I

CONTRIBUTED PAPERS - I

Chairman, Om Aggarwal, Iowa State University

Page

The Effect of Mismatching on the Measurement of Response Errors - John Neter, University of Minnesota, E. Scott Maynes, U. S. Bureau of the Census and University of Minnesota and R. Ramanathan, University of Minnesota.......... 2

Sampling Variation of Age-Adjusted Rates - Regina Loewenstein and Jack Elinson, Columbia University.................................... 9

Multiple Frame Sample Surveys - Robert S. Cochran, University of Wyoming.......................... 16

The Alphabetic Array Sampling Technique - Abraham Frankel and Patricia Wright, U. S. Office of Education.................................... 20

Methodological Procedures in a Very Large National Survey of Physicians - T. Donald Rucker, National Association of Blue Shield Plans and Martin Taitel, Chicago, Illinois............... 25

THE EFFECT OF MISMATCHING ON THE MEASUREMENT OF RESPONSE ERRORS*

John Neter, University of Minnesota,
E. Scott Maynes, U. S. Bureau of the Census and University of Minnesota
and R. Ramanathan, University of Minnesota

1. INTRODUCTION

A major approach to the measurement of response errors is by record checks, or validation studies, in which survey responses are compared on a case-by-case basis with more-or-less accurate records. An important, though often unrecognized, obstacle to the usefulness and correct interpretation of record checks is the existence of matching errors. Matching erors arise when responses pertaining to one person (or family, organization, etc.) are incorrectly associated with and compared with record data pertaining to a different person. The impact of such matching errors on the measurement of response errors constitutes the core of this article.

An example will clarify the meaning of both record check and matching error. In the financial area one might undertake a record check study to measure the accuracy with which bank account balances are reported by their owners in sample surveys. The design of such a study -- and the following description conforms roughly to a study actually being carried out by the Bureau of the Census -- might be quite straightforward: (1) Do a probability sample of bank accounts from bank records; (2) Interview owners of sample accounts, asking them to provide complete information regarding each of the bank accounts they own; (3) Compare information obtained in response to survey questions on an account-by-account basis with information in bank records.

For this design, matching errors (also called "mismatches") may arise in several ways: (1) the wrong person (not the owner of the sample account) is interviewed; (2) a bank clerk records the wrong bank balance (the balance of Account #53402 rather than of Account #53401); (3) analysts mistakenly match the owner's report about his Account A in the sample bank with bank records pertaining to his Account B in the same bank. This list is by no means exhaustive.

In this paper the implications of matching errors are spelled out for two simplified models. The first model assumes that each account in the sample has the same probability of being matched correctly. In addition, if a mismatch occurs, the model assumes that each sample account has an equal probability of being mismatched with any other account in the parent population, regardless of the Account Number, the number of accounts possessed by this owner in the sample bank, the size of balance, the "uncommonness" of the owner's name (e.g., Smith), or other factors realistically related to the probability of mismatching.

The second model retains the assumption that mismatches occur according to a random mechanism, but permits mismatches to occur only within subsets of the parent population, and permits the probability of mismatching to differ among subsets. Because of these two features, the second model is more realistic and flexible. For example, if accounts owned by multiple-account holders (in the sample bank) are more likely to be mismatched, then "number of accounts owned in the sample bank" might be one of the criteria for defining population subsets in the second model. Subsets could be similarly defined to fit many other plausible hypotheses regarding sources of mismatching.

In actual record check studies, it is frequently found that it is impossible to match a sample account with any account in the parent population, thus giving rise to "nonmatches." Neither of our models deals with nonmatches. We have elected here to sacrifice realism in favor of simplicity.

2. RECORD CHECKS: ENCOUNTERS WITH MISMATCHING

An Example -- Consider Table 1, taken from a carefully conducted record check study by Horn [4]. For a sample of purportedly identical savings accounts, the table presents mean balances, as shown by bank records (Col. 2) and as reported by respondent-owners (Col. 3). Assuming perfect execution, accurate bank records, and that observed differences in means are statistically significant, the conclusion emerges inescapably from Col. 4 that respondents with large balances tended to underreport and respondents with small balances tended to overreport -- in short, a regression-toward-the-mean effect.

But supposing that mismatches occurred -- so that in some cases the respondent report and the bank report used to evaluate the accuracy of that particular respondent report do not refer to the same account. What would be the consequence? The answer is that mismatching could yield the same regression-toward-the-mean effect even in the case of zero response errors. Whether mismatching can explain the particular results of Table 1 or whether -- by elimination -- these results must be attributed to response errors will be discussed later.

Before we turn to the formal analysis, however, we must deal with three important questions: (1) What factors give rise to mismatching? (2) What has been the frequency of mismatches in various matching studies? (3) How have the consequences of mismatching been dealt with analytically?

Sources of Mismatching -- Mismatching may occur through either (1) inadequacy of items available for matching, or (2) errors in the

---

*The authors happily acknowledge helpful comments by I. Richard Savage, Gad Nathan, and Robert Ferber.

TABLE 1

ACTUAL VS. REPORTED SAVINGS ACCOUNT BALANCES, OCTOBER, 1958
NETHERLANDS VALIDATION STUDY
(GUILDERS)

(1)	(2)	(3)	(4)
Groups of 100 Observations Ranked in Descending Order by Actual Balances	Mean Actual Balance	Mean Reported Balance	Difference in Means: Reported Less Actual Balance
1	6110	5180	-930
2	4310	3820	-490
3	3530	3220	-310
4	3080	2610	-470
5	2730	2510	-220
6	2470	2280	-190
7	2130	2080	- 50
8	1790	1670	-120
9	1490	1410	- 80
10	1220	1170	- 50
11	1010	1160	+150
12	740	810	+ 70
13	480	550	+ 70
14	320	450	+130
15	180	270	+ 90
16	90	140	+ 50
17	10	170	+160

execution of the matching operation. We discuss each in turn.

Ideally, an item (or set of items) suitable for matching should define the thing being matched (a bank account, person, organization, etc.) uniquely, be available in both sources of data, and be measured or recorded accurately.

There are several ways in which record checks can be designed so as to improve the probability of correct matches. The first way is to maximize the number and detail of items being used in matching. For instance, the number of "Robert Johnson's" in Minneapolis (telephone book count) is 224 (out of 334,000). Reduce the size of this subset by obtaining information about middle initials and you get 29 "Robert W. Johnson's." Finally, obtain (say) the name of wife and address, and the identification of a "Robert W. Johnson" in Minneapolis approaches uniqueness.

A second device to achieve uniqueness in matching -- and the best if it is feasible -- is by specifying items for matching which are in fact unique and provide a one-to-one mapping from one list to another, e.g., a bank account number in a particular bank or social security number (excepting the case where an individual maintains "aliases").

A third means of seeking uniqueness is by minimizing the size of lists in which persons are identified. In the bank account record check mentioned earlier, it would be better, _ceteris paribus_, to draw a sample from a small rural bank (with, say, 15,000 accounts owned by people in small towns or rural areas) than to draw a sample from a New York City bank (with, say, 1-2 million accounts owned by people living mainly in a large metropolitan area).

Finally, it is desirable to locate the record check in a place (or list) which is as heterogeneous as possible with respect to the items used in matching. For example, it would be highly undesirable to match on surnames in Copenhagen with its abundance of Andersen's, Hansen's, etc.; by contrast, surnames may be more nearly unique if used, say, for UN personnel in New York City.

Practical considerations have prevented many record check studies from employing optimal matching items. In some cases, a desire to protect the anonymity of respondents has forced some investigators to undertake matching without using the names of sample individuals [7]. Other studies have not asked respondents to supply their social security number or bank account numbers, either for fear of jeopardizing cooperation or because it was felt that the respondent could not or would not provide bank account numbers accurately.

As noted above, mismatching may also arise through errors in the execution of matching procedures. In general, mismatching from this source may be reduced by (1) minimizing the extent to which subjective judgments must be made, (2) replicating matches independently, and (3) utilizing consistency checks to detect errors due to carelessness.

Frequency of Mismatches -- Unfortunately, few data on frequency of mismatches are available. A number of studies have provided information on nonmatches, which may be considered a proxy variable for mismatches in the sense of indicating the difficulty of matching. A nonmatch occurs when, using items available for matching, there appears to be no case in the parent population whose description conforms to a particular sample case.

Past matching studies have varied considerably with respect to reported rates of mismatch or nonmatch. As Table 2 shows, reported mismatch or nonmatch rates vary from an inconsequential 0.4 percent in the Horn study, to a rather large 35.5 percent in the Sirken study.[1] It should be noted that the reported rate of mismatch or nonmatch may differ in either direction from the actual rate: characteristics of actually identical persons may be recorded erroneously in either of the two sets of records on which matches are based; alternatively, persons with apparently identical characteristics (e.g., the same name, same age, same sex, etc.) may in fact not be the same individuals.

Analytical Treatment of Mismatching -- In most matching studies, investigators have taken great pains to accomplish accurate matching. Due to differing underlying circumstances, their success in this has varied. What efforts were made analytically to take account of either detected or undetected mismatches? In some studies, particularly where the outcome of the matching procedures was obviously imprecise, analysts have designated various classes of matching, e.g., "positive matches," "probable matches," etc. When this procedure has been followed, it has been typical to confine most of the analysis to the "best" match class [11]. This has the possible undesirable effect of introducing bias, and also reduces sample size. On the other hand, it has the virtue of recognizing that mismatching may vitiate the statistical analysis unless corrective action is taken.

As far as residual, undetected mismatches go, this factor has not been explicitly dealt with in any of the studies with which we are familiar. Such mismatches are our primary concern.

3. MODEL 1: MATCHING ERRORS OCCURRING THROUGHOUT POPULATION

Nature of Model Studied -- We begin the study of the effects of matching errors on the measurement of response errors by considering a highly simplified model. As a vehicle for discussion, we shall use an example concerning the study of response errors in reporting of bank balances by household respondents. Suppose that the population consists of bank accounts A_1, A_2, ... A_N. The balance of the j-th account according to the bank records is denoted by Y_j (j = 1, 2, ... N). These balances according to bank records are taken as the "true" values. Thus, the true mean balance per account in the population is:

$$\bar{Y} = \frac{1}{N} \sum_{j=1}^{N} Y_j \qquad (1)$$

and the population variance of the account balances is:

$$\sigma_Y^2 = \frac{1}{N} \sum_{j=1}^{N} (Y_j - \bar{Y})^2 \qquad (2)$$

We suppose now that the respondent for account A_j will report a balance W_j which is not subject to random errors. In other words, the simple model investigated here does not involve random response errors. Thus, the "true" response error R_j for the j-th account is:

$$R_j = W_j - Y_j \qquad (3)$$

For reasons mentioned in the previous section, matching errors may occur in the record check study, so that R_j may not be observed directly. Thus, the reported balance for the j-th account may not be compared with the correct balance Y_j but with some other balance Y_j $(k \neq j)$. We therefore introduce a random variable Z_j for the j-th account, which is defined as follows:

$$Z_j = \begin{cases} Y_j \text{ with probability } p \\ Y_k \text{ with probability } q \ (k \neq j) \end{cases} \qquad (4)$$

Z_j represents the bank balance against which the reported balance W_j is compared. According to the simple model, the comparison is made against the correct balance with probability p, but may be made against any other bank balance in the population with probability q for any specific alternate account. It follows therefore that:

$$p + (N - 1)q = 1 \qquad (5)$$

This model has two important restrictive properties:

a. A match against some account must be made; thus, there is no provision for nonmatches in doubtful cases.

b. If any mismatching occurs, any other bank balance is equally likely to be the mismatched balance.

---

[1] We have neglected the 80.1 percent mismatch rate in the Phillips study (line 5) since the computer match was viewed as but one of two stages of matching.

TABLE 2

FREQUENCY OF NONMATCHES OR MISMATCHES: SELECTED MATCHING STUDIES

Reported Rate of Nonmatches or Mismatches	Sample Description	Variables Being Matched: First Source	with Second Source	Information Used in Matching	Reference
1. 0.4% mismatch[a]	3321 savings acct. owners in 3 metrop. areas in Netherlands	Savings accounts and owners as reported in survey interviews	Accounts and owners as shown in bank records	Name, address, age family composition	Horn [4,5]
2. Two stages:					
a. 1% nonmatch	1491 persons in NC and NE U. S. who had been hospitalized	Persons as identified by interviewers	Persons as shown by hospital records	Name, address, age, sex, race	U. S. NHS [12]
b. 3.6% nonmatch	"	Hospitalization episodes as reported in interview	Hosp. episodes as shown in hospital records	Subjective matching by two persons	
3. 21% nonmatch	206 workers in single plant	Persons, as identified from answers to pencil-and-paper tests	Persons, as shown by plant records	Age, sex, section of plant where works, shift, etc.	Kahn [6]
4. 35.5% nonmatch	National sample of 1500 families	Families and related individuals identified by one survey organization	Families and related individuals identified by a second survey organization	Head's sex, age, occupation, veteran status, family size, no. of children, no. born in 1949-50 (special weight given to unusual characteristics)	Sirken [11]
5. Two Classes of computer matches[b]	22,869 psychiatric case records, incl. in some cases more than 1 record per person	Case records of particular persons (1961)	Any other case records among the 22,869 pertaining to the same person	Sequential comparisons on soundex code, surname, first name, address, birth year range, soc. security no., maiden name, sex-race, birth month and day, birth year	Phillips [9]
a. 3.7% mismatch	627 "positive matches" as determined by computer	"	"		
b. 80.1% mismatch	1,011 "possible matches"	"	"		

---

[a] Actually, matching with respect to family composition and age was carried out subsequent to the initial matching on name and address. The family composition-age check disclosed 60 actual mismatches (1.8% of the sample) which had been incorrectly accepted as matches. Of these, 47 cases were deleted from the sample before analysis; the remaining 0.4% were detected after analysis.

[b] The object was to eliminate duplicate records (finally determined to be 805) from the 22,869 records. This was achieved in two stages, first by a computer check which identified "positive matches" and "possible matches" and second by a careful clerical check which produced the mismatch rates shown in the table. Some of the records were incomplete with respect to the items used for checking.

The second limitation is relaxed in the following section. It is a serious limitation since mismatching is probably more likely to occur within a small subset of the population accounts (for instance, within the accounts held by a family or by persons of the same name). We consider the case of possible mismatching throughout the population first because it is a simple case which provides considerable insights into the effects of matching errors, and because it serves as the foundation for the next model where matching errors are restricted within mutually exclusive subsets of the population.

The measured response error for the j-th account is denoted by M_j, defined as follows:

$$M_j = W_j - Z_j \tag{6}$$

where M_j is a random variable since Z_j is a random variable. It follows from (4) that:

$$M_j = \begin{cases} W_j - Y_j = R_j & \text{with probability } p \\ W_j - Y_k & \text{with probability } q \ (k \neq j) \end{cases}$$

Thus, M_j provides the "true" response error only with probability p.

To summarize our basic notation in one location, we have:

Y = true bank balance
W = reported bank balance
R = true response error
Z = matched bank balance
M = measured response error

When an account is selected from the population at random, we denote the random variable corresponding to the measured response error as $\underline{m}$, and similarly denote the random variables corresponding to the true balance and to the reported balance as $\underline{y}$ and $\underline{w}$ respectively.

A simple random sample of accounts with replacement is defined as one such that the $\underline{w}$'s are independent and the $\underline{z}$'s are independent. The condition that the $\underline{z}$'s are independent implies that the same account could be matched against several responses. If the survey matching procedures preclude duplicate matching, then the model may be appropriate only for larger populations where the probability of duplicate matching according to the model would be very small. On the other hand, if duplicate matching is possible - and this is the case in the matching studies with which we are familiar - the model permitting duplicate matching may be appropriate even for smaller populations.

<u>Results</u> -- We shall now state the major results, without giving any of the derivations:

1. With the model assumed, matching errors do not affect the study of mean response errors. It can be shown that:

$$E(m) = \bar{R} \tag{7}$$

where $\bar{R}$ is the mean of the "true" response errors for the population. Hence, if a simple random sample of accounts is selected with replacement and the response errors measured, the mean measured response error of the sample is an unbiased estimator of $\bar{R}$ even though matching errors are present.

2. It also follows for this model that:

$$\sigma_m^2 = \sigma_R^2 + 2Nq\,\sigma_{WY} \tag{8}$$

where σ_R^2 is the variance of the true response errors for the population and σ_{WY} is the covariance between the reported bank balances and the corresponding true bank balances in the population. Thus, the variance of the measured response errors is in general different from the variance of the true response errors. For instance, if σ_{WY} is positive, σ_m^2 would then exceed σ_R^2.

3. If a linear regression between the measured response error $\underline{m}$ and the matched bank balance $\underline{z}$ is calculated, then it can be shown that for this model, we have:

$$\beta_{mz} = \beta_{RY} - qN\beta_{WY} \tag{9}$$

and:

$$\alpha_{mz} = \alpha_{RY} + qN\beta_{WY}\bar{Y} \tag{10}$$

Thus, if the correlation between W and Y is positive:

$$\beta_{mz} < \beta_{RY}$$

and, assuming $\bar{Y}$ is also positive:

$$\alpha_{mz} > \alpha_{RY}$$

In other words, for the typical case where σ_{WY} is positive and $\bar{Y}$ is positive, the regression between the measured response error $\underline{m}$ and the matched bank balance $\underline{z}$ involves a smaller slope and larger intercept than the regression of the true response error R on the true bank balance Y.

4. MODEL 2: MATCHING ERRORS RESTRICTED TO SUBSETS OF POPULATION

<u>Nature of Model Studied</u> -- In many cases it may not be realistic to assume that matching errors can occur throughout the population. Rather, such errors may be limited to subsets of the population, such as persons in a household, persons at the same address with the same name, or persons with the same name and age. The subsets within which matching errors can occur depend on the specific matching techniques that are employed, and will vary from problem to problem.

The model considered in this section assumes that:

a. The population is divided into K mutually exclusive and exhaustive subsets.

b. Matching errors can occur only within a subset.

c. Within the i-th subset, containing N_i elements, the probability of a correct match for any element is p_i, and the probability that any other particular element in the subset is used for the match is q_i. Thus we have:

$$p_i + (N_i - 1)q_i = 1 \qquad (11)$$

given that an element from the i-th subset is selected.

It is thus clear that the conditions within any subset correspond to those utilized in Section 3. Consequently, the derivations of results for the model in this section are an extension of those obtained earlier.

The limitations of the model discussed in the previous section still apply, namely that a match must be made and that mismatches against other elements are equally likely (but here only within the subset). In addition, Model 2 requires the subsets within which mismatches may occur to be mutually exclusive. This latter restriction often may be met approximately, as when the probability of a mismatch against elements outside the subset is very small compared to the probability of a mismatch within the subset.

To illustrate the nature of these subsets, we shall consider a record check study of bank balance reports. Here, for instance, mismatches may occur only within the group of accounts for persons with the same surname living at the same address. If, however, the mismatching probabilities depend also on the bank balance, subsets meeting the requirements of the model discussed would have to be defined on three dimensions: surname, address, and size of bank balance.

Results -- Again, we simply present results without showing derivations:

1. As in the case of Model 1, the expectation of the measured response error $\underline{m}$ is $\bar{R}$.

2. Model 2 yields the same conclusions concerning the variance of $\underline{m}$ and the regression of $\underline{m}$ on $\underline{z}$ as Model 1, provided that the correlations between true and reported values are in the same direction in each subset.

5. APPLICATION OF THE MODELS

We shall now apply the earlier results to the data obtained from the Horn record check study [4] in order to examine the possibility that matching errors alone could account for the regression-toward-the-mean effect noted in Table 1. With model 1, the regression of measured response errors on the matched balance is, from (9) and (10):

$$E(m|z) = \alpha_{RY} + qN\beta_{WY}\ \bar{Y} + (\beta_{RY} - qN\beta_{WY})z$$

If there were no response errors, but only matching errors, $\alpha_{RY} = \beta_{RY} = 0$, $\beta_{WY} = 1$, and the regression equation would reduce to:

$$E(m|z) = qN\bar{Y} - qN\ z$$

Horn calculated for grouped data the unweighted regression of the measured response errors on the matched balances as:

$$\hat{m} = 202.6 - 0.178\ z$$

We can get estimates of q, assuming no response errors, from matching each of the two equation constants. Matching the slope terms, we have:

$$-qN = -\ 0.178$$

or:

$$p = 1 - \left(\frac{N-1}{N}\right) 0.178$$

Since N in this study was large, we obtain:

$$\hat{p} \cong 0.82$$

Thus, if no response errors were present in the Horn study, the probability of a correct match would have had to be in the vicinity of .8 in order to account for the observed regression-toward-the-mean effect. Is this a reasonable probability for a correct match for this study? We believe not. The conductors of the Netherlands Validation Study took a variety of steps to minimize the possibility of mismatches.

Their matching procedures were so thorough that it is our judgment that the probability of a correct match for this study would be about .95 or higher. Thus, it appears to us highly unlikely that the negative slope of measured response errors on matched balances found by Horn is due to matching errors only, but rather reflects the behavior of response errors.

REFERENCES

[1] Hansen, Morris H., Hurwitz, William N., and Madow, William G., Sample Survey Methods and Theory, Volume II. New York: John Wiley, 1953.

[2] Hauser, Philip M., and Kitagawa, Evelyn, "Social and Economic Mortality Differentials in the U. S., 1960: Outline of a Research Project," Proceedings of the Social Statistics Section, American Statistical Association, 1960, 116-21.

[3] Health Insurance Plan of Greater New York, Annual Statistical Report, 1962.

[4] Horn, W., "Reliability Survey, A Survey on the Reliability of Responses to an Interview Survey," Reprint of an article appearing in Het PTT-bedrijf, 10 (1960).

[5] Horn, W., "Non-Response in an Interview Survey," Reprint of an article appearing in Het PTT-bedrijf, 12 (1963).

[6] Kahn, Robert L., A Comparison of Two Methods of Collecting Data for Social Research: The Fixed Alternative Questionnaire and the Open-Ended Interview. Ann Arbor: University of Michigan, Ph.D. Dissertation, 1952.

[7] Lansing, John B., Ginsburg, Gerald P., and Braaten, Kaisa, An Investigation of Response Error, Studies in Consumer Savings, No. 2. Urbana, Illinois: Bureau of Economic and Business Research, 1961.

[8] New York Stock Exchange, Department of Research and Statistics, Methodology and Sample Design of 1962 Census of Shareowners. New York: New York Stock Exchange, 1962.

[9] Phillips, William Jr., and Bahn, Anita K., "Experience with Computer Matching of Names," paper presented at the September, 1963 Meetings of the American Statistical Association, Cleveland.

[10] Shapiro, Sam, and Densen, Paul M., "Research Needs for Record Matching," paper presented at the September, 1963 Meetings of the American Statistical Association, Cleveland.

[11] Sirken, Monroe G., Maynes, E. Scott, and Frechtling, John A., "The Survey of Consumer Finances and the Census Quality Check," in National Bureau of Economic Research, An Appraisal of the 1950 Census Income Data, Studies in Income and Wealth, Volume 23. Princeton: Princeton University Press, 1958, pp. 127-68.

[12] U. S. National Health Survey, Reporting of Hospitalization in the Health Interview, A Methodological Study of Several Factors Affecting the Reporting of Hospital Episodes. Washington: U. S. Department of Health, Education, and Welfare, Public Health Service, 1961, Publication No. 584-D4.

SAMPLING VARIATION OF AGE-ADJUSTED RATES 1/

Regina Loewenstein and Jack Elinson
Columbia University School of Public Health and Administrative Medicine

Introduction

Analysts of data from health surveys frequently encounter indices of medical care that vary widely by age. It is often desired to compare these indices among classes, e.g. income classes, which have different age compositions. Reports of health surveys describe many methods of analyzing this type of data, including rates for specific age groups within classes 2/, multiple regression or analysis of variance with age as one variable 3/, and comparison of age-adjusted rates 4/.

This paper will discuss methods of studying the sampling variation of age-adjusted means and proportions. The procedures include estimation of confidence intervals of age-adjusted means and tests of hypotheses about their homogeneity.

This research was done as part of a collaborative health interview survey covering 5344 persons in the Washington Heights Health District of New York City, conducted by the Columbia University School of Public Health and Administrative Medicine, 1960-1961. Questions about medical care in the year prior to the interview were included for the Patterns of Medical Care Study of the New York City Department of Health 5/.

The index of medical care used for illustrative purposes here is the reported number of physician visits per person per year. The distribution of this variable is not normal for any age group. For example, about half of all persons covered in these interviews had no reported visits, one-fourth had one through three visits, one-fifth had four through 14 visits, and the remaining 5 percent reported 15 through 100 visits.

Estimating Variances for Specific Age Groups

Variances of age-adjusted means are functions of variances of specific age groups which are estimated, of course, with consideration of the sample design.

In this survey, a self-weighting sample of housing units with families to be interviewed was selected by a two-stage stratified cluster sampling plan with varying first-stage sampling rates among strata, and with uniform first and second stage sampling rates within each stratum 6/. The ten strata were defined by geographical location, racial composition in 1957 and rent in 1950.

From this sample design an equal number of observations was expected from each cluster within the same stratum. When this sampling plan is used, the expression 7/ frequently suggested for the variance of a mean per sampling unit includes terms that are variances of cluster totals within each stratum. These variances of cluster totals would be meaningful only if there was an approximately equal number of observations from each cluster in the same stratum.

But, the number of interviews within clusters from the same stratum varied widely for several reasons. First, clusters within each stratum had been estimated to have approximately the same number of housing units, based on Block Statistics from the 1950 Census and from maps showing the number of floors, but not the number of apartments, in each building. Second, even if the number of housing units in clusters had been equal, response rates among clusters varied. Third, there was more than one family in almost 10 percent of the housing units in the sample.

In addition, the number of persons per family varied among clusters, and response rates were lower among smaller families 8/. Consequently, the number of persons covered in interviews from clusters within the same stratum varied; for example, from 6 to 58 in one stratum.

The number of persons of specific age groups covered in interviews from clusters in each stratum varied even more, because clusters had different age distributions and because there were higher refusal rates among older persons 8/. In one stratum, for example, 20 of 75 clusters had no persons 65 years and older, and four clusters had 6 to 16 persons in this age group. Thus, variances of age-specific means could not be determined by a method that assumed an approximately equal number of observations per cluster.

Random Group Method

Therefore, the random group method was used to estimate variances of means for specific age groups 9/. It was decided to use thirty random groups for computations.

The 5344 persons were randomly assigned to 31 groups of equal size, each group with approximately the same distribution of persons among the ten strata. Cases in the 31st group were randomly distributed to other groups when necessary 10/. All persons in the same family were assigned to the same random group.

For computation of the variance of a specific age group, the number of persons from that age group in each of thirty random groups was to be equal. The machine first distributed all persons in an age group who had been assigned to random groups 1 through 30 to 30 different locations. Persons in the 31st group were randomly distributed to locations with the smallest number of persons.

For example, there were 1531 persons 35 through 54 years old covered in the interviews. After the 31st group had been distributed among the smallest groups, there were five groups with 46 persons and 25 groups with 47 through 63 persons. The program instructed the machine to find the sum of visits reported for all 46 persons in each of the five groups, and the sum for 46 randomly selected persons in each of the other 25 locations. Thus, in the random group method, computation of the variance for this age group was based on 30 times 46, that is, 1380 randomly selected persons, or 90 percent of all persons in this age group.

The sum of the number of physician visits reported for an equal number of persons in each random group from each age group was used to compute variances as follows 9/:

If = number of reported physician visits for a person in one year,

K = number of persons in each random group used for computations,

T = number of random groups = 30

and X_g = total number of visits for K persons in group g,

then, the variance of X was estimated to be:

$$\frac{\sum_{g=1}^{T} X_g^2 - \left(\sum_{g=1}^{T} X_g\right)^2 \Big/ T}{K(T-1)} \qquad (1)$$

The variance of mean number of visits for each age group was estimated to be $\mathrm{Var}\,(\bar{X}) = \frac{\mathrm{Var}\ (X)}{30K}$, since 30K was the number of persons in each age group used to estimate Var (X). The standard error of the mean was estimated to be

$$\mathrm{S.E.}\ (\bar{X}) = \sqrt{\mathrm{Var}(\bar{X})}\ .$$

Table 1 summarizes estimates by the random group method for all persons and for persons in each of six age groups.

Comparisons of Variances and Means of Age Groups

Because the frequency distribution of visits was not normal, usual tests of homogeneity of variances might indicate differences that do not really exist 11/. Since the estimated variances for these six age groups ranged from 16 to 108 (Table 1), it seemed reasonable to reject the hypothesis that variances of the six age groups were homogeneous.

Therefore, tests of means that do not assume equal variances were needed. An approximate test of homogeneity of means of large samples allowing for unequal variances is a chi square test of homogeneity 12/. The least squares estimate of the mean is the weighted sum of observed means with the weights for each mean equal to the reciprocal of its estimated variance. The weighted sum of the squares of the deviations of the observed means from this estimate, using the same weights, has a chi square distribution. This test is expressed mathematically below.

If $\bar{X}_i$ = observed mean number of physician visits of persons in age group i,

and $\mathrm{Var}(\bar{X}_i)$= estimated variance of mean of age group i using the random group method (equals square of S.E. $(\bar{X}_i)$ in Table 1),

then $$\bar{X}_s = \sum_i \left[\bar{X}_i \ / \ \mathrm{Var}(\bar{X}_i)\right] \Big/ \sum_i \left[1 \ / \ \mathrm{Var}(\bar{X}_i)\right]$$

= least squares estimate of the mean

and the statistic $\sum_i \left[(\bar{X}_i - \bar{X}_s)^2 \ / \ \mathrm{Var}(\bar{X}_i)\right]$ is approximately distributed as chi square with the number of degrees of freedom equal to one less than the number of groups.

Applying this test to mean visits of six age groups in Table 1, one rejects the hypothesis that the six means are equal. ($P<.001$)

One may also want to compare pairs of these means. Since these means were based on sufficiently large independent samples, and since the variances were estimated with sufficient degrees of freedom, i.e. 29, one may assume that the following ratio is approximately distributed as a normal deviate 13/:

$$\begin{aligned} Z &\div \left|\bar{X}_1 - \bar{X}_2\right| \Big/ \sqrt{\mathrm{Var}(\bar{X}_1 - \bar{X}_2)} \\ &\div \left|\bar{X}_1 - \bar{X}_2\right| \Big/ \sqrt{\mathrm{Var}(\bar{X}_1) + \mathrm{Var}(\bar{X}_2)} \\ &\div \left|\bar{X}_1 - \bar{X}_2\right| \Big/ \sqrt{\frac{\mathrm{Var}(X_1)}{N_1} + \frac{\mathrm{Var}(X_2)}{N_2}} \end{aligned}$$

where N_1 and N_2 are sizes of two independent samples from distributions with unequal variances and $\mathrm{Var}(X_i)$ are from Table 1.

Applying this test, one infers that the mean number of visits for persons 65 years and older is significantly higher than the mean for persons in each of the four youngest age groups, but not significantly different from the mean for persons 55 to 64 years old. (Tests done on 5 percent level).

Because of this wide variation of mean visits with age, comparison of means of classes with different age compositions might obscure variations due to factors other than age.

Age-Adjusted Means and Their Variances

Comparisons of age-adjusted means is one method to study differences of means of classes with different age compositions. Means of each of six age groups within an income class of size n, for example, were first computed. It was assumed that these six means were observed for a sample of size n with the same proportions in the six age groups as the total 5344 persons covered in interviews.

This procedure is expressed mathematically.

If n_i = number of persons of income class in age group i,

$\sum_i n_i = n$ = total number of persons in income class,

X_{ij} = number of physician visits for j'th person in i'th age group,

and $\text{Var}(X_i)$ = estimated variance of visits for persons in age group i of income class,

then $\bar{X}_i = \sum_{j=1}^{n_i} X_{ij} / n_i$ = mean number of visits of persons in age group i of income class

and $\text{Var}(\bar{X}_i) = \text{Var}(X_i) / n_i$ = variance of mean of age group i of income class.

If W_i = proportion of 5344 persons in age group i,

then $\bar{X}_a = \sum_i (W_i \bar{X}_i)$ = age-adjusted mean of income class.

The variance of the age-adjusted mean is

$$\begin{aligned} \text{Var}(\bar{X}_a) &= \text{Var}\left(\sum_i W_i \bar{X}_i\right) \\ &= \sum_i W_i^2 \text{Var}(\bar{X}_i) \; \text{[14]} \\ &= \sum_i W_i^2 \left(\text{Var}(X_i) / n_i\right) \qquad (2) \end{aligned}$$

The estimated standard error of the age-adjusted mean is S.E. $(\bar{X}_a) = \sqrt{\text{Var}(\bar{X}_a)}$, and the 95 percent confidence interval is estimated to be $\bar{X}_a \pm 2$ S.E. $(\bar{X}_a)$.

The example to be discussed compares age-adjusted mean number of visits per person in each of four income classes with different age distributions. For example, 32 percent of persons in the lowest income class was 65 years or older in contrast with 7 to 10 percent in the three other income classes.

Since estimated variances of age-adjusted means of each income class are functions of estimated variances for each age-income class, methods of estimating these variances will be described.

In the compromise method; the variance of X found by the random group method for all persons in each age group was used as the estimate of $\text{Var}(X_i)$ for persons of the corresponding age group in each income class. More specifically, the $\text{Var}(X_i)$ in Table 1 were substituted in equation (2) to estimate variances of age-adjusted means of each income class.

In the optimum method; more precise estimates of $\text{Var}(X_i)$ could be obtained by applying the random group method to each of 24 age-income classes. Because of the small number of persons in some of these age-income classes, the allocation to random groups by the procedure described above resulted in some random groups with no or very few persons. Therefore, modifications of this procedure would be needed, including the use of fewer than 30 groups and/or random groups without homogeneous distributions among the strata used for sample selection.

In the direct method; $\text{Var}(X_i)$ would be estimated by $\sum_j (X_{ij} - \bar{X}_i)^2 / (n_i - 1)$ for each age-income class. This method would be expected to underestimate variances because it ignores within cluster correlation.

For example, the variance estimated for all persons 35 to 54 years by the random group method was 108. (Table 1) Variances estimated by the direct method for persons of this age group in each of four income classes were considerably smaller; ranging from 33 to 74.

It can be shown mathematically that the variance for each age group estimated by the compromise method is the sum of variances within four income classes plus a weighted sum of differences of income class means from the overall mean [15].

Thus, the first method is a compromise between the optimum method that gives more precise estimates but requires a great deal of time and money, and the direct method that underestimates variances.

The results of computations by the compromise method are shown in Table 2. Estimated 95 percent confidence intervals of age-adjusted means for four income classes were very similar. A chi square test of homogeneity of these means indicated that the age-adjusted mean number of reported physician visits per person per year did not differ significantly among four income classes. (It should be noted here that these reported visits included paid, prepaid and free visits in homes, offices and clinics.)

This finding of homogeneity of age-adjusted mean visits among income classes suggests further study. Since the average family size was smallest in the lowest income class (Table 2), future analyses might preferably use an index of family income per person. Other analyses might include comparisons of age-adjusted means of income-ethnic classes, for example.

Age-Adjusted Proportions

Age-adjusted proportions of persons with characteristics that vary with age can also be used to compare classes with different age compositions.

One method of studying sampling variation of age-adjusted proportions will be discussed.

Let n_i = number of persons of income class in age group i,

$\sum_i n_i$ = n = total number of persons in income class,

W_i = weight = proportion of 5344 persons in age group i,

p_{ir} = proportion of n_i persons in category r, for example, persons with at least one visit,

and $q_{ir} = 1 - p_{ir}$.

Then, $P_a = \sum_i (W_i p_{ir})$

= age-adjusted proportion of income class in category r,

and $Var(P_a) = \sum_i Var(W_i p_{ir}) = \sum_i W_i^2 \; Var\;(p_{ir})$.

The standard error of the age-adjusted rate is S.E. $(P_a) = \sqrt{Var(P_a)}$, and the 95 percent confidence interval is estimated to be

$$P_a \pm 2\;S.E.\;(P_a).$$

Estimates of Var (p_{ir}) that consider the sample design can be obtained by the random group method. Let the variable X_{ir} be defined as 1 for persons of age group i in category r and as 0 for persons not in category r. The mean of X_{ir} for age group i is an estimate of p_{ir}, and the $Var(X_i)$ estimated by expression (1) is an estimate of $Var(p_{ir})$. Comparisons of estimated confidence intervals and tests of homogeneity of $\bar{X}_a$ would be equivalent to analyses about P_a.

This method can be used, for example, to compare age-adjusted proportions of persons in each income class who had visits to an outpatient department.

Appropriate extensions would be needed to develop significant tests of age-adjusted distributions.

Summary

A health survey included questions about physician visits, which varied widely with age. Age-adjusted means were used to compare mean visits among classes with different age compositions.

Because of the complex sample design, the random group method was used to estimate variances for each age group.

Variance of the age-adjusted mean of each class was estimated as $\sum_i W_i^2 \;(Var(X_i)\; / \;n_i)$, where $Var(X_i)$ was the variance estimated for all persons of the i'th age group by the random group method. Using these means and variances in a chi square test of homogeneity, one accepts the hypothesis that the age-adjusted mean visits of four income classes were equal.

Implications

This paper described the application of well-known statistical techniques to the study of sampling variation of age-adjusted means and proportions. Implications for other research work are:

1. The random group method is appropriate to estimate variances when one or more of the following circumstances obtain:

   a. When the method frequently suggested for estimating variances for the specific sample design cannot be applied because there are vastly different numbers of observations in sampling units expected to have approximately the same number of cases.

   b. When the frequently proposed method of estimating variances for that sample design would require a prohibitive amount of time and/or money if used for many variables and many types of classes.

2. The optimum method to estimate variances of age-adjusted means would use variances for each age group within each class, estimated by a procedure that considers the sample design, including the random group method.

   In a compromise method that requires less time and/or money, variances estimated for all persons of each age group by a method that considers the sample design can be used as estimates for variances of persons of corresponding ages within each class.

