.

K.W. WACHTER (1976c). "The Limiting Empirical Measure of Multiple Discriminant Ratios," Research Memo W-76-2, NS-335, Harvard University Department of Statistics.

ACKNOWLEDGEMENT

This work was facilitated by Grant SOC-75-15702 from the National Science Foundation.

"GETTING IT ALL TOGETHER": THE DEVELOPMENT OF APPROPRIATE RELATIONSHIPS BETWEEN FEDERAL AND STATE GOVERNMENTS FOR STATISTICAL PROGRAMS

Katherine K. Wallman, Office of Management and Budget

The Importance of Federal-State Cooperation in Obtaining Subnational Data

The requirement for subnational data on the composition and characteristics of the population, the economic condition of small areas, and the nature, scope and effects of programs designed to meet the needs of the nation and its people is shared by governments at all levels. At the Federal level, information is needed to formulate legislation which will respond to critical needs, to allocate resources available for distribution to State and local governments, and to assess the degree to which federally supported programs are meeting specified goals. At the State level, small area data must be used in the sub-state allocation of federally available resources, and in the planning, development and implementation of State initiatives to complement those sponsored by the Federal government. Local governments, which play the critical role in ensuring that goods and services are appropriately and equitably distributed, must use similar, though frequently more detailed, information to allocate available support and to determine the nature and scope of needs which are not or can not be met with existing resources.

The use of subnational statistics in the allocation of Federal funds alone serves to indicate the importance of complete and accurate State and local data. In Fiscal Year 1975, fourteen Federal agencies[1/] used data on total population as the primary element in distributing the benefits of 75 programs. The obligations under these programs, which resulted in the provision of almost $33.7 billion to State and local governments, were distributed as matching categorical grants, fully funded categorical grants, or bloc/revenue sharing grants.[2/] In addition to data on total population, information on per capita income, specific age groups, low income, and/or unemployment must be used to complete the allocation of resources under the majority of the Federal programs providing benefits to State and local governments.[3/] When the data required to perform governmental functions beyond the allocation of Federal funds are added to this base, the breadth and depth of information needs expand significantly.

Why should the Federal, State and local governments cooperate in the production of statistics? At a minimum, because it is clearly in the interest of the benefit recipients to provide complete and accurate population estimates. Exhibit I to OMB Circular A46 requires the use of standard data on total population for all Federal programs which make use of total population data in the distribution of Federal benefits. The data on total population must be the most current and comprehensive published by the Bureau of the Census. Under the Census Cooperative Federal/State Program on Local Population Estimates, the State and Federal governments are finally working together to produce the single set of county estimates required for Federal as well as State planning and funding purposes.

But the scope of mutual data needs extends well beyond the production of population statistics. The Congress through legislation, and the Federal executive agencies through program regulations, have increasingly emphasized the responsibility and role of State agencies in the treatment of national problems. At each level of government, both general statistics and program specific information are needed to plan, administer, and evaluate the allocation of resources. While less detailed data may be needed as we move from the local to the State to the Federal level, consistent, comparable, and accurate information is needed across the board. Federal and State governments alike have recognized the need to work jointly on information collection and production activities to reduce respondent burden and to improve the scope, uniformity and quality of statistical data. Cooperative statistical programs represent a necessary and viable approach to meeting needs in the most economical and efficient manner.

The Objectives and Status of the Federal-State Cooperative Statistical Programs

The Federal-State Cooperative Systems of Data Collection include those federally initiated or sponsored statistical programs in which State agencies participate in the collection, processing or utilization of routinely needed information. Such initiatives are characterized by joint efforts of representatives at all governmental levels to produce or use federally standardized information about State or local activities or functions. The cooperative systems are undertaken for the mutual benefit of the participants, and contain data of a recurrent nature which is intended to have broad applicability. Several ongoing Federal-State cooperative statistical programs are described below.

The Crop and Livestock Reporting Program-Statistical Reporting Service, Department of Agriculture.

The objective of The Crop and Livestock Reporting Program is "to avoid duplication of effort and confusion of conflicting reports and to promote economy and efficiency of operations." The system is predicated on the fact that the "Federal service is primarily concerned with national estimates and estimates for the individual States" while the State agency "is particularly interested in the collection, development, and publication of more detailed information than is provided in the Federal program of reports."

A key feature of the Agriculture program is the provision by the Federal government of an agricultural statistician who directs the program in the State. The terms of the agreement establish a formal framework within which the national government retains control of the statistical techniques and operations used in each survey by making them the responsibility of its own employee as "Statistician in Charge." The Statistical Reporting Service supports approximately 400 field-based professionals responsible for the crop and livestock reporting program in 47 participating States.

A unique feature of this program is the provision of funds by the States to the Federal Government. These funds, paid under the terms of the individual State cooperative agreements, are used by the Statistical Reporting Service to cover the additional cost of collecting and processing the sub-State (county-level) data required by the State over and above the national need.

The Employment, Hours and Earnings System-Bureau of Labor Statistics, Department of Labor.

The Cooperative Employment, Hours, and Earnings System is designed to eliminate duplication of collection efforts at the State and Federal levels, to increase the comparability and timeliness of information collected, and to extend the scope of coverage of the survey to the broadest possible number of establishments. Like the Department of Agriculture, the Department of Labor supports personnel at the State level as the key strategy for accomplishing its objectives.

In cooperation with the Employment and Training Administration, the Bureau of Labor Statistics funds approximately 400 positions in State Employment Security Offices. These individuals are responsible for editing data and completing follow-up activities with non-respondents. In addition, BLS supports an average of two to three professionals in each of eight Federal regional offices. The regional personnel are responsible for training new State employees, answering technical questions relating to surveys, and monitoring the application of survey methodology by the States. Staff at the national level complement the efforts of field personnel by:

(1) Producing survey manuals to ensure comparability,
(2) Sponsoring training seminars, and
(3) Developing and disseminating new analysis techniques.

The Cooperative Federal/State Program on Local Population Estimates-Bureau of the Census, Department of Commerce.

The Census Cooperative Federal/State Program on Local Population Estimates was initiated in 1967 to develop an annual series of county population estimates prepared jointly by the Bureau of the Census and the State agencies designated by their Governors to work with the Bureau. This activity results in the production of a single set of county estimates for potential State and Federal planning and funding purposes, and concurrently increases the timeliness of reporting through the direct involvement of a single State agency designated by the Governor.

Although no funds are specifically allocated to the States under the Census Cooperative Program, the budget for the local population estimates effort has proven adequate for the additional activities, including the conduct of regional conferences to determine the estimation methods most appropriate to each State as well as the publication and provision of annual county population estimates to each of the States.

The Medicaid Management Information System-Social and Rehabilitation Services, Department of Health, Education and Welfare.

The Medicaid Management Information System was authorized by legislation enacted in 1972 (P. L. 92-603). Development of the general systems design, including specification of the data elements to be maintained in the automated systems at the State level, was completed under contract at the Federal level. State personnel are, in turn, being trained by various systems design and development contractors, and are responsible for submitting individual State plans detailing the State's proposed system development procedures.

Upon apprcval of a State's plan, the agency becomes eligible to receive 90 percent matching funds for the execution of system development activities. Although funds were not available for grants to the States until 1974, the law allows States to receive matching dollars retroactive to 1972, if they have developed systems meeting the approved design specifications. The Medicaid Management Information System is expected to improve data quality and increase the availability of information for program management, monitoring, and planning. In addition to the matching developmental funds, the law allows for provision of 75 percent matching funds to the States for costs of operating the system. This assistance is made available to the State following approval of its operational system design and products by the Social and Rehabilitation Service.

The Cooperative Health Statistics System-National Center for Health Statistics, Department of Health, Education and Welfare.

The Cooperative Health Statistics System is designed to establish a coalition among the various levels of government for the purpose of developing an economical and effective system to assess the health status of the nation. Simultaneously, the implementation of the Cooperative System has the goal of eliminating various problems of earlier health statistics activities, including:

(1) Duplication of effort at the Federal, State and local levels;
(2) Inequities in the support provided to the States for data system development and operation;
(3) Lack of small area data in sufficient detail and with large enough samples to meet needs within States;
(4) Inadequate supplies of trained statistical manpower and computer capabilities; and
(5) Unmet State needs for data systems assistance, and information for the planning and administration of State and local health programs.