   Variances of each age group estimated by the compromise method would always be larger than corresponding variances estimated by the optimum method by an amount related to the difference between class means.

   Tests using variances estimated by the compromise method would be more conservative. The probability of accepting a false hypothesis would be greater with the compromise method than with the optimum method.

Table 1

Mean, Variance and Standard Error of Reported Physician Visits per Person per Year, by Age Group, Estimated by Random Group Method

Age Group (years) i	Number of persons covered in interviews N	Mean number of visits $\bar{X}$	Number of persons used to estimate variance 30K	Estimated variance Var(X)	Estimated standard error of mean S.E.($\bar{X}$)[2]
Total	5344	3.1	5340[1]	110.7	.14
Under 5	386	2.2	270	16.1	.24
5-14	593	1.5	480	32.2	.26
15-34	1358	2.8	1290	49.2	.20
35-54	1531	3.0	1380	107.9	.28
55-64	743	3.9	570	87.0	.39
65 and older	733	4.9	540	103.5	.44

---

[1] The number of persons used to estimate variance for persons of all ages was greater than the total persons used to estimate variances for each age group.

[2] The standard error of the mean was estimated to be: $\text{S.E.}\ (\bar{X}) = \sqrt{\frac{\text{Var}(X)}{30K}}$.

Table 2

Estimated Standard Error and 95 Percent Confidence Interval of Age-adjusted Mean Reported Physician Visits per Person per Year, by Family Income Class

	Total	Family Income in 1960 or 1961			
		Under $3000	$3000-4999	$5000-7499	$7500 & over
Size of Family					
Number of persons of all ages	5344[1]	922	1221	1468	1125
Number of families interviewed	2216[1]	551	510	514	380
Number of persons per family	2.4	1.7	2.4	2.9	3.0
Mean visits per person	3.11	3.49	3.05	3.09	3.09
Age-adjusted mean visits per person					
Age-adjusted mean[2]		3.11	3.23	3.36	3.05
Estimated standard error[3]		.30	.26	.24	.26
Estimated 95 percent confidence interval[4]		2.5-3.7	2.7-3.8	2.9-3.8	2.5-3.6

---

[1] Includes 608 persons in 261 families with family income not reported.

[2] Weights (W_i) used to compute age-adjusted means were the proportions of 5344 persons in six age groups.

[3] The estimated variances for each age group used to estimate the variance of age-adjusted means were Var(X_i) from Table 1. The standard error of the age-adjusted mean was estimated to be

$$\text{S.E. } (\bar{X}_a) = \sqrt{\sum_i W_i^2 \left(\text{Var}(X_i) \,/\, n_i\right)} \text{ .}$$

[4] The 95 percent confidence interval was estimated to be $\bar{X}_a \pm 2$ S.E. (X_a).

FOOTNOTES

1/ This research was supported by the Health Research Council of New York City under Contract U-1053, and was conducted at the School of Public Health and Administrative Medicine, Columbia University.

2/ United States Department of Health, Education and Welfare, Public Health Service, National Health Survey, PHS Publication No. 584: Series B-19, Volume of Physician Visits, August 1960.

United States Department of Health, Education and Welfare, Public Health Service, National Center for Health Statistics, PHS Pub. No. 1000: Series 10, No. 9 - Medical Care, Health Status, and Family Income. May 1964.

3/ McNerney, W. J. and Study Staff. "Hospital and Medical Economics; A Study of Population, Services, Costs, Methods of Payment, and Controls," Vol. 1. Hospital Research and Educational Trust. Chicago, 1962.

4/ Kitagawa, Evelyn. "Standardized Comparisons in Population Research." Demography, Vol. 1, No. 1, 1964.

Levene, Howard. "Interrelations of Genetic and Environmental Factors in Disease." Annals of the New York Academy of Sciences. Vol. 84, Art. 17, 1960.

5/ This research was supported by Public Health Service Grant No. CH 00010-05 from the Division of Community Health Services, under direction of Edward A. Suchman.

6/ Elinson, J. and Loewenstein, R. "Community Fact Book for Washington Heights, New York City, 1960-61." School of Public Health and Administrative Medicine, Columbia University, 1963.

Glasser, M. "Master Sampling Plan for Columbia-Washington Heights Community Mental Health Project." School of Public Health and Administrative Medicine, Columbia University. Memorandum, December 15, 1958.

7/ In the original description of this sampling plan, Marvin Glasser suggested that variances per housing unit be computed by using Equation 5.14, page 318 of Hansen, M.; Hurwitz, W.; and Madow, W.G. "Sample Survey Methods and Theory; Methods and Applications." Vol. 1. John Wiley and Sons, Inc., 1956.

8/ Loewenstein, R.; Colombotos, J. and Elinson, J. "Interviews Hardest-to-Obtain in an Urban Health Survey." Proceedings of the Social Statistics Section, American Statistical Association, 1962.

9/ Hansen, M.; Hurwitz, W.; and Madow, W.G. "Sample Survey Methods and Theory; Methods and Applications." Vol. 1, pp. 440-444. John Wiley and Sons, Inc., 1956.

10/ Dr. Norman Matlin conceived of the concept of an additional group, and also wrote the program for this operation.

11/ Eisenhart, C. "The Assumptions underlying the Analysis of Variance." Biometrics, Vol. 3, No. 1, March 1947.

Scheffé, H. "The Analysis of Variance." John Wiley and Sons, 1959.

12/ Fertig, J.W. Mimeographed notes for Biostatistics 204, School of Public Health and Administrative Medicine, Columbia University, 1963.

13/ Walker, H. and Lev, J. "Statistical Inference." Henry Holt and Co., 1953.

14/ Based on Theorem: $\mathrm{Var}(\sum_i a_i X_i) = \sum a_i^2 \mathrm{Var}(X_i)$ if a is a constant.

15/ We are grateful to Professor John W. Fertig, School of Public Health and Administrative Medicine, Columbia University, for his contribution to this part of the paper, as well as his review of the entire manuscript and many helpful suggestions.

MULTIPLE FRAME SAMPLE SURVEYS

Robert S. Cochran, University of Wyoming

I. INTRODUCTION

In the theory and practice of sampling finite populations two concepts are carefully distinguished in the literature. They are

(1) the population of units, and
(2) the frame for sampling the units.

Concept (1) does not involve much more than a clear definition of the units deemed to belong to the population. However, concept (2) goes beyond a mere definition of an aggregate of units.

In many situations it is not possible to designate a unique reference frame for some reason. It then becomes necessary to supplement the original frame with an additional frame or frames in order to obtain full coverage of the population, and the investigator must design a survey based upon a multiplicity of sampling frames. In other situations it is possible to designate one frame that will give complete coverage but it may be possible to use another frame to cover a subset of the original frame. Here, again, it may be advantageous to the investigator to consider his problem as being one of multiple frame sampling.

Historically, most cases of multiple frames have been concerned with a "master" frame with 100 coverage and a "cheap" frame not possessing 100 per cent coverage. Usually a sample design had been designated for both frames, but whenever a unit sampled from the "master" frame was encountered which was also in the "cheap" frame it was discarded.

At the 1962 meetings of the American Statistical Association, H. O.Hartley presented a paper entitled "Multiple Frame Surveys." (1962) In his paper he suggested a weighting system whereby it would not be necessary to discard any information obtained in order to arrive at an estimate of the population total or the population mean. In this paper the basis for Hartley's work will be reviewed and a comparison will be presented between using the multiple frame estimator and the screening estimator outlined above.

II. MULTIPLE FRAME TERMINOLOGY AND NOTATION IN THE TWO FRAME CASE

Consider the general situation of two interlocking frames, A and B, where their union gives 100 per cent coverage to the populationof interest. In this situation there are three subsets consisting of those units that are only in A, (domain a), those units that are only in B (domain b), and those units in both frames (domain ab).

Also assume a simple random sample is to be drawn from each frame and that the number of population units in the three domains are known. The notation introduced by Hartley will be used for this discussion. It is presented in Table I.

III. ESTIMATION OF THE POPULATION TOTAL

Hartley's method of weight coefficients actually creates non-overlapping strata out of the overlapping frames. Letting y_i be the value of the Y characteristic attached to the i^{th} sampling unit define

$$u_{Ai} = y_i, \quad \text{if the } i^{th} \text{ unit is in } \underline{a}$$
$$= py_i, \quad \text{if the } i^{th} \text{ unit is in } \underline{ab}$$

when sampling from the A frame. Also define

$$u_{Bi} = qy_i, \quad \text{if the } i^{th} \text{ unit is in } \underline{ab}$$
$$= y_i, \quad \text{if the } i^{th} \text{ unit is in } \underline{b}$$

when sampling from the B frame. Thus, the number of units in the intersection, ab, have been artificially doubled, creating two mutually exclusive strata with U characteristic attached to the units of each strata. Clearly the population total of the original Y characteristic is equal to the population total of the newly constructed U characteristic whenever p + q = 1.

The estimate of the population total of the Y characteristic is

$$\hat{Y} = N_a\bar{y}_a + N_{ab}p\bar{y}_{ab}' + N_{ab}q\bar{y}_{ab}'' + N_b\bar{y}_b \qquad (1)$$

This estimator is in the form of a post-stratified sampling estimator for the U characteristic. Because of this the variance of $\hat{Y}$ when the usual finite population correction can be ignored and the sample is sufficiently large from each frame is approximately

$$V(\hat{Y}) = \frac{N_A}{n_A}\left\{\sigma_a^2 N_a + \sigma_{ab}^2 N_{ab} p^2\right\} + \frac{N_B}{n_B}\left\{\sigma_b^2 N_b + \sigma_{ab}^2 N_{ab} \ q^2\right\} \qquad (2)$$

where σ_a^2, σ_b^2, and σ_{ab}^2 are the wihin post stratum variances.

After some algebra this becomes

$$V(\hat{Y}) = \frac{N_A^2}{n_A}\left\{\sigma_a^2(1-\alpha) + \alpha\ p^2\ \sigma_{ab}^2\right\}$$
$$\frac{N_B^2}{n_B}\left\{\sigma_b^2(1-\beta) + \beta\ q^2\ \sigma_{ab}^2\right\}$$

where $\alpha = N_{ab}/N_A$, and $\beta = N_{ab}/N_B$. (3)

Assuming a simple cost function

$$C = c_A\ n_A + c_B n_B \qquad (4)$$

TABLE I
HARTLEY'S NOTATION FOR TWO FRAME DESIGNS AND ESTIMATES

Category	Frame			Domain		
	A	B	a	b	ab	
Population number	N_A	N_B	N_a	N_b	N_{ab}	
Sample number*	n_A	n_B	n_a	n_b	n'_{ab},	n''_{ab}
Population total	Y_A	Y_B	Y_a	Y_b	Y_{ab}	
Population mean	$\bar{Y}_A$	$\bar{Y}_B$	$\bar{Y}_a$	$\bar{Y}_b$	$\bar{Y}_{ab}$	
Sample total*	y_A	y_B	y_a	y_b	y'_{ab}	y''_{ab}
Sample mean*	$\bar{y}_A$	$\bar{y}_B$	$\bar{y}_a$	$\bar{y}_b$	$\bar{y}'_{ab}$	$\bar{y}''_{ab}$
Cost of sampling unit*	c_A	c_B				

*Applies to case of drawing random samples from both frames

where C is the total cost of sampling, c_A is the cost of obtaining an observation from the A frame and c_B is the cost of obtaining an observation from the B frame, it is possible to contemplate finding the values of n_A, n_B, and p that will give a minimum value for the variance whenever the cost is fixed or vice versa.

After some labor, the optimum value of p is found to be one of the solutions of

$$p^2 \rho \quad [\emptyset_B(1-\beta\ ) \ + \beta \ \ q^2] = q^2 \ [\emptyset_A(1-\alpha)+ \alpha p^2] \tag{5}$$

where $\rho = c_A/c_B$, $\emptyset_B = \sigma_b^2/\sigma_{ab}^2$, and $\emptyset_A = \sigma_a^2/\sigma_{ab}^2$. Once the value of p has been determined from the above the values of n_A and n_B can be found from

$$\frac{n_A}{N_A} = \theta \left\{(\sigma_a^2 \ (1-\alpha) + \alpha \ p^2 \ \sigma_{ab}^2) \ / \ c_A\right\}^{1/2}$$
$$\frac{n_B}{N_B} = \theta \left\{\frac{(\sigma_b^2 \ (1-\beta) + \beta \ q^2 \ \sigma_{ab}^2)}{c_B}\right\}^{1/2} \tag{6}$$

where θ would be determined by the budget available.

IV. ONE FRAME IS 100 PER CENT

An important special case of the above exists whenever one of the frames, say A, gives 100 per cent coverage to the population of interest and the other frame covers only a subset of the population. An example of such a situation is a survey of households in a city which would be sampled by the block sampling plan while it may be possible to sample most of the population by using the telephone directory. However, there are households that do not have a telephone and can only be sampled by more expensive personal interviews.

When using simple random sampling from both frames we have a special case $\beta = 1$ and $\sigma_B^2 = \sigma_{ab}^2$. With this simplification the estimator becomes

$$\hat{Y} = N_a \bar{y}_a + p \ N_{ab} \ \bar{y}'_{ab} + q \ N_{ab} \ \bar{y}''_{ab} \tag{7}$$

and the variance becomes

$$(\overline{Y})= \frac{N_A^2}{n_A} \ [\sigma_a^2(1-\alpha) + \alpha \ p^2 \ \sigma_{ab}^2]$$
$$+ \ \frac{N_B^2}{n_B} \ \sigma_{ab}^2 \ q^2. \tag{8}$$

Using the same cost equation as above the optimum p now becomes

$$p= \sqrt{\frac{\emptyset_A \ (1-\alpha)}{-\alpha}} \tag{9}$$

An alternate sampling plan for such a situation is the one mentioned in the introduction as the plan that has been usually used when such situations have developed historically. This alternate is actually a special case of the above with p = 0 and q = 1. Therefore, the estimator is

$$\hat{Y}_0 = \frac{N_a}{n_a} \ y_a + \frac{N_B}{n_B} \ y_B \tag{10}$$

and the approximate variance is

$$V(\hat{Y}_0) = \frac{N_A^2}{n_A} \ (1-\alpha) \ \sigma_a^2 + \frac{N_B^2}{n_B} \ \sigma_{ab}^2$$
$$= \frac{N_A^2}{n_A} \ \sigma_{ab}^2 \ [(1-\alpha) \ \emptyset_A + \frac{n_A}{n_B} \alpha^2] \tag{11}$$

For this procedure the cost equation would be

$$C = n_a \ c_a + n'_{ab} \ \ c'_A + n_B \ c_B \tag{12}$$

because n_a and n'_{ab} are random we have

$$E(C) = \ (1-\alpha) \ c_A + \alpha c'_A \ \ n_A + c_B \ n_B$$
$$= c^*_A \ n_A + c_B \ n_B$$

and

c_A = the cost of interviewing sampled units from the 100 per cent frame

c_B = the cost of interviewing sampled units from the list

c'_A = the cost of determining that units sampled from the 100 per cent frame are also on the list (screening cost)

Using this cost equation the optimum values of n_A and n_B yield

$$\frac{n_A}{n_B} = \frac{1}{\alpha} \sqrt{\frac{\emptyset_A \ \ (1-\alpha)}{\rho^*}} \tag{13}$$

where $\rho^* = \frac{c_A^*}{c_B}$

Using this expression for n_A/n_B in $V(\hat{Y}_0)$ yields

$$V(\hat{Y}_0) = \frac{N_A^2}{n_A} \sigma_{ab}^2 \left[(1-\alpha) \phi_A \quad \alpha \sqrt{\frac{\phi_A (1-\alpha)}{\rho *}} \right] \qquad (14)$$

Likewise in C it yields

$$C = n_A c_A^* \left(1 + \frac{\alpha}{\sqrt{\rho *} \sqrt{\phi_A (1-\alpha)}} \right) \qquad (15)$$

Thus the variance of the screening estimator is

$$V_0 = \frac{N_A^2 \sigma_{ab}^2}{C} \quad c_A^* \left(1 + \frac{\alpha}{\sqrt{\rho *} \sqrt{\phi_A (1-\alpha)}} \right)^2 \phi_A (1-\alpha) \qquad (16)$$

On page 4 equation (3), the corresponding multiple frame estimator is given to be

$$V(\hat{Y}) = \frac{N_A^2}{n_A} [(1-\alpha) \sigma_a^2 + \alpha p^2 \sigma_{ab}^2] + \frac{N_B^2}{n_B} q^2 \sigma_{ab}^2 . \qquad (17)$$

The optimum values for p and q given values of n_A and n_B lead to

$$\frac{p}{q} = \frac{n_A}{n_B} \frac{N_B}{N_A} = \frac{n_A}{n_B} \alpha. \qquad (18)$$

Substitution of the above into $V(\hat{Y})$ leads to

$$V(\hat{Y}) = \frac{N_A^2}{n_A} \sigma_{ab}^2 \quad [(1-\alpha) \phi_A + \alpha p] \qquad (19)$$

Also substituting p/q above into the cost equation on page 4, equation (4), leads to

$$C = c_A n_A \quad (1 + \frac{q}{\rho} \frac{\alpha}{p})^2 \qquad (20)$$

Therefore, the variance of the multiple frame estimator is

$$V(\hat{Y}) = \frac{N_A^2 \sigma_{ab}^2}{C} \quad c_A p^2 \rho \quad (1 + \frac{q}{p} \frac{\alpha}{\rho})^2 \qquad (21)$$

When the total budget C is the same for the two types of investigations the ratio of $V(\hat{Y})$ to $V(\hat{Y}_0)$ will give an indication of the relative efficiency of the screening estimator as compared to the multiple frame estimator.

After some algebra this ratio becomes

$$\frac{V(\hat{Y})}{V(\hat{Y}_0)} = \frac{V_P}{V_0} = \frac{c_A}{c_A^*} \frac{\rho}{\rho-\alpha} \frac{(1 + \frac{q \alpha}{p \rho})^2}{(1 + \frac{\alpha}{p \sqrt{\rho^*} \sqrt{\rho-\alpha}})^2}$$

Letting

$$w = \frac{c_A}{c_A^*}$$

ρ^* becomes

$$\rho^* = \frac{\rho}{w}$$

and the variance becomes

$$\frac{V_p}{V_0} = w \frac{\rho}{\rho-\alpha} \quad \frac{(1 + \frac{q \alpha}{p \rho})^2}{(1 + \frac{\alpha \sqrt{w}}{p \sqrt{\rho} \sqrt{\rho-\alpha}})^2} \qquad (24)$$

The screening estimator will have the lower variance whenever this ratio is greater than 1. In order to determine parametric conditions that will result in such a situation let

$$A = \frac{\rho}{\rho-\alpha}$$
$$B = (1 + \frac{q \alpha}{p \rho})^2 \qquad (25)$$
$$C^{-1} = \frac{\alpha}{p \sqrt{\rho} \sqrt{\rho-\alpha}}$$

With these definitions $V/V_0 > 1$ becomes

$$\frac{w AB}{(1 + \frac{\sqrt{w}}{C})^2} > 1 \qquad (26)$$

From this it can be shown that

$$\sqrt{w} > (\sqrt{AB} - C^{-1})^{-1} \qquad (27)$$

and because ρ is usually greater than α it can be shown that

$$\sqrt{AB} > C^{-1}. \qquad (28)$$

Therefore,

$$w > (\sqrt{AB} - C^{-2}) \qquad (29)$$

which yields

$$w > \left[(1 + \frac{q \alpha}{p}) \sqrt{\frac{\rho}{\rho-\alpha}} - \frac{\alpha}{p \sqrt{\rho} \sqrt{\rho-\alpha}} \right]^{-2} = \frac{\rho}{\rho-\alpha} > 1 \qquad (30)$$

Thus w will be greater than $\frac{\rho}{\rho-\alpha}$ whenever V/V_0 is greater than 1. It can also be shown that w greater than $\frac{\rho}{\rho- \alpha}$ implies that V/V_0 will be greater than 1. (31)

In terms of the cost conditions in its definition

$$w = \frac{c_A}{c_A^*} = \frac{c_A}{(1-\alpha) c_A + c_A' \alpha} = \frac{1}{(1-\alpha) + w'\alpha}$$

where

$$w' = \frac{c'_A}{c_A} .$$

Since the advantage is to the screening estimator whenever $w > \frac{\rho}{\rho - \alpha}$ using the definition of w in (31) we find

$$\frac{1}{(1-\alpha) + w'\,\alpha} > \frac{\rho}{\rho - \alpha} \qquad (32)$$

or

$$\frac{\rho-\alpha}{\rho} > 1 - \alpha + \alpha w'$$

and

$$1 - \frac{1}{\rho} > w'. \qquad (33)$$

Using the definitions of ρ and w' this can be written as

$$1 - \frac{c_B}{c_A} > \frac{c'_A}{c_A} \quad \text{or} \quad c_B < c_A - c'_A \qquad (34)$$

This indicates that on the average the screening estimator will have the lower variance whenever the cost of sampling from the supplementary frame is less than the difference between sampling from the 100 per cent frame and screening members of the 100 per cent frame in the supplementary frame. For example, if it does cost less to ask informative questions of a person on the telephone than it does to ask them face to face, the screening estimator will have the lower variance.

Some illustrations of the relationship between the variance ratio V/V_0 and the screening cost ratio w' under various parametric conditions are given in the following tables and graphs. These are set up for a low (.25) and a high (.75) values of $\emptyset_A = o^2_a/o^2_{ab}$, for a low (.20), a medium (.50), and a high (.90) values of $\alpha = N_{ab}/N_A = N_B/N_A$, and for a low (1), a medium (2), and high (10) values of $\rho = c_A/c_B$. The range presented for $w' = c'_A/c_A$ is from 0.0 to 1.0 in increments of 1.

TABLE II

$\emptyset_A = .75$

		α = .50			α = .20			α = .90		
	ρ	1	2	10	1	2	10	1	2	10
w'	p	.87	.50	.20	.87	.58	.25	.87	.26	.09
.0		1.00	1.26	1.63	1.00	1.10	1.20	1.00	1.35	2.12
.1		.95	1.20	1.52	.99	1.08	1.18	.99	1.24	1.82
.2		.92	1.14	1.43	.97	1.06	1.15	.89	1.16	1.58
.3		.88	1.09	1.34	.95	1.04	1.12	.86	1.10	1.44
.4		.84	1.04	1.27	.93	1.02	1.11	.83	1.05	1.32
.5		.82	1.00	1.20	.92	1.00	1.08	.80	1.00	1.23
.6		.79	.96	1.14	.90	.98	1.06	.78	.96	1.15
.7		.76	.92	1.09	.88	.96	1.04	.76	.93	1.09
.8		.74	.90	1.04	.87	.94	1.02	.74	.91	1.03
.9		.72	.87	1.00	.86	.93	1.00	.72	.88	.98
1.0		.70	.84	.91	.84	.91	.98	.71	.85	.93

TABLE III

$\emptyset = .25$

		α = .20			α = .50			α = .90		
	ρ	1	2	10	1	2	10	1	2	10
w'	p	.5	.33	.14	.50	.29	.11	.50	.15	.05
.0		1.00	1.09	1.19	1.00	1.20	1.52	1.00	1.20	1.67
.1		.99	1.07	1.17	.96	1.14	1.44	.96	1.14	1.50
.2		.97	1.05	1.14	.94	1.11	1.36	.93	1.09	1.38
.3		.96	1.03	1.12	.91	1.07	1.29	.91	1.06	1.30
.4		.94	1.02	1.10	.88	1.03	1.23	.89	1.02	1.23
.5		.93	1.00	1.07	.86	1.00	1.18	.87	1.00	1.17
.6		.92	.99	1.06	.83	.96	1.13	.85	.97	1.12
.7		.90	.96	1.04	.81	.94	1.08	.84	.95	1.08
.8		.89	.95	1.02	.81	.92	1.05	.83	.94	1.04
.9		.88	.94	1.00	.79	.91	1.00	.81	.91	1.00
1.0		.87	.92	.98	.77	.86	.97	.80	.90	.97

THE ALPHABETIC ARRAY SAMPLING TECHNIQUE

Abraham Frankel and Patricia Wright

U. S. Office of Education

In many of the surveys conducted by the U. S. Office of Education, it is often necessary to select samples of elementary reporting units such as students and teachers. Because listings of these units are usually not available, a two-stage sample design has been used in such situations. In order to facilitate the selection of samples of these elementary units, an alphabetic array sampling technique was designed.[1] Informal studies have shown that the first letter of the surnames of the faculty and students in educational institutions is generally distributed in a non-random fashion; that is, some of the letters appear more often than others. On this basis, in the alphabetic array technique the letters of the alphabet may be grouped so that selections may be made with known probabilities. Instructions are then issued to respondents in the first stage of the sample design to select for the sample of elementary units a group or a combination of groups of letters. For example, some university deans might be asked to select in their sample all faculty members whose surname began with the letters A, P, R, or T. Other deans might be requested to select other combinations of letters in order to provide the required sample size.

The sampling design of the alphabetic array is based on the fact that the relative frequency of the first letters of the surnames is quite stable for large populations. Hence, the technique involves arranging the 26 letters of the alphabet into groups of letters representing the first letter of the surname of individuals in such a way that each group of letters includes about the same proportion of all names of the large population. That is, for an n line alphabetic array,

$$\Sigma p_{i1} = \Sigma p_{i2} = \ldots = \Sigma p_{ij} = \ldots = \Sigma p_{in}$$

and

$$\Sigma p_{i1} + \Sigma p_{i2} + \Sigma p_{i3} + \ldots + \Sigma p_{in} = 1.00$$

where

$j = 1, 2, 3, \ldots, n$

n = the number of lines in the array

p_{ij} = the probability of the occurence of the i-th letter which is grouped in the j-th line

Σp_{ij} = the probability of the occurence of all the letters which are grouped in the j-th line

The relative frequency of the occurence of the letters that is used in this technique is based on the surnames in the Social Security records which are fairly precise and surely representative of all surnames in the United States. The percentage distribution of surnames by initial letter from the Social Security records and from other listings are shown in Table 1.

Five and six line alphabetic arrays were developed from this distribution. In the five line array the sum of the probabilities of the occurence of the letters in each line should equal .20 or 1/5th of the population; in the six line alphabetic array the sum of the probabilities should equal .1667 for each line. The groupings of the letters to form these two arrays are shown below.

5 Line Alphabetic Array

Line	Letters	Σp_{ij}
1	J K O S V	.1980
2	B G L U Y	.1992
3	I M T W Z	.2011
4	C E F H X	.2023
5	A D N P Q R	.1994
Total		1.0000

6 Line Alphabetic Array

Line	Letters	Σp_{ij}
1	F J S X	.1678
2	B E L U Y	.1670
3	D M N Z	.1657
4	C G I K	.1670
5	H O Q V W	.1662
6	A P R T	.1663
Total		1.0000

The alphabetic array sampling technique was used in the sample selection procedure for the survey of the "Status and Career Orientation of College Faculty (COLFACS)" which was conducted by the U. S. Office of Education in the spring of 1963. The purpose of this study was to obtain basic data (personal characteristics, position and assignment, educational background, work experience, economic status and occupational plans) about college faculty in the aggregate United States. The universe of inquiry was defined as the full-time faculty with the rank of instructor or above, who taught at least one degree credit course in the spring of 1963 in universities, liberal arts colleges, teachers colleges or independent technological schools.

The sample design provided for a two-stage stratified sample (see Table 2). In the first stage, the universe of institutions was stratified by public or private control, by type of institution, and by size of faculty. The institutions that were included in the sample were selected with varying probabilities of 1:1, 1:2 or 1:5 depending upon the number of institutions in the stratum universe. In the second stage, the number of faculty that were sampled was selected in such a manner that the overall sampling fraction in each stratum was 1:10. In other words, where the institution

---

[1] Philip Desind, now with the Post Office Dept., developed the alphabetic array sampling technique in 1962 when he was with the U] S] Office of Education.

TABLE 1--PERCENTAGE DISTRIBUTION OF SURNAMES BY INITIAL LETTER FROM VARIOUS LISTINGS

Initial Letter	Social Security[1]	Who's Who in America (Vol. 30)	Who's Who in American Education (19th Ed.)	Washington,D.C. Metropolitan Telephone Dir. (Fall 1961)	ASA Directory 1961
A	3.051	3.223	3.536	3.327	3.310
B	9.357	9.947	9.491	9.620	9.172
C	7.267	7.406	7.072	7.594	7.162
D	4.783	4.803	4.901	4.846	4.279
E	1.888	2.138	2.171	1.917	1.978
F	3.622	3.936	4.032	3.653	4.081
G	5.103	4.648	5.025	4.991	5.372
H	7.440	8.119	8.561	7.594	6.839
I	.387	.434	.558	.362	.552
J	2.954	2.107	2.543	2.821	1.853
K	3.938	3.874	4.280	3.616	5.236
L	4.664	4.741	4.839	4.484	5.070
M	9.448	9.482	8.933	9.475	9.109
N	1.785	1.735	2.047	1.700	2.238
O	1.436	1.394	1.365	1.302	1.426
P	4.887	4.307	4.032	4.774	4.310
Q	.175	.155	.124	.217	.167
R	5.257	4.896	4.529	4.991	4.914
S	10.194	10.381	10.361	10.306	10.639
T	3.450	3.223	2.854	3.544	3.310
U	.238	.279	.248	.325	.354
V	1.279	1.209	1.365	1.085	1.249
W	6.287	6.694	6.017	6.582	5.944
X	.003	.019	.000	.007	.000
Y	.555	.465	.620	.506	.666
Z	.552	.403	.496	.361	.770
	100.0	100.0	100.0	100.0	100.0
Total Names	117,358,888	3,227	1,612	690,000[2]	9,606

[1] In 1957 the social security record keeping operations were converted to an electronic data processing system. At that time the Division of Accounting Operations, BOASI, SSA, HEW, compiled the material shown in this column. It represents all account numbers issued from the beginning of the social security program in 1936 to the middle of 1956.

[2] This represents the total listings and not the individual names in the telephone directory.

sampling fraction was 1:1, the faculty sampling fraction was 1:10; where it was 1:2, the faculty sampling fraction was 1:5; and where it was 1:5, the faculty sampling fraction was 1:2.

The overall 1:10 sampling rate was obtained by the use of the alphabetic array sampling technique in the following manner. If the institution in the sample was selected with a sampling fraction of 1:1, a line was selected at random from the five line alphabetic array; for example, line #3. Then that institution was told to include in their sample all faculty members meeting the specifications stated above whose surname began with the letters I, M, T, W or Z. Since this provided a 20 percent (1/5th) sample of the faculty under study, it was necessary to select, after a random start, every other name in order to obtain the desired 1:10 sample.

For those institutions that were selected with a 1:2 sampling fraction, a line was selected at random from the five line alphabetic array and the institution was asked to include in their sample those faculty members whose surname began with the letters of the selected line. No subsamples were taken as the required sample take of 1:5 was obtained.

For those institutions that were selected with a 1:5 sampling fraction, a sample of 1:2 of the faculty was required. This was obtained by taking three lines at random from the six line alphabetic array and forwarding these letters to the institution to be used in selecting their faculty to be included in the sample.

The sample selection plan for the COLFACS survey may be summarized as follows:

	Sampling Fraction		
First stage:			
Select institutions	1:1	1:2	1:5
Second stage:			
Select faculty:			
a. Use 1 line of five line array -	1:5	1:5	-
b. Use 3 lines of six line array -	-	-	1:2
c. Subsample -	1:2	-	-
Overall sampling fraction -	1:10	1:10	1:10

TABLE 2--SAMPLE DESIGN USED IN THE SURVEY OF STATUS AND CAREER ORIENTATION OF COLLEGE FACULTY, SPRING 1963

Type of Institution	Faculty Size	Public Control				Private Control			
		M	m/M	m	n/N	M	m/M	m	n/N
All Inst.		360		289		782		304	
University	Total	82		82		58		58	
	1,000 & over	15	1:1	15	1:10	4	1:1	4	1:10
	500--999	24	1:1	24	1:10	22	1:1	22	1:10
	Under 500	43	1:1	43	1:10	32	1:1	32	1:10
Liberal Arts	Total	85		85		668		190	
	150 & over	31	1:1	31	1:10	27	1:1	27	1:10
	75--149	27	1:1	27	1:10	115	1:2	58	1:5
	Under 75	27	1:1	27	1:10	526	1:5	105	1:2
Teachers College	Total	167		96		31		31	
	150 & over	23	1:1	23	1:10	0		0	
	75--149	69	1:2	35	1:5	1	1:1	1	1:10
	Under 75	75	1:2	38	1:5	30	1:1	30	1:10
Technological Schools	Total	26		26		25		25	
	150 & over	7	1:1	7	1:10	6	1:1	6	1:10
	75--149	9	1:1	9	1:10	6	1:1	6	1:10
	Under 75	10	1:1	10	1:10	13	1:1	13	1:10

M = number of institutions in the universe (1,142)
m = number of institutions in sample (593)
N = number of faculty in the universe
n = number of faculty in sample
m/M = sampling fraction for selecting institutions
n/N = sampling fraction for selecting faculty members

All institutions that were selected in the sample responded. A total of 15,494 names was received and questionnaires were mailed to these individuals. The response breakdown is shown below.

Number mailed out	15,494
Number of useable forms received	13,017
Number out of scope	1,694 1/
Refusals	56
Nonresponse	727

The number of faculty in 1962-63 meeting the survey specifications was estimated from the preliminary reports on "Faculty and Other Professional Staff in Institutions of Higher Education, 1961-62" (OE: 53014-62, Nov. 1963) and for 1963 - 64. These reports provide a count of the faculty as of the first semester of the 1961-62 and the 1963-64 academic years, respectively, classified by types of duty. The estimated number in 1961 was 129,700 and in 1963, 148,000. The number in 1962 is estimated at 138,500. The sample take, using the alphabetic array technique was 9.94 percent of the estimated universe.

II

In the belief that it may be of interest to others we discuss in this section of the paper a method of selecting elementary sampling units through the use of an n x 1 matrix where, by use of combinatorial methods, it is possible to include a fixed number of elementary units in the sample with a lesser or a larger number of primary sampling units (psu's) as desired. A larger number of psu's usually has the advantage of reducing the variance of estimates if the between psu variance is large. Conversely, when the within psu variance is large, a smaller number of psu's with an increased number of elementary units will tend to reduce the overall variance.

Let us consider a universe of M primary sampling units that has a total of T elementary units. A sample of t elementary units is to be selected from this universe. The following method may be used for selecting a sample of the primary sampling units from the universe and for determining the sampling fractions to be used in selecting the elementary units for a fixed sample size.

If the k-th row of an n x 1 matrix is arbitrarily selected to divide the n rows of a matrix into two groups, then we define those rows in the matrix from 1 through k as letter rows and those rows from k + 1 to n as number rows. This is shown in the

---

1/ This includes members of the faculty who were not employed full-time or who were serving in administration, personnel services, research, etc., and were not teaching.

figure below where

k = the number of letter rows
ϕ = the number of number rows
n = k + ϕ

$$\begin{bmatrix}1\\2\\3\\ \cdot\\ \cdot\\ k\\ k+1\\ k+2\\ \cdot\\ \cdot\\ n\end{bmatrix} \begin{matrix}\left.\begin{matrix}a_1\\a_2\\a_3\\ \cdot\\ \cdot\\ a_k\end{matrix}\right] \text{— letter rows}\\ \left.\begin{matrix}1\\2\\ \cdot\\ \cdot\\ \phi\end{matrix}\right] \text{— number rows}\end{matrix}$$

The n rows of the matrix are then combined p rows at a time. The total number of combinations is

$$\binom{n}{p} = \frac{n!}{p!(n-p)!} \qquad (1)$$

This equation may be rewritten as

$$\binom{n}{p} = \binom{k}{p}\binom{\phi}{0} + \binom{k}{p-1}\binom{\phi}{1} + \ldots + \binom{k}{p-(p-j)}\binom{\phi}{p-j} + \ldots + \binom{k}{0}\binom{\phi}{p} \qquad (2)$$

where $k \geq p$ and $\phi \geq p$. For clarity, we call that part of the combination that has the letter k in it the letter row component of the combination and the part that has ϕ in it the number row component.

Equation (1) provides the number of combinations into which the universe (M) has been divided and equation (2) the composition of the combinations according to the number of letter and number rows.

As the total number of psu's in the universe is M, the average number of psu's in each combination is

$$\bar{M} = \frac{M}{\binom{n}{p}} \qquad (3)$$

and the total number of psu's associated with the different combinations in equation (2) is

$$M = \bar{M}\binom{k}{p}\binom{\phi}{0} + \bar{M}\binom{k}{p-1}\binom{\phi}{1} + \ldots \bar{M}\binom{k}{0}\binom{\phi}{p} \qquad (4)$$

In order to determine the number of elementary sampling units that is to be taken from the psu's in the sample a weighting factor (w) is used in each of the different combinations. The value of the weighting factor w_i for the i-th combination is the value of p in the letter row component of the combination. For example, in equation (4), w equals p in the first combination; w equals (p-1) in the second combination; and w equals 0 in the last combination.

For the desired sample size (t) the following equation is set up and evaluated for x.

$$t = \bar{M}\binom{k}{p}\binom{\phi}{0}w_p x + \bar{M}\binom{k}{p-1}\binom{\phi}{1}w_{p-1}x + \ldots + \bar{M}\binom{k}{0}\binom{\phi}{p}w_0 x \qquad (5)$$

Then

$$\frac{w_i x}{\frac{T}{M}} \qquad (6)$$

is the sampling fraction to be used to select the elementary sampling units from those psu's that fall in the combination having the value of p_i in the letter row component of the combination.

An illustration of the above method is shown in the following example. A sample of 10,000 (t) is to be selected from a universe of 50,000 (T) teachers. The number of schools in the universe is 1,000 (M). It is believed, from other sources, that the between psu variance for the characteristic under study is high so that we would want a large number of schools in our sample. As our first trial we set up a 10 x 1 matrix with 5 letter rows (k) and 5 number rows (ϕ) and take the 10 rows in combination 2 at a time (p).

Using equations (1) and (2), we get

$$\binom{10}{2} = \binom{5}{2}\binom{5}{0} + \binom{5}{1}\binom{5}{1} + \binom{5}{0}\binom{5}{2}$$

$$45 = 10 + 25 + 10$$

This means that the 45 combinations are identified in the following manner: 10 combinations having 2 letter rows; 25 combinations having 1 letter and 1 number row; and 10 combinations having 2 number rows.

Note 1. Table 3 shows the number of combinations that result from various n x 1 matrices and row combinations.

The average number of schools in each combination is, by equation (3),

TABLE 3--NUMBER AND COMPOSITION OF COMBINATIONS FOR N x 1 MATRICES WITH 5 LETTER ROWS AND DIFFERENT ROW COMBINATIONS

Total no. of rows (N)	5	6	7	8	9	10	15	20	25
No. of letter rows	5	5	5	5	5	5	5	5	5
No. of number rows	0	1	2	3	4	5	10	15	20
2 row combinations									
Total combinations	10	15	21	28	36	45	105	190	300
2 letters (ij)	10	10	10	10	10	10	10	10	10
1 letter (i0)	0	5	10	15	20	25	50	75	100
0 letters (00)	0	0	1	3	6	10	45	105	190
3 row combinations									
Total combinations	10	20	35	56	84	120	455	1140	2300
3 letters (ijk)	10	10	10	10	10	10	10	10	10
2 letters (ij0)	0	10	20	30	40	50	100	150	200
1 letter (i00)	0	0	5	15	30	50	225	525	950
0 letters (000)	0	0	0	1	4	10	120	455	1140
4 row combinations									
Total combinations	5	15	35	70	126	210	1365	4845	12650
4 letters (ijkl)	5	5	5	5	5	5	5	5	5
3 letters (ijk0)	0	10	20	30	40	50	100	150	200
2 letters (ij00)	0	0	10	30	60	100	450	1050	1900
1 letter (i000)	0	0	0	5	20	50	600	2275	5700
0 letters (0000)	0	0	0	0	1	5	210	1365	4845

$$\bar{M} = \frac{M}{\binom{n}{p}} = \frac{1{,}000}{45} = 22.2$$

Substituting in equation (5)

$$10{,}000 = 22.2\binom{5}{2}\binom{5}{0}2x + 22.2\binom{5}{1}\binom{5}{1}1x + 22.2\binom{5}{0}\binom{5}{2}0x$$

and solving for x, we find that x = 10.

As T/M = 50,000/1,000 = 50, the sampling fraction to be used in those schools associated with the first type of combination (2 letter rows) is, by equation (6), (2 x 10)/50 = 2/5; for the second type (1 letter row and 1 number row), (1 x 10)/50 = 1/5; and for the third type (2 number rows), (0 x 10)/50 = 0; that is, schools in the combination identified by 2 number rows will not be included in the sample.

The sample selection procedure then involves the selection of 222 schools at random from the 1,000 schools in the universe and denoting them as 2 letter row schools and selecting 556 schools at random and denoting them as 1 letter and 1 number row schools. A 40 percent sample (2:5) of the teachers will be selected in the former schools and a 20 percent sample (1:5) of the teachers in the latter schools.

The results of the above example along with those obtained by using other n x 1 matrices are shown in Table 4. In all of these matrices the number of letter lines was equal to 5 and the rows were taken in combination 2 at a time.

TABLE 4 - RESULTS OBTAINED BY USING DIFFERENT n x 1 MATRICES

	n x 1 matrix			
	n = 5	n = 10	n = 20	n = 25
Number of combinations	10	45	190	300
Number of schools in sample	1,000	778	448	366
Number of schools (2 letter rows)	1,000	222	53	33
Sampling fraction for teachers	1:5	2:5	4:5	1:1
Number of schools (1 letter row and 1 number row)	0	556	395	333
Sampling fraction for teachers	-	1:5	2:5	1:2

METHODOLOGICAL PROCEDURES IN A VERY LARGE NATIONAL SURVEY OF PHYSICIANS

T. Donald Rucker, National Association of Blue Shield Plans
and
Martin Taitel, Chicago, Illinois

Introduction

This paper discusses some of the methodological considerations that arose during a large survey of physicians in the United States. The purpose of the study was to determine the relationship between Blue Shield payments and the cost of physician care for selected types of medical services under various contractual arrangements.

The survey was conducted by 54 (out of 75) Blue Shield Plans and covered nearly 90 percent of the payments made by these organizations under their basic contracts. Questionnaires were prepared from selected claims paid during the six-week period ending June 26th, 1964 and they served as the basis for the study. Four hundred seventy thousand forms were mailed to physicians and over 366,000 usable returns (78 percent) had been received by the middle of October. Before referring to the methodological problems which are the subject of this report, we will present a brief outline of the steps which were followed in implementing the project.

Preliminary work was undertaken by the staff at the National Association of Blue Shield Plans. This consisted of the preparation of a prospectus, design and printing of a one-page questionnaire, pilot-testing the survey in two states, and evaluation of the findings. In addition, the staff prepared a detailed Manual, including sample design and selection, which indicated the procedures that all Plans should follow. Finally, the staff conducted regional orientation meetings in five cities to ensure that uniform and correct methods were used by Plan personnel throughout the nation.

The various Blue Shield Plans then performed the following functions: (1) Mailed an orientation letter to all physicians in their areas about ten days before sending the first questionnaires. (2) Selected claims from surgical, maternity, anesthesia, and medical services according to random sampling techniques. Small Plans, generally those with an enrollment under 240,000, participated on a census basis. Plans whose enrollment exceeded this figure selected claims using the terminal digit of the claim control number as specified by a table of random numbers. (3) Prepared questionnaires by inserting information at the top of each form which furnished the doctor with the name of the patient, date, type of service, procedure code, description, and amount paid by Blue Shield for the given service. (4) Prepared an 80-column punch card for each selected claim. This included 22 items such as age, sex-relationship, county and type of physician, amount paid, type of service and so on. (5) Mailed the forms, as imprinted with the identifying information, at intervals of either two or three weeks. As a result of this accumulation process, some doctors received 20 or more questionnaires in a single mailing. This problem arose because a large proportion of their business involved patients with Blue Shield contracts, and because of the distribution of claims by size.

All of the punched-cards, and except for four Plans, all of the questionnaires were sent directly to the trade association headquarters in Chicago. (The four Plans subsequently shipped their returned forms to Chicago en mass.) At this point, the staff at NABSP assumed responsibility for editing the forms (4500 man-hours) and having the information keypunched. These cards, and those submitted by the Plans, were placed on magnetic tape with admissable-code controls specified for each field. The final step included matching the basic data cards with their corresponding informational cards and calculating the weights. From the resulting 115-column records, numerous reports showing Blue Shield's performance have been and will be prepared.

The above overview should provide sufficient background to enable us to proceed with a discussion of selected methodological problems encountered during the survey.

Variates Used to Measure Plan Performance

A measure of Plan "performance" was required that met a number of minimum requirements. These were deemed to be (a) effective differentiation between different levels of performance, (b) application to different types of certificates and services within Plans and between Plans, (c) capable of calculation in terms of operational constraints, (d) simple enough to be understood by a variety of interested parties. Performance, therefore, was subsumed under a single ratio which was obtained by dividing the Blue Shield payment for a given service by the cost of the doctor's care for the same service.

Such a measure is satisfactory if all other things are equal. Yet the survey was conducted in the real world and the variates were influenced by a number of factors. Among these were differences in fee schedules, benefit levels, supply and type of physicians, number of hospital beds per community, subscriber incomes, proportion of premium dollar available for benefit payments, age and sex distributions, proportion of premium contributed by the employer, frequency with which patients in Service Plan areas visited Participating physicians, and so forth. In undertaking the survey, there was no practical way to control for such items. Research will be conducted subsequently, however, to assess the impact of a number of variables on the performance results which were obtained.

The Statistical Unit

Selection of the statistical unit seemed to be confined to three choices - the claim, the illness or the patient. Focusing on either of the latter two would have required access to a number of related records. From an administrative point of view, there was no economical and quick way to pull all the claims together for a particular illness (assuming terminal points could have been defined) or patient. Moreover, some of the financial obligations of the patient might not have been a Blue Shield responsibility. Nor was there any way to overcome the fact that certain claims pertaining to a case may have required a number of months to process. Further, the time lag with which doctors (or a given doctor) submit claims to Blue Shield varies considerably.

The above reasons, when combined with the difficulties inherent in a short time period of six weeks, militated against using either the "illness" or "patient" as the unit. Consequently, it was concluded that the least-worst approach to Plan performance was by means of the "claim."

Sampling Design

The prime consideration in the determination of sample size was to obtain results which would be useful in the operations of individual Blue Shield Plans. The precision of national averages was not critical; nor was the precision of Plan averages. Rather, each Plan had to be considered as being partitioned into cells, and it was the precision of each individual cell which was the critical factor. Thus, there was not one overall national or Plan sample but a large number of samples. The exact number was dependent upon the extent and nature of the stratification used for sampling purposes.

One extreme of the possible sampling designs was to have sampling strata and analytical cells in one-to-one correspondence; the other was to take a single sample for a Plan large enough to ensure satisfactory precision for cells within the Plan. Actually, an intermediate design was used.

As indicated above, the design incorporated stratification. But, instead of setting the sample size to provide the required precision for a stratum as a whole, it was increased by an amount estimated to be sufficient to ensure reasonable proportionate representation for the smaller, but sizeable sub-cells within the stratum. In effect, this yielded more than the required precision for the sampling stratum as a whole, and about the required precision for the larger sub-cells within the stratum. Moreover, it ensured satisfactory precision for many averages of marginal distributions of variables not used for the sampling stratification.

The original sampling design indicated a potential of some 2500 different samples - the exact number being unknown because provision was made for Plans to sub-stratify if necessary for operating needs and because the number of null cells was not known. A minimum sample size of 225 was arbitrarily set for small sized populations (about 300 or less). The theoretical maximum sample size for a sampling stratum was computed at about 700, though in certain instances larger samples did actually occur.

Stratification and The Sampling Unit

Within Plan stratification was clearly necessary because of widely different groups and widely different sizes of the groups. It was not possible, however, to achieve the optimum benefits of stratification. To do so would have required an identity between the variables of classification and the variables of stratification. As noted above, there was a potentially large number of variables of classification; at the same time, though, administrative consideration limited the variables of stratification to a very few.

Stratification was based upon the contractual relation between the subscriber and the Plan (with one exception) and upon the generic type of medical service rendered by the doctor. Thus, one variable of stratification was the Basic Certificate which is the major means of differentiating between a great variety of contractual relations; the other was the Type of Service (surgery, anesthesia, etc.) which is the major means of differentiating between a large number of diverse services rendered by physicians. These two had the desirable aspects of being related (or expected to be related to) the performance ratio and of being easily identified and readily available in Plan records. (Some variables - doctor's charge, doctor's specialty - were not generally available at the time samples were selected.)

One feature of the Basic Certificate variable should be noted. It was not the same from Plan to Plan. In order to satisfy the condition of usefulness in Plan operations, it was defined with respect to the Plan rather than uniformly for all Plans. Thus, "Best Certificate" and "Most-Widely-Held Certificate" categories differed from Plan to Plan. But doing so, though, ensured comparisons of the most meaningful kind. The Type of Service variable was uniform over all Plans.

Stratification by size of claim (a substitute for cost of physician care which was not available beforehand) was considered because of the importance of total dollar cost. In addition, the pilot survey had indicated some variation of performance with differences in cost. Further, the proportion of large claims tends to be small; moreover, there was the question of measuring performance on a "per dollar of cost" basis instead of, or as well as, on a "per claim" basis. Obviously, the precision of the regression of "payment per dollar of cost" on "actual cost" and the precision of the "relative frequencies of dollars or claims" by size of claim would have increased without significant change in the sample size. It was not administratively feasible, however, to stratify further than by Certificate and Type of Service.

A suggestion made after survey operations were under way may be of interest. Professor Nathan Keyfitz suggested that using dollars rather than claims as the sampling unit might have been a better solution. Such a procedure does not necessarily gain the advantages of stratification, and perhaps gain them with smaller samples, without actually stratifying.

Random sampling of dollars, of course, gives claim-selection probabilities proportional to claim-size in place of equal probabilities when sampling is by claim. This is an inherent rigidity not present in stratification. In addition, the number of times a claim is selected becomes a random variable; the comparable item under stratification is a combination of the sampling ratio and the size of the claim - neither being subject to sampling errors. Finally, it may be noted that observations for claims are independent; those for dollars from the same claim are not; hence, it is the number of claims, rather than the number of dollars, which determines the precision of the sample statistics. And to set the sample size, therefore, requires an error element - average size of claim - to be introduced which is not present in stratification.

Systematic sampling of dollars from randomly ordered claims (the assumption used in the survey) would lessen the relative shortcomings of sampling dollars. Let R equal sampling ratio and X equal claim size, then every claim for which $X \geqq (1/R)$ would be selected at least once. Other claims would essentially be sampled at random.

It appears, therefore, that random sampling of dollars would have advantages only if (a) it was appropriate to have claim-selection probabilities proportionate to claim size, and (b) frequencies of large claims - say those greater than 1/R - were relatively small.

From the administrative viewpoint, both stratification by size and sampling by dollars would have required large increases in cost. Stratification would have involved them at the point of sample selection, primarily, but also later, though it would have reduced the total cases in the survey. Sampling of dollars would have necessitated sorting and collating of "dollars" selected in order to reduce the data to unduplicated claim records which included a count of the number of times selected.

Sample Size: Specified Precision

The initial precision specification for the survey was in terms of the average performance per claim (ratio of the Blue Shield payment to the actual charge made by the doctor) for a sampling stratum. A 95 percent confidence interval of plus or minus 0.05 was specified when the average was considered as a measurement of the true average performance for the actual finite population (and not for a hypothetical infinite population).

The upper limits for sample size to achieve this precision on the assumption of 100 percent response were as indicated below. Such a routine application of the finite population sampling formula does not, of course, take account of various deviations from the assumption underlying such an application.

Assumed Probability Distribution	V(X/Y)	Upper Limit For Sample Size
$P(X/Y = 1) = P(X/Y = 1) = \frac{1}{2}$	$\frac{1}{4}$	400
$P(X/Y) = d(X/Y)$	1/12	135
$B(3/2, 3/2)$	1/16	100

Sample Size: Response Rate

In this survey, expected non-response was the easiest element for which to adjust sample size. Assuming non-responses were at random, division of the theoretical sample size by the expected response rate provided the necessary adjustment. The estimate of an 80 percent response did, in fact, turn out to be of the same order of magnitude as the actual response rate which was achieved. It reflected the results of the pilot survey for which a higher estimate was used (and again achieved).

While the validity of the assumption of randomness of non-response cannot be fully determined, the survey will provide at least some significant indications. These will be derived from comparisons of characteristics of the non-response and the response parts of each sample. (As noted above, a punch-card exists for each non-response. It contains 22 variables, many of which will be invaluable for analytical purposes.) It may also be possible to derive additional indications from between-sample comparisons.

The possibility of biases resulting from differences between the response and non-response parts of a sample was recognized, and provision for offset was made through an increase in the variance estimate. Thus, the precision specification was not applied to the sampling error itself, but rather applied as an asymetrical range for the sampling error plus an estimate of bias.

Sample Size: Estimates of Population Size

Population sizes were not known, but had to be estimated beforehand. These estimates were subject to large errors, because the number of claims submitted or approved during relatively short periods is subject to wide fluctuations. Moreover, the varying ability of Plan personnel also contributed to the estimating problem.

To prevent serious errors in sample sizes arising from this source, two steps were taken. First, in making the estimate of population variance, this element was given consideration; such consideration was designed to offset only the smaller errors resulting from errors in population estimates. Second, and much more important when needed, was provision for modification of the sampling rate during the course of the survey, mailing period by mailing period. Only very marked deviations were covered by this provision; and, when used, each mailing period became a sampling sub-stratum within the original sampling stratum (with independently computed weights in the subsequent tabulations). Fortunately, only a few occasions developed in which sampling rates were modified during the course of the survey.

Sample Size: Non-Homogeneity of Sampling Frame

It was not possible to establish a sampling frame which included only the population to be surveyed. Subsequently, however, it was possible to delete claims which did not belong to the proper population. Two considerations were involved in the provisions made to offset this condition.

First, population size estimates were discounted by the expected proportion of deletions. This by itself could be relied upon to provide the sample sizes necessary to meet the individual sampling stratum precision requirement. However, there was the further consideration that the sample then provided the only estimate of the population size. This second consideration, under conditions where deletions would be substantial, was the more important one.

A calculation made on the assumption of deletions running about 15 percent indicated some 200 cases would be necessary to give a 95 percent confidence interval of 5 percent for deletions, in contrast to some 100 cases in the absence of deletions. Further, reduction of the size of the confidence interval could only be achieved by relatively large increases in sample size.

Sample Size: Tabulating Cells Different From Sampling Strata

Precision for tabulating cells which were subclassifications of a sampling stratum would, of course, be less than for the sampling stratum itself (except possibly under very unusual conditions). No direct provision was made on this account, because such large increases of sample size would have been required. Thus, assuming a sub-cell to be 1/10th of the sampling stratum, and illustrative calculation indicated that an increase from around 100 to 1200 would be necessary to provide the specified precision for this size sub-cell. It may be noted that the number in the sub-cell sampling, under these conditions, is subject to sampling error as well as the averages.

Precision for tabulating cells which cut across sampling strata might, of course, be greater or less than for the sampling strata, depending upon the size of such cells and other factors. Again, no direct provision was made on this account because of uncertainty with regard to the terms of the problem. Instead, reliance was placed upon the provision made for deletions because of non-homogeneity in the sampling frame.

Sample Size: Final Estimation

Based in part upon judgment, the final formula used in actual calculations was:

$$1/n = [(1/N) + (3/1600)](4/5)$$

where n is the size of the sample and N is the size of the finite population. The "4/5" factor represents a discounting for reasons of non-response; and the "3/1600," the ratio of

$V(\overline{X/Y})$ = Estimated variance of sample average performance necessary to achieve the precision specified

to

$V(X/Y)$ = Estimated variance of the parent finite population

Calculations based upon this formula are shown in Table 1.

Adaptation to actual operating conditions required one further modification; i.e., the population sizes to which the sampling ratios were to be applied were rounded up to round figures and the sampling ratio was applied to the minimum for the range. The final figures are reported in Table 1. This adaptation added an extra safety factor, of varying importance from place to place, on the population-size scale.

The above procedure wherein the sampling ratio was based on the lower limit of the expected number of claims in a given class interval tended to inflate the size of the sample. To take hypothetical illustration, the required n where the number of claims fell between 35,000 but under 70,000 was equal to 700. If, however, 50,000 claims occurred, 1,000 sample observations would have been generated by applying the 2 percent rate as called for in the table used by Plan personnel in drawing the sample. Because the last two digits of the claim number were used to select sample cases in the larger cells, (like the one in our example) there was no simple way to write a program which incorporated a selection rate of 1.4 percent which was necessary to yield the desired 700 claims. Thus the survey specified a 2 percent rate and thereby inflated the n in our illustrative cell by about 43 percent.

Table 1

Sampling Ratio	Calculated		Modified	
	N	n	N	n
0.01	66,000	660	70,000 & over	700-
0.02	32,667	653	35,000 but under 70,000	700-1400
0.03	21,556	647	24,000 but under 35,000	720-1050
0.04	16,000	640	18,000 but under 24,000	720-960
0.05	12,667	633	14,000 but under 18,000	700-900
0.06	10,444	627	12,000 but under 14,000	720-824
0.07	8,857	620	10,000 but under 12,000	700-720
0.08	7,667	613	8,000 but under 10,000	640-800
0.09	6,741	607	7,000 but under 8,000	630-720
0.10	6,000	600	6,000 but under 7,000	600-700
0.15*	3,773	566	4,000 but under 6,000	600-900
0.20	2,667	533	3,000 but under 4,000	600-800
0.25*	2,000	500	2,000 but under 3,000	500-750
0.30	1,556	467	1,500 but under 2,000	450-600
0.40	1,000	400	1,000 but under 1,500	400-600
0.50	667	333	700 but under 1,000	350-500
0.60	444	267	500 but under 700	300-420
0.70	286	200	400 but under 500	280-350
0.80	168	133	300 but under 400	224-320
0.90	74	67	275 but under 300	248-270
1.00	1-73	1-73	Under 275	N

* These optional steps were included so that the Plan, if it chose to do so, might limit the number of claims drawn in the survey. The disadvantage from an operational point of view was that these cells, like those with more than 7,000 claims, required that selection be based upon the last two digits of the claim number rather than the terminal digit as indicated for all other conditions.

Because the number of cells where N exceeded 3,000 claims was not great, and because there was no *a priori* reason to indicate that, on the average, the actual N would tend to exceed the midpoint of the class interval, it is estimated that the inflationary factor increased the over-all sample size by, at best, 10 percent. This was construed to be a salutary feature since precision would be increased and administrative considerations suggested no better alternative.

Sampling Variance and Teleology

The survey provides an illustration of situations in which there is not necessarily a unique sampling variance. Thus, depending upon the definition of the parent population or, stated otherwise, upon the use of the statistic involved, the sampling variance may be upon the basis of a finite population or upon the basis of an infinite population of one specification or another.

The observations actually obtained were "random samples from *a finite* population." Considering the sample $\overline{X/Y}$ as a measure of the true average performance of that finite population, the sampling variance is appropriately computed according to the usual finite population formula. However, the sample may also be considered as the result of a two-stage sampling process so that it is selected, not from a population, but from a first-stage sample generated under given conditions from an infinite population. In this latter case, assuming random generation of the sample, the sampling variance is appropriately computed according to the usual infinite population formula, so that it provides a measure of error when the sample results are used as measures of the hypothetical infinite population.

Finally, it may be noted, sample results may also be considered in other ways; for example, there may be an absence of randomness in the first-stage sample, even though the second-stage sample is taken at random; or, again, the hypothetical population for which random sampling may be assumed appropriate is not the hypothetical population with reference to which the sample results may be used, thus raising the question of measuring biases. In such cases as the latter ones, the usual formulae have, of course, to be modified in terms of special considerations.

In the survey under discussion, sampling variances have been computed both on the finite and infinite population basis using the assumption of random selection. It is recognized that, in transferring from the finite to the infinite population assumption, more than a formula change is ncessary. Such items as differences in relevant conditions over time, including fee schedule adjustments, benefit levels, medical risk characteristics, months of the year, and so forth, must be considered before the extent of sampling and other errors may be judged.

Validation By Means of a Patient Survey

Formal validation of the survey was attempted to obtain a general indication of the reliability of the results. Some of the factors that might have contributed to respondent bias in the project were as follows: Many physicians have a heavy patient load and the doctor, in his haste to complete the form(s),[1] might have inadvertently supplied incorrect information. A secondary source, such as a nurse, could have been responsible for completing all or part of the questionnaire. Moreover, recorded data, either in the doctor's files or furnished on the form, could have led to an erroneous response. Finally, deliberately biased answers could not be discounted since one of the purposes of the survey was to test whether Blue Shield fee schedules were realistic in terms of contemporary costs of physician care.

Consequently, in order to indicate the magnitude of any such bias, the Plans were requested to draw a second sample for the purpose of determining which patients would be contacted. This was done by taking a systematic sample from those claims previously drawn in the doctor survey. One of the pilot-study Plans selected claims at the rate of 1/11 while most of the others used a rate of 1/20. Theoretically, the former seemed more desirable, but from an administrative point of view, the taking of only one out of every twenty was close to the maximum that could be achieved under existing conditions.[2]

The Blue Shield Plans, therefore, prepared a questionnaire similar to the one sent to the physician which included the appropriate identifying data needed by the patient. These were mailed about 45 days after the last doctor forms had been forwarded. This timing was specified so that the doctor would have time to bill the patient, if contemplated, and still minimize the possibility that patients would misplace their health care records, move away, expire, etc. Some of the Plans sent a second, duplicate questionnaire to those patients who had not replied within 40 days (as indicated by NABSP records in Chicago). No follow-up, however, was undertaken in the physician survey.

It is not anticipated that every returned patient questionnaire will confirm the cost information on the similar form submitted by the physician. Among other reasons, many patients may find it difficult to isolate the cost of the procedure covered in the survey from related expenses for the same illness. It is expected, nevertheless, that a sizeable majority will confirm the results obtained in the primary survey, that the differences will be randomly distributed and, to a large extent, offsetting.

Summary

This paper has outlined some of the methodological questions that arose during a large, national survey of physicians. The discussion dealt with statistical techniques that were inherent in such a project and attempted to focus on the relationship between theoretical and operational considerations which formed the basis of the survey. In none of the sections was it intended that the treatment should be characterized as "exhaustive." It is hoped, however, that the issues covered will be of interest to some who labor in the field of applied statistics.

---

[1] Although the distribution of doctors according to the number of questionnaires received, in total and by mailing period, is not yet known, it is true that a large number of physicians did get more than ten forms. A study of the propensity to response under such conditions should be of great interest.

[2] Substantial resources, financial and personnel, already had been committed by the Plans to this project. Moreover, professional relations considerations militated against a large survey of patients.

II

SAMPLING PROBLEMS IN SOCIAL SURVEYS - I: COMPARATIVE COST ANALYSIS OF SURVEY OPERATIONS

Chairman, Frederick F. Stephan, Princeton University

Page

An Experimental Attempt to Reduce Field Costs by Limiting Callbacks and Increasing Cluster Size - Morris Axelrod, Combined Jewish Philanthropies.................................. 32

Remarks on Dr. Axelrod's Paper - W. Edwards Deming... 40a

Time Allocation of Survey Interviewers and Other Field Occupations - Seymour Sudman, National Opinion Research Center......................... 41

AN EXPERIMENTAL ATTEMPT TO REDUCE FIELD COSTS BY LIMITING CALLBACKS AND INCREASING CLUSTER SIZE

Morris Axelrod, Combined Jewish Philanthropies

PURPOSE

This study attempts to test some procedures which might effectively reduce the cost of interviewing without increasing the sampling error or non-response error beyond acceptable limits. At the outset may I note, however, that any such savings should not be completely offset by an increase in variance, or by the introduction of non-response error.

Interviewing expenditures loom large on most survey research budgets, and so are a prime target for cost-cutting. Field costs often account for a third or more of the total research budget. This is particularly true when the research entails the use of a strict probability model, in which only the designated person within a selected dwelling unit can be interviewed. Much effort and expense goes into getting such an interview because people vary considerably in the amount of time, and the particular hours they spend at home. It is not unusual for an interviewer to make as many as ten calls without being able to contact the designated respondent; thus, any attempt to approach the ideal of achieving a 100% response rate means correspondingly greater costs as the number of callbacks increase. Any lesser percentage, of course, increases the sampling error and may introduce a non-response bias. Actually, a 100% rate is rarely approached and need not be. Cost and other considerations dictate that we accept less than the perfect rate, and we must take this into account in determining a tolerable level of sampling error. When we go beyond a certain point, however, the marginal cost of additional interviews becomes prohibitively high and the percentage gain in response rate is out of proportion to the reduction in variance. For a sample of 2, 000, an increase in response rate from 85% to 86% will reduce the variance by only a small fraction of a percentage point yet these twenty or so interviews will be tremendously costly ones.

Even if we assume optimum performance by the interviewer, there are still two practical means by which we may reduce costs and increase the efficiency of contacting households. One is to increase the number of addresses which are clustered together so that an interviewer's travel time is reduced. Another is to limit the number of callbacks an interviewer is permitted to make at a single address. (Presumably there is always a limitation imposed by calendar or money which translates into fewer calls - either selectively or systematically limited.)

CLUSTER SIZE AND EFFICIENCY

In a typical cross-section sample survey, selected addresses are clustered so that the interviewer can contact several on one trip. This increases the probability of his getting an interview on any single trip. But the more easily obtainable interviews tend to be picked off early. This means that clustering diminishes rapidly as more and more callbacks are made. Subsequent trips must then be made to individual addresses rather than to clusters of addresses. Although increasing the cluster size tends to reduce travel and interviewing costs, it also increases the sampling error and may thus negate or vitiate the saving in cost.

NUMBER OF CALLBACKS AND EFFICIENCY

Obviously a reduction in the number of callbacks will reduce per/interview field costs, since the marginal cost of each callback beyond three tends to be high. On the other hand, however, fewer calls will result in a lower response rate. This lower rate can lead to two undesirable consequences. First, insofar as the reduction in calls excludes persons who are least likely to be at home -- the single, the married without children, the more socially active -- it becomes selective and thus introduces non-response error. Second, it has a damaging effect on the variance. The smaller the number of interviews obtained, the larger the variance.

THE DESIGN OF THE EXPERIMENT

It was felt that the research design should try to capitalize on the effect of increased cluster size as well as the effect of reduced callbacks in combination, rather than to study the two independently. We made this decision for two reasons: one - the relatively small number of sample areas; and two - the normal variability in average cost per interview among interviewers and among Primary Sampling Units. [1] Therefore, two situations were set up as follows:

(1) A control group of PSUs with the "usual" cluster size and the "usual" number of callbacks.

(2) An experimental group of PSUs with twice the usual cluster size and a maximum of three calls to be made at an address.

The Survey Research Center's national sample, which at the time consisted of 66 PSUs, was used. This sample was divided into two halves, and was stratified by size of the PSU and the average cost per interview on previous studies. Each procedure was randomly assigned to the two halves. The four largest metropolitan areas were exceptions. There each PSU was split into matched halves, and both procedures were carried out in each half.

The decision to use no more than twice the usual cluster size was based primarily on the desire to minimize the effect on variance and at the same time to reduce the number of clusters to be visited. A second reason stemmed from the suspected "contamination" effect, which may result from preliminary and uncontrolled discussion between respondents who are neighbors. This effect shows up in two ways. It may sensitize the respondent to the subject matter and thus alter his responses. It may also crystallize his incipient resistance to the interview and increase the risk of getting a refusal.

A maximum of three calls was decided upon because, in a standard study, roughly 85% of all completed interviews are usually obtained by the third call, and the marginal costs are presumed to rise rapidly with the fourth and subsequent calls.

By way of review, then, the first set of PSUs was the control group in which the usual procedure was followed. The University of Michigan's Survey Research Center generally selects clusters of approximately four addresses and seeks a prescribed minimum response rate of 85% with no limit on callbacks, usually achieved by having the interviewer return to <u>all</u> unresolved addresses.

The second set of PSUs was the experimental group. The average cluster size was about eight addresses which is twice the usual size. The interviewer was instructed to make a maximum of three calls at each address which had not been resolved as an interview or a terminal non-interview on the first or second call. Any visit to the dwelling unit -- whether or not the visit required a special trip -- was considered to be a call.

The experiment was carried out in the spring of 1961. The vehicle was a national study concerned mainly with economic matters.

THE RESULTS

EFFECT ON INTERVIEWING COSTS

As indicated earlier, cost reduction was the main concern. If nothing was saved there would be no need to study further the effects of the experimental variables on variance and non-response error.

The experimental group showed an average savings of one hour per interview, as compared with the control group (Table 1). This represents a savings of about 12% in total interviewing costs.

The size of the savings was found to vary with degree of urbanization. The PSUs were divided into three population categories: (1) large metropolitan areas, (2) other PSUs containing an urban place of at least 50,000,

---

[1] Hereafter referred to as PSUs

and (3) all other PSUs - small towns and rural areas. These will be referred to as "high urbanization", "moderate urbanization", and "low urbanization" areas respectively.

High urbanization areas showed the largest gain -- about 1 3/4 hours per interview. Moderate urbanization areas showed a negligible gain -- just a few minutes per interview -- and low urbanization areas showed a gain of about one hour per interview.

The sizeable gain seems reasonable for high urbanization areas, where, on the average, more calls per interview are necessary to produce an 85% response rate.

As you can see from Table 2, only 65% of the interviews ultimately obtained in these areas are obtained on the first three calls, as compared with 79% for moderate and 88% for low urbanization areas. Thus, any reduction in the number of calls required should have its greatest impact on the high urbanization areas, where such a reduction tends to eliminate the marginal and most costly calls.

The failure to show a large gain for the moderate urbanization areas suggests that the marginal cost of calls may not rise significantly for those areas after the first three. The small effect of clustering in these areas compared with the other two can probably be explained by their differing degrees of "friction of space".

Movement in high urbanization areas is hampered by dense traffic, narrow streets, pedestrian traffic, and such regulatory devices as signal lights, stop signs, and caution signs. Parking space can be obtained only at considerable cost in time or fees. Distances are so great that the interviewer often resorts to public transportation. Under these conditions, of which New York and Los Angeles afford prime examples, clustering has obvious advantages.

In low urbanization areas, which are relatively sparsely populated and rural, the interviewer is not faced with congestion, but is confronted with larger areas and greater distances. (Partly because of the larger areas and partly because of natural barriers such as mountains, rivers, hills, poor roads, and lack of roads, distances are greater to the cluster and within the cluster itself. Sample addresses will fall off the beaten path, away from the main routes. The unreliability of secondary roads and heightened susceptibility to seasonal hazards - mud, fog, snow and ice, floods, washouts -- compound the difficulties. Here also the advantages of clustering are obvious.)

The moderate urbanization areas, on the other hand, consist more typically of middle-size cities in which there is concentration of population without congestion. Sample addresses tend to fall in or near the central city. The traffic flow is smooth and rapid compared with both the metropolitan and rural areas. Such middle-size cities are easy to get around in; one can usually traverse them in ten or fifteen minutes. Since little time is spent in travel, the marginal cost of additional trips is minimal. Thus clustering is not likely to play as significant a part in reducing trip costs.

This view is supported by Table 3, which shows that travel costs are reduced substantially in high urbanization areas, reduced modestly in the low urbanization areas, and not at all reduced in moderate urbanization areas.

Paradoxically, moderate urbanization areas do not have the cheapest interviews, apparently because the first calls do not produce as many interviews as they do in the rural areas. Each call is less costly, however. In other words, it costs less to reach the cluster in a moderate urbanization PSU, but once the cluster has been reached a call is less likely to yield an interview.

RESPONSE RATE AND NON-RESPONSE ERROR

The reduction in cost is large enough to warrant further examination of the procedure, especially in regard to the consequence for non-response bias. The response rate for the control group was 10% higher than for the experimental group. This poses the question whether the 10% who have been sacrificed for the saving in field costs differs significantly from the total population.

Table 4 compares several attitudes and background characteristics for the control and experimental groups. The experimental group is composed of a higher per cent of persons who are married and have very young children (Table 4h). This is to be expected since their movement is relatively hampered and they are tied more to the home. Not unrelated is the very slight tendency for families with no children under 18 years of age to be excluded from the sample when calls are limited. (Table 4j) The experimental group is composed of a few more married persons than single persons. (Table 4b) This group may be slightly more selective of rural households (Table 4g) but there is no significant difference in sex or race between the two groups (Tables 4e and 4f). When calls are limited, two-adult households yield interviews relatively more frequently at the expense of one-adult households (Table 4a).

The educational distribution is somewhat weighted in favor of more education for the control group (Table 4d). College graduates apparently are more elusive than those with less education.

Despite the fact that a reduced response rate seems to be slightly more selective of certain types -- married persons, persons with children, persons with less education-- the attitudes which were studied, on the whole, do not vary greatly between the two groups. The experimental group may be slightly less well off financially than the control group a year ago (Table 4o) and perhaps slightly less affluent (Table 4s). More persons in the experimental group are pessimistic in their attitude toward the car market while more in the control group are unsure (Table 4k). The experimental group also tends more often to see business conditions as worse now compared with a year ago (Table 4n). However, in general, differences are very small, and are significant only by the application of most rigorous standards.

To the extent that there are differences between the two groups which are deemed to be significant, a correction factor is implicit in the control group. The two groups can be combined and all interviews obtained during the fourth or later call can be given a weight of two.

CONCLUSION

In general, a limitation on the number of calls, in combination with larger clusters, has effectively reduced the average cost of each interview while introducing what appears to be relatively small non-response error. The results are promising; however, further analysis of the differential effects for subgroups within our population and of the separate effects of varying the cluster size and limiting the number of callbacks is indicated. Equally important is analysis of the impact on variance.

Limiting the number of calls should be avoided whenever the subject matter under study is correlated with socio-economic characteristics which are underrepresented in the interviews completed on earlier calls.

We chose to experiment with a limitation on the number of calls on this study because it is one in a series of similar continuing studies for which we have information about the relationships which do exist among the variables being studied.

Let me summarize briefly. We clearly cannot proceed to limit the number of callbacks on a general basis. There are risks entailed where there is little previous experience with the relationship.

No gross limitation on the number of calls should be attempted unless it is preceded by clear experimental evidence that the biases introduced are minimal and the savings in cost are significant.

TABLE 1

Difference in Interviewing Cost Per Interview Between Sample Areas Which Differ in Number of Calls Made and Clustering Size

Type of PSU	Difference in Cost Expressed in Minutes (Control minus experimental)
High Urbanization	103
Moderate Urbanization	7
Low Urbanization	65
All Places	60

TABLE 2 [a/]

Of all Interviews Taken, Per Cent Completed on First Three Calls, by Degree of Urbanization

Type of PSU	
High Urbanization	65%
Moderate Urbanization	79%
Low Urbanization	88%

a/ Source: Two large-scale typical national surveys which preceded this experiment.