The establishment of the Cooperative Health Statistics System has been planned in two overlapping stages: a research and development phase and an operational phase. The research and development phase, instituted late in 1971, is designed to examine alternatives for the content, structure, and methods of the Cooperative System in order to identify prototypes for the several statistical components which would include comparable definitions, standards, and procedures to collect, process, analyze and present the needed vital and health statistics through a Federal-State-local data network. Activities under the research phase of the program are being conducted through the funding of six to eight developmental contracts for each of the system components. The results of the independent investigations are then reviewed and analyzed, leading to the selection of a single recommended model for the particular component.

Under the operational phase of the system, contracts are awarded for planning the individual State's program for collection of a particular statistical component, and subsequently for the implementation of individual system components. To encourage coordination among the several State-level health agencies which may be responsible for implementation of various statistical components, in some cases the National Center for Health Statistics has funded a position at the State level for providing liaison among the contributing agencies.

When fully designed and implemented, the Cooperative Health System will provide statistics on health manpower, health facilities, hospital care, household health status, ambulatory care, long term care, and vital records. To date, 39 of the States have participated in the planning phase of the program for at least one component, and more than 80 contracts have been awarded to the States for implementation of various components.

In addition to the research, planning, and implementation contracts for system components, the Cooperative Health Statistics System provides for:

(1) Training of health statisticians for systems implementation and data analysis;
(2) Expansion of technical assistance by NCHS; and
(3) Extension of services provided by the Data Use and Analysis Laboratory to develop, test, and demonstrate the application of special tabulations and analyses of data generated by the Cooperative System's several components for community use. Further, this activity supports special research and development in the problem-oriented analysis of a variety of data useful for local planning purposes.

The Common Core of Data Program-National Center for Education Statistics, Department of Health, Education and Welfare.

The Common Core of Data Program represents the first comprehensive effort to improve the scope, quality, and timeliness of information on the nation's educational system. Through the review of educational issues and an assessment of current data collection activities, NCES expects to specify the new types of information which are needed and to eliminate the collection of data which is no longer relevant. Following completion of the design effort, NCES will modify its surveys as necessary, using standard terminology which has been developed cooperatively over a period of years by the Federal, State, and local education agencies.

For the near-term, limited manpower and dollar resources have made it necessary for NCES to concentrate the Common Core of Data efforts on meeting the recurring statistical data requirements of Federal legislators and administrators. Implementation of the revised data sets, however, is expected to benefit both producers and users of education statistics. The producers can anticipate a reduction in duplicative reporting, for the Common Core program is intended to increase the compatibility of data now sought by numerous components of the Education Division. The users can expect increased timeliness of data, as well as improved quality, based on the adoption of standard definitions by the recordkeeping and reporting agencies.

Plans for a longer term, integrated Federal/State/local system of educational data are also included under the Common Core of Data effort. Implementation of a more ambitious program, however, will not be undertaken until additional resources are available for use at the State and local levels. Under the Education Amendments of 1974 (P.L. 93-380), the National Center for Education Statistics has been specifically mandated "to assist State and local agencies in improving and automating their statistical and data collection activities." NCES is currently studying alternative modes of providing assistance to data producers and users. In cooperation with the State agencies, NCES will develop a multi-year plan for the provision of statistical systems assistance. Implementation of the assistance program is expected to pave the way for an expanded cooperative program of education statistics.

The Criminal Justice Comprehensive Data System-Law Enforcement Assistance Administration, Department of Justice.

The Comprehensive Data System is designed primarily to provide State and local governments with the mechanisms to produce and analyze statistical data for planning, management, and evaluation. As a byproduct, the program provides to the Federal government the national data needed on crime and the administration of justice. Prior to the establishment of this program, almost no data was available on the administration of justice at the State and local levels.

To participate in the Comprehensive Data System, a State must establish a Statistical Analysis Center, develop an Offender-Based Transaction Statistics program, assume responsibility for Uniform Crime Reports, institute a Management and Administrative Statistics program, and design a program to provide technical assistance to participating local agencies.

The Statistical Analysis Centers are responsible for coordinating all of the State's activities in the Comprehensive Data System, and are expected to provide analytical capabilities for policy makers within their respective States. These Centers, which are now operational in 36 of the States, provide input to the Criminal Justice State Planning Agencies.

The Offender-Based Transaction Statistics program is designed to identify persons arrested for serious offenses, and to follow the individual as he passes through the criminal justice system. The original Comprehensive Data System concept required that the Offender-Based Transaction Statistics be developed in cooperation with a computerized criminal history program under development by the FBI; subsequent experience with this concept has argued against its continuation.

The Uniform Crime Reports System, the oldest of the existing criminal justice statistical programs, was incorporated into the Comprehensive Data System in 1972. With this shift, the State agencies became responsible for collection and quality control, including the application of Standardized Audit Systems, of data on crimes known to the police, arrests, offenses cleared by arrest, law enforcement employees, law enforcement officers killed, and other aspects of law enforcement.

The Management and Administrative Statistics Component of the Comprehensive Data System is expected to provide a mechanism for the States to examine their internal needs for data to describe the State systems and to develop the programs to provide the needed information. The Organization of Directors of State Statistical Analysis Centers is attempting to develop a common set of Management and Administrative Statistics data requirements. It is anticipated that many of the LEAA programs could be modified in the future to use State-provided data in place of information now collected by the Bureau of the Census.

Funding for State and local participation in the Comprehensive Data System is provided under three sections of the Omnibus Crime Control Act of 1973. Grants are awarded to the States for the development and operation of the Statistical Analysis Centers from resources available for National Criminal Justice Information Statistical Services. These funds are used to support core staff at the State level to produce federally required data, and, in combination with other resources, are used to fund contracts for the development of program components as well as State participant conferences. Operation of the Comprehensive Data System components at the State level is supported through discretionary monies available under two additional sections of the Omnibus Act. These funds, which are provided to the States under matching grants, are used to implement the various program components of the Comprehensive Data System.

An Assessment of the Current Federal-State Cooperative Statistical Programs

Consistently, the Federal-State Cooperative Systems of Data collection include the specification of federally required information and the definition of data elements included in the programs. At varying levels, attempts are made to ensure the quality and comparability of data within and across the States. And, by definition, each of the cooperative programs involves coordination and joint participation of the Federal and State agencies, as well as involvement of other relevant data producers.

Outnumbering these common features are the cooperative systems' characteristics which vary from program to program. For example, some of the cooperative systems in existence today involve all States equally (e.g. the Cooperative Employment, Hours and Earnings System of the Bureau of the Labor Statistics), while others, such as the Cooperative Health Statistics System, involve the States at variable levels. Reasons for the variability in State participation in the several cooperative systems include lack of interest on the part of the States, inadequate resources at the Federal level to support nationwide participation, and inability of the States to meet the requirements of joint participation.

Likewise, the scope of the several Cooperative Systems varies substantially. Some programs, such as the Census Cooperative Federal/ State Program on Local Population Estimates, deal with a single, clearly defined problem, while others, such as the Comprehensive Data System of the Law Enforcement Assistance Administration, attempt to meet a number of diverse needs.

In addition to variations in scope and coverage, there are also differences in the types of data which are collected through the cooperative programs. The major thrust of cooperative Federal-State efforts has been on the direct production of general purpose, baseline statistics. Two programs of long standing, the Crop and Livestock Reporting Program at Agriculture and the Employment, Hours and Earnings Program at Labor are clearly in this category. Among the newer cooperative endeavors, the Census Cooperative Federal/State Program on Local Population Estimates, the Cooperative Health Statistics System, and the Common Core of Data for Elementary and Secondary Education all have a primary emphasis on the production of general statistics. The Medicaid Management Information System, and the Comprehensive Data System of the Law Enforcement Assistance Administration, on the other hand, are designed primarily to produce administrative data required for the operation and evaluation of programs at the State level. In addition, several of the newer cooperative systems contain components to provide services over and above data collection and processing *per se*. For example, the Comprehensive Data System of the Law Enforcement Assistance Administration provides funding for State Statistical Analysis Centers, and the Cooperative Health Statistics System includes training of health statisticians for systems implementation and data analysis, as well as provision of direct problem solving assistance to States through the production of special tabulations and analyses.