TABLE 3

Difference in Travel Costs Per Interview Between Sample Areas Which Differ in Number of Calls Made and Cluster Size

Type of PSU	Difference in Cost Expressed in Minutes (Control minus experimental)
High Urbanization	49
Moderate Urbanization	-4
Low Urbanization	20
All Places	20

TABLE 4

Experimental and Control Group by Selected Characteristics

	PERCENTAGE DISTRIBUTION	
a. NUMBER OF ADULTS IN FAMILY	Control	Experimental
One	21	17
Two	63	68
Three	12	12
Four or more	4	3
Not ascertained		*
	100%	100%
b. MARITAL STATUS OF RESPONDENT		
Single	9	6
Married	74	77
Divorced	2	3
Widowed	13	11
Separated	2	3
	100%	100%
c. HOME OWNERSHIP		
Owns home	65	65
Pays rent	30	32
Other	5	3
Not ascertained	*	
	100%	100%
d. EDUCATION OF HEAD OF FAMILY		
Grade School	30	35
Some High School	17	18
Completed High School	23	24
Some College	13	10
Completed College	15	11
Not ascertained	2	2
	100%	100%
e. SEX OF RESPONDENT		
Male	45	44
Female	55	56
Not ascertained		*
	100%	100%

* Less than one-half of 1%

(Table 4 - continued)

	Control	Experimental
f. RACE		
White	86	87
Negro	10	9
Other	1	1
Not ascertained	3	3
	100%	100%
g. SIZE OF PLACE		
Metropolitan areas	12	11
Cities with 50,000 and over	17	16
Cities with population 2,500 - 49,999	30	29
Under 2,500	41	44
	100%	100%
h. LIFE CYCLE		
Under 45 and Single	5	5
Under 45, married with no children	5	5
Married, with children under 5	19	28
Married, with children 5 to 15	21	16
Married, with children 15 to 18	5	4
45 and over, married with no children	23	23
45 and over, and never married	16	13
Other (divorced, widowed, separated)	5	6
Not ascertained	1	1
	100%	100%
i. AGE OF HEAD OF FAMILY		
18 - 34	30	23
35 - 49	30	33
50 years or older	40	43
Not ascertained		1
	100%	100%
j. NUMBER OF CHILDREN UNDER 18 YEARS		
None	49	46
One	17	17
Two or three	26	28
Four or more	7	8
Not ascertained	1	1
	100%	100%

* Less than one-half of 1%

(Table 4 - continued)

k. ATTITUDE TOWARD CAR MARKET	Control	Experimental
Good	43	45
Pro-con	10	12
Bad	12	16
Depends	1	*
Don't know	32	26
Not ascertained	2	1
	100%	100%
l. EXPECTATIONS OF BUSINESS CONDITIONS A YEAR FROM NOW		
Better	49	47
About the same	38	42
Worse	6	4
Don't know or depends	6	6
Not ascertained	1	1
	100%	100%
m. EXPECTATIONS OF BUSINESS CONDITIONS DURING NEXT TWELVE MONTHS		
Good times	41	40
Good, with qualifications	21	19
Pro-con	8	8
Bad, with qualifications	5	4
Bad	8	10
Don't know, uncertain	15	18
Not ascertained	2	1
	100%	100%
n. BUSINESS CONDITIONS NOW COMPARED WITH A YEAR AGO		
Better	30	25
About the same	35	35
Worse	32	36
Don't know, or depends	3	3
Not ascertained	*	1
	100%	100%
o. FAMILY'S FINANCIAL SITUATION NOW COMPARED WITH A YEAR AGO		
Better	30	29
Same, pro-con	48	43
Worse	21	27
Uncertain	1	1
Not ascertained	*	*
	100%	100%

* Less than one-half of 1%

(Table 4 - continued)

p. WHETHER GOOD TIME TO BUY LARGE HOUSEHOLD ITEMS	Control	Experimental
Good	46	45
Pro-con	13	16
Bad	16	19
Uncertain	22	17
Not ascertained	3	3
	100%	100%
q. EXPECTATIONS REGARDING FAMILY'S FINANCIAL SITUATION		
Better off	38	38
Same	45	45
Worse off	6	6
Uncertain	11	11
Not ascertained	*	*
	100%	100%
r. WHETHER PLAN TO BUY NEW CAR OR USED CAR		
New	11	8
Used	8	9
Uncertain	1	1
Not ascertained	*	1
Inapplicable	80	81
	100%	100%
s. ANNUAL FAMILY INCOME		
Less than $3000	24	27
$3000	20	21
$5000	24	26
$7500	12	11
$10,000 or more	15	12
Not ascertained	5	3
	100%	100%
Number of cases	703	577

* Less than one-half of 1%

REMARKS ON DR. AXELROD'S PAPER

W. Edwards Deming

Consideration of costs is basic to the statistician in sample-design. Cost-functions enter formulas for optimum allocation of interviewers, optimum number of call-backs, as well as formulas for optimum size of sampling unit, number of secondary sampling units per primary unit, and optimum probabilities of selection. Any paper that promises reduction in costs is therefore of interest.

The problem of how many call-backs to make raises first of all the question of just what is a call-back? Actually, the word can mean almost anything. It can mean productive call-backs, or it can mean call-backs made at the convenience of the interviewer. Some organizations pay for call-backs but do not get them. Call-backs require rules on when and how to call back, and they require supervision: otherwise costs soar, without improvement of quality of data.

Astonishing differences show up between interviewers when one lays out the work so that comparison of interviewers is possible. Interpenetrating samples offer the possibility of detecting faults and need of retraining or need of better supervision. One may allot two interviewers to random parts of two subsamples, and calculate from the results the variance between and within interviewers.

Significant differences between interviewers, in the number and type of responses obtained, will indicate need of retraning, and need of different rules for calling back. They may also show that the supervision is not doing its job.

To show how astounding such comparisons may be, I may cite from my own experience the fact that some interviewers have found significantly more people at home at first call than other interviewers found in the same area. That some interviewers will find more people 65 and over, or more young children, is of course well known, as well as the fact that some interviewers have significantly more or fewer successes, and significantly more or fewer refusals.

One system that increases the effectiveness of call-backs is to allow the first interviewer only 3 call-backs. If the 3d trial is unsuccessful, she turns the case back to the supervisor, who allots the incompleted case to a special interviewer. Interviewers are careful not to build up too many failures, and in consequence they use judgment about when to call. (The number 3 is not important: the principle of a limited number is.)

Nothing is so hazardous as an experiment: it is too easy to draw wrong conclusions. I am glad that Dr. Axelrod is cautious and does not draw conclusions from his experiment. Actually, if he would continue the experiment, and increase the number of call-backs, he would, I believe, come to the conclusion that 5 or 6 call-backs are desirable if one wishes to maximize the amount of information per unit cost.

One can always reduce the cost on anything, by reducing quality. One can cut out call-backs entirely and save money, but what does it do to the quality?

I have known of research organizations that claim to do probability sampling, but which in actuality ask an interviewer to start at a designated point and to ring door-bells until she has completed a quota of interviews. This system requires no call-backs, but it has 3 faults: (1) It is not probability sampling, as there is no way to ascertain anyone's probability of selection. (2) What this system does is to load the sample with people that stay at home. (3) It fails to utilize the investment already placed in the survey to locate the random starting-points.

Miss Sybil Carof has been doing some work with me at Young & Rubicam on call-backs. We find that tabulations on general attitudes, and even on important demographic characteristics, are about the same with 3 attempts as with 6. I think that people forget that one does not carry out a survey to acquire information that you can buy from the Census for $2.50. One carried out a survey to learn about specific characteristics that are not in the Census.

Interviews			Women working outside the home	
Call-back	Number	Cumulated	Number	Cumulated
0	520	520	86	86
1	334	854	116	202
2	218	1072	101	303
3	87	1159	50	353
4	36	1195	22	375
5	24	1219	11	386

For example, take female home-makers that work outside the home for pay.

-- Initial call 17% female home-makers
-- After 2 call-backs, 28%
-- After 5 call-backs, 32%

The nearest Census figure on this characteristic is 34% for women in households that work outside the home for pay.

One can see from this table how the results change with call-backs. Any marketing policy aimed at households with income from working women, or aimed at working women, might go off the track if based on a survey that did not carry call-backs through the 5th or 6th recall.

Here is another example. How often do you use coconut?

-- Initial call, 1 to 3 times per month, 41%
-- After 2 call-backs, times per month, 32%
-- After 5 call-backs, times per month, 29%

Another point. Any interview, be it the initial call, or a call-back, costs money. With proper rules and supervision, the interview obtained on a call-back need be but little more expensive than the initial call. I question Dr. Axelrod's statement that the marginal cost of call-backs increases considerably. I am not sure that it need be this way.

Moreover, I would return to my statement that call-backs should be directed, not at random times; and that they should actually be made, not just billed.

I remarked a while ago that an experiment can be a hazard. It is better, I believe, to depend so far as possible on theory. Theory, if we have the right theory, gives the answers as well as the limitations on the range of validity of the answers that the theory provides.

For example, some years ago, I showed by theory and arithmetic that the amount of information per dollar increases steadily through the 6th call-back for any characteristic that varied from 50 to 100, or from 100 to 50, between people hard to find at home, and people that stay at home most of the time. The computations assumed that a call-back cost 67% more than the initial call. Any other realistic costs, and any other reasonable allowance for variability between people seldom at home and people much at home must reinforce these conclusions.

Theory is usually much cheaper than an experiment, and not so hazardous. You can correct a theory: but you can not correct a bungled or incomplete experiment except by repeating it. Moreover, people are more adept at finding faults in theory than in finding mistakes in experimental data.

TIME ALLOCATION OF SURVEY INTERVIEWERS AND OTHER FIELD OCCUPATIONS

Seymour Sudman, National Opinion Research Center

INTRODUCTION

The survey research interviewer and the job which she does should be of particular interest to social scientists. In the first place, the interviewer is the chief collector of the raw data which are used in social analysis. Her work influences both the quality and cost of social research. Hyman and others of the National Opinion Research Center staff have discussed in detail the effects of interviewers on the interviewing situation in Interviewing in Social Research.[1] Cost data, however, have not been generally available, but have become even more necessary as survey costs have risen precipitously over the past two decades, largely due to increases in interviewing costs. Before costs can be reduced, it is necessary to recognize how they originate. This is the first aim of the present paper.

Secondly, the occupational role of the interviewer is of intrinsic interest in itself. Interviewers spend most of their time in the field under very little supervision. In this respect, they are similar to salesmen, social workers, and public health nurses. The pay method for interviewers differs from that of the other occupations, since interviewers are paid on an hourly basis while the others work for either a fixed salary or commission. This might be expected to influence the interviewer's shaping of her job. A comparison of interviewing with these other field occupations is given in the second part of this paper.

The two parts of this paper are not really disjointed. A better understanding of the interviewing role leads to hypotheses about methods for controlling or reducing interviewing costs. These will be discussed in subsequent papers as part of the National Opinion Research Center project which deals with the control and reduction of survey costs.

Methods of Data Collection

Each of the tables to be presented in this paper will be described in detail so that the differences in the methods used will be clear. In general, the results are based on analysis of time sheets submitted either for the special study or routinely. These figures are certainly subject to memory or clerical errors by the person submitting them, and even the likelihood of deliberate distortion should not be overlooked. Nevertheless, they appear to be of sufficient accuracy for the types of analysis which are attempted here.

Somewhat more troublesome is the fact that time records are not always kept the same way, so that some of the differences observed may be artifactual. For example, to anticipate the detailed discussion below, it is often difficult to separate the actual interview from time spent in the home in introduction and in general conversation.

The nature of the task, and of the organization conducting the field work also have an influence on the time allocations. This will be discussed when the tables are analyzed, but a discussion of individual interviewer differences is left to a subsequent paper. Finally, however, it should be noted that even with all the possible reasons for non-comparability, there do appear to be great similarities between survey organizations and between interviewing and other field occupations, and it is these similarities, rather than the differences which are the most significant findings presented.

National Opinion Research Center Interviewer Time Allocation

Table 1 presents the actual time and percentage of interviewer time spent on various tasks for six National Opinion Research Center studies conducted during the period 1958-64, and for an earlier 1947 study. Since the methods used for obtaining these results were generally similar for all studies, they need only be fully described once. The differences which are observed are not due to different methods of data gathering, but to the peculiarities of the particular studies.

The two main sources of information on time spent by National Opinion Research Center interviewers are the questionnaire itself, and the Interviewers' Time Report. At the beginning and end of each questionnaire the interviewer records the time so that length of interview is known. While interviewers are instructed to enter the times concurrently with the interview, there is some indication that there are two possible sources of error in these figures; some interviewers record the times that they enter and leave the house, while some interviewers forget to enter the times during the interview and fill them in by recall when they edit their questionnaires. In both of these cases, the tendency is for the interviewer to overstate the length of the interview by including non-interviewing time in the household such as waiting and post-interview conversation. In addition, this method does not account for any interruptions in the middle of an interview. For interviews which average about an hour or longer, these errors do not appreciably change the percentages shown in Table 1, but for short interviews (such as Census Enumeration) these errors could be large.

The other source of information on interviewer time allocation, the Interviewers' Time Report, is the form routinely used by interviewers to report their time so they can get paid. The Time Report is divided into three sections: Travel, Interviewing and Other Time. "Other Time" includes study, clerical and editing time.

Study time is defined as the time spent by the interviewer in reading the instructions and specifications for a study and in doing any practice interviews required. It does not include any time spent in personal training by a field supervisor or in group sessions. When these personal training sessions are held, the interviewers involved are generally paid a fixed amount.

Clerical time is time spent filling out forms, including the Time Report, and in sending and receiving mail in connection with a study. It includes trips to and from the Post Office to pick up packages or mail completed questionnaires. It may also include the mailing of special letters explaining the purposes of a survey to respondents, if this is done.

Editing time is the time spent by the interviewer after the interview to insure that her writing is legible, that no questions have been erroneously omitted, and that any ambiguous answers are clarified. There is great variability between interviewers on this category, since some interviewers use shorthand during the interview and transcribe later. There is some difficulty in separating out editing time from travel time within a segment. Many interviewers do their editing while waiting for the next respondent to become available, and thus their time sheets show a combined category of editing and waiting. In these cases, the time spent editing a questionnaire is estimated from those questionnaires of the interviewer which were edited when no waiting time was involved.

Travel to Segment is derived from the Travel column on the Time Report. It includes time to the segment from the interviewer's home and return. It also includes any travel time from one segment to another. It is generally not too difficult to separate this time from the time spent by the interviewer within the segment.

Travel in Segment is defined as all time in a segment not spent on the actual interview. Travel in segment includes all waiting time, and time in a respondent's home spent in conversation not part of the interview, as well as time spent locating the proper house in the segment and knocking on doors. Also included here is the time the interviewer spends on the telephone making appointments for interviews. This type of travel time is not always directly noted by interviewers filling out the present time sheet. It is sometimes included under travel time, sometimes under interviewing time and sometimes under other time. In coding the Time Reports, cross-checks are made with questionnaires. If the interviewer combines waiting time or other time within the segment with the interview, the length of the interview as obtained from the questionnaire is subtracted from the total time shown and the balance is called "travel in segment." Even where the interviewer has separated her time, cross-checks still are made to the questionnaire to insure that dates and times agree. If not, the normal procedure is to adjust the Time Report to the questionnaire since times in the questionnaire were presumably filled out immediately while the Time Report is generally filled out later.

Special Interviewer Records

Segment Call Record

Because of difficulties encountered with the Time Report a new method of accounting for interviewer time was used on the last probability sample study (Study 3) described in Table 1. For this study, a very much simplified time sheet was used in combination with a segment call record sheet. The segment call record, which is kept by the interviewer while she is in the segment, records the times for each of the following steps in the interviewing process:

Travel to and from segment
Travel within segment
Waiting for respondent
Seeking or talking with respondent
Actual interviewing.

Naturally, a cost analysis of interviewer time using this form is far easier and more exact than one which uses recall on time records. On the other hand, some interviewers found the record keeping of this form to be burdensome. Currently, the segment call record is being used for those studies where detailed cost analyses are required, but is not used routinely.

Interviewer Log

The results of the 1947 quota sample shown as Study 7 in Table 1 were based on an interviewer log which was developed especially for that study. Data are not available separately on the amounts of study, clerical, editing, and travel time to segments. The analysis of the field operations on this study were done by Stephan and McCarthy and are found in their book, Sampling Opinions.[2]

Interviewer Time Allocation at the Census Bureau and Survey Research Center

1960 Census

Table 2 gives the percentage of enumerator time spent on various tasks, both for the 1960 Census and the Current Population Survey. The figures have been re-worked from the Census documents to make them as comparable as possible to the data in Table 1. Naturally, different methods make full comparability impossible. Thus, the training for the 1960 Census was done on a personal basis, so there is no Study item included. The Current Population Survey also has no provision for Study since this is not measured on CPS Time Sheets. CPS interviewers are paid a fixed amount for studying any special instructions sent them.

The 1960 Census Results are found in Enumeration Time and Cost Study.[3] During Stage 1 of the enumeration, information was obtained on five characteristics for each member of the household and for ten characteristics of the housing unit. If the household had filled in an Advance Census Report form, the enumerator transcribed the information from the form to the enumeration book; if not, he obtained the responses by questioning the household member.

At one-fourth of the housing units, enumerators left additional forms to be filled out and mailed. During the Stage II enumeration, the enumerator received all the individual questionnaires which were mailed in, and made additional visits or phone calls to obtain missing information.

Information on enumerator activities was obtained by having a Records Clerk accompany randomly selected enumerators and record what they did and how long it took. The Census Bureau made no attempt to estimate what the effect of the Records Clerk was on the enumerator. Clearly, it led to more accurate recording of time than would a time sheet, but it could also have influenced the enumerator's work habits.

Transcription to the various FOSDIC schedules was not measured in the field, but was estimated by the Census Bureau from established standards. In general, this work was done after the canvass was completed. Editing time (Field Review) was defined to include the quality control inspection of the enumerator's work by a crew leader or field reviewer, the time it took the interviewer to travel to the field review, the time the interviewer waited for the reviewer and the time spent on payroll computation.

Travel time to the segment was an insignificant part of the enumerator's task and is not even shown for stage II. For stage I it averaged 17 minutes one way or 34 minutes round trip per average assignment of 32. Thus the average travel time to segment per household was about 1 minute.

Current Population Survey

Interviewer Allocation of time on the Current Population Survey is found in Miscellaneous Statistical Data Memorandum No. 3 of the Bureau of the Census.[4] In general, the definitions used there are comparable to those of the National Opinion Research Center. The data for the CPS are obtained from interviewer time records which were kept to obtain this detailed information. The standard CPS time sheet asks only for starting and finishing times each day.

Survey Research Center

The data in Table 3 are from a paper by Goodman and Cannell of the Survey Research Center, University of Michigan.[5] The data were compiled from detailed time and expense reports submitted regularly by interviewers. As at the National Opinion Research Center, these were the same records which were used to compute the pay an interviewer received. Both surveys were national probability samples, with the interview lasting 45 minutes to an hour. On the first study the interviewer took notes and transcribed these notes afterwards. On the other survey, answers were written on the questionnaire during the interview.

Comparison of the Various National Opinion Research Center Studies

This section will discuss the reasons for the differences in interviewer allocation of time on the various National Opinion Research Center studies. No attempt is made here to examine differences between interviewers on the same study. This analysis is left for a later paper.

Table 1 suggests that while actual interview times vary considerably from survey to survey, percentages are more stable. Study and editing time would normally be expected to increase with the complexity of a study, as would the length of interview. Travel time remains fixed with respect to length of interview, but is larger for probability samples. Some more specific comments on reasons for variability for each task may indicate how the nature of the assignment determines interviewer time allocation.

Study Time

Study time for a survey depends mostly on the complexity and length of the specifications. There is a correlation of .94 between length of specifications and actual study time required for the six studies. Roughly, each page of specifications requires on the average about five minutes of study time with an additional fixed time of an hour regardless of size. While these figures are crude, since they are based on only six studies, they do give some basis for suggesting to the interviewers how much time should be allocated to studying.

The size of specifications for the six studies were:

Study	1	2	3	4	5	6
Pages of specifications	118	97	42	14	14	7
Total study time (minutes)	704	424	172	199	178	85

Clerical Time

Interviewer time spent on clerical tasks can be almost completely determined by the field department.

The variability in the clerical times seen in Table 1 is due to the different tasks required of interviewers. For example, on the first study in Tables 1 and 1A, interviewers were required to use stamps to mail packages of completed interviews back to the office instead of using business reply envelopes which have since become standard. This meant frequent trips to the post office to have the packages weighed and to buy stamps as well as additional entries on the time sheet. On the second and seventh studies in Tables 1 and 1A, clerical times included the filling out of special records showing how interviewers spent their time.

Generally, it is more efficient to have clerical jobs done in a central office. This suggests that whenever possible, questionnaire kits and other interviewer material should be assembled before mailing, rather than by the interviewer, and that interviewer trips to the Post Office be avoided by putting stamps on mailing envelopes in advance or by paying postage when questionnaires are returned.

If one were only concerned with reducing clerical time, the use of detailed logs of interviewer time should be avoided. These records are very valuable, however, in the kinds of studies discussed in this paper.

Editing Time

Editing time is directly related to the length and difficulty of the questionnaire. There is no direct measure of this, but the length of time required to conduct the interview is a good indication of this. Editing time is correlated .96 with interviewing time, and generally requires about one-third as much time as does the interview for the six National Opinion Research Center studies in Table 1.

Travel to and in Segment

Travel to and in segment depends on the number of trips required which is partly a function of the cluster size and call-back instructions. Naturally the location of the interviewing staff in relation to the segments is also important, but this is usually unchanged from survey to survey. Since travel costs form a large part of the total costs of an interview, they will be discussed in detail in a later paper which will examine the effects of location and size of primary sampling unit as well as call-back instructions and cluster sizes.

The comparisons between probability samples and quota samples in Tables 1 and 1A indicate some differences, but the magnitude of these differences is smaller than might be expected. For Travel in Segment, there is hardly any difference between the probability and quota samples. For Travel to Segment quota samples require somewhat less time since call-backs are not required and the number of trips is reduced. The quota samples in Studies 4-6 are those which specify the starting address for an interviewer, and the path she must follow, but do not require her to return to a dwelling unit if no one is available. In addition, these studies all specified the proportion of employed and unemployed women to be interviewed, as well as the proportion of men over and under 30 years of age.

In contrast, the 1947 quota sample imposed no geographic limitations, but required the interviewer to obtain a specified number of respondents in each of several rent levels. The high proportion of travel time on that study was probably due to the fact that the rent quotas used at the time were out of date, requiring a long search by the interviewer to find respondents in the lowest rent levels. In addition, the search procedure of interviewers at the time was inefficient, since many tried to fill their lower rent levels by searching in higher rent neighborhoods.

Interviewing

The length of time spent interviewing depends on the length and complexity of the questionnaire. So far as is known, there is no good way to estimate how long it will take to administer a questionnaire without actually pre-testing it. Table 1 shows, however, that the percentage of time spent interviewing is fairly stable for the first six National Opinion Research Center surveys, varying only from 32 to 41 per cent. Only for the 1947 quota study does the percentage of interviewing time drop to 21 per cent. There are two reasons for this. The first is the large percentage of time spent traveling which was discussed above. Even more important is the fact that this 1947 interview took only about 15 minutes in contrast to the other studies where the interview was three or four times longer. It is clear that short interviews of a half hour or less result in less time spent interviewing, but there is no increase in this percentage as interviews get longer than about 45 minutes. The very long interviews make it hard for the interviewer to complete more than a single interview per trip, and this balances the increase of the ratio of interviewing time to travel time per trip.

Comparison of National Opinion Research Center, Census and CPS

The chief difference between the National Opinion Research Center studies and the CPS and Census enumeration is the length of interview. The average National Opinion Research Center studies in Table 1 are about an hour long while the Census enumeration was less than ten minutes and the CPS interviews run about 15 minutes. This explains why the National Opinion Research Center interviewers spend slightly more of their time on actual interviews. Thus, for the 1960 Census the actual time spent interviewing was only 5.4 minutes in Stage I and 2.8 minutes in Stage II. With such short interviews, the amounts of time spent in the house before and after the interview were large relative to the actual interview. If one included all time in the house as interviewing time, then the Stage I percentage of interviewing time would be 45 per cent instead of the 30 per cent shown in Table 2 and the Stage II percentage would be 28 per cent instead of 23 per cent. Perhaps these percentages as well as those in Table 2 should be considered when making comparisons to the National Opinion Research Center and Survey Research Center results.

Of greater significance than the differences are the similarities. Note the percentage of time spent interviewing on the CPS which follows a strict probability sample design as compared to the three National Opinion Research Center studies using probability samples. The CPS figure of 31 per cent of time spent interviewing is quite close to the three National Opinion Research Center percentages of 32 per cent, 33 per cent and 38 per cent. This close agreement between different survey organizations on percentage of time spent interviewing is also confirmed by the data of the Survey Research Center on Table 3.

It can be seen that for Survey B, which is the more usual type of survey, the Survey Research Center percentage of interviewing time, 32 per cent, is in good agreement with the CPS figure of 31 per cent and the National Opinion Research Center percentages of 32 per cent, 33 per cent and 38 per cent. For Survey A, if interviewing and editing are combined, they account for 49 per cent of the time of Survey Research Center interviewers. Similarly on the three National Opinion Research Center probability sample studies, interviewing and editing combined account for 43 per cent, 44 per cent and 49 per cent of the total time.

To summarize these results: Although there are substantial differences between the requirements for different studies, and although different survey organizations have different requirements and measure interviewer time allocation in different ways, there is a surprising uniformity in the percentage of time which interviewers spend on their chief task--interviewing. For probability samples, it is a safe generalization that interviewers spend about one-third of their time interviewing and two-thirds of their time on less critical tasks.

How this compares to other occupations is the subject of the next section of this paper. It will be seen that these results for interviewers are remarkably similar to those of other field occupations.

Other Field Occupations

Sources of Data

Salesmen

Tables 4 and 5 show how salesmen, social workers, probation officers and public health nurses allocate their time to various tasks. Table 4 summarizes five different reports on salesmen. The data on wholesale drug salesmen is from Davis' book Increasing Wholesale Drug Salesmen's Effectiveness.[6] The time study was conducted by having an observer spend a complete day with a salesman from the time he left his house or hotel in the morning until he returned at night. Since Davis felt that the work of the salesman would be conditioned by the knowledge he was being timed, the salesman was not told of the time-study. Rather the salesman was told that notes were being taken on the methods he used for selling, and how effective they were. A total of 38 country salesmen and 32 city salesmen were observed.

Selling Time, which corresponds to Interviewing Time, included promotional selling, dealer assistance, want book selling, sales promotion and collection and adjustment. Travel In is the time in the store spent waiting at the start of an interview or if interrupted, general conversation and idle time. Travel To includes travel and meals. Clerical Time is the time spent in writing up orders and phoning orders into the wholesale house.

The second study which deals with oil company salesmen is from a study by the Atlantic Refining Company reported in Salesweek.[7] Details of how this study, and the other salesmen studies in Table 4 were conducted are not available. Ordinarily, one would not be willing to give much credence to these studies, except that they all seem to say about the same thing. The study of carpet salesmen is cited by Brown, England and Matthews in their Problems in Marketing,[8] while the study of miscellaneous salesmen is also in the Salesweek article mentioned above.[9] It is based on a study of 255 salesmen in 19 different fields. The data on steel salesmen are from the personal files of Allen Jung of the University of Chicago who obtained them while working in the steel industry.[10]

Social Workers

Table 5 shows the time allocation of social workers and public health nurses. Three different studies of social work occupations show great stability in the percentage of time spent interviewing.[11] The first study deals with probation officers in Contra Costa County, California.[12] Individual deputies kept daily logs for a seven week period, and the tasks were coded using the following classifications: Interviewing included personal and phone contacts with the probationer; Study included conferences with supervisors and with other deputies, psychologists, etc; Clerical included all office paper work; Travel appears to be what would be called travel to segment by interviewers; Miscellaneous most closely corresponds to travel in segment for interviewers. The probation officers are the only male group of social workers in Table 5, but they are no different than the other two groups.

The second study concerns 37 caseworkers of the Jewish Child Care Association of New York.[13] They kept tally sheets for 12 working days recording meetings and conferences, telephoning, paper work, dictation and travel. The actual interviews and record reading in preparation for them were not recorded separately, but were obtained by subtraction. Thus, there is no way to separate out interviewing time from what we would call study time. In Table 6, Conference Time is treated as Study Time. Dictation and Clerical work are both included under clerical, although Dictation from notes which accounts for 13.5 per cent of the total time worked could correspond to Editing shorthand questionnaires for interviewers. The 9 per cent of the time spent on the telephone was classified as Miscellaneous, although it might also be compared to the interviewer's travel time in segment, since it involved making appointments for visits.

The final study by the Bureau of Management Analysis of the State of California Department of Social Welfare concerns independent adoptions caseworkers.[14] It is based on returns of a questionnaire to seven agencies asking them to estimate time spent on various tasks. Since adoption is a complicated process, each of the items in Table 5 is the sum of many individual steps. The study and editing tasks are combined because the record of one step becomes the material to be studied for the next; dictation, however, is classified as a clerical task.

Study time includes pre-petition activity, preliminary steps, review of case material after interviews, and conferences with supervisors, attorneys and other professionals. Clerical includes all dictation and preparation of files. Interviewing includes interviews with the adopting parents, the natural parents, the child, and with other family members if needed. Travel time includes actual time spent traveling and time spent telephoning to arrange for appointments.

Public Health Nurses

Two studies are available on how Public Health Nurses spend their time. The first study is by far the more comprehensive one.[15] It is a nation-wide study of 11 public health nursing agencies conducted by the Department of Public Health Nursing of the National League for Nursing. Each agency did two analyses 5 years apart using special forms kept by the nurses. Although no averaging is done in the report, the figures in Table 5 are the simple averages of the 22 numbers. Total home visiting time is divided into three parts: Actual time in the home, travel time, and preparation or postactivity. Staff education is classified as study time, while community activities are put into miscellaneous.

The second study is from Nursing Outlook and presents information on a study of Georgia Public Health Nurses.[16] One hundred eight Nurses in five local health departments kept daily time records for one week. For the visiting nurses, time was divided into actual time in the home, travel time, and preparation and postactivity.

Time Allocation

Salesmen

Table 4 shows that about 37 per cent of a salesman's time is spent in actual selling with only small variation around this average. Only the steel salesmen are substantially below average, and while it is not clear why this is the case, it may be due to the fact that their customers are more widely separated.

What is surprising is that salesmen do not appear to be much different than survey interviewers in the way in which they allocate their time. This would suggest that method of payment, commission vs. hourly rate, probably does not have a very large effect on the percentage of time either interviewers or salesmen spend on their main task. The difference of five percentage points between the time spent selling and the time spent interviewing is probably a maximum estimate of the effects of changing the compensation system for survey interviewers.

Social Workers

The time spent in interviewing on all three social work occupations averages 37 per cent and varies only from 35 to 39 per cent. It is also striking to note that this is exactly the same average percentage of time spent selling by salesmen, and is very close to the percentage of time spent interviewing by survey interviewers. Before speculating as to why these percentages are so close, data will be presented for Public Health Nurses who show a sharply different pattern.

Public Health Nurses

Public Health Nurses spend a substantially greater part of their time on in-home care (which corresponds to interviewing or selling) than do any of the other occupations studied.

It can be seen that nurses spend better than half their time (54-55 per cent) in their chief function as compared to the other occupations which average about one-third time. Table 6 provides a concise summary of the results of the earlier tables. Certainly one is led to speculate as to reasons why interviewing, selling, and social work show such strong similarities and why nursing differs. These speculations are presented in the final section.

Similarities in Various Field Occupations

In considering why interviewing, selling, and social work show such similar patterns certain reasons can probably be rejected. It might be argued that the agreement is coincidental, but this seems extremely unlikely given the fact that 20 different studies are compared. While the argument that this is a chance occurrence can never be fully discarded, there does appear to be a reason which has a more rational appeal.

Since there is some ambiguity in the data for all these studies, it might be thought that this agreement is artifactual--that the summarization of the data was done in such a way as to bring them into line with a preconceived hypothesis. This does not seem to be the case. The greatest ambiguity in the data are in categories other than interviewing. While there is often a question as to whether something should be classified as study, clerical or miscellaneous it is generally easy to separate the actual interviewing or selling from travel or waiting time in the reports analyzed, although this does not insure the initial accuracy of these reports. In addition, the results shown above differed substantially from the initial hypotheses. Prior to data collection, it was felt that there would be real differences between interviewers, social workers, and salesmen. Using a monetary reward framework, it was felt that salesmen would spend the most time in actual selling since their commissions depended on the number of contacts they made while interviewers would spend the least time in actual interviewing since the longer it took them in noninterviewing activities such as travel and study, the more they received. Clearly, this indicates that method of payment is not the reason for the similarities.

Nor does it seem likely that the characteristics of the persons in these occupations are enough alike to cause these similarities. Sex is not important since interviewers are mostly women, salesmen are men, and social workers are both (at least, in this analysis). Education is not an important variable since social workers generally have some graduate work, interviewers some college, and salesmen are generally high school graduates.[17] Neither age nor family status are identical--interviewers tend to be middle-aged women with children in or through with school, while social workers tend to be younger.

The reason for the similarities seems to lie in the job situation itself. The three occupations, interviewing, selling, and social work are all highly stressful in their most crucial component--the personal contact with the respondent. The interview has generally been arranged at the request of the interviewer rather than the respondent, and there is always the possibility of a slammed door or a curt refusal. Even when the interview has started, the interviewer is always conscious of the effort to keep it flowing smoothly to a successful conclusion. The process can be so wearing emotionally that the interviewer needs time to recuperate and so other activities are included as part of the job, be they traveling, waiting, studying or clerical tasks.

It may be argued that social workers are not in the same fix as are interviewers and salespeople, but for the three examples given in this paper there do appear to be real reasons for tension between the social worker and the respondent, and the social workers have come to expect this tension. The data on the visiting nurses seem to confirm this conclusion by contrast. The visiting nurse is almost always invited into the home to give medical care, and does not expect to overcome any resistance in getting into a home or during the treatment. Since there is less stress on her, she is able to spend a larger part of her time in the home.

If this is a valid conclusion, it has this consequence. It suggests that manipulation of compensation or of details of the job would have very little effect on the percentage of time which the interviewer spent on the interview. The only way to increase time spent interviewing would be to reduce tension, but this may not be possible. It may be that certain individuals are less sensitive to this tension, and are thus able to spend greater parts of their time on the actual interview. These people may not make the best interviewers, however, since this lack of sensitivity could result in more interviews of lower quality. This area too is the subject of research by National Opinion Research Center and will be discussed in a later paper.

This analysis is not intended to suggest that each day will be allocated the same way by workers in field occupations. Some days may be spent entirely in interviewing, while on other days no interviews may be conducted. It is suggestive, however, that a majority of National Opinion Research Center interviewers never spend more than four hours per day interviewing, either on probability or quota sample studies.

Further Research Suggested

The generalizations presented above suggest several areas of additional study. It is not clear what part of the tensions are due to the efforts required to keep the interview going, and what part to the initiation of the interview with a possibly unwilling respondent. It should be possible to obtain records or devise experiments where appointments have been made for the interviewer. If interviewer time allocation did not then change, one would conclude that the tensions were primarily due to interpersonal contact. On the other hand, there are cases which require an initial contact with a respondent, but no additional interactions. Such tasks as store auditing and leave and pick up questionnaires are examples. Again one would look for changes in interviewer time allocation as indicating effects of interpersonal contacts.

It would be extremely useful to obtain data on other occupations where a great deal of interpersonal contact is required, but where the meetings are not initiated by the interviewer. Thus, employment interviewers, sales clerks, and school teachers come to mind as groups worth investigating. The time allocation of people in occupations with little interpersonal contacts such as scientists and engineers would also be illuminating.

FOOTNOTES

[1] Herbert Hyman, William J. Cobb, Jacob J. Feldman, Clyde W. Hart, and Charles Herbert Stember, Interviewing in Social Research (Chicago: The University of Chicago Press, 1954).

[2] Frederick F. Stephan and Philip J. McCarthy, Sampling Opinions (New York: John Wiley & Sons, 1958).

[3] United States Censuses of Population and Housing 1960, Enumeration Time and Cost Study (Washington: U. S. Bureau of the Census, 1963).

[4] I am grateful to Mr. Dean Webber, the author of this memo, for making it available to me, and also for his useful comments and suggestions.

[5] Roe Goodman and Charles F. Cannell, "Sampling Errors and Components of Interview Costs in

FOOTNOTES--Continued

Relation to Sample Design" (mimeographed, Ann Arbor: Survey Research Center, University of Michigan, no date).

[6] James H. Davis, Increasing Wholesale Drug Salesmen's Effectiveness, (Columbus, Ohio: Bureau of Business Research, Ohio State University, 1948).

[7] Salesweek, December 12, 1960, pp. 12-13.

[8] Milton P. Brown, et al., Problems in Marketing (New York: McGraw-Hill, 1961).

[9] Salesweek, December 12, 1960, pp. 12-13.

[10] Personal communication from Allen Jung, Graduate School of Business, University of Chicago.

[11] I am indebted to Edward Schwartz of the School of Social Service Administration, University of Chicago, for bringing this data to my attention and making it available to me.

[12] Contra Costa Probation Department, "Time Study" (typewritten, Martinez, California: 1959).

[13] Jewish Child Care Association of New York, "Time Study, May, 1952" (mimeographed, New York: October 21, 1952).

[14] Department of Social Welfare, State of California, Independent Adoption Yardsticks (Sacramento: September, 1956).

[15] Department of Public Health Nursing, National League for Nursing, A Comparative Study of Costs in Eleven Public Health Nursing Agencies (New York: National League for Nursing, 1956).

[16] Katharine Akin, "Time Study of Georgia Public Health Nurses," Nursing Outlook, X (1962), pp. 544-46.

[17] Davis, pp. 41-53.