The variability in the types of data collected through the Cooperative Systems is accompanied by inconsistencies among the programs in the degree of coordination within the respective Federal agencies between the particular cooperative program and other statistical and data collection activities of the agency. More specifically, there is no consistent policy or procedure followed by agencies having cooperative programs to ensure that data needs of agency components other than the system sponsors are considered or met in the design and implementation of the cooperative program. Furthermore, there is virtually no coordination among Federal agencies, and among the respective sponsors of cooperative statistical programs, to ensure that cross-agency needs are addressed through the statistical systems sponsored.

Coupled with the inconsistencies in the degree to which the cooperative statistical systems serve Federal agency statistical needs are variations in the extent to which federally-sponsored cooperative programs are interdependent with State statistical efforts, and serve State information needs. Programs such as the Medicaid Management Information System are clearly designed to provide data needed by the States, with a byproduct being the capability to produce a more limited set of statistical information required by the Federal government. The Common Core of Data Program of NCES, on the other hand, has been primarily designed to standardize and produce data required by the Federal government, with the service of specific state information requirements left as a secondary, though useful, byproduct.

Finally, in the development and operation of the Federal State Cooperative Systems of Data Collection there are considerable differences in the types of administrative and financial arrangements which are employed. In the Crop and Livestock Reporting Program of USDA, a Federal employee actually directs the program in each of the States; in most of the other programs, the Federal government contracts with the States to produce the required information. The Census Cooperative Program, however, provides neither manpower nor dollars to the participants, but relies on technical coordination and consultation with State representatives to produce the needed results. The extent to which State personnel are involved in the design of the cooperative systems also varies from program to program. In the case of the Medicaid Management Information System, for example, the system design was completed primarily at the Federal level, while alternative designs for the components of the Cooperative Health Statistics System were initially developed through contracts with State agencies. Similarly, in the design and implementation phases of the cooperative programs some Federal agencies have worked essentially one-on-one with their State counterparts while others, such as the National Center for Education Statistics, have sponsored multi-state and national conferences as an integral part of the systems development phase.

Proposed Policies for the Improvement of Federal-State Cooperative Statistical Programs

From the foregoing discussion, it is clear that the current Federal-State Cooperative Systems of Data Collection are, individually and collectively, serving a variety of statistical needs of the State and Federal governments. Equally evident, however, is the fact that the programs now in existence have been developed by the agencies in isolation from one another and in the absence of any policy guidelines at the Federal level. Considerable differences exist among the programs in the extent and types of Federal agency data needs which are served, the control which the agency exercises over systems' operations and standards, the mechanisms employed for coordinating systems' design and implementation activities, and the level and type of support to participants.

In spite of the efforts of the individual Federal agencies to work cooperatively with State and local governments in the production of statistics, the demands for information at all levels have proliferated in response to the requirements of statutory allocation formulae, program targeting needs, and decentralization of decision making. Further, the lack of coordination in data collection activities among the various agencies has led to duplication of effort, non-transferability of data, and insufficient attention to particular data sets which may be needed by one or more agencies.

At OMB, we are placing renewed emphasis on the improvement of the Federal-State Cooperative Statistical Systems. While our chief role will be to bring greater order and coordination to these programs, we believe our efforts will also impact upon the individual agency initiatives. Our primary vehicle for instigating change will be "A Framework for Planning U.S. Federal Statistics, 1978-1989." The framework, which is being drafted by OMB and will be developed in cooperation with data users and producers across the nation, will recommend policies and steps to be taken to improve the Federal statistical system. The policies which are proposed in the balance of this paper concerning the Federal-State Cooperative Statistical Programs are among those which may be included in the Framework. They will be subject to review and modification during the coming year. We would welcome your comments on our proposals.

Based on the information now available, it appears necessary, first and foremost, to define the appropriate role of the Federal government and the limits of Federal responsibility for the cooperative statistical programs. The definition of the role and responsibility of the Federal government must include, at a minimum, policies with respect to:

- the degree to which the cooperative systems should focus on data production in contrast to data utilization;
- the extent to which non-Federal needs for State and local area data should be incorporated, and the associated costs for their production and use underwritten, by the Federal government; and
- the appropriate division of labor and costs among cooperative program partic-

ipants at the Federal, State and local levels.

Given the likelihood that the Federal-State Cooperative Systems of Data Collection will be required, individually and collectively, to operate in an environment of limited budgetary resources, it will be necessary to establish priorities within the virtually limitless activities which could be sponsored or supported from the Federal level. Thus, we recommend that the responsible Federal agencies place the highest priority on efforts which will enable all States to produce the minimum data sets prescribed by their respective agencies. Coupled with this emphasis should be continuous initiatives by the Federal agencies to work with their State counterparts in the design of data systems so that States may capitalize on the Federal development in establishing the more elaborate systems which may be required to meet their own needs. Non-Federal data needs should be incorporated by the States individually; the costs for their production and use should be borne by the States, whether such costs are incurred by the State for its own internal processing or for the additional expense of using federally sponsored systems to process State information not needed at the national level. The Federal sponsors of cooperative systems should be responsible for:

- the design of data bases required to meet Federal statistical reporting requirements;
- the development and dissemination of data standards and definitions of terminology to ensure inter-State comparability;
- the sponsorship of seminars and training sessions necessary to ensure the consistent application of the data system; and
- the provision of matching funds to support systems development and operation to the States willing to meet the federally prescribed requirements and standards.

The balance of the burden for systems operation should be borne by the participating states.

Second, our review suggests that each Federal agency should establish mechanisms to ensure that its Federal-State Cooperative System of Data Collection is systematically integrated with the agency's overall statistical program. Too often, the cooperative programs have been developed and implemented by separate, specially established units of the agencies' statistical centers. The result has been that the individual cooperative programs have not taken into account, let alone been designed to serve, the relevant statistical needs of the sponsoring agency. For the States, this pattern has caused substantial difficulty as they have attempted to meet the requirements of the ongoing programs as well as those of the so-called "cooperative program." A further problem has been the almost systematic exclusion of federally-required administrative data from the scope of the cooperative systems. At the State and local levels, data producers have thus been required to collect and provide sometimes duplicative and often differently defined data sets to the same Federal agency. Each of the Federal agencies sponsoring a cooperative statistical system must take the steps required to eliminate this condition, including, if necessary, administrative and program realignments.

Third, it appears that general and agency specific standards and guidelines should be developed for and applied to the collection of data through the cooperative statistical programs. In each of the Federal-State Cooperative Systems of Data Collection, some attempt has been made to prescribe the definitional, quality, and timeliness standards which should be followed in the production of the required data by the participating States. But the completeness of these standards, and their application in the operation of the various programs, have been spotty. The development of such standards should be the responsibility of the Federal agency sponsoring the program; adherence to such standards should be a necessary condition of the States' eligibility to receive Federal support for the production of statistical information. Staff should be placed in the Federal regional offices to provide the training and technical coordination necessary for the States to meet the prescribed standards, and to monitor the States' operations to ensure that standards are applied and met.

Fourth, our investigation suggests that each of the Federal agencies should review the status of its cooperative statistical program as a basis for determining the nature and scope of technical and financial assistance needed by the States. Recent reviews of the cooperative statistical systems by the Office of Management and Budget have revealed substantial variations in the ability of the States to participate in the several federally sponsored programs. In order to bring all of the States to a point where they can provide the minimum data sets required, it will be necessary to shift priorities and establish assistance and funding initiatives to address individual State problems.

Finally, it is proposed that the Statistical Policy Division of the Office of Management and Budget establish mechanisms to coordinate the activities of the several Federal agencies sponsoring Cooperative Statistical Programs, and, in addition, establish a focal point in each of the States to coordinate State-level input to the Federal level on the Cooperative Systems. Many of the problems of the Cooperative programs are specific to the individual agencies, and are best handled by the respective Federal and State participants. Other difficulties in program development and operation have resulted from a lack of guidance from OMB on the appropriate scope and respective responsibilities of participants for these efforts. Even if all of these needs were met, however, certain roadblocks to the successful operation of joint participation in the collection, processing and utilization of federally standardized information about State and local activities and functions would remain. These problems arise from the lack of coordination at the Federal level of the several cooperative programs, and the parallel absence of centralized efforts in most States to review and evaluate the demands and operations of the Federal-State Cooperative Systems of Data Collection. In those States where efforts have been made to coordinate and centralize statistical activities, progress has been stymied by the lack of coordination at the Federal level. The result at the Federal level is a general lack of knowledge on the part of the several agencies concerning the content and structure of other federally sponsored programs. At the State level, there is evidence that the lack of Federal and State coordination among the several programs has resulted in inconsistent and duplicative demands for data by various Federal agencies.

OMB, the sponsoring Federal agencies and the State participants must join together to address these problems. The Statistical Policy Division of OMB is ready to undertake the leadership and coordination activities which will be required for the participants in the Cooperative Programs to "get it all together." The responsible Federal agencies must cooperate in assessing their current status, preparing action plans for addressing needs, implementing necessary changes, and working with OMB and other agencies to arrive at optimum solutions. The joint participation of data producers at the State and local levels will be a necessary precondition to the success of any policy development or implementation.

FOOTNOTES

[1]The fourteen Federal agencies include: the Departments of Agriculture, Commerce, Defense, Health, Education and Welfare, Housing and Urban Development, the Interior, Justice, Labor, Transportation, Treasury, the Environmental Protection Agency; the Appalachian Regional Commission; the Civil Service Commission; and the Water Resources Council.

[2]$22.5 billion were distributed to State agencies; $11.2 billion were distributed directly to local agencies. Approximately one-third of the total funds were made available through bloc/revenue sharing grants, while two-thirds of the resources were provided under full or matching categorical grants. Data on specific age groups were used to allocate an additional $3.1 billion in Federal resources, primarily for education programs. (Grants-in-aid in the form of loans and direct transfers to individuals are excluded from all figures cited).

[3]For further details, see "The Use of Data on Population in Federal Grants-in-Aid to State and Local Governments in Fiscal 1975" prepared by Charles A. Ellett, Statistical Policy Division, U.S. Office of Management and Budget.

A STOCHASTIC DIFFERENTIAL MODEL FOR THE EFFECTS OF FAMILY CONFIGURATION ON INTELLIGENCE

Richard L.W. Welch and Martin W. Denker, Department of Psychiatry
Chris P. Tsokos, Department of Mathematics
University of South Florida

I. Introduction

The model discussed in this study is a refinement and extension of another proposed by the authors in an earlier report (7). As such, it continues the overall process of this research group, that is, refining existing models and developing new ones for the family as a social-psychological system. Implicit in our work is the attempt to improve definition and quantification of the fundamental variables, and the processes interrelating the variables, used in the study of social systems. Also included in this effort is the definition of the variables as random variables.

The variable chosen for this investigation was that of adult intelligence and the development of intelligence in children from birth to age 19 in the context of their family system. The choice was made because of the existence of large and well-conducted studies of the subject (1,8,9), including an early attempt to develop a mathematical model which explained the observed phenomena. In this model, proposed by R.B. Zajonc and G. Markus in 1975 (8), it was speculated that the process proposed to describe the development of individual intellectual capability in a familial context would hold for many other psychological and psychosocial variables. These variables might include other aspects of problem-solving ability in adults, such as creativity, or the development and expression of truly new concepts, affiliative behavior, or the numbers and types of interpersonal relationships formed by the adult, and stability, or the "success" of interpersonal relationships among adults. Other investigators, approaching social systems from different perspectives, have arrived at descriptions of growth processes which are analogous to that proposed by Zajonc and Markus; see (3).

The study of intelligence within a social system also provides the beginnings of a quantitative model for the interaction of biological and sociological variables. This includes the traditional interaction of "nature" and "nurture" to produce individual human capabilities and behaviors. Intelligence, as measured by a specific instrument, is a variable which indicates the biological, or "nature" component, and family configuration, as measured by family size and birth order of the siblings, is a group of variables representing the psychological or "nurture" side.

As mentioned in the earlier report, this type of work had been generally called birth order research, and had been in somewhat of a state of confusion and mild disrepute (3,4) until the publication of the Dutch study of Belmont and Marolla in 1973 (1). Since then, many investigators have agreed that family size and birth order have been definitively shown to be correlated with intelligence, as measured at maturity. Zajonc and Markus proposed a model for a developmental process which could produce the observed outcomes, and which also stressed the importance of spacing between adjacent siblings (in terms of age). The present investigators reported certain refinements in that model (7).

It is the purpose of the present report to indicate how further refinements will improve the model. The refinements include the re-definition of the variates as random variables and the treatment of the interactions among the variables as continuous stochastic processes. The proposed refinements of the model will need to be tested against other sets of data to confirm that it is indeed an improvement in predictive accuracy. But if the improvement can be shown, the model would then have a number of implications for theories of the general function and structure of social systems which will be described in the concluding section of the present work.

II. Formulation of the Model

The authors have previously proposed (7) a deterministic differential equation for the Zajonc and Markus model. This is given by:

$$\dot{M}(t) = \{ \dot{\ln}\, \alpha(t) - 2k^2 t \} M(t) + 2k^2 t \alpha(t) \tag{2.1}$$

where the dot denotes the derivative with respect to time, t is the age of an individual in years, $M(t)$ is the intellectual level of that individual, and k is an arbitrary rate constant. The function $\alpha(t)$ is a known, continuously differentiable function of time which takes into account the effects of the individual's environment on his development. The initial condition for Equation 2.1 is $M(0) = 0$. The major assumption in this model concerning the psychological process is that the rate of change of the intellectual development (or the rate of growth of intelligence) is linearly proportional to the intellectual level itself. This formulation, based as it is on the studies of Belmont and Marolla (1) and the work of Zajonc and Markus (9) seems to offer a reasonable and realistic approach.

In our earlier paper, we pointed out that the quantities k and $\alpha(t)$ should, however, more realistically be considered as random variables with certain probability distributions. Treating them as physical constants does not seem to reflect the psychological situation accurately. Thus, some of the uncertainty in measuring intelligence and some of the genetic and environmental differences, both within and between individual families, are expressed in the randomness of these variables in the model. A stochastic version of Equation 2.1 may be expressed by:

$$\dot{M}(t) + A(t)M(t) = Y(t) \tag{2.2}$$

The solution of Equation 2.2 is a stochastic process, M(t), which gives the intellectual level of an individual at age t.

We will assume that both A(t) and Y(t) are random variables with finite second moments. According to Soong (5), the stochastic process M(t), $0 \le t \le T$, is a mean square solution of Equation 2.2 if : (1) M(t) is mean square continuous on the interval {0, T }, that is, $M(t + h) \overset{ms}{\to} M(t)$ as $h \to 0$ for each $t \ge 0$; (2) M(0) = 0 with probability one; and (3) A(t) × M(t) + Y(t) is the mean square derivative of M(t) on 0, T . The method of Soong (5, chapter 8) may be utilized to compute the mean square solution of Equation 2.2, which is given by:

$$M(t) = \int_0^t Y(u) \exp\left\{ - \int_u^t A(s)ds \right\}du \tag{2.3}$$

For a more sophisticated discussion concerning the stochastic structuring of such systems which include the nonlinear cases the reader is referred to the recent monograph of Tsokos and Padgett (6).

In the Zajonc and Markus model, the variable $\alpha(t)$ is essentially flat, almost constant; furthermore, its range is small when compared to the total magnitude of the process M(t). Thus, it would be useful as a first approximation to regard it as constant. In other words, we will take $\alpha(t) = \alpha_0$, where α_0 is a random variable.

We have that:

$$A(t) = 2k^2t \tag{2.4 -i}$$

and:

$$Y(t) = 2k^2t\alpha_0 \tag{2.4 -ii}$$

Therefore, the mean square solution is

$$\begin{aligned} M(t) &= \int_0^t 2k^2\alpha_0 u \exp\left\{ - \int_u^t 2k^2 s ds \right\}du \\ &= \alpha_0 \{ 1 - \exp(- k^2t^2) \} \end{aligned} \tag{2.5}$$

which is symbolically identical to the deterministic solution. Knowledge of the probability distributions of the random variables α_0 and k will enable one, through the use of standard techniques such as derived distributions, to calculate the probabalistic behavior of the solution (2.5).

Another logical choice for $\alpha(t)$ would be an exponential function of the form;

$$\alpha(t) = a_0 + a_1 \exp(-c_1t) + a_2 \exp(c_2t), \tag{2.6}$$

where some or all of the coefficients a_0, a_1, a_2, c_1, and c_2 can be treated as random. The exponential sum is an appropriate form for modelling the behavior described by Zajonc and Markus in (9). They postulated that rather than remain constant, the family process variable $\alpha(t)$ would shift subtly over time, due to the effect of later births, deaths, and other changes in the family's structure. Note that (2.6) is based on the sum of a constant and two other terms, of much lesser magnitude, which reflect deviations from the constant level. Equation 2.6 represents a continuous fit to the step-function, discrete model (7).