TABLE 1

PERCENTAGES OF INTERVIEWER TIME SPENT ON VARIOUS TASKS AND ACTUAL TIMES FOR SEVEN NORC STUDIES

Task	Percentage of Time Spent							Actual Times/Interview in Minutes (Study and Clerical Time is Total/Study)						
	Probability Samples			Block-Quota			1947 Quota							
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Study	17	8	4	13	12	9	47	704	424	172	199	178	85	-
Clerical	8	7	1	2	10	4		307	366	73	26	150	39	-
Editing	11	11	11	13	9	12		39	11	24	23	14	15	-
Travel to Segment . . .	21	22	29	17	17	18		74	21	67	29	27	23	-
Travel in Segment . . .	11	19	17	15	14	16	32	40	19	39	27	22	20	35
Interviewing	32	33	38	40	38	41	21	114	32	86	70	60	52	23
Total	100	100	100	100	100	100	100							
Total Interviews . .	2,115	15,690	2,563	1,470	1,449	1,688	1,223							
Total Interviewers .	186	295	119	161	160	231	88							

TABLE 2

PERCENTAGES OF ENUMERATOR TIME SPENT ON VARIOUS TASKS 1960 CENSUS[a] AND CURRENT POPULATION SURVEYS[b]

Task	1960 Census		CPS
	Stage I	Stage II	
Clerical (Transcription). .	21	38	20
Editing (Field Review). . .	6	7	-
Travel to Segment	6	-	35
Travel in Segment	30	12	14
Interviewing	30	23	31
Miscellaneous	7	20	-

Source: [a] Tables 1, 14, 24; pp. 27, 32, 36, Enumeration Time and Cost Study.

[b] Bureau of the Census, Miscellaneous Data Memorandum No. 3, dittoed September 17, 1958.

TABLE 3

PERCENTAGES OF INTERVIEWER TIME SPENT ON VARIOUS TASKS ON TWO SURVEY RESEARCH CENTER STUDIES

Task	Survey A	Survey B
Study, Clerical and Miscellaneous . .	21	24
Editing	26	-
Travel To and In Segment.	30	44
Interviewing.	23	32

Source: Roe Goodman and Charles F. Cannell, "Sampling Errors and Components of Interview Costs in Relation to Sample Design," (mimeographed, Ann Arbor: Survey Research Center, University of Michigan).

TABLE 4

PERCENTAGES OF SALESMAN TIME SPENT ON VARIOUS TASKS

TASK	Wholesale Drug[a]		Oil[b]	Steel[c]	Carpet[d]	Miscellaneous[e]
	City	Country				
Study (Preparation). . .	-	-	-	-	-	19
Clerical	4	2	-	6	-	-
Travel To.	33	32	37	72	-	-
Travel In.	26	25	16		-	45
Selling.	37	41	45	22	40	36
Miscellaneous.	-	-	2	-	60	-

[a] James H. Davis, Increasing Drug Salesmen's Effectiveness (Bureau of Business Research, Ohio State University, Columbus, Ohio, 1948), p. 59.

[b] Survey of Atlantic Refining Company Salesmen, reported in Salesweek, December 12, 1960, p. 13.

[c] Personal communication from Allen Jung, Graduate School of Business, University of Chicago.

[d] Brown, England and Matthews, Problems in Marketing (McGraw-Hill, 1961).

[e] Salesweek, December 12, 1960, pp. 12-13.

TABLE 5

PERCENTAGES OF TIME SPENT ON VARIOUS TASKS BY SOCIAL WORKERS AND PUBLIC HEALTH NURSES

Task	Social Workers			Public Health Nurses	
	Probation Officer[a]	Foster Home Placement[b]	Independent Adoption[c]	National[d]	Georgia[e]
Study (Conference)	9	13	24	2	-
Clerical (Record Keeping).	22	24	20	23	13
Travel	15	16	21	20	32
Interviewing (In-Home Care)	39	38	35	54	55
Miscellaneous	15	9	-	1	-
Total	100	100	100	100	100

[a] Time Study (Martinez, California: Contra Costa County Probation Department, 1959), typewritten.

[b] Jewish Child Care Association of New York Time Study, May, 1952 (mimeographed, New York: October 21, 1952).

[c] Department of Social Welfare, State of California, Independent Adoption Yardsticks (Sacramento: September, 1956).

[d] Department of Public Health Nursing, National League for Nursing, A Comparative Study of Costs in Eleven Public Health Nursing Agencies (New York, 1956).

[e] Katherine Akin, "Time Study of Georgia Public Health Nurses," Nursing Outlook, X (1962), pp. 544-46.

TABLE 6

COMPARISON OF TIME SPENT INTERVIEWING, TRAVELING, AND IN OTHER TASKS BY SURVEY INTERVIEWERS AND OTHER FIELD OCCUPATIONS

Occupation	Interviewing	Traveling	Other	Total
Interviewers:				
National Opinion Research Center:				
Probability Samples	34	40	26	100
Block-quota Samples	40	32	28	100
Census	26	24	50	100
Current Population Survey	31	49	20	100
Survey Research Center	28	37	35	100
Salesmen	37	52	11	100
Social Workers	39	17	44	100
Public Health Nurses	55	26	19	100

III

SAMPLING PROBLEMS IN SOCIAL SURVEYS - II: FURTHER DEVELOPMENTS IN PROBABILITY SAMPLING SURVEYS

Chairman, Frederick F. Stephan, Princeton University

Page

The Redesign of the Canadian Labour Force Survey - I. P. Fellegi, G. B. Gray and R. Platek, Dominion Bureau of Statistics.................. 54

Generalizations for Complex Probability Sampling - Leslie Kish, University of Michigan............ 63

THE REDESIGN OF THE CANADIAN LABOUR FORCE SURVEY

I.P. Fellegi, G.B. Gray, R. Platek
Dominion Bureau of Statistics, Ottawa

The Canadian Labour Force Survey was established in 1945 and designed, at that time, to provide quarterly estimates of labour force characteristics at the national level. Since 1952 it has been conducted monthly and estimates have been published also for five regions (a region being a province or a group of provinces). The method of enumeration is through interviewing during a specified week each month (the survey week) and collecting information pertaining to the previous week (reference week).

The sample design of the survey was originally based on the Census of 1941, and although it has been revised several times since, many of its features are still based on the Census of 1941. We are now in the process of redesigning the survey, introducing the new design province by province.

Table 1 of the handout provides a somewhat crude summary description of the main features of the old design. All cities with a population of more than 30,000 persons were selected with certainty and they constituted the so called self-representing areas of the survey. In each of these cities a two-stage sample of segments (mostly city blocks) and households was selected, each selected segment yielding an expected 5 households. The remaining parts of the country were divided into primary sampling units (p.s.u.'s) and within each province p.s.u.'s with similar overall characteristics were grouped into strata. The p.s.u.'s in a stratum were not necessarily contiguous and there was considerable variation in the number of p.s.u.'s per stratum as well as the total population of the stratum. The population of strata was about 100,000-150,000 persons. One p.s.u. was selected with p.p.s. in each stratum within which in successive stages a sample of segments, clusters and households was selected.

The following comparisons between the Labour Force Survey and the CPS should be emphasized:

1) Our p.s.u.'s were not natural units as the counties are in the CPS. Had we used counties as p.s.u.'s we would have had much too few p.s.u.'s due to the relatively small population of Canada.
2) Cities of 30,000 population with our 1% sampling ratio provide a sample of about two enumerator assignments. In the CPS, cities of more than 250,000 persons are made self-representing and this yields a sample of about one small enumerator assignment.
3) Our strata were much smaller than those in the CPS (about one-third as large) yet even so our strata yielded more than 4 enumerator assignments while the CPS strata usually yield about 1 enumerator assignment. In both surveys 1 p.s.u. is selected in each stratum.

With the creation of a sampling research group, a more systematic investigation of the old design has begun a few years ago. The desirable objectives of a new design have been developed partly through the observation of several serious shortcomings and imbalances in the old design:

1) An exceedingly important lack of balance was noticed between the self-representing areas and the non-self-representing areas with respect to cost and variance. Table 3 of the handout indicates that in Alberta the cost of the survey in the non-self-representing areas accounted for about one third of the cost but three quarters of the variance.
2) A study of the components of variance revealed that about 50% of the variance in the non-self-representing areas is accounted for by the first stage of sampling. This clearly indicated that we had too few p.s.u.'s in the sample and in comparison, too large a sample within. In fact, the number of p.s.u.'s in the sample from the non-self-representing areas, was 78 in total, employing 288 enumerators. By selecting 288 p.s.u.'s and a sample of one enumerator assignment in each, we could increase the number of p.s.u.'s in the sample while maintaining the same number of enumerators and the same overall sample size. In fact, the number of p.s.u.'s in the sample has but a small affect on the total cost, as long as the enumerators reside in the p.s.u. they enumerate, and the sample take in the p.s.u. provides at least one enumerator assignment. Consequently one of the early decisions was to increase the number of p.s.u.'s in the sample and to reduce the take within to one enumerator assignment. In line with this decision we have also decided to lower the population level above which the cities will be self-representing, again to the point where a self-representing city yields at least one enumerator assignment. This includes in the new design all cities of more than 15,000 population removing most of the industrialized urban population from the non-self-representing areas. This, of course, makes the task of stratification in the non-self-representing areas considerably easier.
3) The measures of size used for p.p.s. sampling at the various stages which are based on 1951 Census information are also outdated. This results in a considerable variance in the sample take which, of course, is reflected in the variance of the main estimates as well.
4) For some time, now, we have been concerned

with the bias of the so-called collapsed stratum variance estimates due to selecting only one p.s.u. in each stratum. Paradoxically, improving the stratification in the new design would have meant increasing this bias (which basically is a between strata variance). The selection of two p.s.u.'s per stratum, for a given set of sampling ratios, would necessitate the creation of twice as large strata. We knew that for the contemplated sampling ratios even under the two p.s.u. stratum scheme our strata would not have a population larger than 40,000-80,000 persons. The 1961 Census information was utilized in an empirical study investigating the relative efficiency of the two schemes. We have found that there is very little to be gained by creating strata with populations less than 40,000-80,000 persons and so we decided to have the luxury of unbiased variance estimates by selecting two p.s.u.'s from each stratum. It should be emphasized that this decision was taken in light of the fact that even so our strata would be rather small; 40,000-80,000 persons is about the size of a county. In the CPS strata have a population of more than 300,000 persons and only one p.s.u. is selected in each stratum. Again, the quantitative difference between the higher overall sampling ratio in Canada and the substantially smaller one, less than one-tenth as large in the U.S.A. resulted in a qualitative difference in the design.

5) Studies based on 1961 Census data indicated that our stratification (essentially unchanged since 1945) is hopelessly outdated. The socio-economic map of Canada has changed a great deal in the meantime. In the new design we wanted to form strata which had a good chance of staying internally homogeneous even in a few years' time. If a stratum with a large population is to be formed, the practise is usually to put into the same stratum areas which are sometimes separated by considerable distances, because one seldom finds enough neighbouring municipalities with similar characteristics to make up a stratum. Thus local developments which take place after the stratification affect different parts of the stratum differently thereby contributing to the gradual deterioration of the original stratification. We have found that due to the relatively small size of our strata and due to the fact that most of the industrialized urban population is removed from the non-self-representing areas and is sampled directly, it was possible to create strata which are made up of geographically contiguous areas. Therefore, we decided that wherever possible we will form geographically compact strata.

The procedure of stratification was carried out within Economic Regions. These are subdivisions of provinces, defined as areas of "structural homogeneity" according to such factors as soil characteristics, production and marketing possibilities, commercial and industrial potential. It has been decided to carry out the stratification process in each Economic Region separately for two main reasons. Firstly, estimates may be required in the future, for areas smaller than a province, which may well be an Economic Region or a group of them. Secondly, the Economic Region as an area, is more conveniently manageable than the whole province at the redesign stage as well as in future revisions of the Labour Force Survey. The immediate problem in stratifying an Economic Region was to select the characteristics which should be used in the stratification. Ideally these characteristics should fulfill the following conditions:

a) The characteristic is relatively stable over time

b) The characteristic is related to some Labour Force characteristic

c) The number of persons having this characteristic varies from area to area so it discriminates between areas

d) The number of persons having this characteristic accounts for a sizable part of the population of the region.

Conditions a) and b) were taken into account by exercising our judgement in selecting characteristics (i.e., employment by industry, average wages and salaries, etc.). As for c) and d) the characteristics were evaluated in a particular Economic Region by computing the so-called "Importance Factor" defined as

$\frac{n\sigma^2}{N}$ where

n=total number of persons in the Labour Force in an Economic Region having the particular characteristics

N=total number of persons in the Labour Force in a particular Economic Region

σ^2=variance of the characteristics between areas (municipalities) in the Economic Region.

This measure was calculated for a number of characteristics in each region and the relative size of the Importance Factor determined the importance of a given characteristic. Then three or four of the most important characteristics were used for stratification purposes.

Considering the repetitive character and the volume of the operations involved in stratification, it was essential to have a uniform method, which could be followed by clerks without much supervision, to ensure both speed and efficiency. The building blocks used in setting up strata were enumeration areas which are the smallest units for which Census data are available. Each enumeration area was compared with the whole Economic Region with respect to the "Important" characteristics. This comparison might show, for example, that a given E.A. is higher than average with respect to one characteristic, lower than average with respect to the second, etc. Plotting these comparisons a pattern is obtained for each E.A. The enumeration areas whose pattern was similar (on the basis of visual inspection by clerks) were grouped into one stratum with due regard to the desired objectives of geographic contiguity and approximately equal size of strata. Within each stratum p.s.u.'s were formed. These again were geographically compact and were, of course, representative of the entire stratum: a maximum deviation of 5% from the corresponding stratum average of the important characteristics was allowed. The proportion of urban and rural population in the primary sampling unit was even more strictly controlled. Due to the variation in the size and geographical distribution of urban centres it was often necessary to split urban centres between two or more primary sampling units. This splitting of urban centres was not made geographically but on a ratio basis. Throughout the whole operation a set of punched cards was maintained (one for each E.A.) containing the values of the important characteristics as well as room for strata and p.s.u. codes. This enabled us to utilize the computer and clerks to maximum efficiency and statisticians were only used to polish the draft strata-p.s.u. frame prepared by clerks. Table 4 contains a summary description of the frame of the new design.

In each province, institutions (such as Hospitals, Schools, Hotels, Military Establishments) and also Remote Areas were designated as special areas. Special sampling and enumeration procedures were used for these areas.

6) Concerning the self-representing areas preliminary cost and variance analysis indicated that the old design was not far from the optimum. Here the main problem seemed to be that in some parts of a city due to its local development (such as urban renewal, new apartment buildings, etc.) the original household counts become outdated quite fast and quite drastically. Any revision of the original household counts would alter the probabilities of selection and would therefore disrupt the whole systematic sample, spread throughout the entire city. For example, the selection of a new sample of segments in a city like Montreal where almost 500 of them are in the sample, would have been a considerable job. A feature of the new design to which we attach considerable importance, will be the division of the larger cities into strata, we call them sub-units, and the independent selection of a systematic sample of segments in each sub-unit. A regular programme of checking building permits will then indicate if a large development occurs in any of the sub-units. Whenever this occurs we will revise the measures of size but only in the affected sub-units. Other features of the design in the large cities include special treatments for institutions and large apartment houses. Time does not permit to go into details.
7) The old design of the Labour Force Survey was established to provide national estimates. Later on separate estimates were also published for five regions where some of the regions were the larger provinces. The estimates for the smaller provinces, though unpublished, were made available to the provincial governments with due warnings concerning their reliability. There was an increasing demand, however, for strengthening the provincial estimates. In the new design a compromise allocation of the sample into the ten provinces was worked out in such a way that the estimates of unemployed in each province would have a coefficient of variation of not more than 13% (with the exception of Prince Edward Island whose total population is only 100,000). Like all compromises this one is also difficult to explain rigorously. It was arrived at by considering various alternatives and choosing the one which did not deviate too far from the optimum design and yet brought the provincial estimates within sight of being publishable. In comparison the CPS publishes estimates only for four regions of the U.S.A. but then the Bureau of the Census is fortunate that there are 50 states in the U.S.A. rather than 10.

Having made the broad decisions outlined above the details of the new design emerged on the basis of guidance provided by the cost and variance study. I shall attempt to outline our general approach. As usual in studies of this kind a mathematical model was established for both the cost of the survey and the variance of the resulting estimates. Each of these was to be a function of the same variables with respect to which the design was to be optimized. The mathematical model for the cost was designed to reflect field costs only since Head Office costs were assumed to be given and fixed.

Since a preliminary study indicated that in the self-representing areas the old design was already close to the optimum, the cost function set up to represent the cost of enumeration in the self-representing areas was rather simple and shall be omitted here. The cost in the non-self-representing areas was split into two main parts:

1) enumeration costs - This is that part of the cost which may be thought to be proportional to the number of households in the sample. It includes, therefore, the cost of the actual interviews, the travel from household to household within clusters and also the cost of training the enumerators, since this latter is proportional to the number of enumerators which in turn is directly proportional to the sample size.
2) travel cost - The travel cost itself can be broken into various components corresponding

to travel between sampling units at the various stages of sampling with the exception that there is no travel between p.s.u.'s since one enumerator enumerates each p.s.u. and that there is an additional component of travel from home to area (that is the travel of the enumerator from his residence to the first household to be enumerated each day and back to his residence at the end of the day).

A detailed description of the cost function appears in the Appendix. It was assumed that the total field cost may be represented as the sum of the cost in the rural and urban areas.

The variance function pertaining to either the urban or the rural portion of a province appears in the second part of the Appendix. It is split up into four components, each corresponding to a stage of sampling. In each selected p.s.u. the urban and rural areas are subsampled independently. Therefore, the variance components are additive over the urban and rural parts except at the p.s.u. level where an additional covariance term occurs between the urban and rural subsamples.

The symbols appearing in the functions are either constants or variables belonging to one of the following five categories:

1) constants based on a detailed study of the enumerators' records of time and mileage. The information was supplemented by map studies. These constants are averages referring to a particular province (i.e. average time spent in enumerating a household).
2) constants derived from other sources, for example intercensal population estimates or information pertaining to variance components derived under the old design.
3) constants for which a provincial average was not satisfactory. Population density is an example: it varies within a province to such an extent that the average is hardly useful. Several different values were substituted for these constants and their affect on the optimum allocation was examined.
4) the basic variables in terms of which the optimization was carried out. In the present study these were the weights at the various stages of sampling.
5) variables which were functions of the basic ones.

The operation of optimization was an exercise in non-linear programming. Various boundary conditions had to be satisfied to take into account known restrictions. Some of these restrictions were imposed somewhat arbitrarily, because we felt that the empirical formulae were applicable only within certain ranges of the variables. Also the assumption that certain quantities are constants is valid only within a limited range of the variables. For example it was thought that the average time spent in enumerating a household is not sensitive to changes in the design and was therefore regarded as a constant. The details of the operation of optimization will be omitted here. It was carried out by a high speed computer (IBM 7074).

In closing, we would like to say a few words about the results. At this time, the survey is operating under the new design in three of the ten provinces. A full-scale field test had been completed in another two provinces. By the end of 1965 the survey will operate entirely under the new design. The results quoted in Table 3 refer to Alberta, because it was the first province to be redesigned and hence the only one with a few months of data under the new design. It is not an entirely typical province, in that the difference under the old design between the variance in the self-representing and non-self-representing areas is substantially greater in the other provinces (those outside the Prairies). Since the gain in the new design is made in the non-self-representing areas, the reduction of the variance in Alberta is smaller than what we expect in the other provinces. Even under these relatively unfavourable conditions Table 3 shows that under the new design the variance of the estimated employed under the new design has been reduced by 20% in the self-representing, 70% in the non-self-representing areas for an overall reduction of 55%. The variance of the estimated unemployed has been reduced by almost 25%. All of this variance reduction was effected while also reducing the field cost of the survey by 14%. The amount of information per unit cost ([1/variance]/cost) has increased by 155% for employed and 56% for unemployed.

TABLE 1
OLD DESIGN

Stage of Sampling	SELF-REPRESENTING AREAS			NON-SELF-REPRESENTING AREAS		
	Nature of Units	Size of Units (pop.)	Method of Selection	Nature of Units	Size of Units (pop.)	Method of Selection
Stratum	Metropolitan area or Special area	30,000+	Certainty	Group of similar p.s.u.'s (generally) not contiguous	80,000-200,000	Certainty
First Stage (P.S.U.)	None	None	None	Groups of Heterg. municipalities	10,000-22,000	One unit selected with p.p.s.
Second Stage (Segment)	City block(s)		Systematic with p.p.s. to household count	Census enumeration area(s)	Rural approximately 500 Urban 1,000 to 20,000	Systematic p.p.s. area sub stratification
Third Stage (Cluster)	None	None	None	Small area with recognizable boundaries	Multiple of 4-8 H.H.'s	Random (p.p.s. for multiple clusters)
Fourth Stage (Household)	Household	3-4	Random systematic to yield expected 5 H.H.'s as of design	Household	3-4	Random in multiple clusters

TABLE 2
NEW DESIGN
(in the Province of Alberta; some variation from province to province)

Stage of Sampling	SELF-REPRESENTING AREAS			NON-SELF-REPRESENTING AREAS		
	Nature of Units	Size of Units (pop.)	Method of Selection	Nature of Units	Size of Units (pop.)	Method of Selection
Stratum	Metropolitan area or Special area	15,000+	Certainty	Group of similar p.s.u.'s (geographically contiguous)*	35,000-55,000	Certainty
First Stage (P.S.U.)	Census Tracts	15,000	Certainty	Rural enumeration areas and nearby small urban*	3,200-5,500	Two units selected with p.p.s.
Second Stage (Segment)	City block(s)		p.p.s. systematic	Rural enumeration area and small urban or part of it	Rural 500 Urban approximately 800	Systematic p.p.s. within urban and rural
Third Stage (Cluster)	None	None	None	Small area with recognizable boundaries	Multiple of 3 or 4 H.H.'s	Random systematic (p.p.s. for multiple clusters)
Fourth Stage (Household)	Household	3-4	Random systematic	Household	3-4	Random systematic in multiple clusters

* Note: In the old design p.s.u.'s were formed first and then combined into strata; in the new design strata were formed first and then divided into p.s.u.'s.

TABLE 3

Cost, Variance and Information per Unit Cost of the Estimated Employed and Unemployed in Alberta by Type of Area and Design

		Cost in thousand dollars	EMPLOYED		UNEMPLOYED	
			Variance (4) in millions	Information per cost (5) in 10^{-10}	Variance (4) in millions	Information per cost (5) in 10^{-10}
Old design (1)	Self-representing	2.35	20.84	0.20	2.39	1.78
	Non-self-representing	1.33	45.46	0.17	6.94	1.09
	Total	3.68	66.30	0.04	9.33	0.29
New design (2)	Self-representing	1.55	16.59	0.39	3.24	2.00
	Non-self-representing	1.63	13.47	0.46	3.67	1.67
	Total	3.17	30.07	0.10	6.91	0.46
Ratio (New/Old) (3)	Self-representing	0.66	0.80	1.91	1.36	1.12
	Non-self-representing	1.23	0.30	2.75	0.53	1.54
	Total	0.86	0.45	2.55	0.74	1.56

(1) The old design in Alberta utilized a 1% overall sampling ratio in the self-representing areas and a 0.67% sampling ratio in the non-self-representing areas.

(2) The sampling ratio in the new design is 0.8% in both self-representing and non-self-representing areas.

(3) The ratios in the last three rows are in units.

(4) For the sake of making the variances comparable, they were adjusted to refer to the same level of employment and unemployment under both designs.

(5) The information per unit cost here refers to (1/Variance)/Cost.

Appendix to the "Redesign of the Canadian Labour Force Survey"

The Cost Function

The cost function for non-self-representing units is split into enumeration and travel costs while the travel cost is further split into three components. In each case, the time spent is given in terms of hours spent and this is followed by the cost in dollars. The selection of two p.s.u.'s per stratum is assumed here.

(1) $T_e = \left(\frac{P_U}{W_U} \cdot \frac{1}{P_{4U}} + \frac{P_R}{W_R} \cdot \frac{1}{P_{4R}}\right) t_H$

= Enumeration Time (1)

$C_e = T_e(r_h+r_mS_H)$ = Enumeration Cost (2)

and (2)(i) $T_{HA} = 2N\ d_{HA}\ \frac{1}{S_1}$ = Travel time between home and area. (3)

$C_{HA} = T_{HA}\ (r_h+r_mS_1)$ = Cost of travel between home and area. (4)

(ii) $T_{SS} = (M_U+M_R-N)\ d_{SS}\ \frac{1}{S_2}$

= Time required for segment to segment travel. (5)

$C_{SS} = T_{SS}\ (r_h+r_mS_2)$

= Cost of segment to segment travel. (6)

(iiiJ) $T_{CC:J} = (C_J-M_J)\ d_{CC:J}\ \frac{1}{S_{3J}}$

= Time required for cluster to cluster travel (broken down by J=U (urban) and J=R (rural)). (7)

$C_{CC:J} = T_{CC:J}\ (r_h+r_mS_{3J})$ (8)

The Variance Function

The variance function for non-self-representing units is split up into two main parts, urban and rural and within each it is split up into four components as follows: (a) between p.s.u.'s, (b) between segments, (c) between clusters, and (d) between households. In addition to these four components of variance there exists a covariance between urban and rural p.s.u. totals given by (e). The four components of variance for either the urban or rural areas (denoted by J) and the covariance in (e) are given by:

(a) $V_{1J} = B_J\ (W_{1J}-1)\ \frac{1}{1-\frac{P_{1J}}{P_{0J}}}$

$[1 + (P_{1J}-1)\ \delta_{1J}]$ (9)

(b) $V_{2J} = B_J\ (W_{2J}-1)\ W_{1J}\ \frac{1}{1-\frac{P_{2J}}{P_{1J}}}$

$\{[1+(P_{2J}-1)\delta_{2J}] - \frac{P_{2J}}{P_{1J}}\ [1+(P_{1J}-1)\ \delta_{1J}]\}$ (10)

(c) $V_{3J} = B_J\ (W_{3J}-1)\ W_{1J}W_{2J}\frac{1}{1-\frac{P_{3J}}{P_{2J}}}$

$\{[1+(P_{3J}-1)\delta_{3J}] - \frac{P_{3J}}{P_{2J}}\ [1+(P_{2J}-1)\delta_{2J}]\}$ (11)

(d) $V_{4J} = B_J\ (W_{4J}-1)\ W_{1J}W_{2J}W_{3J}\ \frac{1}{1-\frac{P_{4J}}{P_{3J}}}$

$\{[1+(P_{4J}-1)\delta_{4J}] - \frac{P_{4J}}{P_{3J}}\ [1+(P_{3J}-1)\delta_{3J}]\}$ (12)

(e) $(CV)_{1:UR} = r_{UR}\ \sqrt{V_{1U} \cdot V_{1R}}$ (13)

The symbols in formulae (1)-(13) are either constants or variables. They refer to averages in a particular province and may be classified into one of the following five categories.

Category 1: Constants based on a detailed study of enumerators' records of time and mileage. The information was supplemented by map studies.

t_H = average time spent in enumerating a household, including travel between households within clusters (but not between clusters or between segments).

S_H = average speed of travel between successive households within a cluster.

t = average time per enumerator per round trip (time spent by an enumerator during a day starting from his place of residence, travelling to the first sample household, enumerating and travelling during his work and finally back to his place of residence).

$\beta_1, \beta_2, \beta_{3J}$ = constants of proportionality estimated from empirical results under the assumption that the distances travelled from home to area, from segment to segment and from cluster to cluster are proportional to the square root expressions of formulae (3), (5) and (7) respectively.

Category 2: Constants derived from sources other than enumerators' records or map studies.

P_J = population, 14 years of age and over (excluding non-enumerable persons such as inmates, armed forces personnel, etc.) within the urban or rural part of a province (estimated using the intercensal estimates of the population and the estimated proportion of the population living in urban or rural areas).

P_{4J} = average size of household in type of area J (urban or rural)

r_h = hourly rate of pay

r_m = rate of pay per mile

H = size of enumerator assignment (number of households)

r_{UR} = correlation coefficient between the estimates (generally unemployed in our studies) derived for the urban and rural parts of a p.s.u.

p_J = proportion of persons with a certain specific labour force characteristic (generally unemployed in our study) in type of area J.

$B_J = P_J p_J(1-p_J)$; (binomial variance).

Category 3: Constants for which one provincial average was not satisfactory. Several different values were substituted for these constants and their effect on the optimum allocation was examined.

ρ_J, ρ = population density (urban or rural or overall)

P_{2J} = average population in urban or rural segments

P_{3J} = average population in urban or rural clusters

Note: there are natural limitations concerning the values P_{2J} or P_{3J} may assume. Segments are to be formed by combining Census Enumeration Areas and so the average population of a segment must be a multiple of the average population of Enumeration Areas. Similarly, the availability or lack of natural boundaries impose restrictions on the size of clusters.

Category 4: The basic variables, in terms of which the optimization was carried out.

W_J = overall weight (inverse of overall sampling ratio) in type of area J (urban-rural)

W_1 = p.s.u. weight (inverse of sampling ratio at first stage)

W_{2J} = segment weight (inverse of sampling ratio at second stage in urban-rural areas)

W_{3J} = cluster weight (inverse of sampling ratio at third stage in urban-rural areas)

Note: in the final analysis both the cost and the variance functions were expressed in terms of constants and the seven basic variables listed above.

Category 5: Functions of the basic variables.

$K = \frac{1}{2H}\left[\frac{P_U}{W_U P_{4U}} + \frac{P_R}{W_R P_{4R}}\right]$ = number of strata in province (the selection of two p.s.u.'s per stratum is assumed to have been decided)

$P_{1J} = \frac{P_J}{2K\ W_1}$ = average population per p.s.u. living in urban or rural areas.

$d_{HA} = \beta_1 \sqrt{\frac{P_{1U}+P_{1R}}{\rho}}$ = average distance from home to area. (β_1 is a regression coefficient)^1

$d_{SS} = \beta_2 \sqrt{\frac{P_{1U}+P_{1R}}{\rho} / \left[\frac{P_{1U}}{W_{2U}P_{2U}} + \frac{P_{1R}}{W_{2R}P_{2R}}\right]}$ = average distance between sampled segments within a p.s.u. (β_2 is a regression coefficient)^1

$d_{CC:J} = \beta_{3J} \sqrt{\frac{P_{2J}}{\rho_J} / \frac{P_{2J}}{W_{3J}P_{3J}}}$ = average distance between sampled clusters within urban or rural segments. (β_{3J} is a regression coefficient)^1

$S_1 = a_1+b_1\ d_{HA}$ = speed of travel between home and area (a_1, b_1 are regression coefficients)^2

$S_2 = a_2+b_2\ d_{SS}$ = speed of travel between sampled segments (a_2, b_2 are regression coefficients)^2

$S_{3J} = a_{3J}+b_{3J}\ d_{CC:J}$ = speed of travel between sampled clusters within urban or rural segments (a_{3J}, b_{3J} are regression coefficients)^2

$n_{SJ} = \frac{P_J}{W_1W_{2J}P_{2J}}$ = number of urban or rural segments in the sample in the province.

$n_{CJ} = \frac{P_J}{W_1W_{2J}W_{3J}P_{3J}}$ = number of urban or rural clusters in the sample in the province.

$v_{SJ} = C_{1J}+d_{1J}\frac{P_{2J}}{W_{3J}W_{4J}P_{4J}}$ = average number of visits per urban or rural segment during survey week (C_{1J}, C_{2J} are regression coefficients)^3

$v_{CJ} = C_{2J}+d_{2J}\frac{P_{3J}}{W_{4J}P_{4J}}$ = average number of visits per urban or rural cluster during survey week (C_{2J}, d_{2J} are regression coefficients)^3

$M_J = n_{SJ}v_{SJ}$ = total number of visits to all urban or rural segments in the sample.

$C_J = n_{CJ}v_{CJ}$ = total number of visits to all urban or rural clusters in the sample.

$P_{OJ} = \frac{P_J}{K}$ = urban or rural population per stratum

$\delta_{rJ} = e_J\ (P_{rJ}^{-f_J} - P_{OJ}^{-f_J})$ = intraclass correlation between pairs of persons within r-th stage units within stratum (e_J, f_J are regression coefficients)^4

N = number of "round trips" by all enumerators during a survey week. An enumerator may make one or more round trips during a day travelling from his home to sampled households, between sampled households and back home. N is estimated as the solution of the (linear) equation

$$Nt = T_e + T_{HA} + T_{SS} + T_{CC:U} + T_{CC:R}$$

after substituting for the quantities on the right hand side expressions (1), (3), (5) and (7).

Note 1: distances travelled between sampled r-th stage units are assumed to be directly proportional to the square root of the area in which they are located (i.e. the area of the (r-1)-st stage unit) and inversely proportional to the square root of the number of sampled r-th stage units in the area. The area of an (r-1)-st stage unit was estimated as the ratio of population over density of population. The home to area distance was assumed to be proportional to the square root of the area of the p.s.u. since an enumerator generally resides in his assigned p.s.u.

Note 2: the average speed of travel, within the range of distances involved, was assumed to be linearly dependent on the distance travelled.

Note 3: the average number of visits to be made to a sampled unit was assumed to be linearly dependent or the number of households to be enumerated in the unit. The effect of callbacks is hoped to be incorporated at this point.

Note 4: the intraclass correlation model applied here has been used by many authors (i.e. Hansen, Hurwitz, Madow: Sample Survey Methods and Theory, vol. 1, pp.307) but without the term involved P_{OJ}. This correction term was added to make the intraclass correlation between persons within stratum reduce to zero. It has little effect if P_{rJ}/P_{OJ} is small, however it has a noticeable effect at the p.s.u. level. This correction improved the fit between the curve and computed values of the intraclass correlation.

GENERALIZATIONS FOR COMPLEX PROBABILITY SAMPLING*

Leslie Kish, Survey Research Center, The University of Michigan

Summary

Most statistical theory is based on assuming simple random distribution of the sample elements. However, that assumption is only justified when either the population distribution is random or the sample design is simple random. Generalizations are needed for broader classes of samples in practical use. A broad class of design is *epsem*: each population element Y_i, ($i = 1, 2, \ldots, N$), has the same selection probability f. A broader class of designs is *probability sampling*: each population element has a known selection probability fP_i--with f and P_i known and positive. We show that *for all epsem* $E(y) = fY$, where $y = \sum_j y_j$, ($j = 1, 2, \ldots, n$) is the sample total. Similarly $E(\sum y_j/p_j) = fY$ for all probability samples. We also have $E(y/n) = Y/N$ for all epsem, if n is a fixed constant. If n is a variate, with $E(n) = fN$, y/n is a ratio mean; similarly for $E[(\sum y_j/p_j)/(\sum 1/p_j)]$. These results also apply to powers $Y_i = X_i^k$, and other functions $Y_i = g(X_i)$ of the population variables, and to vectors $Y_i = g(X_i, \ldots, Z_i)$. For example, we have $E(\sum y_j^2) = f\sum Y_i^2$. From this we also obtain $E[\sum y_j^2/n - (y/n)^2] = \sigma_y^2 - \mathrm{Var}(\bar{y})$ *for all epsem* (exactly if n is fixed).

1. General Statement

Consider a population of N elements, and a variable Y_i associated with each element ($i = 1, 2, \ldots, N$). The population total is $Y = \sum Y_i$, and the population mean is $\bar{Y} = Y/N$. *Probability sampling* is a selection method, an operation which insures that the *expected appearance* in the sample for the ith element is P_i', a known positive number. When sampling without replacement of elements, the ith element appears either once or not at all; then P_i' also represents the probability of selection of the ith element.

It is convenient to represent P_i' by its equivalent fP_i, where f is a known positive constant, a selection factor common to all elements. The P_i are positive values known for each element in the population. Thus *probability sampling without replacement of elements is an operation which assigns a known positive probability fP_i to each element*. Often the P_i are simple numbers, perhaps integers, associated with the elements.

For example, British households may be selected by applying the $f = 1/1000$ to the electoral list; the P_i denote the number of electors from the ith household on the electoral list. Or, we may select area segments with 1/1000, then subselect large, medium, and small farms with 1/1, 1/5, and 1/20; here $f = 1/20{,}000$, and the P_i take values of 20, 4, and 1. The selection factor f may be applied in different ways; the definitions and their consequences will hold. One may select every Fth listing, where $f = 1/F$; one may select n listings at random, where $n = fN$; or one may assign the probability f independently to each of the N listings. The selection is often more complex; multistage and multiphase designs may be used, with stratification and other techniques introduced at each stage. Nevertheless, the overall selection factor f is carefully controlled in probability samples.

A selection is *epsem* (equal probability selection method) when the selection probability is the same known constant f for all elements; that is, $P_i = 1$ for all i. When sampling elements with replacement, an element may appear more than once in the sample, and we substitute expected appearance for selection probability in the definition.

The number of sample elements is n, and the simple total of a variable present in a sample is $y_c = \sum_j y_j$ in epsem ($j = 1, 2, \ldots, n$). The analogous simple total for other probability samples is $y_c = \sum_j y_j/p_j$; here y_j and p_j denote the values Y_i and P_i where the jth sample element is the ith population element. This simple total is not the only possible estimator, but it is the one most frequently used in practice. It might be called the simple or symmetrical estimator, and it has theoretical justifications beyond its simplicity and naturalness. For all probability samples we have directly the simple relation

$$E(y_c) = fY. \tag{1}$$

This relationship follows directly from the definition of probability sampling, and from the basic relationship for a sum of random variables: $E(\sum_j y_j) = \sum_j E(y_j)$, (see section 5). Most statistical theory also assumes independence between the j selections, but that independence is lacking in complex selections. Contrariwise our aim is to find some useful results based on (1) alone, without assuming the independence of observations.

---

*This research was supported by grant G-7571 of the National Science Foundation. Another version appears in section 2.8 of my book, [Kish, 1965]. I am grateful for the suggestions of Morris H. Hansen, James G. Wendel, Bruce M. Hill, and William Ericson.

This and related results have wide utility. It is usually easy to state for a sample design if it is an epsem or a probability sample. The above relations, and others similar to them, then follow immediately, without having to derive them separately for the great variety of specific designs that are widely utilized for selecting samples. Many of these designs are complex, involving several stages with stratification at each stage, with random or systematic selection; multiphase sampling or controlled selection may also be used. With all their complexity, the desired overall selection probability of elements is maintained at fP_i in probability samples, with operations based typically on tables of random numbers. These operations are readily specified with practical office and field procedures.

Several other relations can be also derived easily from the above. When the sample size n is a fixed constant, the sample mean $\bar{y} = y/n$ is an unbiased estimate of $\bar{Y}$, because $n = fN$, and

$$E(y/n) = E(y)/n = fY/fN = \bar{Y}. \qquad (2)$$

Epsem with fixed n occurs in many varieties of element sampling, and in the sampling and subsampling of equal clusters. For all of these varieties of sampling the unbiasedness of y/f and y/n follows immediately without having to derive them laboriously and separately for the many types current in theory and practice.

Epsem with variable n occurs with unequal sized clusters; also when dealing with subclasses. For these, y/f is still unbiased, but y/n is generally not. With a fixed sampling fraction f, we have $E(n) = fN$, and $E(y/f)/E(y/f) = Y/N = \bar{Y}$. However, $\bar{y} = y/n$ is a ratio mean (the ratio of two random variables) and $E(\bar{y}) \neq \bar{Y}$ in general. Nevertheless, the ratio mean is widely employed and preferred.

The situation is similar for probability samples which are not epsem; when the selection probabilities are fP_i. Note that $E(\Sigma\, 1/p_j) = fN$; this is but a special case of (1), when $Y_i = 1$ for all i. The commonly preferred mean is a ratio mean similar to the mean above

$$\bar{y} = \frac{\Sigma\, y_j/p_j}{\Sigma\, 1/p_j} \text{ and } \frac{E(\Sigma\, y_j/p_j)}{E(\Sigma\, 1/p_j)} = \frac{fY}{fN} = \bar{Y}, \qquad (3)$$

but $E(\bar{y}) \neq \bar{Y}$ generally. The ratio mean y/n in epsem with variable n may be considered a special case of (3). Ratio means generally are accepted either as adequate approximations or as preferred statistics. My concentration on unbiased estimates and expectations is forced by the limitations of my capabilities and of the development of survey sampling literature. Although unbiasedness is given a prominent role, it is usually abandoned for the most important designs of survey sampling, such as unequal clusters. It is possible to make the analysis conditional on the denominator (n or $\Sigma\, 1/p_j$) found in the sample. This may also be done within the frame of likelihood functions instead of sampling distributions [Birnbaum, 1962; Raiffa, and Schlaifer, 1961].

2. Some Applications

Other important relations, similar to $E(y) = fY$ may be derived for all probability samples. The nature of Y_i was not and need not be restricted. It may represent $Y_i = X_i^2$, or $Y_i = X_i^k$, or $Y_i = X_i^k Z_i^m$. It may represent some function of the element values, or a function of several variables: $Y_i = g(Y_{1i}, Y_{2i}, \ldots, Y_{ki})$; this should be confined, I suppose, to real, finite, single-valued functions of the vector of variables defined on individual elements. Since Y_i may also represent the binomial variable $X_i < K$, where K is any fixed constant, it follows that the cumulative distribution function of the sample maintains the proportionality f to that of the population.

The importance of the principle can be illustrated by obtaining a much-needed result: estimates of the population variance σ_y^2 from any epsem or other probability sample. From the sample we construct n, $y = \Sigma\, y_j$, and $\Sigma\, y_j^2$, either self-weighted or properly weighted, $(\Sigma\, 1/p_j$, $\Sigma\, y_j/p_j$, and $\Sigma\, y_j^2/p_j)$. Since $E(n) = fN$, and $E(y) = fY$, and $E(\Sigma\, y_j^2) = f\Sigma\, Y_i^2$, we get

$$E(\sum_j^n y_j^2 - \frac{y^2}{n}) = E[(\sum_j^n y_j^2 - \frac{fY^2}{N}) - (\frac{y^2}{n} - \frac{f^2Y^2}{fN})]$$

$$= f(\sum_i^N Y_i^2 - \frac{Y^2}{N}) - E(\frac{y^2}{n} - \frac{f^2Y^2}{fN}).$$

Thus

$$E(nv_y^2) = fN\sigma_y^2 - E(\frac{y}{n} \cdot y - \frac{f^2Y^2}{fN}), \qquad (4)$$

where $v_y^2 = (\Sigma\, y_j^2/n - \bar{y}^2) = (n - 1)s_y^2/n$. We should like also to express the expectation of this element variance in the sample. When n is fixed at fN for the sample, we have directly that

$$E(v_y^2) = \sigma_y^2 - E\frac{(y^2 - f^2Y^2)}{(fN)^2} = \sigma_y^2 - \text{Var}(\bar{y}),$$

and

$$E[v_y^2 + \text{var}(\bar{y})] = \sigma_y^2, \text{ when } E[\text{var}(\bar{y})] = \text{Var}(\bar{y}). \qquad (5)$$

When n is not actually fixed, the analysis may be made conditional on a fixed n, and arrive at essentially the same result. Furthermore, it can be shown that the bias is bound to be usually small for nv_y^2/n considered as a ratio mean.

Hence,

$$E(v_y^2) \doteq \frac{E(nv_y^2)}{E(n)} = \sigma_y^2 - E(\frac{y}{n} \cdot \frac{y}{fN} - \frac{Y}{N} \cdot \frac{Y}{N})$$

$$\doteq \sigma_y^2 - R_{\frac{y}{n}\frac{y}{fN}} \sigma_{\frac{y}{n}} \sigma_{\frac{y}{fN}} . \qquad (6)$$

The second term becomes Var(y/n) for fixed n, and it should approach the mean-square-error of $\bar{y}$ when n is not fixed. Generally then $v_y^2 + \text{mse}(\bar{y})$ computed from the sample will be a good estimate of σ_y^2 (or S_y^2) among the population elements.

$\text{Var}(\bar{y})$ is roughly σ_y^2/n for many designs, and then $s_y^2 = v_y^2 n/(n - 1) = (\sum y_j^2 - y^2/n)/(n - 1)$ may be employed to estimate σ_y^2. For the special case for simple random sampling, when $\text{Var}(\bar{y}) = (1-f)S_y^2/n$, $E(s_y^2) = S_y^2$ follows immediately.

This result has great practical utility--and provided my chief motivation for this effort. Survey samplers find it useful to compute the ratio of the actual variance of a complex sample to the variance that a simple random sample based on the same number n of elements would have had. I called [Kish, 1965, 8.2] this the "design effect":

$$\text{deff} = \frac{\text{var}(\bar{y})}{(1-f)S_y^2/n} . \qquad (7)$$

Here then we may estimate $\hat{S}_y^2 = v_y^2 + \text{var}(\bar{y})$. Often $s_y^2 = nv_y^2/(n - 1)$ will serve well enough. Errors in estimating the second term of the denominator are smaller by the factor 1/n than the errors in estimating the numerator.

Methods for estimating the population covariance σ_{yx} between two variables Y_i and X_i are similar to those for estimating the population variance σ_y^2. Hence $[v_{yx} + \text{cov}(\bar{y}, \bar{x})]$ should be a good estimate, and s_{yx} often an acceptable approximation.

Then it should follow that the correlation coefficient $R_{yx} = \sigma_{yx}/\sigma_y\sigma_x$ can be well estimated from similar statistics. We may use $[v_{yx} + \text{cov}(\bar{y}, \bar{x})]/\sqrt{v_y^2 + \text{var}(\bar{y})}\ \sqrt{v_x^2 + \text{var}(\bar{x})}$ and often merely $v_{yx}/v_y v_x$, computed from any probability sample. These kinds of analytical statistics are frequently computed from complex probability samples, but without adequate (or any) justification--so far as I know.

3. Some Questions

Justification can be found, I believe, in the symmetry and proportionality of the sampling distribution of any epsem selection to the distribution of population elements. Those symmetries also hold for probability samples properly weighted with the $1/p_j$ values. I hope that others will obtain derivations of needed statistics for probability samples which are now available only for simple random samples.

Simple random sampling, with or without replacement, is often called "random sampling," or merely "sampling" in the statistical literature. "Complex" in the title refers to other types of probability sampling. Other epsem methods represent suppression of most of the $\binom{N}{n}$ combinations equally probably in simple random sampling. Yet they preserve the symmetry of equal selection probability of n/N for each element through equal numbers or expectations of the combinations in which it appears. Analogous properties of probability sampling, when not epsem, are more complex because they require the weights P_i. This symmetry is well known and exploited for simple random sampling, but not for other selection methods.

The symmetries of probability samples resemble those of simple random samples, and the sample moments will be similar to those of the population. What is missing from nonrandom samples is the independence of individual observations of random samples. In complex samples the observations are not independent, and the correlation between sample values may have serious effects. Hence, although $\bar{y}$ is a good estimate of $\bar{Y}$, and v_y^2 of σ_y^2, the variance of $\bar{y}$ may be much greater than v_y^2/n. Similarly, the sampling variabilities of other statistics computed from complex samples may differ greatly from those of simple random selections.

I believe these results are important for three reasons. First, to prove that $E(y_c) = fY$ consumes a half-page or page, for each of several types of the selection designs described in sampling textbooks. Second, these proofs imply that the reader must behave similarly when faced with other, perhaps more complicated designs. For example, a multistage design with stratification and systematic sampling at each stage may look formidable to the unwary; a "controlled selection" even worse. Instead one may say directly that "$E(y_c) = fY$, because y_c was based on a probability sample." Third, results are needed and can be had for other valuable statistics--similar to s_y^2 as estimator of S_y^2.

If the results are so important, and the method is so basic and simple, why has it been missed by mathematical statisticians? Because in their derivations they assume the independence of sample observations needlessly and too early. They ignore other methods of probability sampling, and assume simple random sampling without further thought. For example, instead of deriving a result in terms of $\text{Var}(\bar{y})$, they write it in terms of $\sigma_y/\sqrt{n}$. When simple random selection (independence of selection or observation) is unstated, a necessary condition has been omitted. When that assumption is stated, although it is sufficient, it may be an unnecessarily narrow restriction on the results.

For example, the variance of the function of several random normal variates is stated in terms of the covariance matrix $\sum_i \sum_j \frac{\sigma_{ij}}{n} \frac{\partial \bar{y}_i}{\partial \bar{Y}_i} \frac{\partial \bar{y}_j}{\partial \bar{Y}_j}$ in the best books [Rao, 1952]. This result follows also for complex samples, if stated in terms of $\mathrm{Cov}(\bar{y}_i, \bar{y}_j)$ instead of $\sigma_{ij}/\sqrt{n}$ [see Kendall and Stuart, 1958, 10.6].

To assume that the sample observations are random when the selection is complex, amounts to assuming that the population distribution is random. This is never exactly true, and--more important--it is often far from true.

4. Some Extensions

For an extension of probability sampling we coin the term *randomized sampling*: when the P_i are known, but the selection constant may be either known f or unknown f'. The selection probabilities of elements may be known (P_i) proportionately to each other, but with the common unknown factor f'. If $P_i = 1$ we have an *equal choice* selection method (*escem*), an extension of epsem. Many results for probability (and epsem) sampling can be extended to randomized (and escem) sampling.

The selection constant f' may be unknown in two important classes of problems: "urns" of unknown sizes and hypothetical universes. First, consider a list containing N + B listings; the presence of a number B of "blank" (empty) listings may render the number of elements unknown even if the list total N + B is known. For example, on the British electoral list families can be associated with uniquely defined family heads, but their numbers will remain unknown. Most lists assume this aspect when we analyse a subclass whose size N is unknown. If f is fixed and known for the entire list, it will be also for the subclass, but the subclass size n becomes a random variable. On the other hand, if n is fixed (with N unknown) f' becomes unknown. Of the three quantities involved in $f = n/N$, if two are fixed, so is the third, but a pair may remain not fixed together. Second, f' becomes unknown when inference from the results of a sampled population are extended to a larger, hypothetical, and indefinite universe with an assumed similar distribution of the variable Y_i. When the unknown f' may be considered an unknown constant, and when it must be treated as a random variable, we shall not consider here.

Note the vital fact that *subclasses inherit* from the entire sample the four broad classes of selection we discussed: epsem, equal chance, probability or randomized sampling. Therefore, the results above for an entire sample also holds for its subclasses. Fixed sample size, however, is not inherited generally by subclasses. But for this exception, simple random sampling is also inherited by the subclasses. On the other hand, other, complex, selection types--such as stratified, systematic selection, or equal sized clusters--are not generally inherited by subclasses.

5. Derivation of $E(y_c) = fY$.

Some colleagues, skeptical about the direct validity of (1) for all probability samples, demand some proof. Several others have pointed to as many distinct proofs, each with some claim for preference and reference. Still others say that no proof is necessary, since the basic rule about the expectation of a sum of random variables established $E(y_c) = fY$ for probability sampling directly after its definition. The following brief derivation may clarify the situation, and distinguish the established truth from other things which only resemble it.

We define C, the set of all possible samples c under a specified sampling plan applied to a population, and assume that C is finite. W_c is the probability of obtaining sample c; hence $\sum_{c \in C} W_c = 1$, summed over the entire distribution.

Let δ_{ic} represent the number of times the ith element appears in sample c; $\delta_{ic} = k$ when the ith element appears k times in sample c. When sampling with replacement, k may be 1, 2, ..., n. When sampling elements without replacement $\delta_{ic} = 1$ or $\delta_{ic} = 0$, if the sample does or does not contain the ith element. Thus δ_{ic} is a random variable associated with the ith population element representing its number of appearances. The expected number of appearances of the ith element is

$$E(\delta_{ic}) = \sum_c W_c \delta_{ic} = fP_i, \text{ for all probability samples} \quad (8)$$

and

$$E(\delta_{ic}) = f, \text{ for all epsem.}$$

Best known of epsem methods is simple random sampling: when each of the $\binom{N}{n}$ possible combinations receive the same selection probability, $W_c = 1/\binom{N}{n}$. Each of the N elements appears in $\binom{N-1}{n-1}$ combinations; thus the selection probability of each element is perceived as $\binom{N-1}{n-1}/\binom{N}{n} = n/N$. There are $\binom{N-1}{n-1}$ combinations which contain the ith element, when $\delta_{ic} = 1$. In the other possible combinations $\delta_{ic} = 0$, and their number is $\binom{N}{n} - \binom{N-1}{n-1} = (\frac{N}{n} - 1)\binom{N-1}{n-1}$. The expectation of δ_{ic} is $E(\delta_{ic}) = [\binom{N-1}{n-1} \cdot 1 + (\frac{N}{n} - 1)\binom{N-1}{n-1} \cdot 0]/\binom{N}{n} = \binom{N-1}{n-1}/\binom{N}{n} = n/N$, for all $\underline{i}$.

The random variable associated with the ith element is the number δ_{ic} of its appearances in the sample; whereas its value remains constant at Y_i. The contribution of the ith population element to the sample is $\delta_{ic}Y_i$, the product of its constant value Y_i with the random variable δ_{ic} that represents its appearance in the sample. The expected contribution of the ith element is

$$E(\delta_{ic}Y_i) = \sum_c Y_i\delta_{ic}W_c = Y_i \sum_c \delta_{ic}W_c$$

$$= Y_i\, E(\delta_{ic}) = Y_i fP_i, \text{ for all probability sampling,} \qquad (9)$$

and

$$E(\delta_{ic}Y_i) = Y_i f \text{ for all epsem.}$$

The sample total represents the sum of contributions for all those population elements which appear in the sample:

$$y_c = \sum_j y_j/p_j = \sum_i (Y_i/P_i)\delta_{ic}. \qquad (10)$$

Its expected value is

$$E(y_c) = \sum_i (Y_i/P_i)E(\delta_{ic})$$

$$= \sum_i (Y_i/P_i)(P_i f) = fY. \qquad (11)$$

References

[1] Birnbaum, A. [1962], "On the foundations of statistical inferences," *JASA*, 57, 297-306.

[2] Kendall, M. G. and Stuart, A. [1958], *The Advanced Theory of Statistics*, Vol. II., London: Griffin and Company.

[3] Kish, L. [1965], *Survey Sampling*, New York: John Wiley and Sons.

[4] Raiffa, H. and Schlaifer, R. [1961], *Applied Statistical Decision Theory*, Boston: Graduate School of Business Administration, Harvard University.

[5] Rao, C. R. [1952], *Advanced Statistical Methods in Biometric Research*, New York: John Wiley and Sons.

IV

MODELS OF POPULATION DYNAMICS

Chairman, Paul Demeny, Princeton University

Page

Models of Population Dynamics - Nathan Keyfitz, University of Chicago.......................... 70

Fertility Models with Social Parameters - James M. Beshers, Massachusetts Institute of Technology.. 76

A Strategy of Analysis of Variations in Family Structure: Actual Convergence and Ideal Patterns - Marion J. Levy, Jr., Princeton University.................................... 80

MODELS OF POPULATION DYNAMICS

Nathan Keyfitz, Population Research and Training Center
University of Chicago

SUMMARY

A population changing under the operation of a given set of mortality and fertility rates may be represented mathematically in at least three different ways. Each of these shows, among other things, how the age distribution changes in a trajectory towards ultimate stability. If the trajectory for the nine age groups from 0 to 45, say for females, is seen as a point moving in nine dimensional space, then all three methods may be interpreted as showing the reduction of the space from nine dimensions to three. The characteristic mortality and fertility pattern of real human populations is such that even at the outset of the projection nearly the whole movement is embraced in the three dimensions corresponding to the first three latent roots of the projection matrix. Just as factor analysis rotates a cluster of points in order to view it in as small a space as possible, this portion of the analysis of populations rotates a rigid twisted wire with the same purpose.

Partial fractions

The number of births which will occur during her lifetime to a girl child now born, under a given regime of mortality and survivorship, may be expressed in the form of a series of probabilities

$$f_1 = 0,\ f_2 = 0,\ f_3 = \frac{5\ell_{15}}{\ell_0} m_{15},$$

$$\dots\ f_9 = \frac{5\ell_{45}}{\ell_0} m_{45}. \qquad (1)$$

The ℓ's are the life table number living; the m's are the chance of having a child at the age specified. Translation into the more usual five-year age groups will be easily carried out later in the calculation; for the moment m_{15} may be thought of as the age-specific birth rate for the age group $12\frac{1}{2}$ to $17\frac{1}{2}$. If u_t is the number of births at time $\underline{t}$ to this girl or her descendants, and including as u_o this initial birth, then

$$u_o = 1;\ u_1 = 0;\ u_2 = 0;$$

$$u_3 = u_of_3 + u_1f_2 + u_2f_1 + u_3f_o = u_of_3; \qquad (2)$$

$$u_4 = u_of_4 + u_1f_3 + u_2f_2 + u_3f_1 + u_4f_o$$

$$= u_of_4 + u_1f_3;$$

and so on for as long as the regime of fertility and mortality persists.[1] If now we multiply the members of the set (2) by 1, s, s^2, s^3, etc., and add up the set as so multiplied, we have

$$u_o + u_1s + u_2s^2 + \dots$$

$$= 1 + (u_of_1 + u_1f_o)s +$$

$$(u_of_2 + u_1f_1 + u_2f_o)s^2 + \dots,$$

an equation which may be represented by

$$U(s) = 1 + U(s)F(s), \qquad (3)$$

U(s) being the generating function of the u's and F(s) of the f's. The solution of (3) for U(s) is

$$U(s) = \frac{1}{1 - F(s)} \qquad (4)$$

In order to base calculations on (4) we may expand it in partial fractions, a technique recommended by Feller[2] which has become popular in recent years. If s_1, s_2 . . . s_9 are the roots of the ninth degree polynomial equation

$$F(s) = f_o + f_1s + f_2s^2 + \dots = 1,$$

then the expansion of (4) in partial fractions gives for u_t the coefficient of s^t, assuming the roots distinct,

$$u_t = \frac{1}{s_1^{t+1}F'(s_1)} + \frac{1}{s_2^{t+1}F'(s_2)} + \dots + \frac{1}{s_9^{t+1}F'(s_9)}, \qquad (5)$$

where $F'(s_i)$ is the derivative of F(s) evaluated at $s = s_i$. But the s_i are approximated by $1/\lambda_i$, the reciprocals of the latent roots of the projection matrix. Knowing the λ_i we can calculate (5) by

$$u_t = \sum_{i=1}^{9} \frac{\lambda_i^{9+t}}{3\ell_{15}m_{15}\lambda_i^6 + \dots + 9\ell_{45}m_{45}}, \qquad (6)$$

which is a convenient form if the latent roots and their powers are at hand. It is found that after a very brief initial period, at most 20 or 25 years, the omission of all but the first three latent roots (when the roots are arranged in order of absolute value) makes little difference to the calculation; after about 100 years the dominant root suffices for a description. Between 20 and 100 years from the starting point

---

[1] A.J. Lotka, "Application of Recurrent Series in Renewal Theory," Annals of Mathematical Statistics, XIX (1948), 190-206.

[2] W. Feller, An Introduction to Probability Theory, Vol. I (2nd. ed.; New York: John Wiley & Sons, 1950), pp. 257-261.

considerable departures from stability are caused by the second and third roots, which are harmonic functions with period equal to one generation, gradually attenuated as the projection continues.

To extend the procedure to a more usual age distribution than one girl child just born, suppose now the initial population of an actual country to be K_0, K_5, K_{10}, etc., the number of persons in the population, but think of them as centered at exactly age 0, exactly age 5, etc. This concentration at points five years apart simplifies the formulae, and can be readily adapted to calculation at the final stage. We make the unit of the calculation K_0; in the new units we have $k'_0 = K_0/K_0 = 1$; $k'_5 = K_5/K_0$, etc. The stable population distribution which corresponds to this--ℓ_0, $\ell_5 e^{-5r}$, $\ell_{10}e^{-10r}$, $\ell_{15}e^{-15r}$--may be similarly divided by the number at age zero, which is the same as assuming that the radix of the life table is 1. Finally we find the departures of the k' from the stable population, and write the departures as k. Thus

$$k_0 = 1 - 1 = 0; \; k_1 = k'_1 - \ell_5 e^{-5r}; \qquad (7)$$
$$k_2 = k'_2 - \ell_{10}e^{-10r}, \text{ etc.}$$

A new generating function may be defined which will give the numbers of children to be expected from the population born before time t = 0, insofar as it departs from the stable age distribution:

$$B(s) = b_0 + b_1 s + b_2 s^2 + b_3 s^3 + \dots + b_8 s^8 \qquad (8)$$

in which

$$b_0 = \frac{k_{15}}{\ell_{15}}\ell_{15}m_{15} + \frac{k_{20}}{\ell_{20}}\ell_{20}m_{20} + \dots + \frac{k_{45}}{\ell_{45}}\ell_{45}m_{45}$$
$$b_1 = \frac{k_{10}}{\ell_{10}}\ell_{15}m_{15} + \frac{k_{15}}{\ell_{15}}\ell_{20}m_{20} + \dots + \frac{k_{40}}{\ell_{40}}\ell_{45}m_{45} \qquad (9)$$
$$b_2 = \frac{k_5}{\ell_5}\ell_{15}m_{15} + \frac{k_{10}}{\ell_{10}}\ell_{20}m_{20} + \dots + \frac{k_{35}}{\ell_{35}}\ell_{45}m_{45}$$

. . . .

$$b_8 = \frac{k_5}{\ell_5}\ell_{45}m_{45}; \; b_9 = 0$$

The set (9) may be expressed more compactly:

$$b_n = \sum_{i=1}^{9-n} \frac{k_{5i}}{\ell_{5i}}\ell_{5(i+n)}m_{5(i+n)} \qquad (10)$$

where n = 0, 1, 2, . . . 8.

On these definitions the relation of the generating functions is

$$U(s) = 1 + U(s)F(s) + B(s) \qquad (11)$$

and the solution

$$U(s) = \frac{1 + B(s)}{1 - F(s)}, \qquad (12)$$

which in partial fractions becomes

$$U(s) = \frac{1 + B(s_1)}{-F'(s_1)(s-s_1)} + \frac{1 + B(s_2)}{-F'(s_2)(s-s_2)} + \dots + \frac{1 + B(s_9)}{-F'(s_9)(s-s_9)} \qquad (13)$$

The coefficient of s^t in a form suitable for computation is

$$u_t = \sum_{i=1}^{3} \frac{[(1+b_0)\lambda_i^8 + b_1\lambda_i^7 + \dots + b_8]\lambda_i^{t+1}}{(3\ell_{15}m_{15}\lambda_i^6 + \dots 9\ell_{45}m_{45})} \qquad (14)$$

and evaluation for the first three roots is all that will be necessary. The only additional consideration for computation is that one calculates the b'_0, b'_1, etc. for the usual age groups 0-4, 5-9, etc., and then interpolates back half a period or $2\frac{1}{2}$ years to the points 0, 5, etc. required by (14): $b_0 = \frac{1}{2}b'_{0-4}$; $b_1 = \frac{1}{2}(b'_{0-4} + b'_{5-9})$, etc., and similarly for the $\ell_{15}m_{15}$ in the denominator.

The Matrix

A population projection through a five-year period may be represented as the premultiplication of the distribution of ages, treated as a vertical vector $\{K'_0\}$, by a suitably constructed matrix M. The form of M is

$$M = \begin{bmatrix} 0 & 0 & m_{15} & . & . & m_{40} & m_{45} \\ \frac{{}_5L_5}{{}_5L_0} & 0 & 0 & . & . & 0 & 0 \\ 0 & \frac{{}_5L_{10}}{{}_5L_5} & 0 & . & . & 0 & 0 \\ . & . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & 0 & \frac{{}_5L_{40}}{{}_5L_{35}} & 0 \end{bmatrix}, \qquad (15)$$

and the projection operation is then represented by $M\{K'_0\} = \{K'_1\}$. The properties of this way of representing the changes in a population were investigated in detail by Leslie.[1]

With M as a starting point we define the latent roots as the values of λ which satisfy the determinantal equation $| M - \lambda I | = 0$, and which may be called $\lambda_1, \lambda_2, \ldots \lambda_9$, when arranged in order of their absolute values (see Table I). Corresponding to each of the roots is a modal vertical vector $\{K_i\}$ satisfying

$$M \{K_i\} = \lambda_i \{K_i\}, \quad i = 1,2, \ldots 9. \qquad (16)$$

TABLE I. LATENT ROOTS FOR THREE COUNTRIES, FEMALES, 1961

	Latent root	Absolute value
	England and Wales, 1961	
λ_1	1.0503	1.0503
λ_2, λ_3	.3593 ± .7727i	.8521
λ_4, λ_5	-.3902 ± .4338i	.5835
λ_6, λ_7	-.0096 ± .5113i	.5114
λ_8, λ_9	-.4846 ± .1454i	.5059
	United States, 1961	
λ_1	1.1081	1.1081
λ_2, λ_3	.3043 ± .7931i	.8495
λ_4, λ_5	-.4109 ± .3895i	.5662
λ_6, λ_7	.0264 ± .5421i	.5427
λ_8, λ_9	-.4739 ± .1617i	.5007
	Hungary, 1961	
λ_1	.9763	.9763
λ_2, λ_3	.2663 ± .7004i	.7493
λ_4, λ_5	-.3676 ± .3780i	.5273
λ_6, λ_7	.0748 ± .4882i	.4939
λ_8, λ_9	-.4616 ± .1721i	.4926

There are also 9 horizontal vectors $[H_i]$ satisfying

$$[H_i]\, M = \lambda_i \, [H_i], \quad i = 1,2, \ldots 9. \qquad (16a)$$

The first of the stable vertical vectors, $\{K_1\}$, is well known as the usual stable population. The first of the stable horizontal vectors, $[H_1]$, was shown by R.A. Fisher to represent in a certain sense the reproductive value of a woman from age 0 to 45.[2] For our purpose the important property is that of orthogonality:

$$\begin{aligned} [H_i]\,\{K_i\} &\neq 0, \quad i = 1,2, \ldots 9; \\ [H_i]\,\{K_j\} &= 0, \quad i \neq j = 1,2, \ldots 9. \end{aligned} \qquad (17)$$

This enables us to evaluate the c_i in the equation

$$\{K'\} = c_1\{K_1\} + c_2\{K_2\} + \ldots c_9\{K_9\}, \qquad (18)$$

since premultiplying both sides of (18) by $[H_i]$ and transposing gives

$$c_i = \frac{[H_i]\,\{K'\}}{[H_i]\,\{K_i\}}, \qquad (19)$$

the right hand side of (19) being the ratio of two scalar quantities, which except for c_1 will be complex. As in the partial fractions approach, it will turn out that only the first three terms of the expansion (18) are needed in practical work. This arises from the relative absolute values of the latent roots, shown for three countries in Table I.

The use of (18) to expand $\{K'\}$, an actual age distribution, in terms of the unprimed $\{K_i\}$ which are stable vectors, appears from the premultiplication of (18) by the powers of M. By virtue of (16), applied *t* times, the projection of the actual population to time *t* is

$$M^t\{K'\} = c_1\lambda_1^t\{K_1\} + c_2\lambda_2^t\{K_2\} + c_3\lambda_3^t\{K_3\}, \qquad (20)$$

where we stop with the third root. This result may be thought of as providing a kind of graduation for the trajectory of the population, age by age, as long as the life table and fertility rates hold. It is nine separate scalar equations, one for each element of the 1 × 9 vector. The first of these, representing the top element corresponding to age 0-4, would be practically identical with the value for u_t given in (14).

On the other hand, if we start from (14) we could work towards (20) by the consideration that the population at time *t* aged 35, say, would be that in (14) at time t - 7 (the unit of time here is five years) as multiplied by the appropriate survival factor. The fitting to United

---

[1] P.H. Leslie, "On the Use of Matrices in Certain Population Mathematics," *Biometrika*, XXXIII(1945), 183-212; "Some Further Notes on the Use of Matrices in Population Mathematics," *Ibid.*, XXXV (1948), 213-245.

[2] R.A. Fisher, *The Genetical Theory of Natural Selection* (2nd rev. ed.; New York: Dover Publications, 1958), pp. 27-30.

TABLE II. ANALYSIS OF UNITED STATES FEMALES, 1961, BY MATRIX APPROACH

(000's)

Age	Actual Population 1961 $\{K'\}$	Components of Fitted Stable Populations		
		Dominant root $c_1\{K_1\}$	Principal pair of complex roots $c_2\{K_2\} + c_3\{K_3\}$	Other roots plus error $c_4\{K_4\} + \dots + c_9\{K_9\}$
0-4	10,122	10,022	74	26
5-9	9,424	9,003	601	-180
10-14	8,784	8,114	610	60
15-19	6,781	7,308	-317	-210
20-24	5,725	6,576	-1,076	225
25-29	5,489	5,912	-494	71
30-34	5,965	5,308	1,106	-449
35-39	6,394	4,755	1,603	36
40-44	6,045	4,245	-168	1,968

15-year Projection on 1961 Fertility and Mortality Schedules

Age	Population Projected to 1976 $M^3\{K'\}$	Components of Fitted Stable Populations		
		Dominant root $c_1\lambda_1^3\{K_1\}$	Principal pair of complex roots $c_2\lambda_2^3\{K_2\} + c_3\lambda_3^3\{K_3\}$	Other roots plus error $c_4\lambda_4^3\{K_4\} + \dots + c_9\lambda_9^3\{K_9\}$
0-4	13,782	13,633	202	-53
5-9	11,975	12,248	-238	-35
10-14	10,595	11,038	-479	36
15-19	10,044	9,942	-75	177
20-24	9,364	8,945	597	-178
25-29	8,707	8,043	605	59
30-34	6,701	7,221	-313	-207
35-39	5,632	6,469	-1,090	253
40-44	5,361	5,774	-483	70

30-year Projection on 1961 Fertility and Mortality Schedules

Age	Population Projected to 1990 $M^6\{K'\}$	Components of Fitted Stable Populations		
		Dominant root $c_1\lambda_1^6\{K_1\}$	Principal pair of complex roots $c_2\lambda_2^6\{K_2\} + c_3\lambda_3^6\{K_3\}$	Other roots plus error $c_4\lambda_4^6\{K_4\} + \dots + c_9\lambda_9^6\{K_9\}$
0-4	18,739	18,547	-192	24
5-9	16,734	16,662	34	38
10-14	15,294	15,017	294	-17
15-19	13,676	13,525	201	-50
20-24	11,899	12,170	-236	-35
25-29	10,502	10,942	-574	134
30-34	9,925	9,824	-73	174
35-39	9,212	8,801	588	-177
40-44	8,504	7,855	590	59

States females for 1961, along with the projection over 15 and over 30 years, is shown as Table II. By the end of 30 years the effect of the fourth to ninth roots has become small.

The Integral Equation

The integral equation, being better known,[1] may be treated more briefly than the two preceding methods. With a continuous formulation $B(t)$, the number of births at time t, is

$$B(t) = \int_0^\beta B(t-a)\ell(a)m(a)da, \tag{21}$$

where $\ell(a)$ is the number surviving to age a on the same life table of radix unity. If one tries for the solution

$$B(t) = Ae^{rt},$$

(21) is converted into an equation for r

$$\int_0^\beta e^{-ra}\ell(a)m(a)da - 1 = 0. \tag{22}$$

Clearly (22) has the same role here that $F(s) - 1 = 0$ has in the partial fraction formulation, and that $| M - \lambda I | = 0$ has for the matrix. Unlike the previous approaches, however, there are here an infinite number of solutions, rather than nine, and here only one can be real, while previously only one could be positive but real negative roots were possible. While (22) could be solved by fitting moments to the net maternity function, and solving a quadratic or higher polynomial, the point made by Coale[2] that a direct iterative solution not dependent on moments is preferable applies to the complex as well as to the real root. In terms of the real root ρ and the series of complex roots r_1, r_2, etc. the solution to (21) is expressible as a series

$$B(t) = A_o e^{\rho t} + A_1 e^{r_1 t} + A_2 e^{r_2 t} + \ldots \tag{23}$$

Since (21) is linear in $B(t)$, the sum of the solutions, each with an arbitrary weight, is also a solution.

Just as before the s_i, the λ_i, and the $\{K_i\}$ represented the pattern of fertility and mortality and had nothing to do with the actual age distribution, so this is true of the ρ and the r's in (23). The A's, on the other hand, correspond, like the b's of the partial fractions and the c's of the fitting to the stable vectors, to the matching to initial conditions--the age distribution of the real population that we are attempting to analyze. If $G(t)$ is the births at time t to that portion of the population itself born before $t = 0$, which starts the system off, so to speak, the coefficient in the term $A_i e^{r_i t}$ is

$$A_i = \frac{\int_0^\beta e^{-r_i t}G(t)dt}{\int_0^\beta ae^{-r_i a}\ell(a)m(a)da}, \tag{24}$$

where again $\ell(a)$ is the number living at age a in the life table of radix 1. The denominator of (24) is the derivative with respect to r, evaluated at the root r_i, of the quantity on the left hand side of (22), just as the denominator of (5) and (6) is the derivative of $1 - F(s)$ evaluated at s_i.

* * * * *

When the three methods for which computing procedures are given above are applied to the usual five-year age groups they turn up different finite approximations. The numerical discrepancies could be made smaller by the methods of finite differences. Some work has been done on the theoretical reconciliation of the matrix and recurrence approaches of equations (3) and (21) above, especially by Feller,[3] and Lopez[4] has helped to relate the matrix formulation to the integral equation, as well as the recurrence equation to both.

There are more important issues for demographers than reconciliation of numerical differences or bridging of the theories which underlie them. They would like to see the methods used to study and classify actual populations; analysis of the two sexes treated together rather than each in abstraction from the other; extension to the recognition of parities and, in general, rates specific for variables other than age and sex; the use of the methods to treat sets of data representing real cohorts, rather

---

[1] A.J. Lotka, *Theorie Analytique des Associations Biologiques*. Part II (Paris: Hermann, 1939); A.Lopez, *Problems in Stable Population Theory* (Princeton University Press, 1961); E.C. Rhodes, "Population Mathematics I, II, and III," *Journal of the Royal Statistical Society*, CIII (1940), 61ff, 218ff, and 362ff; A.J. Coale, "How the Age Distribution of a Human Population is Determined," *Cold Spring Harbor Symposia on Quantitative Biology*, XXII (1957), 83-89.

[2] A.J. Coale, "A New Method for Tabulating Lotka's r," *Population Studies*, XI (1957), 92ff.

[3] W. Feller, "On the Integral Equation of Renewal Theory," *Annals of Mathematical Statistics*, XII (1941), 243-267.

[4] A.Lopez, *Op. cit.*

than the period or cross-section data applied in the foregoing.[1] A function that would generate probabilities of various numbers of prospective population, rather than the expected values generated by all three of the above methods, would escape the disadvantage of the deterministic model pointed out by Feller exactly a quarter of a century ago.[2] Each of the above directions of extension of the work here summarized presents difficulties, but none of these are insuperable.

---

[1] Norman B. Ryder, "The Process of Demographic Translation," Demography, I (1964), 74-82.

[2] W. Feller, "Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung," Acta Biotheoretica, V (1939), 11-40.

FERTILITY MODELS WITH SOCIAL PARAMETERS

James M. Beshers, M.I.T.

I

In this paper we represent three aspects of fertility behavior as functions of social variables: first, the diffusion of family planning across a society; second, the determination of an initial ideal family size for a planning couple; and third, the achievement of a given family size throughout the child bearing cycle of a planning couple.

The first formulation is drawn from an earlier paper;[1] it is empirically suggested by research by Keyfitz.[2] The second formulation represents fairly conventional theories of the transmission of values in sociology; much American survey data on "ideal family size" provides the empirical substance.[3] The third formulation is an effort to represent the argument of J. A. Banks as presented in Prosperity and Parenthood.[4]

The social parameters will be defined over an n-class social system.[5] The set of classes C_1, C_2, C_n will be arranged from high to low with order determined by the matrix of probabilities of inter-marriage among classes.[6] An unambiguous order will be assumed to exist for this paper.[7] Further, it will be assumed that the distribution of other relevant contacts among classes will be approximately indexed by the distribution of marriages.[8]

II

Thus we shall define a concept of social distance over the inter-marriage matrix. Up to a rank order social distance will be defined by the subscripts of the classes. We shall further assume that a metric can be defined over the classes in two vectors, one representing upward distance and one representing downward distance. The elements of both vectors will be represented as d_{ij} with the relative magnitude of i and j determining the direction of the distance and $d_{ij} \geq 1$ if $i \neq j$. (The technical issues involved in the definition and estimation of such vectors are discussed elsewhere.)[9]

A similar analysis over a system of areas would utilize the volume of interchanges of migrants as indices of the "social distance" among inhabitants. Thus a village that has a high rate of interchange with a city will be "closer" to the city than a village with a low rate of interchange. Ideally we would like to combine social classes and areas in a single set of equations. It is just such a comparison that is made by Keyfitz.