Substituting (2.6) in (2.1) and (2.2), we have;

$$A(t) = 2k^2 \frac{-a_1c_1 \exp(-c_1t) + a_2c_2 \exp(c_2t)}{a_0 + a_1 \exp(-c_1t) + a_2 \exp(c_2t)} t \tag{2.7 -i}$$

and;

$$Y(t) = 2k^2t \{a_0 + a_1 \exp(-c_1t) + a_2 \exp(c_2t)\} \tag{2.7 -ii}$$

After putting (2.7) in (2.3) and integrating, the mean square solution is, as before,

$$M(t) = \alpha(t)\{ 1 - \exp(-k^2t^2)\}.$$

This section closes with a comment about the domain 0, T of the problem and the level at maturity of M(t). It is easy to see that M(t) is dominated above by $\alpha(t)$. Also, $\alpha(t)$ itself is unbounded for large t in (2.6). The model is defined, however, only for $t \le T$. The bound T represents the point at which the individual leaves the family setting. This does put a limit on the growth of M(t); it is assumed that for $t > T$, the intellectual level is essentially stable and the period of its rapid increase is over.

III. An Example

This example makes use of the notation and discretized version of the model (2.1) in (7) and (9). Some familiarity on the part of the reader with at least the former reference must be assumed. Because the Dutch data was measured at age 19 on each subject, and was not measured longitudinally through time, it is not possible to estimate the coefficients for $\alpha(t)$ directly. It is possible, on the other hand, to approximate them in the following fashion.

The data consists of 45 scores (transformed) on the Raven Progressive Matrices test, representing the mean score for the ith child in a family with j siblings, $i \leq j$, $i,j \in \{1,2,\ldots,9\}$. These mean values are substituted into the discrete model, which is then solved for 45 coefficients α_{ij}. We form the step function;

$$\alpha'(t) \equiv \alpha_{ij_0} \tag{3.1}$$

for an ith child, where j_0 is the number of siblings in the family when child i is t years old ($i \leq j_0$, $i,j_0 \in \{1,2,\ldots,9\}$). This representation (3.1) involves the assumption that $\alpha'(t)$ changes only at the birth of each later sibling. The desired coefficients are then estimated by fitting a regression curve to the points where $\alpha'(t)$ jumps, that is, by choosing $\hat{a}_0$, $\hat{a}_1$, $\hat{a}_2$, $\hat{c}_1$, and $\hat{c}_2$ to minimize

$$\sum_{n=1}^{j} \{\alpha'(t_n) - \hat{a}_0 - \hat{a}_1 e^{-\hat{c}_1 t_n} - \hat{a}_2 e^{\hat{c}_2 t_n}\}^2$$

where

t_n = age of child i at the birth of child n.

By choosing k = .1 and assuming the average spacing between successive births to be 2 years, the mean traces in Figure 1 were computed for each sibling in a 9 child family. This graph displays the typical shape of the solution (2.5).

We have thus far only found the mean square solution of the problem, which is identical to the deterministic solution. In an actual problem where more data was available, this would not be of interest. The next step is to study the behavior of the higher order moments of the solution, especially that of the second moment. It is of vital interest to discover the effects of these moments. If, for example, the variance is small relative to the mean, it is not so important to insist upon the stochastic model. The deterministic version could be used. On the contrary, if the variance is large, then the use of the deterministic model could result in a large discrepancy between the predicted behavior and an experimental realization. This might even lead one to discard the basic model as inappropriate.

It is also possible to simulate various outcomes by specifying particular distributions for each of the coefficients, thereby investigating the model under a wide range of conditions. Such broad studies could be of great help in applying this or a similar model to social system variables other than intelligence. In summary, the major objective of this paper has been to discuss, in general terms, how an existing discrete social or psychological model could be treated as a differential equation. This allows the use of well-developed techniques for handling both deterministic and stochastic equations. We have specifically shown how the Zajonc and Markus model might be naturally stochastized. In addition, we proposed a formulation for the process variable $\alpha(t)$ which attempts to reflect the behavior that they postulated for it.

References

1. Belmont, L., and Marolla, F.A., Birth Order, Family Size, and Intelligence , Science, 1973, 182, 1096-1101.

2. Breland, H.M., Birth Order, Family Size, and Intelligence., Science, 1974, 184, 114.

3. Hamblin, R.L., Jacobsen, R.B., and Miller, J.L.L., A Mathematical Theory of Social Change, New York: Wiley, 1973.

4. Schooler, C., Birth Order Effects: Not Here, Not Now!, Psychological Bulletin, 1972, 78, 161-175.

5. Soong, T.T., Random Differential Equations in Science and Engineering, New York: Academic Press, 1973.

6. Tsokos, C.P., and Padgett, W.J., Random Integral Equations with Applications to Life Sciences and Engineering., New York: Academic Press, 1974.

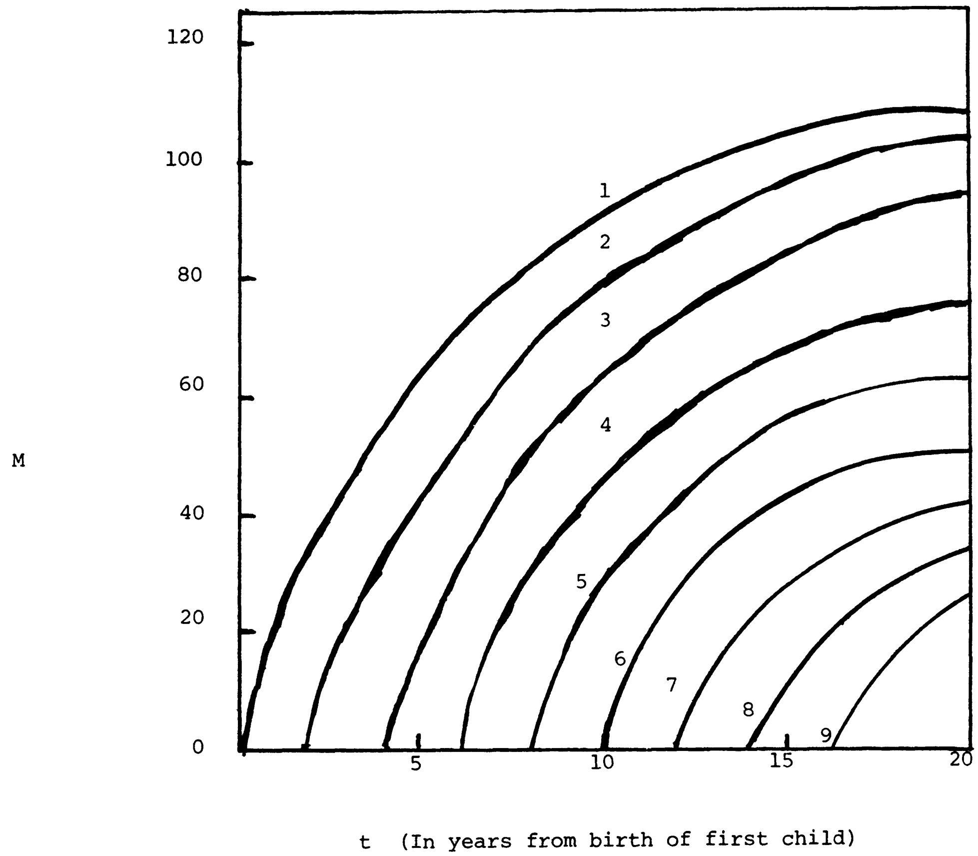
7. Welch, R.L.W., Denker, M.W., and Tsokos, C.P., A Statistical Model for Intellectual Development., American Statistical Association Preceedings of the Social Statistics Section 1975, 1976, 700-704.

8. Zajonc, R.B., Family Configuration and Intelligence., Science, 1976, 192, 227-236.

9. Zajonc, R.B., and Markus, G., Birth Order and Intellectual Development., Psychological Review, 1975, 82, 74-88.

Figure 1: Mean Intellectual Development

For 9 Child Family (Exaggerated)



Δt = Gap between successive births

= 2 years

k = .1

TRENDS IN PUBLIC ATTITUDES ON THE LEGAL PROFESSION

Kenneth R. White, Florida Technological University
R. Thomas Stone, Jr., University of Illinois, Urbana

On at least three occasions prior to the so-called "Watergate era" lawyers have conducted public opinion surveys in an effort to learn what the public thinks of them. In the states of Missouri in 1960, and North Dakota and Texas in 1970 such surveys have been conducted with strikingly similar results.[1] The Florida survey that is the basis for this article was conducted in the fall of 1975.