III

Suppose family planning enters the social system in a given class, then spreads. For simplicity let us suppose that family planning enters at one "end" of the system, say C_1. Let us define the proportion planning in C_1 as $\mathcal{P}_1$, in C_2 as $\mathcal{P}_2$, and so on to C_n as $\mathcal{P}_n$. Then at time t we have $\mathcal{P}_1(t) > 0$, $\mathcal{P}_2(t) = \mathcal{P}_3(t) = - - - = \mathcal{P}_n(t) = 0$.

Given that the order of classes is in the order defined by the "social distance" among them, then we expect that the "barriers" between classes will be broken down in the order of the subscripts.

In order to further specify the values of $\mathcal{P}_i(t)$, we must supply within class diffusion parameters, we must supply a functional form for the rate of diffusion and we must specify a parameter that indexes the magnitude of the barrier.

One such formulation that has the merit of simplicity is the following:

$$\text{If } \left(t - \sum_{i=2}^{n} \rho_{i-1,\, i}\right) \geq 0$$

$$\text{Then } \mathcal{P}_i\left(t - \sum_{i=2}^{n} \rho_{i-1,\, i}\right) =$$

$$1 - e^{-k_i\left(t - \sum_{i=2}^{n} \rho_{i-1,\, i}\right)}$$

$$\text{If } \left(t - \sum_{i=2}^{n} \rho_{i-1,\, i}\right) \angle 0$$

$$\text{Then } \mathcal{P}_i = 0$$

where k_i is a parameter indexing the within class rate of diffusion and ρ is a parameter indexing the barrier between adjacent classes. We can define ρ as follows: $\rho_{i-1,\, i} =$

$$\frac{d_{i-1,\, i} + d_{i,\, i-1}}{2}$$

IV

For the proportion of a class that plans we shall define an ideal family size ξ_j in the class C_j. How shall we determine $\xi_j(t)$?

Let us assume that $\xi_j(t)$ is initially defined at the marriage of a couple and that we wish to date the marriage by the year of birth of the woman (the cohort of birth in the data of P. K. Whelpton). We shall use T as the time index for year of birth.

Thus we have ξ_j^T for each married couple. We shall allow this number to vary in successive cohorts, but it shall be a constant for the life cycle of any particular couple, independent of t.

We shall represent ζ^T as dependent upon the ascribed status of the couple, the achieved status of the couple, and prestige effects specific to the class membership of the couple.

To summarize: ideal family size for a given cohort of married women is determined by the successive application of three functions: f_1, f_2 and f_3.

The function f_1 specifies the influence of the values of the parents of the couple upon the values of the couple, the function f_2 specifies the influence of social mobility of the couple upon their ideal family size values, and the function f_3 specifies a prestige drift throughout the social system that changes ideal family size in the direction of the highest social classes.

Thus we have, aside from linear constants,

$$f_1: \zeta_k^T = \frac{\zeta_i^{T-G_i} + \zeta_j^{T-G_j}}{2}$$

where C_k is the class expected for the couple conditional that the wife's parents are in Ci and the husband's parents are in Cj, where $\zeta_i{}^{T-G_i}$ is the ideal family size of parents of wife and $\zeta_j{}^{T-G_j}$ is the ideal family size of parents of husband, G stands for generation. Next,

$$f_2: \zeta_h^T = \frac{1}{2} \left(\zeta_k^t + \frac{\zeta_h^{T-1}}{d_{k,h}} \right),$$

where ζ_h^{T-1} is last year's cohort ideal family size and $d_{k,h}$ is distance from C_k implied by parental status to C_h achieved by couple.

$$f_3: \zeta_g^T = \zeta_g^{T-1} + \frac{\zeta_g^{T-1} - \zeta_{g-1}^{T-1}}{d_{g-1,g}} +$$

$$\frac{\zeta_g^{T-1} - \zeta_{g-2}^{T-1}}{d_{g-2,g}} + \ldots\ldots\ldots$$

$$\zeta_m^T = \frac{\zeta_h^T + \zeta_g^T}{2}$$

where ζ_g^{T-1} is last year's cohort value for the class achieved by the couple and the succeeding terms represent the influence of higher classes attenuated by a distance divisor. Thus ζ_m^T depends upon i, j, k, h and g.

If we include linear constants as weights for the influence of the three types of effect, then we have

$$f_1: \zeta_k^T = a_1 \zeta_i^{T-G_i} + (1-a_1) \zeta_j^{T-G_j}$$

$$f_2: \zeta_h^T = a_2 \zeta_k^T + (1-a_2) \frac{\zeta_h^{T-1}}{d_{k,h}}$$

$$f_3: \zeta_g^T = \zeta_g^{T-1} + \frac{\zeta_g^{T-1} - \zeta_{g-1}^{T-1}}{d_{g-1,g}} +$$

$$\frac{\zeta_g^{T-1} - \zeta_{g-2}^{T-1}}{d_{g-2,g}} + \ldots$$

$$\zeta_m^T = a_3 \zeta_h^T + (1-a_3) \zeta_g^T$$

where
$$\begin{pmatrix} 0 \leq a_1 \leq 1 \\ \\ 0 \leq a_2 \leq 1 \\ \\ 0 \leq a_3 \leq 1 \\ \end{pmatrix}$$

By suitably choosing the values of the constants we may express alternative theories for the influences of parents, present class, and prestige drift or fashion.

Consider the vector (a_1, a_2, a_3), in which we allow the elements to be either 0 or 1. Clearly a_2 and a_3 differentiate the theories alone, a_1 provides alternative theories for the relative influence of parents:

If $a_2 = 0$, then no parental influence;
" " $= 1$, then no mobility influence;
" $a_3 = 0$, then no parental or mobility influence; and
" " $= 1$, then no prestige influence.

In general we will allow the elements to range from 0 to 1 inclusive. We now have ζ_m^T defined - the cohort X class ideal family size.

Next let us consider the achievement of the ideal family size throughout the life of the cohort of planners.

In each year we have $p_{i,\ i+1}^{X}$, the probability of an additional child in a year at age x to a woman with i previous births. We shall define $p_{i\ i=1}^{x,\ldots} = 0$ if the woman is never married. At first marriage $p^x_{i,i+1} > 0$. We wish to make $p^x_{i\ i+1}$ depend upon the ideal family size τ, the expected future income I(x) and biological capacities declining with age h(x). One way to do this is to table h(x) for non-planning populations as a complete set of $p_{1,i+1}$. Then the effect of τ and of I(x) is to reduce a subset of values of $p^x_{i,i+1}$ to zero.

Let us first express p^x_i, i+1 as a function of τ and i. Then we will consider I(x), and then h(x).

$$g_1:\ p^x_{i,\ i\ +\ 1} \begin{matrix} > 0 \text{ if } \tau & - i > 0\ ) \\ = 0 \text{ if } \tau & - i \leq 0\ ) \end{matrix}$$ Note i is a monotonically increasing function of x.

In order to specify g_2 we must introduce new definitions. For each cohort X class combination we have two constants expressed in dollars, $S_j \equiv$ standard of living for self and $M_j \equiv$ standard of living for a child. These two constants may be scaled to the annual income or to the total career income of the household. We further have a variable $I_j(x) \equiv$ expected future income which may be scaled either to annual or total income. We should define a total career income curve and integrate for annual income, then do the same for S and M. In this way intermittent spacing to meet costs of college education can be represented in the model. Further, since we allow I(x) to be reestimated annually the fluctuations can be represented in the model.

It seems desirable that, at marriage, the values of I, S, M, and τ are compatible. This will be achieved over the career if, for any household,

$$I\ -\ S \geq M\tau$$

In order to preserve this relationship in the annual values we define

$$g_2:\ p^x_{i,i+1} > 0 \text{ if } (I_j(x)-S_j) \geq iM_j$$

$$p^x_{i\ i+1} = 0 \text{ if } (I-S) < iM.$$

Strictly speaking, Banks argues that S can be hedged to preserve M in the face of a drop in I(x) but this is incidental to our problem. Note that the difference between self-employed and salaried can be introduced in estimates of I(x).

The function g_3 comes from the analysis of non-planners

$$g_3:\ p^x_{i,i+1} = h\ (x)$$

where h (x) is a function representing the biological constraints of age and spacing upon birth probabilities assumed fixed for all cohorts. Thus,

$$p^x_{i,i+1} = 0 \text{ if } g_1 \text{ or } g_2 \text{ imply zero,}$$
$$= h(x) \text{ otherwise.}$$

Therefore we use h(x) as g_3 to introduce biological constraints into our birth projections for the planners, and we also use h(x) to express the births of the non-planners directly.

There are a number of sticky issues in the estimation of the parameters defined in this paper. However, in most cases the parameters may be directly estimated and indirectly estimated. For example, the proportion planning and the ideal family size may be directly estimated by appropriate surveys. These parameters may be indirectly estimated by the use of the model itself. With minor adjustments the model will deduce a whole series of demographic parameters, such as birth rates, average completed family sizes and child spacing that may be compared with such measures in a population. The implied values of some parameters are then indirectly estimated under the hypothesis that the other parameters are correctly estimated by calculating the values of the parameters in question that would be consistent with the demographic data; the degree of precision of such indirect estimation will, of course, depend upon the number of degrees of freedom that are available in turn the number of degrees of freedom are in part dependent upon one's confidence in his assumptions and direct estimates of other parameters.

Let me consider the estimation of two kinds of parameters in greater detail. The expected future income I(x) is a subjective phenomenon. Good direct estimates can only be made in carefully designed panel studies that utilize detailed theories of cognitive processes (the Bayesian concept of subjective probabilities is relevant here - we wish to predict the distribution of such numbers in a population). Ultimately we should make I(x) depend upon such numbers as income measured objectively in the population. However, if we recall that this parameter has two purposes: (1) to enable us to represent planned spacing; and (2) to enable us to represent hedging against down turns in expected future income, then we can test the model against data that we believe strongly reflect these two phenomena and attempt indirect estimation.

The biological function h(x) bears some relation to a number of computations that have been made by demographers on "non-contraceptive" populations. Ideally we would like to have cohort birth probabilities of the type that can be reconstructed from Whelpton's tables for the U.S. We would need these values for several different non-contraceptive populations so that we could take account of the effects of customs such as sexual abstinence or the sporadic use of birth control.

An alternative procedure might involve the use of a continuous stochastic process model, such as a semi-Markov process, fitted to demographic parameters of a non-contraceptive population. A discrete table of probabilities might then be calculated by integration.

This paper clearly leaves much unfinished business. Yet I hope that its main goal is accomplished, namely that some of you are now convinced that social and psychological parameters can be mathematically defined and introduced into birth projection models.

FERTILITY MODELS WITH SOCIAL PARAMETERS
Footnotes

1. "Birth Projections with Cohort Models" by James M. Beshers, paper read at the annual meeting of the Population Association of America, July 1964, forthcoming in Demography.
2. "A Factorial Arrangement of Comparisons of Family Size," by Nathan Keyfitz, American Journal of Sociology, Vol. 58 (March, 1953), pp. 470-479.
3. In particular we have the influence of parental status as expressed by Goldberg, upward mobility as expressed by Westoff, and a prestige drift that represents fashion. See D. Goldberg, "The Fertility of Two-Generation Urbanites," Population Studies, 12 (1959): 214-44, and Westoff, Charles, "The Changing Focus of Differential Fertility Research: The Social Mobility Hypothesis" in J. J. Spengler and O. D. Duncan (eds.) Population Theory and Policy (Glencoe, Ill.: Free Press, 1956).
4. Banks argues that expected future income, standard of living desired for self, and standard of living desired for children all must be taken into account to determine the influence of income upon patterns of child spacing as well as the influence upon total number of children. See J. A. Banks, Prosperity and Parenthood. (Humanities Press, N. Y., 1954).
5. See James M. Beshers and Stanley Reiter "Social Status and Social Change" Behavioral Science, Vol. 8, No. 1 (Jan. 1963).
6. Note that race, religion, nationality, occupation and other social variables may enter into the definition of these classes.
7. James M. Beshers "Urban Social Structure as a Single Hierarchy," Social Forces, Vol. 41, No. 3 (March, 1963) and Urban Social Structure (New York: Free Press, 1962).
8. Ibid.
9. Paper in progress in collaboration with Edward Laumann.

A Strategy of Analysis of Variations in Family Structure: Actual Convergence and Ideal Patterns

Marion J. Levy Jr., Center of International Studies, Princeton University

Ladies and Gentlemen, it is beyond the everyday call of my arrogance for me to be here at all since I am in no sense expert or even literate in either statistics or demography on which the work I shall propose depends. In these matters I am fortunate in being able to rely on the help of scholars such as your chairman, Paul Demeny, Carl E. Helm, Roger S. Pinkham, Frederich F. Stephan, and Charles F. Westoff. Without their help and past help from Ansley J. Coale I would not continue this work.

This work was undertaken using statistics, demography and computer facilities but not in the spirit or hope that serendipity was a major method of analysis instead of a by-product of science. Rather I believe that it is possible to combine these elements and come to conclusions about the range of variations of actual family structures having to do with numbers of members, generations represented, age distribution, sex distribution, marital spouse pair distribution, sibling distribution, etc. that are in important respects independent of many other myriad variations in the patterns of the societies presently, previously, or likely to be extant.

This is an attempt to arrive at extremely general knowledge of both theoretical and practical significance in terms of an explicit strategy of analysis. It seeks to do this in terms of a small number of considerations [variables and constants] chosen in such a way that their implications for the empirical phenomena to which they are presumed to apply are not likely to be erroneous or misleading even though no attempt at descriptive treatment is made.

The analysis of family phenomena is the focus because: 1. The family as that term is used here is a type of organization found as a sub-system of all known societies, 2. The vast majority of individuals in all known societies have roles in some family organization throughout their life histories, and, 3. The behavior of individuals in all other context ideally and/or actually is extremely likely to be affected by what happens to them in a family context. For these reasons any highly generalized knowledge about family contexts is likely to have general implications for all social phenomena.

The guiding principle of this strategy of analysis is that failure to distinguish systematically between ideal and actual patterns has led us to overlook many actual uniformities, while perceiving important variations in ideal patterns.

The second element in this strategy is to stick as closely as possible to the implications of biological and psychological factors or aspects for social analysis--to abandon the overreaction to the past excesses of biological determinists, etc.

The third principle of this strategy is that the findings of demography constitute our most reliable data-tool when we seek hypotheses whose applications by their very generality cut across wide varieties of language and other social distinctions. That we are less likely to go wrong asking how many died and how many were born and then scouting the implications of this for what interests us than by asking most other questions. Furthermore, the work already in hand by demographers is of a high order of sophistication, and no other set of findings of such general implications or subjected to such high order critical scouting is available.

One of the most obvious examples of this strategy is the certainty with which we can dispose of the implication that a polygynous society is one whose male members upon reaching maturity generally acquire two or more wives. It cannot be so unless peculiar circumstances converge, and it has probably never been so in history. These are societies whose members have polygyny as an ideal pattern--probably realized by a small minority of males who by that alone would be considered an elite set. The reasoning is simple. Everywhere the ratio of males to females at birth is very slightly in excess of unity--varying say from 103/100 to 107/100. Under these circumstances only mortality rates of a peculiar sort would make it possible for the great majority or even a bare average of men to have two or more wives. Realization of this of course, expands the implications of the presence of polygyny as an ideal pattern.

The present project grew out of an essay pending publication. [Aspects of the Analysis of Family Structure, Princeton University Press, Spring `65.]

In that essay the following distinctions were made:

1. Societies were distinguished on the basis of the presence or absence of modern medical technology and modernization in general.

2. The maximum range of size of family membership is covered by the range of patterns from nuclear to extended families.

3. Patterns were distinguished as ideal [i.e., believed good and proper by the persons referred to] and/or actual [i.e., extant as an observed pattern of behavior rather than of belief].

Given the demographic findings and probabilities the hypothesis was constructed that well over 50% of all family sizes would fall under the same curve of distribution characteristics of societies with ideal patterns calling for nuclear families. The index for family size only varied between 1 and 1.75 save in cases of late comers to modernization who had imported modern technology but not yet the rest of modernized patterns. That state was predicted to be unstable.

The line of reasoning employed [e.g., in the absence of modern medical technology few grandparents long survive the birth of grandchildren--ideal patterns to the contrary not withstanding] has implications for other considerations than size of family membership. Specifically it has implications for: 1. age distributions, 2. sex distributions, 3. marital pair distribution, 4. sibling distribution (male siblings, female siblings and mixed), 5. generational distributions, etc. These distributions plus that of size of membership certainly do not determine all other family variations and possibilities--but they are certainly not without implications for many or most of them. By use of the computer and the strategy of analysis indicated we seek to show what the actual variations possible are in any of these matters given any possible combinations of ideal patterns. If the strategy of analysis is effective, any possible use must fall at or close to one of the points here identified in terms of pau~~s~~ce variables.

Without attempting to use a computer to "run everything against everything," I hope with the aid of my colleagues to test the actual implications of varying ideal patterns [from nuclear family structures through stem family structures, to extended family structures] of various factors chosen as setting parameters within which all actual cases must fall. For example ratios of birth to death rates starting at say 80/80 and going down in intervals of 5 [e.g., 80/75, 80/70 . . .; 75/80, 70/80 . . .] and running to 10/10 will be used. Marital ages set at 11, 13, 15, . . . 39 will be used. Sub-routines, to test whether maximum possible error [one year in the case of ideal marital ages] can make much difference to the various outcomes predicted, can be set up.

If this simulation can be carried out--and the program for it in IPL is being written now, we hope to have all possible societies within our sets of boxes. The more accurately anyone can describe the ideal and actual patterns of a given society, in terms of the small number of variables called for here the more accurately we can locate it in all of these respects.

No claim is made of the relevance of these matters to everything, but the arguments about the relevance of family phenomena to other social contexts makes total irrelevance unlikely. Certainly these findings would have implications for any others that could be hooked up to such elements [e.g., if a given type of character structure does vary with birth order or number of generations represented in the family unit, these findings would predict the actual and possible incidence of those character traits for any society].

Finally there is another, hope, belief, prediction--call it what you will--even more extravagant. If this simulation can be effected, there must be the possibility of concise mathematical statement of it.

V

STATISTICAL STUDIES OF THE DEMAND FOR MEDICAL CARE

Chairman, Monroe Lerner, Blue Cross Association

Page

The Demand for General Hospital Facilities: A Principal Components Analysis - Gerald Rosenthal, Harvard University.................. 84

The Effect of Income on Medical Care Spending - Paul J. Feldstein, University of Michigan and W. John Carr, American Hospital Association..... 93

Discussion - Herbert E. Klarman, Johns Hopkins University........................... 106

Discussion - Jerome Rothenberg, Northwestern University........................... 109

THE DEMAND FOR GENERAL HOSPITAL FACILITIES: A PRINCIPAL COMPONENTS ANALYSIS

Gerald Rosenthal - Harvard University

In an earlier study,[1] the use of general hospital facilities and its components, the admissions rate and the average length of stay in terms of patient-days per 1000 population were expressed as a function of the influence of the following social, demographic and economic characteristics:

X_1	Price	Average 2-bed room rate
X_2	Income	% incomes over $5999
X_3	"	% incomes under $2000
X_4	Insurance	% hospital insurance coverage
X_5	Age	% over 64 years of age
X_6	"	% under 15 years of age
X_7	Marital Status	% Females 14 and over who are married
X_8	Sex	% 14 years and over that is male
X_9	Urbanization	% residing in urban area
X_{10}	Education	% 25 years and over with 13 or more years of education
X_{11}	Race	% non-white
X_{12}	Family size	Population per dwelling unit

The relationship was estimated by means of a cross-sectional multiple regression analysis with the individual states in the Continental United States as units of observation.[2] The model was estimated for the census years 1950 and 1960 and displayed corrected multiple correlation coefficients (R^2's) of .7195 and .6832 respectively for patient—days per 1000 population. The model proved to be quite useful for the purposes of the study, but certain difficulties arise in the use of twelve characteristics. An ultimate goal of this research is to develop time-series analyses. Since there are few data prior to 1948, it seems unlikely that a period greater than 15 or 16 years can be used. This limitation on observations places a high premium on reducing the dimensionality of the model. In addition, there is considerable evidence that many of the twelve characteristics in the original model are collinear and reflect similar influences. To the extent that this statistical redundancy can be reduced, the model can be both statistically and conceptually more useful.

Derivation of Independent Influences

In the current paper a procedure is established whereby interrelated characteristics can be grouped into sets of independent influences on hospital utilization. The procedure used is principal components analysis. By means of this statistical technique, it is possible to derive a set of linear combinations of the twelve variables, each of which is independent of the other. Each of the combinations is known as a principal component. Each component will reflect, to some degree, all of the twelve variables. However, the degree to which individual variables will be associated with each component will differ significantly.

The first step is to ascertain how much of the total variation in the original twelve variables can be explained by each principal component. If there are fewer than twelve independent influences and it is possible to represent each independent influence by a single variable, a basis is provided for the development of a smaller and more meaningful model. The percentage of variation in the original twelve variables which is explained by each principal component is presented for the years 1950 and 1960 in Table 1. In each case the principal components are listed in order of the percentage of total variation they explain. In both years 4 or 5 principal components explain approximately 90% of the variation in the twelve variables as a whole.

Interpretation of Independent Influences

After the principal components have been derived, one can examine the degree to which each of the components is related to each of the twelve original variables. These relationships provide the basis for ascribing meaningful interpretations to the statistically-created principal components. This is done by correlating the observations for each variable with the statistically-created observations for each principal component. The correlation coefficients measuring the relationships between each principal component and each of the original variables are known as factor loadings. They represent the degree to which each individual characteristic is embodied in the general influence represented by each principal component. The factor loadings for each variable in each of the seven principal components are also presented in Table 1.

The factor loadings show that in each principal component certain of the original twelve variables have much higher factor loadings than others. This, then, suggests a way of distributing the original twelve variables into the basic independent influences with which they are most closely related. When this is done, it becomes possible to ascertain whether any meaningful interpretation can be given to these principal components.

This investigation was supported by Public Health Service Research Grant HM00302 from The Division of Hospital and Medical Facilities.

[1] Rosenthal, Gerald, The Demand for General Hospital Facilities, American Hospital Association Monograph Series No. 14.

[2] District of Columbia, Maryland and Virginia were aggregated into a single unit leaving 47 observations. The correlation matrix for all variables is presented in Appendix 1.

Table 1

FACTOR LOADINGS AND % VARIATION EXPLAINED OF TWELVE ORIGINAL VARIABLES ON PRINCIPAL COMPONENTS 1950 and 1960

1950

			1	2	3	4	5	6	7
% Variation Explained		Cumulative	48.33	73.04	83.68	89.07	93.10	95.39	97.17
		Individual	48.33	24.71	10.64	5.39	4.03	2.29	1.78
Characteristics	1		-.7893	-.0833	.2934	-.2090	-.3272	-.3281	-.1477
	2		-.8208	.2156	.3804	.0893	.2260	-.0516	.2323
	3		.8940	-.2262	-.1684	-.2985	-.0348	.0266	.0092
	4		-.7277	-.4987	.1379	.2130	.1728	.1742	-.2384
	5		-.5667	-.2837	-.7527	-.0622	-.0206	-.0255	.0362
	6		.8956	.2496	.1765	.1634	-.1676	.0690	-.1381
	7		.1818	.8814	.0208	-.2090	.2839	-.0019	-.2102
	8		.0207	.9427	-.0256	.1432	.0801	-.1423	.0342
	9		-.8364	-.2416	.3638	-.1922	.0086	.1400	-.0168
	10		-.4514	.7300	.0590	-.0719	-.3936	.2888	.0756
	11		.7139	-.3045	.3976	-.4248	.1138	.0347	.0850
	12		.7776	-.3456	.3263	.3775	-.1061	-.0625	.0431

1960

			1	2	3	4	5	6	7
		Cumulative	43.41	69.12	81.88	88.39	92.18	94.91	96.83
		Individual	43.41	25.71	12.76	6.51	3.79	2.73	1.92
Characteristics	1		.8626	.0242	.1985	.0325	.2610	.3368	.0443
	2		.8612	.2451	.3535	.0554	-.1258	-.0233	.0335
	3		-.8743	-.2383	-.2048	-.2320	.1423	.0570	-.1080
	4		.6675	-.4882	.0820	.3956	-.2234	-.0392	-.2930
	5		.3573	-.6036	-.6529	-.0157	.1321	-.1402	.0525
	6		-.6734	.5932	.1912	.2761	.1354	-.1606	.0363
	7		.0621	.8520	-.2876	-.1833	-.3402	-.1004	.0256
	8		.0406	.8574	-.3607	.0824	-.0510	.2901	-.0534
	9		.7463	.0109	.4909	-.3653	.0303	-.1454	.1083
	10		.4872	.7172	-.0215	-.0878	.3586	-.1931	-.2583
	11		-.7126	-.1739	.4636	-.3953	-.1100	.1053	-.1964
	12		-.7904	.0866	.4307	.3917	.0652	-.0243	.0682

The most important variables in the first principal component in both 1950 and 1960 are the economic variables. In 1950, however, more of the original twelve variables showed up strongly in this first factor. The proportion of population under 15 is not directly an economic variable, but can perhaps be interpreted as indicative of economic ability to raise a family. The low income, urban, high income, and price variables, all of which reflect economic conditions, show up quite strongly in the first principal component. The two income variables and the price variable also show up strongly in the first component in 1960. It seems quite evident that this principal component represents the influence of the general level of economic prosperity, or, perhaps, the ability to pay.

In both years marital status, the sex distribution, education, and the proportion of aged in the population appear to be very unimportant in the first principal component. However, in the second principal component, the sex distribution and marital status show much higher factor loadings than do any of the other variables in both years. In the third principal component, both sex and marital status show little association. The aged variable, a variable which shows little relative association with the other principal components, is most highly associated with this factor. This suggests that the influence of the proportion of population 65 and over is largely independent of both the sex-marital status factor and the economic prosperity factor.[3]

In the fourth factor, race and population per dwelling unit seem to be fairly highly associated in both 1950 and 1960, though in 1960 both the urban-rural distribution and insurance coverage are much more highly associated than in 1950. In the original study, however, the significance of the insurance coverage variable was considered to be quite different in 1950 and 1960 because of differences in benefit levels. It is possible that it can be interpreted as being a different variable in each year. In 1950, the proportion of population with incomes under $2000 also shows up in this factor. This combination of characteristics might be interpreted as representing the "crowdedness" of housing since both the population per dwelling unit and the urban-rural distribution as well as the low-income measure reflect this influence. The influence of the insurance variable in 1960 is difficult to account for.

In the fifth principal component, the level of education seems to show the highest association. At this point, however, the degree of association between the variables and the principal components is rather small and differentiation among various characteristics are relatively insignificant.

The analysis suggests that there are 4 or 5 independent influences which might be related to the utilization of short-term general hospitals: a general economic prosperity factor, a sex-marital status factor,[4] an aged factor, a housing factor, and possibly an education factor.

A New Model of Hospital Utilization

The principal components analysis made it possible to express the total variation of the twelve variables as a set of four or five basic influences. These independent influences, or factors, provided the basis upon which a new model of hospital utilization could be developed. Two aspects of the factors merit attention. First, the individual principal components provide a basis for ranking the influences in the order of their importance and second, the factor loadings within each principal component provide a basis for selection of individual characteristics which can then be used to represent the factor in an analytic model. A model which says that the utilization of short-term general hospitals is a function of the first, second, third, fourth, and fifth influences is desired. Since these influences are statistical creations, however, they cannot be measured directly. Thus it is necessary to represent these influences by a single variable selected from among the variables associated with each principal component.

For the first, third, and fifth variables, the selection was fairly straightforward. In general, the variables with the highest factor loadings also satisfied the requirement of being meaningful on a priori grounds. For the first variable, the low income measure was used, a priori reasoning serving to eliminate the proportion of population under 15 years of age from consideration.[5] For the third and fifth factors, the aged variable and the education variable were used respectively.

For the second and fourth variables, an experimental criterion was added. The marital status variable was, in all cases, preferable to the sex variable for the second principal component, both on grounds of a priori reasoning

---

[3] This finding also suggests that the present model is more meaningful than the original 12-variable models, which yielded results showing no significant influence for this factor.

[4] In the original study it was suggested that the marital status variable and, more tenuously, the sex variable were indicators whether a household existed in which an individual could receive some degree of medical care without consuming hospital facilities. To the extent that this reasoning holds, this influence could be characterized as an alternative-to-hospitalization factor.

[5] Some valid objections were raised concerning the desirability of using a fixed-dollar cutoff for income in two different time periods. However, the results using per capita income or median income do not differ significantly from the results obtained here.

and statistical stability. For the fourth variable, the urban-rural measure was selected. Even though it had a lower factor loading in 1950 it showed more consistent statistical behavior.[6] The case for using population per dwelling unit also had appeal although it seemed to reflect an influence similar to the marital status. The arbitrary selection here is primarily in the interests of brevity. The race variable proved of no usefulness while the insurance variable merits some special, and more detailed attention.

The results of the new regression analyses, using only five variables, are presented in Table 2. As in the original analysis, these are cross-section linear multiple regressions expressing patient-days per 1000 population in short-term general hospitals and its components, admissions per 1000 population and average length of stay, as a function of the five selected independent variables.

Evaluation of Statistical Results

A comparison of the statistical results obtained using the four-and five-variable models with the results obtained using a twelve-variable model is the first way of evaluating the performance of the new models. The basis for comparison is the efficiency of prediction of the models; that is, the amount of total variation in hospital utilization which is "explained" by the model as presented by the corrected R^2's obtained in the original study for 1950 and 1960 which are presented in Table 3.

Table 3

Original Twelve-Variable Models Corrected Multiple Correlation Coefficients

	1950	1960
Patient-Days per 1000 population	.7195	.6832
Admissions per 1000 population	.3742	.5693
Average length of stay	.6348	.7097

It is evident that the models using fewer variables do not "predict" as well as did the twelve-variable model. However, the degree of loss is not so great as to render the five-variable models statistically unsatisfactory. In 1950, the smaller model showed a R^2 corrected close to .6 which is certainly significant and indicates a relatively high degree of prediction. As will be discussed later, ability of the model to predict patient-days per thousand was significantly lower for 1960, although its ability to predict the admissions rate and the average length of stay was similar for both 1950 and 1960. It seems reasonable to suggest that, even though the estimating equations do not perform as well as the original model, they have sufficiently high multiple R^2's to serve as a base upon which to estimate utilization. Far more important is the indication that the degree of multi-collinearity among the variables has been reduced, and, in turn, that coincidently the statistical problems associated with this multi-collinearity have also been reduced (but not eliminated).[7]

If estimation were the sole desideratum, it would be possible to create a model using five of the twelve original variables which would have a higher multiple R^2 than do the models that have been presented here. Such models were estimated by selecting, from the twelve original variables, the five variables which, for each dependent variable, would yield the highest R^2. The variables were selected in order of their contribution to the R^2. The most highly correlated variable was taken first, and then, of those that were left, the next most significant variable, and so on. The results of this exercise are presented in Table 4 and compared to the earlier results in Table 5. The purpose of this experiment was to evaluate the degree of "prediction" loss incurred by insisting on a more meaningful selection of variables.

Table 5

Comparison of Corrected R^2's-Two Alternative Estimating Procedures

	Principal Components	
	1950	1960
Pat.Days/1000	.5838	.3901
Adm./1000	.3106	.3678
Length of Stay	.5409	.5736
	Maximum "Prediction"	
	1950	1960
Pat.Days/1000	.6043	.6182
Adm./1000	.3669	.3846
Length of Stay	.6334	.6888

For the 1950 estimates, no housing variables showed up in the patient-days per 1000 population model, no education variable showed up in the length of stay model. In the latter case, proportion of population under 15 years of age did appear to be strongly related. This variable is highly associated with the economic variables. For 1960 none of the models included the aged variable or an education variable. There is considerable difference between these two statistical approaches with respect to the actual variables used in the estimating process. However, with the exception of the 1960 estimates for patient-days per 1000, the principal component models do not perform significantly worse than the maximum prediction models. The cost of using consistent and meaningful models of utilization based on sound a priori analysis is not,

---

[6] Professor Rothenberg has suggested, rightly, that the urban variable has a significant potential for reflecting supply conditions and is, therefore, a poor choice for a demand equation. This observation alters the merits of the variables and makes population per dwelling unit a preferable variable for this analysis.

[7] Obviously, collinearity is only removed if the principal components are used to estimate demand. However, this requires all 12 variables. The 5 variable models are a compromise and contain some collinearity, albeit less than the original.

Table 2

Regression Estimates - Principal Components Models

Constant	Income		Marital		Aged		Urban		Educ		R^2	R^2 Corr.
1950												
Patient Days per 1000 Population												
2574.19	-11.87	(-3.75)	-24.73	(-2.63)	42.26	(2.74)	-3.94	(-2.06	23.13	(2.53)	6369	.5838
Admissions per 1000 Population												
146.90	-1.24	(-2.26)	.04	(.02)	3.60	(1.35)	-1.00	(-3.03)	3.20	(2.03)	.3986	.3106
Average Length of Stay												
21.42	-.04	(-1.46)	-.22	(-2.99)	.12	(.98)	.03	(2.14)	-.01	(-.14)	.5995	.5409
1960												
Patient Days per 1000 Population												
2950.20	-17.21	(-4.08)	-24.48	(-2.30)	33.61	(2.64)	-4.66	(-2.55)	2.15	(.27)	.4680	.3901
Admissions per 1000 Population												
213.86	-1.50	(-3.36)	-.39	(-.34)	2.28	(1.69)	-.97	(5.03)	1.27	(1.51)	.4485	.3678
Average Length of Stay												
18.30	-.06	(-3.07)	-.16	(-3.27)	.12	(2.06)	.02	(2.06)	-.06	(-1.53)	.6280	.5736

Table 4

Regression Estimates - Maximum "Prediction" Models

	R^2 Corr.
1950	
$Y_1 = - 75 - 6.9X_3\ (-2.9) + 50.6X_5\ (3.4) + 14.1X_{10}\ (1.5) - 36.8X_7\ (-3.2) + 62.4X_8\ (2.6)$	.6043
$Y_2 = -690.0 + 19.1X_8\ (4.1) + 10.0X_5\ (2.8) - 3.0X_7\ (-1.5) + 1.2X_{11}\ (1.7) - .8X_3\ (-1.5)$	.3669
$Y_3 = 51.5 - .19X_6\ (-1.7) - .38X_7\ (-3.6) - .49X_5\ (-2.9) - .03X_{11}\ (-1.5) - 2.6X_{12}\ (-1.5)$	.6334
1960	
$Y_1 = -1475.7 + 9.8X_4\ (5.9) + 134.1X_8\ (5.3) - 52.8X_7\ (-4.5) - 20.1X_1\ (-3.5) - 287.4X_{12}\ (-2.7)$	.6182
$Y_2 = 296.7 - .43X_{11}\ (-1.3) - .80X_9\ (-3.4) - 1.45X_3\ (-2.4) - 1.29X_1\ (-1.4) - 16.02X_{12}\ (-1.1)$	.3846
$Y_3 = 8.69 + .03X_4\ (3.7) - .11X_6\ (-2.3) - .18X_7\ (-3.1) + .02X_9\ (2.4) + .22X_8\ (1.7)$	.6888

where: Y_1 = Pat. Days/1000
Y_2 = Admissions/1000
Y_3 = Average Length of Stay

X_1 = Price
X_3 = Lo Income
X_4 = Insurance
X_5 = Age 65+
X_6 = Age 15-

X_7 = Marital Status
X_8 = Sex
X_9 = Urban
X_{10} = Education
X_{11} = Race

X_{12} = Pop/DU

in this case, very great.[8] The primary purpose of this analysis is to provide some insight into the factors which influence hospital utilization and to find a form in which this knowledge can be used to "predict" utilization. The procedure described in this report has the advantage of providing this meaningful form.

Some Observations On the Variables

The analytic results for the first variable indicate that, in a statistical sense, the low income variable performs well.[9] In both 1950 and 1960 it is significant in almost all cases. There is, however, one interesting point. The first principal component represented something over 40% of the total variation in the twelve original variables, and the degree of association between the low income variable and the first principal component was quite high in both years. The degree of explanation contained in that variable, however, was much lower in 1960 than it was in 1950. Thus, to the extent that the first variable represents the ability to pay component, it can be said that the ability to pay was a less important determinant of overall hospital utilization in 1960 than it was in 1950.

This is true in spite of the fact that the ability to pay factor was almost as strongly represented in the twelve variables in 1960 as it was in 1950. This could lead to the conclusion that by 1960 there were other factors, not represented in the original twelve variables, which had begun to show a significant association with hospital utilization. This would suggest that the specification of the model must be changed and that new characteristics reflecting influences not now contained in the model must be incorporated into it. Although this type of change may arise from many circumstances, conceivably the explanation is that the organizational structure of medical care has become a more important determinant of the utilization of short-term hospitals than it was in the past.

There is, however, another explanation suggested by the data. In 1960, the first principal component no longer represented all of the ability to pay influence. The insurance variable, which is included in the fourth principal component, appeared to be an independent influence in 1960. It is possible that both an insurance and a housing variable should be incorporated in the prediction model.[10] This hypothesis is corroborated by the fact that in the maximum prediction models for 1960, the insurance variable was the first chosen for both patient-days per 1000 and average length of stay. It did not appear in the admissions rate models. In 1950, it was not of any significance. In the maximum prediction models, the 1960 estimates had higher R^2's, relative to 1950, for all utilization measures. This suggests that the twelve original variables do contain the same explanatory value but that the structure of their relationship to utilization has changed during the 10 year period.