In the Florida study, as was the case in each of the three earlier surveys, non-lawyers were quizzed about their attitudes concerning lawyers and lawyers were also asked what they thought the public attitude was toward the legal profession.

The Florida survey consisted of a statewide representative response of 468 out of 2200 laymen selected at random and a statewide representative response of 278 out of 650 lawyers selected at random. Both the laymen and lawyers were asked substantially the same questions.

Public Use of Legal Services

Of the population surveyed 71.8% of the laymen had utilized the services of a Florida attorney within the past five years while 28.2% indicated that they had not used an attorney during the past five years.

The following results were tabulated as to frequency of contact with an attorney during the past five years:

Use of an attorney no times in the last five years	= 28.2%
Use of an attorney one time in the last five years	= 25.8%
Use of an attorney 2 to 5 times in the last five years	= 36.9%
Use of an attorney over 5 times in the last five years	= 9.1%

Satisfaction with Legal Services and General Attitude Toward Lawyers

The public was asked, "Were you satisfied with the service you received the last time a Florida lawyer dealt with a legal problem for you?" The responses were rated on a scale of "completely satisfied" (5) to "not at all satisfied" (1). The results are as follows:

Completely satisfied (5 on the scale)	= 32.2%
4 on the scale	= 16.1%
Average (3 on the scale)	= 29.1%
2 on the scale	= 7.5%
Not at all satisfied (1 on the scale)	= 15.1%

When Florida lawyers were asked "Do you believe that your clients are satisfied with your legal services?", the following responses were given:

Completely satisfied (5 on the scale)	= 44.0%
4 on the scale	= 46.0%
Average (3 on the scale)	= 7.6%
2 on the scale	= 0.0%
Not at all satisfied (1 on the scale)	= 2.3%

From a comparison of the responses of laymen and lawyers, it appears that Florida clients are not as satisfied with the performance of their attorneys as the lawyers think they are. The lawyers have indicated that they believe that 90% of their clients would place their services in the two categories above the position of average whereas the clients placed only 48.3% of the services on the higher end of the rating scale. Likewise, the attorneys only placed 2.3% of their work in the lawer two categories of the scale whereas the clients placed 22.6% of the services in the "less than average" service categories.

In response to a request that laymen check each of the favorable characteristics that they attribute to lawyers, the following results were obtained:

48.2% said "they are helpful"
37.1% said "they are qualified"
31.4% said "they are friendly" (personable)
30.6% said "they are good for the community"
38.5% said "they are intelligent"
14.4% said "they are honest"
15.3% said "they are hard working"

By comparison, the attorneys checked the characteristics they think the public attributes to lawyers as follows:

71.4% said "they are helpful"
84.8% said "they are qualified"
26.2% said "they are friendly" (personable)
36.2% said "they are good for the community"
76.2% said "they are intelligent"
16.7% said "they are honest"
28.5% said "they are hard working"

It appears that the lawyers think that the public sees them as highly qualified, intelligent, and helpful. However, the laymen do not concur as to the magnitude of those observations. Both laymen and attorneys generally agree on the questions of honesty, personality, effort, and community service. Unfortunately, both groups are concurring as to low marks for attorneys in those categories.

When asked to check the unfavorable characteristics of lawyers, the public responded as follows:

42.3% said "their fees are too high"
28.1% said "they are too slow"
23.1% said "they do not keep clients informed"
11.7% said "they are dishonest"
8.4% said "they have no ethics"
4.3% said "they are lazy"
11.2% said "they waste the time of clients"

The attorneys surveyed checked the following points as being criticisms held by the public:

38.7% said "their fees are too high"
18.6% said "they are too slow"
14.2% said "they do not keep their clients informed"
12.4% said "they are dishonest"
3.7% said "they are lazy"
5.6% said "they waste the time of clients"

In ranking lawyers with other community members as to honesty, the public responded with an order of ranking as follows:

1. Ministers
2. Medical Doctors
3. Teachers
4. Bankers
5. Businessmen

6. Lawyers

Lawyers saw the laymen as ranking the groups as to honesty as follows:

1. Ministers
2. Teachers
3. Medical Doctors
4. Lawyers
5. Bankers
6. Businessmen

In ranking lawyers with other community members as to prestige and leadership in the community, the public responded with an order of ranking as follows:

1. Medical Doctors
2. Bankers
3. Businessmen
4. Lawyers
5. Ministers
6. Teachers

Lawyers saw the laymen as ranking the groups as to prestige and leadership in the community as follows:

1. Medical Doctors
2. Lawyers
3. Bankers
4. Ministers
5. Businessmen
6. Teachers

In ranking lawyers with other community members as to interest in helping people, the public responded with an order of ranking as follows:

1. Ministers
2. Teachers
3. Medical Doctors
4. Bankers
5. Businessmen
6. Lawyers

Lawyers saw the laymen as ranking the groups as to interest in helping people as follows:

1. Ministers
2. Teachers
3. Lawyers
4. Medical Doctors
5. Businessmen
6. Bankers

In ranking lawyers with other community members as to interest in making money the public responded with an order of ranking as follows:

1. Businessmen
2. Lawyers
3. Medical Doctors
4. Bankers
5. Teachers
6. Ministers

Lawyers saw the laymen as ranking the groups as to interest in making money as follows:

1. Medical Doctors
2. Bankers
3. Businessmen
4. Lawyers
5. Teachers
6. Ministers

In each of the preceding comparison categories the public held the lawyers in lower esteem than the lawyers perceived the public view as being. Furthermore, the more specific questions of high fees, slowness, informing of clients, honesty, effort and respect for the time of clients all showed the general public giving the attorneys more criticism than the attorneys perceived the public view as being.

How a Client Chooses a Lawyer

Members of the general public were asked how they would select an attorney if they had never hired one. The responded as follows:

64.2% said "they would inquire around and check the lawyer's character and reputation in the community"
18.4% said "they would go to a lawyer they know"
2.1% said "they would consult several lawyers and then select one"
14.8% said "they would consult the phone book"
.5% said "they would consult the lawyers' referral service"

Lawyers saw the general public as choosing an attorney as follows:

62.3% said "they would inquire around and check the lawyers character and reputation in the community"
21.7% said "they would go to a lawyer they know"
4.2% said "they would consult several lawyers and then select one"
8.6% said "they would consult the phone book"
3.2% siad "they would consult the lawyers' referral service"

Both attorneys and clients saw the reputation of a lawyer in the community as being tantamount for lawyer selection. On the other hand, "shopping around" for an attorney and scrutiny of telephone and referral service aid was not labeled as significant intermediaries by attorneys or clients.

Service and Fees

In response to the request of "check all of the legal services for which you required help," the general public responded as follows:

38.2% sought aid for land and/or title work
39.4% sought aid for estate services
24.7% sought aid for family law problems
11.2% sought aid for the recovery of monies
4.8% received service from insurance counsel
4.7% sought income tax counsel
2.4% sought aid in criminal cases
12.6% sought aid in "other" categories

Attorneys were asked to rank in decreasing order the three areas that comprised most of their practice. They responded as follows:

	% Ranked 1st	% Ranked 2nd	% Ranked 3rd
Land &/or title work	26.8	23.6	16.1
Estate services	14.6	20.5	18.2
Divorce services	6.3	12.2	19.7
Suits to recover monies	2.3	8.7	12.6
Insurance counsel	3.9	2.9	4.2
Income tax counsel	7.3	6.4	4.6
Criminal cases	12.2	10.9	14.3
Others	26.6	14.8	10.3

Laymen were asked to rank the propriety of fees they had been charged by Florida lawyers from "very high" to "very low." The results are as follows:

Very high (5 on the scale)	= 38.6%
4 on the scale	= 26.9%
Reasonable (3 on the scale)	= 28.1%
2 on the scale	= 6.4%
Very low (1 on the scale)	= 0%

Attorney perceived the public concept of fees as follows:

Very high (5 on the scale)	= 15.2%

4 on the scale	= 72.5%
Reasonable (3 on the scale)	= 12.3%
2 on the scale	= 0%
Very low (1 on the scale)	= 0%

Publicity and Specialization

The general public and lawyers were asked if the news media had treated lawyers fairly with the following results:

	Public	Lawyers
Treated fairly	72.4%	66.7%
Treated unfairly	27.6%	33.3%

The question of the impact of the so-called "Watergate Affair" on the public attitude toward the honesty and integrity of the legal profession was presented to both laymen and attorneys with the following results:

	Public	Lawyers
A strong impact (5 on the scale)	38.6%	32.5%
4 on the scale	24.2%	17.2%
Mild impact (3 on the scale)	21.3%	41.5%
2 on the scale	11.6%	6.4%
No impact (1 on the scale)	4.3%	2.4%

To the question of whether attorneys should be allowed to advertise their services the following views were expressed by the general public and lawyers:

	Public	Lawyers
Yes	72.4%	24.3%
No	27.6%	75.7%

With specialization of practice becoming a reality, laymen and attorneys were asked whether the general public would prefer specialization to general practice. The following responses were presented:

	Public	Lawyers
Yes	66.4%	60.5%
No	33.6%	39.5%

Conclusions

The most constant theme of the surveys is that lawyers have a higher opinion of what the public thinks of them than the public responses indicate as the view of the layman. Fees are generally considered too high and less than half of the laymen surveyed rated the legal services rendered by their attorneys as being "above average".

If these views could have been labeled as danger signals after the surveys of 1960 and 1970,[2] the harsher criticism of 1975 calls for the organized Bar and individual attorneys to assume a more active role in making the nature, necessity and cost of legal services clear to the public.

Perhaps the skeptical view of the public has been temporarily intensified by the impact of the "Watergate Syndrome." However, with or without that impact the Bar has a duty to educate the public, make its services readily available, and provide efficient and quality service. Such a mandate is essential for all attorneys who wish to maintain and strengthen the American system of jurispudence.

Footnotes

[1]For a summary of the Missouri, North Dakota and Texas surveys, see John Thomason, "What the Public Thinks of Lawyers," 46 New York State Bar Journal 151-157 (April, 1974).

[2]Id., Generally, the Florida surveys indicated somewhat greater criticism of lawyers and their services than did the 1960 Missouri survey and the 1970 North Dakota and Texas surveys.

THE USE OF A MULTIPLE ENTRY DIARY IN A PANEL STUDY ON HEALTH CARE EXPENDITURE

Robert A. Wright, Richard H. Beisel, Julia D. Oliver, Michele C. Gerzowski
National Center for Health Statistics

A major problem in the production of reliable national statistics on the health status of the country is that most of the data gathering techniques rely on recall periods of up to a year. Using a recall period of this length may cause substantial reporting errors to occur.

Review of Related Literature

The use of diaries to obtain health care information has been shown to be an effective way to help reduce this problem in collecting health care information.

Prior to 1970, the diary as a means of securing data had been investigated primarily in the fields of consumer expenditures and food consumption.[1,2] Both journal and ledger diaries have been used in consumer expenditure studies. In journal diaries, the entries are made in time sequence, while in ledger diaries they are made by categories of events such as visits to the doctor, absences from work and purchases of medicines. For consumer expenditure studies, the work of Sudman and Ferber (1971)[3], has shown that ledger diaries obtain more complete information and a higher level of reporting. However, in health expenditure studies a form of diary, "an attractive household calendar," was used in the Canadian Sickness Survey[4] as early as 1952 and a diary was used in the family health survey conducted by the University of California School of Public Health in 1954.[5] In these instances the diary served as a reminder for reporting at the interview. In 1954, Allen _et al_. conducted a study on the characteristics and quality of data obtained by diary keeping. It was found that those maintaining diaries reported higher rates of illness for a calendar month than did those who were interviewed for whatever form of rate was used as an expression of morbidity levels. But rates for the groups maintaining diaries were close to rates for interviewed groups when medically attended illness was considered. Differences between rates for disabling illness were less than differences between rates for nondisabling illness. The difference between diary rates and interview rates was about the same for males, females, preschool children, school children and adults under 65. For older persons, a group of special interest in health surveys, diary rates were much closer to interview rates than for other groups.

Some of the most systematic work on the use of diaries for medical care has been by Sudman _et al_., at the Survey research Laboratory of the University of Illinois. They conducted an experimental study from October 1973 through March 1974 to investigate the use of diaries to obtain health care information. The study focused on two methodological issues: (1) what should be the format (i.e., ledger or journal) for a medical diary?, and (2) should households be compensated for keeping records?

Respondents were chosen from the membership in two prepaid health maintenance organizations (HMO's). An attempt was made to validate all medical events which resulted in a visit to an HMO physician or to a clinic or hospital for outpatient or overnight care. The results of this experiment confirm the results of early studies which found that households that use a diary report a higher number of medical events than do households who recall information on a personal interview. There seemed to be no meaningful differences between ledger and journal diaries.

Medical Economic Research Study Methodology

A multi-entry format was designed for the diary used in the Medical Economic Research Study. This study was designed to test various methods of collecting the medical care expenditures for families and individuals. Because average family out-of-pocket expenditures for medical care does not adequately serve the needs of policy makers for data of this type the research study had two components. The first component consisted of a longitudinal survey of a panel of households. The second component consisted of a survey of all providers and third party payors mentioned by the persons in the household survey.

The research study tested several methods of household data collection, i.e., monthly versus bimonthly interviewing, telephone versus personal interviewing and receiving permission to do the provider survey early in the survey period versus receiving such permission at the last interview. The research study was conducted in a rural area and an urban area and included areas where a high proportion of persons belonged to an HMO.

To assist household respondents in keeping track of health care information during the survey period a diary was designed and distributed to all participating households. The diary incorporated together features used separately by other longitudinal surveys: a calendar portion for recording health care events chronologically; a ledger portion where details of health care events could be recorded in different sections, one section for each type of event; and a pocket for household members to collect bills, receipts, and other information which would aid the household respondent in recalling events. A diary using three distinct methods of entry was used for a number of reasons. First, respondents vary in their ability to understand and use something as complex as a ledger. For reasonably educated respondents with a strong motivation, the ledger provides a place on which to record almost all the information needed about a particular event. For persons who use a calendar, even if they don't, won't or can't use a ledger, the calendar provides a place for recording some details. At a minimum the fact an event took place can be indicated and the date is thus known. As bills come in, the pocket provides a convenient receptable for all respondents to reduce searching for them at interview time.

At the final interview at the household, the diary was picked up by the interviewer. The household was not informed at the initiation of the study that the diary would be collected. Thus, the fact that the diary was to be collected caused no pressure on the respondents to utilize the diary.

The coding of the diary was done by the staff of the Utilization and Expenditure Statistics Branch in the Division of Health Interview Statistics.

Findings and Analysis

The analysis is primarily concerned with how diary use is related to certain demographic variables: urban versus rural residence; age of head of household; education of head of household; and income.

In addition, two variables concerning the actual interviewing technique were used: personal versus telephone interviews; and monthly versus bimonthly interviews.

Before beginning the analysis, those households who reported no medical events during the survey were dropped from consideration. They were not included, because they could not possibly use the diary if they had nothing to record. There were 32 such households. Their elimination dropped the population size to 650.

At the beginning of the study, neither the interviewers nor the respondents were told that the diaries would be picked up at the end. The respondents were told, in fact, that the diaries were to be used as a memory aid for subsequent interviews and were theirs to keep. When the interviewers returned to collect the diaries they picked up 66 percent of them. The remaining 34 percent were either lost, thrown away, or not collected for some other reason (see Table 1).

TABLE 1. Diary Status at the End of the Study

Returned diary		428 (65.8%)
1. used diary	311 (47.8%)	
2. did not use diary	107 (16.5%)	
Did not return diary		222 (34.2%)
1. lost diary	126 (19.4%)	
2. threw diary away	36 (5.5%)	
3. all other reasons for not returning diary	60 (9.2%)	
TOTAL		650 (100.0%)

For the 222 households in Table 1 who did not return the diary, there is no way of knowing if the diaries were used or not before they were lost or thrown away. We, therefore, do not know the overall proportion of diary usage in the population or the demographic characteristics of all the users. There is only information on those 428 households who returned the diary. We also do not know if the group who returned the diary was somehow different from the group who did not.

The demographic characteristics of the two groups are known, however. Table 2 shows these characteristics for the two groups. Table 3 shows the demographic characteristics of the households who used the diary and of those who did not, within the population of those who returned the diary.

TABLE 2. Selected Demographic Characteristics of Those Who Returned the Diary and Those Who Did Not

	Diary returned	Diary not returned
Percent of total population	66%	34%
Number	428	222
Median age of head of household	46 (yrs.)	43 (yrs.)
Median income	$13,900	$12,900
Median education of head of household	12 (yrs.)	12 (yrs.)
Percent urban	48%	64%
Percent rural	52%	36%
Percent personal interviews	48%	47%
Percent telephone interviews	52%	53%
Percent monthly interviews	45%	53%
Percent bimonthly interviews	55%	47%
Percent diary use	70%	???

TABLE 3. Selected Demographic Characteristics of Those Who Used the Diary and Those Who Did Not Use the Diary for All Those Who Returned the Diary

	Diary used	Diary not used
Number	311	117
Percent of population who returned the diary	73%	27%
Median age of heat of household	44 (yrs.)	46 (yrs.)
Median income	$15,100	$10,800
Median education of head of household	12 (yrs.)	12 (yrs.)
Percent urban	41%	62%
Percent rural	59%	38%
Percent personal interviews	48%	50%
Percent telephone interviews	52%	50%
Percent monthly interviews	47%	41%
Percent bimonthly interviews	53%	59%

The 428 households who returned their diaries were the ones used in the analysis. Usage versus no usage was used as a dependent variable in a stepwise regression analysis. The independent variables used were the demographic and interview variables mentioned earlier. For the basic question of diary usage, three regressions were used: diary use versus no diary use; calendar use versus no calendar use; and ledger use versus no ledger use.

In all 3 regressions, income came out as the most significant independent variable with rural residence second (see Table 4). What was surprising was that education was not included as a significant variable for either diary or ledger usage. It was only included for calendar usage, but was number 6 our of 6.

TABLE 4. Independent Variables Kept in Regression Equation by Order

Diary Use	Calendar Use	Ledger Use
1. Income	Income	Income
2. Rural residence	Rural residence	Rural residence
3. Telephone interviewing	Age	Telephone interviewing
4. Age	Monthly interviewing	Age
5. Monthly interviewing	Telephone interviewing	Monthly interviewing
6. —	Education	—

The second phase of analysis dealt with how well the diary was used, what items were recorded most and least often, and what, if any, time trends were observed. Certain items, such as hospital stays, listed on the Summary were matched with those same items recorded on either the calendar or ledger. Correlation coefficients were then determined for certain events as they were recorded on the Summary and on either the calendar or ledger. A problem to note about this approach is that those items listed in the Summary were those which the respondent reported to the interviewer in the normal course of the interview. They were not necessarily all the events which acutally occurred in the household. What the correlations actually represent, then, are comparisons between those items given orally to the interviewer, which were then transcribed to the Summary, and those items which the respondent himself wrote in the calendar or ledger. The correlations then represent the correlations between two methods of recall.

Table 5 shows the correlations for types of events as recorded in the Summary and in the calendar. Similarly, Table 6 shows correlations for types of events as recorded in the Summary and in the ledger. The calendar protion of the diary had only a blank square surrounding the date in which to record an event. The ledger was more formalized. There were specific spaces for specific items; such as doctor and dental visits, prescriptions, and hospitalizations; and check boxes for those visits which included x-rays, lab tests, and visits to the emergency room. There was also a space for "other medical expenses" which included non-medicinal items prescribed or recommended by a doctor, i.e., corrective shoes, crutches, syringes for diabetics. The ledger also provided special space to record not only the encounter, but the date, provider's name, prescription numbers and conditions.

TABLE 5. Ranking of Correlations Between Visits Recorded in the Summary and Visits Recorded in the Calendar

Type of Visit	Correlation
Dental, Feb. - March	.758
Doctor, Feb. - March	.673
Hospital Stays, Feb. - July	.663
Dental, April - May	.654
Dental, June - July	.627
Doctor, April - May	.580
Doctor, June - July	.517
Medicine, April - May	.495
Medicine, Feb. - March	.449
Emergency Room Visits, Feb. - July	.401
X-rays, Feb. - July	.399
Medicine, June - July	.370
Lab Tests, Feb. - July	.340
Other Medical Expenses, Feb. - July	-.461

TABLE 6. Ranking of Correlations Between Visits Recorded in the Summary and Visits Recorded in the Journal

Type of Visit	Correlation
Medicine, April - May	.868
Emergency Room Visits, Feb. - July	.815
Medicine, Feb. - March	.705
Hospital Stays, Feb. - July	.659
Doctor, Feb. - March	.635
Dental, Feb. - March	.632
Dental, April - May	.542
Medicine, June - July	.497
Doctor, April - May	.473
Dental, June - July	.448
X-Rays, Feb. - July	.420
Doctor, June - July	.394
Lab Tests, Feb. - July	.393
Other Medical Expenses, Feb. - July	-.548

TABLE 7. Correlations Between Visits Recorded on the Summary and Visits Recorded in the Diary, by Type of Diary Use

Type of Visit	(1) Correlation between visits recorded in the Summary and visits recorded in the calendar	(2) Correlation between visits recorded in the Summary and visits recorded in the journal	Col. (1) minus (-) Col. (2)
Dental, Feb. - Mar.	.758	.632	+
Dental, April - May	.654	.542	+
Dental, June - July	.627	.448	+
Doctor Feb. - Mar.	.673	.635	+
Doctor, April - May	.580	.473	+
Doctor, June - July	.517	.394	+
Medicine, Feb. - Mar.	.449	.705	-
Medicine, April - May	.495	.868	-
Medicine, June - July	.370	.497	-
Hospital Stays	.663	.659	+
Emergency Room Visits	.401	.815	-
X-Rays	.399	.420	-
Lab Tests	.340	.393	-
Other Med. Exp.	-.461	-.548	+

Table 7 compares correlations for specific encounters between those visits recorded on the Summary with those recorded in the calendar and those visits recorded on the Summary with those recorded in the ledger. The results were mixed with some events, emergency room visits, prescribed medicine, having higher correlation with encounters on the ledger, some other events, dental and doctor visits having higher correlation with encounters on the calendar.

Another aspect of the analysis was to see if there was any trend as the study progressed for changing habits of recording items in either the calendar or ledger. This could be determined for doctor visits, dental visits and prescribed medicines as they were coded by two-month time periods; February - March, April - May, June - July. Tables 8 and 9 show the results.

TABLE 8. Correlations Between Visits Recorded on the Summary and Visits Recorded on the Calendar by Time Period

	Feb.-Mar.	Apr.-May	June-July
Dental Visits	.758	.654	.627
Doctor Visits	.673	.580	.517
Prescribed Medicines	.449	.495	.370

TABLE 9. Correlations Between Visits Recorded on the Summary and Visits Recorded on the Ledger by Time Period

	Feb.-Mar.	Apr.-May	June-July
Dental Visits	.632	.542	.448
Doctor Visits	.635	.473	.394
Prescribed Medicines	.705	.868	.497

Both tables show a decrease in correlation by time period, indicating less usage of the diary as the study progressed with increases in April-May for prescribed medicines.

Conclusions

In retrospect, this data set could have furnished more information on diary usage if we had known more about those households who did not return the diary. If they had been asked, "Before you lost (threw away) your diary, did you use it at all? If so, what sections?", there would have been a better grasp of the differences between diary users and non-users. Also, looking at other demographic variables, particularly race, would have added interest. However, we did find out how households used the diary as a memory aid, what items they used if for and how they used it over time.

References

1 Brady, Dorothy S. and Williams, Fath M. "Advances in the Techniques of Measuring and Estimating Consumer Expenditures." Journal of Farm Economics, 27, No. 2 (May 1945), 315-344.

2 Murray, J.E.C., Blake, D. Dickens, and Moser, A.M. "Collection Methods in Dietary Surveys." Southern Cooperative Series, Bulletin 23. Clemson, S.C.: South Carolina Agricultural Experiment Station, April 1952.

3 Sudman, Seymour and Feber, Robert. "Experiments in Obtaining Consumer Expenditures by Diary Methods." Journal of the American Statistical Association, 66, No. 336 (December 1971), 725-736.

4 Peart, A.F.W. "Canada's Sickness Survey: Review of Methods." Canadian Journal of Public Health, 43, No. 10 (October 1952), 401-414.

5 Allen, George I., et al. "Interviewing Versus Diary Keeping in Eliciting Information in a Morbidity Survey." American Journal of Public Health, July 1954, pp. 919-927.