Both of these hypotheses require additional analysis before the degree to which each can be considered a realistic interpretation of the true circumstances can be determined. When the time series analyses are developed, it is hoped that individual geographic areas can be separated into groups roughly similar in organization of medical care and that the relationship between hospital utilization and an area's characteristics can be examined more explicitly. The differences between 1950 and 1960 may reflect changes in types of practice rather than different responses to the same types of organization. This suggestion is appealing on a priori grounds, but must be subjected to empirical investigation.

The marital status variable, in general, shows a positive association to admission rates, which reflects the impact of child bearing, and a negative association with the average length of stay, in part because the average length of stay for obstetrical admissions is shorter than that for all admissions. To the extent that higher obstetrical admissions yield an increase in the number of admissions of shorter duration, the average length of stay should be diminished.

Even though the association of marital status with each of the sub-components of utilization is consistent with the impact of child-bearing, the overall association of marital status with patient-days per thousand is negative. This suggests that while an increase in the proportion of the population that is married adds to utilization by increasing child-bearing, some other influence, associated with marital status, functions in the opposite direction because the overall net effect is a lowering of utilization. Since we know that the negative influence operates primarily through lowering length of stay, our analysis suggests that a tentative description of the second factor as representing the existence of a substitute for hospital services might not be unreasonable. The existence of a household may permit shorter hospital stays by providing a form of convalescent facility outside of the hospital.

In almost all cases, the proportion of population 65 and over had a positive association with utilization. It is interesting to note that in most cases the association was more strongly operative through the admissions rate than through the length of stay. This suggests that the total incidence of all disease among the aged is more important in affecting total utilization than the fact that the disease mix is more heavily chronic than for other age groups.

---

[8] The corrected R^2's are still significant, although less of the variation in demand is explained.

[9] See Footnote 5.

[10] Subsequent analyses indicate that this is true.

This latter circumstance would serve to increase the average length of stay rather than the admissions rate while the higher incidence of all disease would be reflected in the admissions rate. There is some indication that the influence of the proportion of aged on hospital utilization was greater in 1960 than in 1950. This may stem in part from the dilution over the ten-year period of the impact of economic circumstances. Since there is some basis for assuming that the physiological requirements for hospitalization are greater for the aged, the results suggest that technical rather than economic influences are becoming relatively more important in determining hospital utilization.

The impact of the urban-rural distribution appears primarily in a reduction in the admissions rate which more than offsets the significant, but quantitatively less important, positive association with the average length of stay.[11] In this case the negative association with admissions rate might suggest that the urban-rural distribution reflects proximity to medical care outside of the hospital. Heavily urban areas are likely to have a greater array of substitutes for hospitalization. The behavior of the urban-rural distribution seems to support this interpretation. Its positive association with the average length of stay might indicate that those who go to the hospital are sicker, since less ill patients utilize other sources of medical care, and therefore remain in the hospital longer. It should be noted that this interpretation differs considerably from the usual hypothesized relationship between hospital utilization and the urban-rural distribution. In general, it is held that in a more rural area there will be a higher length of stay since the length of time needed to reach the facility makes it unlikely that two short stays will be used when one single long stay may provide the desired medical services.

The results with respect to the education variable are quite different from those obtained in the original models and they provide an excellent demonstration of the advantages of limiting the interrelationships among the data. In most cases education has a positive association with the admissions rate and a negative association (insignificant in 1950) with the average length of stay. To the extent that the educational level measures the ability to perceive illness and to seek medical care, these findings are quite consistent. This would indicate that a higher level of education indicates a population group likely to seek care more frequently, and that the net result of this early medical treatment may be a reduction in the severity of illness, indicated by the reduced length of stay in 1960.

Another observation might be made with regard to the impact of the educational level variable. Some people have suggested that the increase in the admissions rate which stems from awareness and perception of the usefulness of early entrance into the hospital ought to be compensated for by diminution in the length of stay. The evidence gathered here indicates that the net effect of a high level of education, given other characteristics, is an increase in total utilization as measured by patient-days per thousand. This finding is consistent with what might be called an increasing propensity to consume care, and it has been suggested by at least one researcher in this area that constant exposure to medical attention will, in the long run, merely result in the finding of more disease rather than in an overall diminution in the amount of medical care that is required.

Admissions Rate Influences Vs.Length of Stay Influences

One significant feature of this analysis is the demonstration of the differing impact of the various influences on the two components of hospital use. These influences were summarized rather briefly as the ability to pay or the economic circumstances, the existence of a family, the impact of the aged, the degree of crowdedness in housing, and finally the level of education. Two of these factors seem to have a highly significant positive association with the admissions rate. These factors, the proportion of the aged and the educational level, apparently represent, respectively, technical requirements for medical care and perceptions of the desirability of medical care. Two other factors, the existence of a family and the degree of crowdedness, each of which reflect the availability of a substitute for hospital utilization, operate mainly on the average length of stay. These factors, while not necessarily significantly affecting the likelihood that a patient will go to the hospital do affect the length of time he is likely to remain. The impact of the income variable, or the ability to pay measure, is apparent in both these aspects of hospital utilization, but significantly less so in 1960 than in 1950.[12]

In an effort to test the hypothesis that certain influences operate by affecting the admissions rate and that others operate by affecting the length of stay, one additional experiment was attempted. In the regression analysis that has been presented the order for introducing the variables was determined by the degree to which the principal component that they represented reflected the total variation of all twelve variables from which they were selected. Another set of relationships was derived statistically in an attempt to determine experimentally which variables showed up most strongly in the admissions rate equations and which variables showed up more strongly in the average length of stay equations. In each case the five variables were identical to those which had been chosen on the basis of the principal components analysis. However, the computation of these regression relationships was executed in such a way that each variable would be selected in order of its contribution to the

---

[11] See Footnote 6.

[12] The insurance variable, however, is associated almost entirely, and positively, with the length of stay.

R^2. In one case, the dependent variable was admissions per thousand and in the other the dependent variable was the average length of stay.

The variables showing up most strongly in the average length of stay for 1950 were urban-rural proportion, marital status, and income, with education and age contributing least to the "prediction" ability of the relationship. For the admissions rate, on the other hand, the variable with the largest contribution to the R^2 was education, with the aged variable falling somewhere in the middle and the marital status variable being least contributory to the relationship.

In 1960, the results are similar but not so clear-cut. The degree of total prediction derived from the admissions rate relationship is particularly small, and very little meaningful information can be obtained from it. The educational variable made a greater contribution to estimations of the admissions rate than it did to estimates of the average length of stay. The aged variable was reflected similarly, but the marital status variable was most important in its contribution to explaining the average length of stay and least important in terms of explaining the admissions rate.

The empirical observations support the suggestion that economic constraints and the availability of substitutes, as we have defined them, affect utilization by affecting the average length of stay. Apparently, the characteristics of the population which reflect physiological requirements for hospitals, such as age distribution,or perception of the need for care, such as educational level, operate by affecting the admissions rate, and, once the admissions rate is set, the other characteristics determine the actual duration of stay within a significant range.

Some Implications

The early results justify some tentative observations that might be of some interest. One, while the economic factors have a demonstrably strong effect, they are clearly not overridingly important. Indications are that other factors are of considerable importance in determining utilization. Indeed, the economic circumstances demonstrate relatively little impact on the admissions rate as compared to other population characteristics.

A second point relates to the existence of substitutes for hospitalization. Two of the variables measure to some degree the availability of substitutes for hospitalization. It might be suggested that the marital status variable represents the existence of an individual to perform hospital services outside the hospital, and the population crowdedness variable represents the existence of physical facilities outside the hospital for receiving this care. The behavior of these variables gives additional encouragement to the notion that utilization of general hospitals can be reduced by the provision of certain hospital-type services in other kinds of facilities.

A third and final observation relates to the association between hospital utilization and the characteristics of the population as between 1950 and 1960. There are many indications that the degree of association between the five influences as represented by the five characteristics was significantly less in the later period. This strongly suggests that either the organizational characteristics of medical care or other influences within the population characteristics not included in the analysis have become more important. In some earlier discussions of organizational characteristics it was hypothesized that the physician was the determining factor. The results of the research presented here suggest that there might be a considerable measure of truth in this *a priori* observation, since our ability to predict is poorest when dealing with the admissions rate. In addition, the degree to which the substitute measures were of importance might indicate that measures of substitutes for hospital facilities which reflect the existence of other types of medical care facilities and which are subsumed under our category of organization of medical care also should provide a fruitful area for further research. It is hoped that the observations presented here prove a stimulus to the conduct of research on this aspect of utilization.

APPENDIX 1

CORRELATION MATRIX FOR INDEPENDENT VARIABLES

1960

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}
X_1												
X_2	.78											
X_3	.75	-.91										
X_4	.52	.52	-.56									
X_5	.16	-.06	-.02	.44								
X_6	-.52	-.35	.37	-.64	-.66							
X_7	-.08	.20	-.20	-.39	-.32	.35						
X_8	.05	.12	-.16	-.37	-.31	.38	.79					
X_9	.68	.78	-.66	.37	-.03	-.47	-.02	-.19				
X_{10}	-.14	.53	-.52	-.07	-.19	.13	.55	.57	.40			
X_{11}	-.25	-.49	.66	-.44	-.46	.33	-.22	-.34	-.22	-.45		
X_{12}	-.57	-.49	.50	-.42	-.60	.78	-.19	-.09	-.51	-.35	.57	

THE EFFECT OF INCOME ON MEDICAL CARE SPENDING*

Paul J. Feldstein, University of Michigan
W. John Carr, American Hospital Association

I. Introduction

The percentage of income allocated to medical care by the private sector of the economy has risen substantially in the post World War II period. Private expenditures for personal health services amounted to 4.05 per cent of disposable income in 1948 and remained below 4.20 per cent through 1951 when they began a steady climb to 5.71 per cent in 1962. Stated more dramatically, medical care spending was over $6,000,000,000, or 40 per cent, higher in 1962 than it would have been if the 4.05 per cent rate had been maintained. 1/ What were the causes of this large increase, and what is the prognosis for medical care spending?

The historical trend of expenditure reflects the influence of a multitude of factors which cannot readily be separated and is, therefore, of limited value for analytic purposes. A more promising approach for studies of this type has been found in the use of cross-section analysis. By examining data for a cross-section of families in the same time period, it is possible to assess the effect of factors influencing their expenditure over a wide range of values and also hold approximately constant the state of technology and other conditions which change over time but cannot easily be quantified. The precision of the analysis is thereby increased and the complexities substantially reduced.

The number of factors affecting medical care expenditure is still large, however, and they are of widely differing types--some having to do with state of health and others with social and economic conditions. The techniques best suited to the assessment of the effects of these variables differ somewhat, further complicating the problem. We have, therefore, decided to concentrate our attention on only one of the factors affecting the level of spending--one which the analysis to follow will show is of considerable importance, but which has long been overlooked--namely, family income.

The primary purpose of this paper will be to determine empirically the average net effect of income on private expenditures for personal health care in the United States and to appraise the stability of this relationship over time. We will first consider evidence of the apparent (unadjusted) relationship between income and spending for personal health services derived from cross-section surveys of medical care spending and from more general surveys of consumer expenditure which contain data on medical care. Next, factors which may bias the observed income-expenditure relationships will be pointed out. Various means will be employed to adjust for the influence of these factors in order to determine the net, or independent, effect of income on medical care spending. Finally, a best numerical estimate of this effect will be decided upon and its implications discussed.

II. Review of Survey Data

A. Evidence from Medical Care Surveys

Three nation-wide studies have been conducted on the relationship of family medical care expenditures to demographic and socioeconomic characteristics. The first was a part of the research program of the Committee on the Costs of Medical Care and was based on a survey conducted from 1928 to 1931 which provided data for 8,758 white families. 2/ Information relating to the incidence, causes, and severity of illness, as well as the utilization and costs of medical care, was gathered and presented. This study remained the primary source of statistics on medical care costs until a report on the first of two Health Information Foundation surveys was published in 1956. 3/ These surveys, based on data from 1952-53 and 1957-58, were designed to be comparable, thus making it possible to relate statistics over the five-year span. 4/ Area-probability samples provided information on the experience of 2,809 and 2,941 families in the respective periods. Little data were collected on the incidence of illness, but in addition to information on cost and utilization, evidence was gathered on the use of voluntary health insurance, which by this time had become an important element in medical care financing. 5/

In the 1928 to 1931 survey, it was found that the absolute level of charges for medical care rose with family income (Table 1). However, the percentage of income spent for personal health services remained approximately constant, except that it was higher in the lowest income class. In the two Health Information Foundation studies, the level of charges was also found to increase with income in an absolute sense. The percentage of income allocated to medical care, however, decreased with increased income. 6/

---

*A longer version of this paper, which contains a more complete theoretical discussion and additional empirical evidence, is available from the authors on request.

Table 1

Average Family Medical Care Charges and Expenditures in Relation to Family Income, 12-Month Periods, 1928-31, 1952-53 and 1957-58*

Income class	Charges (1)		Expenditures as a per cent of average income (2)	
	A. 1952-53 and 1957-58			
	1952-53	1957-58	1952-53	1957-58
All income classes	$207	$294	4.8	5.5
Under $2,000	130	165	11.8	13.0
$2,000 to $3,499	152	226	6.1	8.4
$3,500 to $4,999	207	287	5.4	6.4
$5,000 to $7,499	259	336	4.7	5.4
$7,500 and over	353	411	3.0	3.9
	B. 1928-31			
All income classes	$108		4.0	
Under $1,200	49		5.2	
$1,200 to $2,000	67		4.0	
$2,000 to $3,000	95		3.9	
$3,000 to $5,000	138		3.8	
$5,000 to $10,000	249		4.1	
$10,000 and over	503		3.8	

Source: I. S. Falk, Margaret C. Klem, and Nathan Sinai, op. cit., pp. 151 and 206; Odin W. Anderson, Patricia Collette, and Jacob J. Feldman, op. cit., pp. 17-18.

*Charges represent bills incurred during the survey year which may or may not have been paid. Expenditures represent money spent during the survey year and may include amounts paid on bills incurred prior to the survey year and exclude unpaid bills incurred during the year.

B. Evidence from Consumer Surveys

It would be interesting to know if the relationships determined by the medical care surveys were supported by other data. A number of surveys of consumer expenditure conducted by the U. S. Government may be used to verify our knowledge and extend it to other time periods. In many of these, the averages of both income and medical care spending are shown for families classified by income group. This permits us to use regression analysis to calculate measures of the over-all relationships.

Regression lines were fitted to data covering various time periods and types of families. Table 2 shows, for each cross-section, average family income (column 1), average medical care expenditures (column 2), and the amount by which medical care spending changed per $1,000 of income--a measure of the slope of each regression line (column 3). In each case, higher family income was associated with increased average medical care spending; the amount of increase per $1,000 of income ranged from $27.5 in 1944 to $44.8 in 1960. 7/

In discussing the three nation-wide medical care studies, it was noted that the average percentage of family income allocated to spending for personal health services remained nearly constant regardless of income in 1928-31, but declined as we moved up the income scale in 1952-53 and 1957-58. Another way of showing the relative relationship is to calculate the average percentage difference in expenditure for each one per cent difference in income. This number, known as the income elasticity of demand, is shown for each of the ten cross-section series in column 5 of Table 2. 8/ The figure for cross-section 1 (1960: families in metropolitan areas) indicates that the income elasticity of medical care spending was about 0.699. This means that a 10 per cent higher income, for example, would be associated with approximately 7 per cent higher medical care expenditures on the average. The fact that the elasticities are less than one indicates, as before, that families in higher income groups generally allocate a smaller percentage of their incomes to medical care. Here, also, there is evidence of variation in the income-expenditure relationship over time

Table 2

Regression Coefficients of Medical Care Expenditures with Respect to Family Income, and Related Information, 10 Cross-Section Series

Cross Section Number	Year and Type of Family[a]	Average Income[b] (1)	Average Medical Care Expenditures[b] (2)	Regression Coefficient (average difference in expenditures per $1,000 difference in income)[c] (3)	Coefficient of determination (arithmetic)[d] (4)	Income Elasticity[c] (5)	Coefficient of Determination (logarithmic)[d] (6)	Observations used and Excluded[e] (7)	Source of Data[f] (8)
1	1960-61: families in cities	$5,906	$355	$43.5 (1.8)	0.99	0.683 (0.027)	0.99	1-8-1	g
2	1950: families of two or more persons in cities	4,224	215	37.5 (1.3)	0.99	0.706 (0.023)	0.99	1-7-1	G 355 G 367
3	1950: families of city and clerical workers of two or more persons	3,923	200	36.9 (2.3)	0.98	0.735 (0.032)	0.99	1-6-2	G 246 G 258
4	1944: families of two or more persons in cities	3,411	148	27.5 (1.7)	0.98	0.500 (0.072)	0.89	1-8-0	G 375 G 385
5	1941: families of two or more persons in cities	2,672	107	37.8 (3.2)	0.96	0.952 (0.092)	0.96	1-7-1	G 393 G 404
6	1941: farm-operator families of two or more persons	1,163	62	39.7 (2.7)	0.98	0.496 (0.075)	0.90	0-7-1	G 447 G 458
7	1935-36: families of two or more persons in cities	1,971	89	40.9 (0.6)	0.99	0.957 (0.018)	0.99	1-10-1	G 411 G 421
8	1934-36: families of employed city wage and clerical workers of two or more persons	1,518	59	36.2 (1.8)	0.99	0.915 (0.036)	0.99	1-7-2	G 266 G 277
9	1935-36: farm-operator families	1,237	50	30.7 (1.8)	0.98	0.730 (0.034)	0.99	0-9-3	G 466 G 476
10	1917-19: families of city wage and clerical workers with at least one child	1,505	64	34.2 (3.1)	0.98	0.854 (0.065)	0.98	1-5-1	G 285 G 297

[a] The characteristics used as criteria of eligibility for including families in the samples from which the income and expenditure data were obtained varied from survey to survey in some respects not disclosed in the brief descriptions presented here. In addition, there were differences between surveys as to the definition of income and also, no doubt, as to what was classified as medical care expenditure. The effect of this diversity on the income and expenditure figures for cross sections with similar descriptions is probably quite limited. However, the income figures for cross sections 1 through 4 and 10 refer to income after taxes while those for the other cross sections refer to income before taxes, but the effect of this difference is lessened by the fact that individual income taxes were a relatively small proportion of family income in the years for which pre-tax income is given. For a more precise description of the data (except that used to construct cross section 1), see the notes covering the series numbers shown in column 8 of this table in U. S. Bureau of the Census, op. cit., pp. 169-74 and the sources to which reference is made therein. A description of the information used to construct cross section 1 is given in note g to this table.

[b] The averages of income and medical care spending shown are based on the entire cross sections of families before some of the observations representing families at the high and low ends of the distribution were excluded in fitting the regression lines, as explained in note e to this table.

[c] The figures in parentheses under the regression coefficients shown here and throughout this paper are the standard errors of these coefficients.

[d] The values of the coefficients of determination shown here and elsewhere in this paper are not adjusted for sample size. The coefficients of determination in column 6, which refer to the regressions used in finding the income elasticities in column 5, measure the strengths of the linear relationships of the logarithms of the original variables.

[e] The three numbers for each cross section refer, respectively, to the number of observations at the lower end of the income distribution excluded from fitting the regression line, the number used in fitting the line, and the number of observations at the upper end of the distribution which were excluded. Observations were not included if they represented less than five per cent of the total sample size counting from the lower end of the income distribution or less than five per cent of the total counting from the upper end. Observations representing families with very high and very low incomes were excluded in order to make the regression coefficients more comparable. This was necessary because some surveys used income classes which resulted in separate observations for those at the extreme ends of the income distributions while others did not. The inclusion of observations

representing extreme incomes would have substantially altered the slopes of some of the regression lines (thus changing the regression coefficients) while others would have remained nearly unaffected.

[f] The series numbers shown in this column for regressions 2 through 10 refer to those in the tables of U. S. Bureau of the Census, op. cit., pp. 179-80 and 182-84.

[g] Cross section 1 is based on data in U. S. Bureau of Labor Statistics, "Consumer Expenditures and Income, Urban United States, 1960-61," Report No. 237-38 (1964), p. 10.

and between different types of families in the same time period. The income elasticities of farm family expenditures in 1935-36 and 1941 are lower than the corresponding elasticities for city families in the same years. Also, the elasticity of expenditures of city families in 1944, a war year, is unusually low.

III. The Effect of Factors Other Than Income

A. Theoretical Analysis

Throughout this discussion, we have avoided making statements to the effect that measured differences in expenditure were caused solely by differences in family income. This is not the case because families at different income levels are also heterogeneous in other respects. Families with relatively large or small incomes may show great variations with respect to size, age composition, and other factors which have an important bearing on the level of expenditures for personal health services. To the extent that these factors are correlated with income, their effects will be included in the measured income-expenditure relationship. For example, if medical care spending increases with family size, other things being equal, and larger families have higher incomes on the average, then the effect of size will tend to bias the measured income elasticities upward. Other variables which affect medical care spending and are correlated with income will have similar effects. It is thus not possible to say that because there was a seven per cent difference in medical care expenditures among families with incomes differing by ten per cent (as in the previous example), increasing all incomes by ten per cent would increase medical care spending by seven per cent. In order to assess the independent effect of income, the income-related factors which influence expenditure must be taken into account and their effects held constant.

The variables which have an effect on a family's level of medical care spending may be divided, for purposes of analysis, into three main categories: (1) Socio-demographic variables which reflect factors such as physiological condition, considerations which influence the perception of health status, and attitude toward medical care, (2) resources available for spending in terms of income and wealth, and (3) the price of medical care as modified by eligibility for "free" care, tax deductibility, and insurance coverage. The first of these factors may be thought of roughly as a family's "desire" for medical care and is dependent primarily upon the perception of a health deficiency and a belief in the efficacy of medical treatment. In translating this desire to expenditure, the family is limited by the extent of its financial resources as care cannot generally be obtained free of charge. That is, the determination of the amount to be spent for personal health services becomes a part of the problem of allocating scarce financial resources among alternative means of want satisfaction.

We discuss next the socio-demographic and price variables with regard to their effect on the income-expenditure relationship. Later, the adequacy of the income measure itself will be considered.

1. Socio-Demographic Variables: We are interested not in predicting the expenditures of individual families but in predicting the average expenditures of families at different income levels. That is, the incidence of disease and other conditions which affect the level of spending for personal health services may be thought of as partially random and partially dependent on factors correlated with income, such as age, sex, marital status, and family size. It is the effects of these latter variables which may bias the observed income-expenditure relationships and in which we are, therefore, interested. Thus, those socio-demographic characteristics which are indicators of physiological condition, perception of illness, and attitude toward medical care will be considered.

The greater prevalence of chronic and degenerative diseases among older people lends support to the belief that, other things being equal, medical care expenditures may be expected to increase with age. In considering the difference in average expenditure between men and women, both marital status and age must be taken into account. Of course, there is not necessarily a direct correspondence between physiological condition and the desire for medical care. The need for care may not be perceived or, alternatively, a non-existent or imaginary "need" may be perceived. Also, a recognized health deficiency may not be translated directly into expenditure because of variations in disposition toward risk-taking and differences in belief in the effectiveness of medical treatment. Variables we might expect to explain these factors include family size, education and geographic area of residence.

We have now related medical care spending to various demographic and social characteristics, using age, sex, and marital status as proxy variables to represent physiological condition; education and family size as measures of perception, attitude, and their effect on health status; and the family size variable also to adjust the income variable.

2. Price: There are a number of price-related factors which may affect our estimate of the income-expenditure relationship and which should be distinguished in order to determine the possible direction and magnitude of the bias they may cause. Four important determinants of price differentials among income groups may be noted.

(a) Differences in prices charged, according to ability to pay, especially by physicians and other practitioners: There is evidence that charges for medical care, especially by physicians, are based to some extent on ability to pay. The so-called sliding-scale practice has ostensibly developed in order to shift the cost of providing free or low cost service for indigent patients to those at higher income levels. 9/ Statistical evidence of the effect of this practice on prices paid by those at different income levels is lacking, no doubt because of the extreme difficulty of separating differences in price from variations in the amount of service and amenity received. 10/ If those in higher income groups pay higher prices for the same service and if higher prices result in greater expenditures, then estimates of income elasticity based on expenditure data will be biased upward to some extent.

(b) "Free" care and subsidies made available to those in low income groups by government and charitable organizations: Third party payments for or provision of care may serve to either increase or decrease family expenditures. If medical care is provided directly, the quantity utilized will be greater but family spending will be lower because the cost of care provided is not included in family data. On the other hand, if subsidies are paid to the family, both the quantity utilized and reported level of spending may be higher. Government welfare aid in the medical care field is generally made directly to the providers of care and benefit primarily those with low incomes. They would thus tend to raise the income elasticity from what it would otherwise be.

(c) Another factor which may alter the income-expenditure relationship is the partial tax deductibility of medical care spending. Deductions are more likely to be taken by those in low income groups because they spend a larger proportion of their reported incomes on medical care, but the value of the deduction is likely to be higher to those in the upper income categories because they pay higher marginal tax rates. In balance, it would be difficult to determine the net effect of tax considerations on the income distribution of medical care spending.

(d) The effect of health insurance enrollment: One of the possible effects of health insurance is that it may cause total medical care expenditures to differ among persons with and without coverage. This is because to those persons having insurance the price of the covered service is, in effect, lowered. To the extent that enrollment in insurance plans and the degree of coverage is correlated with income, this would tend to raise the income elasticity.

Possibly more important in its effect on the income-expenditure relationship, however, is that among those enrolled in group plans, the expenditures of individual families will tend to be equalized. This is the case because the coverage and cost to individual families under such plans is relatively uniform. Another factor, which should be taken into account in empirical analysis, is the fact that employer and other third-party contributions to health insurance premiums are not normally included in family expenditure data.

From our brief discussion on the effects of price-related factors, it is apparent that price differences related to income and welfare aid tend to raise the measured income elasticity. The over-all effects of the tax deductibility of medical care spending and health insurance enrollment are not clear. However, all sources of payment of insurance premiums should be included in estimating the income-expenditure relationship.

IV. The Adequacy of the Income Measure

A. Permanent-Income Theory of Consumption

The nature of the income variable must be understood if the multiple regression technique is to be applied meaningfully. A family's income in any given year may be abnormally low or high because of temporary loss of employment, windfall gains, and other similar causes. Empirical evidence suggests that total consumption (that is, the use of products and services) is not generally raised or lowered to correspond with temporary changes in income. Rather, a family's level of consumption is determined primarily by its expected normal, or "permanent" income. 11/ It is very likely that nearly all categories of consumption, as well as the total, exhibit this unresponsiveness to temporary income fluctuations.

In the general case, transitory income is hypothesized to have little or no effect on expenditure levels. 12/ Thus, if all income differences were transitory, a regression of spending on income would have nearly zero elasticity. If, on the other hand, incomes differed but contained no transitory components, the regression line would approximate the effect of normal income. Empirical observations of incomes of individual families are composed of mixtures of normal and transitory components. Regression lines fitted to them will therefore lie somewhere between these two extremes. 13/

The relationships described in Table 1 and by the regression equations in Table 2 are those between medical care expenditures and total income, which includes both permanent and transitory components. A more useful estimate would be that between expenditures and normal (permanent) income alone. This is the case because we are most interested in the relationship between average income and average expenditure and, in the averages, differences in transitory income peculiar to individual families will cancel out. Since transitory income is included in the income variable but presumably has little effect on the level of expenditure, the indicated relationships presented in the earlier analyses provide a biased estimate of the effect of normal income on medical care spending.

Another income-related consideration which should be taken into account is the fact that physiological condition may affect both expenditure and income. Illness may result in higher medical care spending and, at the same time, reduce family income by causing the disability of a wage earner. This may occur in either the permanent or transitory sense. In both cases, the result will be to lower the measured effect of income on medical care spending. To the extent that physiological condition does not depend upon income, it should be held constant among income groups in calculating the income-expenditure relationship. Estimates based on family survey data are thus also biased downward by this factor.

B. Empirical Study

It is obvious by now that a fairly large number of factors may bias the measured income-expenditure relationships, and although the direction of influence of some of these has been hypothesized, quantitative measures cannot be obtained from a priori reasoning. One major problem which must be handled in some way is that of transitory income. As mentioned earlier, differences in transitory income peculiar to individual families will tend to cancel out in group averages (provided the averages are not correlated with transitory income). Therefore, a regression of average incomes and average medical care expenditures of families grouped by city should provide a useful approximation of the effect of normal income on medical care spending. In addition, extreme values of various other factors which differ among families, such as age distribution, size and insurance coverage, will tend to be averaged out. Multiple regression analysis may then be utilized to hold approximately constant and estimate the effects of differences in the city averages of many of these other factors for which data are available.

1. 1950 Cross-Section of City Averages: The results of regressions of average income, medical care expenditure and other variables in the form of city averages for 1950 are presented in linear form and as elasticities in Table 3. These estimates demonstrate that the effect of using cross-sections of city averages to approximate normal income is to raise the estimated effect of income considerably from the results based on family data described earlier in Table 2. 14/ The regression presented in Table 3, equation 4, indicates that the elasticity of medical care spending with respect to income was approximately 1.065 as opposed to 0.706 for the unadjusted cross-section of families in 1950. 15/

Among the 1950 city average estimates, income, family size and the two insurance variables had effects which may be considered statistically significant. It is interesting to note that the effect of increased family size was a decrease in medical care spending. The other variables were not statistically significant, possibly because there was little spread in their values relative to the unexplained variation in spending.

The insurance variables indicate medical care expenditures of families increased with insurance expenditures and decreased with insurance enrollment. One would expect medical care spending to increase with expenditures for insurance if for no other reason that these are included in the medical care total. The decrease in medical care expenditures with increased insurance enrollment (equation 2 of Table 3) is not surprising in view of the fact that third parties, particularly employers, pay a large proportion of the cost of insurance and such payments are not included in M.

2. 1960 Cross-Section of City Averages: The regressions based on 1960 city averages (summarized in Table 4) yielded considerably lower coefficients than those derived from the 1950 data. The effect of Y on M in 1960 was about $27.9 per $1,000, or, in terms of elasticity, 0.433. A number of possible explanations for this difference may be suggested: (a) There may have been a substantial transitory component in the city averages of income in 1960 due to the recession in that year, (b) higher insurance enrollment, especially in group plans, may have produced a downward shift, (c) the elasticities were calculated over a higher range of real incomes where they may have been lower, and (d) changes in the nature of the medical care "product" or of other goods and services competing for consumer expenditure may have produced a real shift in the elasticity.

Table 3

Relationship of Medical Care and Health Insurance Expenditures to Income and Other Characteristics, 91 City Averages, Expenditure Survey, 1950*

Equation Number	Y	S	A	E	N	I	%I	R^2
	A. Medical Care Expenditures: Arithmetic Estimates							
1.	52.8							0.52
	(5.4)							
2.	54.0	-21.5	0.2	-2.4	3.2	1.2	-1.0	0.59
	(6.8)	(12.0)	(1.4)	(4.7)	(34.7)	(0.5)	(0.4)	
	B. Medical Care Expenditures: Elasticity Estimates							
3.	1.059							0.51
	(0.110)							
4.	1.065	-0.349	0.094	-0.180	-0.012	0.251	-0.291	0.60
	(0.135)	(0.200)	(0.342)	(0.255)	(0.147)	(0.079)	(0.112)	
	C. Health Insurance Expenditures: Arithmetic Estimates							
5.	6.8							0.12
	(1.9)							
6.	3.8	1.8	-0.6	0.8	16.5			0.19
	(2.4)	(4.3)	(0.5)	(1.7)	(11.2)			
	D. Health Insurance Expenditures: Elasticity Estimates							
7.	0.901							0.15
	(0.229)							
8.	0.484	-0.105	-0.949	0.273	0.538			0.23
	(0.286)	(0.429)	(0.721)	(0.550)	(0.281)			
Mean	3.77	3.03	46.5	10.3	0.85	33.4	61.5	
Standard Deviation	0.54	0.29	2.6	0.8	0.12	10.6	14.0	

Source: Study of Consumer Expenditures, Incomes and Savings, Statistical Tables, Urban, U. S.--1950 (tabulated by the Bureau of Labor Statistics for the Wharton School of Finance and Commerce, Philadelphia: University of Pennsylvania, 1956-57), Vol. VIII, Summary of Family Expenditures for Medical Care and Personal Care, pp. 2-3.

*Brief descriptions of the measures used in these analyses are given below. More complete information is available in the Study of Consumer Expenditures volumes.

Symbol	Definition
Y	Annual net money income, in thousands of dollars, after deduction of personal taxes and occupational expenses.
S	Family size, measured by the number of equivalent full-year members.
A	Age of the head of the family in years.
E	Education of the head of the family as measured by the number of years of schooling completed.

N	Number of family members who were gainfully employed 48 weeks or more in 1950.
I	Family expenditures for medical care group plans and insurance in dollars.
%I	Per cent of families in each city reporting expenditures for health insurance. (Note that this measure will not reflect enrollment under which the entire premium is paid by third parties.)
M	Medical care expenditures, including expenditures for health insurance, in dollars.
R^2	Coefficient of determination.

Table 4

Relationship of Medical Care and Health Insurance Expenditures to Income and Other Characteristics, 39 City Averages, Expenditure Survey, 1960*

Equation Number	Y	S	A	E	N	I	%I	R^2
A. Medical Care Expenditures: Arithmetic Estimates								
1.	34.0							0.40
	(6.8)							
2.	27.9	39.9	-1.8	0.6	-74.2	0.2	1.2	0.47
	(10.6)	(45.2)	(3.5)	(11.6)	(92.2)	(0.7)	(1.2)	
B. Medical Care Expenditures: Elasticity Estimates								
3.	0.567							0.39
	(0.116)							
4.	0.433	0.523	-0.164	0.083	-0.201	0.034	0.304	0.48
	(0.182)	(0.416)	(0.469)	(0.314)	(0.242)	(0.179)	(0.258)	
C. Health Insurance Expenditures: Arithmetic Estimates								
5.	4.9							0.10
	(2.4)							
6.	1.0	36.8	1.3	9.2	-59.9			0.35
	(3.3)	(12.2)	(1.1)	(3.3)	(26.9)			
D. Health Insurance Expenditures: Elasticity Estimates								
7.	0.336							0.12
	(0.149)							
8.	-0.005	1.081	0.280	0.809	-0.617			0.32
	(0.209)	(0.436)	(0.549)	(0.340)	(0.263)			
Mean	5.58	3.14	47.3	10.2	0.84	86.6	74.8	
Standard Deviation	0.93	0.27	2.8	1.3	0.10	14.6	8.0	

Source: U. S. Bureau of Labor Statistics, Survey of Consumer Expenditures, 1960-61, Report Numbers 237-1 through 237-27.

*The definitions of variables used in these analyses are very similar to those in Table 3.

The possibility that there was a greater degree of transitory income in the 1960 city averages may be tested by aggregating the data into larger groups. The results of regressions calculated after combining the 39 cities into 8 urban areas are shown in Table 5. Upper-bracket estimates, derived from regressions of income on expenditure, are also presented. These estimates, which indicate a substantially larger income effect, suggest that there may have been a large transitory component of income in the 1960 city averages. 16/

Another factor which must be taken into account in the empirical analysis is the fact that employer and other third-party contributions to health insurance premiums are not normally included in family expenditure data. Although spending for group plans is probably fairly independent of the incomes of individual families within a given plant or company, it appears likely that the average income of the group as a whole exhibits a strong influence on third-party expenditures for health insurance. Not only may the direct effect of income be at work, but there are tax considerations which provide an incentive for increased employer contributions at higher average income levels. Contributions of employers toward the payment of health insurance premiums do not constitute taxable income to the recipients. Thus, the higher the average income tax bracket, the larger the potential tax saving and the greater the incentive to have payments made by the employer. 17/ These employer contributions act as substitutes, to some extent, for family payments. Any estimate of the over-all income elasticity of medical care spending will, therefore, be biased downward if third-party payments of health insurance premiums are not included. We would expect the bias to be greater in 1960 than in 1950 because the percentage of private medical care expenditures accounted for by insurance rose from 12.8 to 29.7 per cent over this period. 18/

Regressions were calculated which related total health insurance premiums to per capita income for a cross-section of 48 states in 1960. The results of these regressions are presented in Table 6. 19/ The average increase in total expenditures for health insurance premiums per $1,000 increase in income was about $25.0. In terms of elasticity, the effect of income on total premium payments was approximately 0.981. 20/

In calculating the over-all income elasticity of medical care spending in 1960, it is necessary to consider the portion of

Table 5

Relationship of Medical Care Expenditures to Income
8 Urban Areas, Expenditure Survey, 1960*

Equation Number	Y	R^2
	A. Medical Care Expenditures: Arithmetic Estimate	
1.	42.8 (11.6)	0.69
	B. Medical Care Expenditures: Upper-Bracket Arithmetic Estimate	
2.	61.6	0.69
	C. Medical Care Expenditures: Elasticity Estimate	
3.	0.676 (0.173)	0.72
	D. Medical Care Expenditures: Upper-Bracket Elasticity Estimate	
4.	0.942	0.72
Mean	5.55	
Standard Deviation	0.79	

*These regressions were calculated by combining the data for the 39 cities of Table 4 into groups of both large and small cities in the Northeast, North Central, South, and West regions as defined by the Bureau of the Census.

Table 6

Relationship of Health Insurance Premiums to Per Capita Income and Degree of Urbanization, 48 States, 1960

Equation Number	Per Capita Personal Income	Degree of Urbanization	R^2
	A. Health Insurance Premiums: Arithmetic Estimates		
1.	26.0 (3.4)		0.55
2.	25.0 (4.9)	0.07 (0.23)	0.55
	B. Health Insurance Premiums: Elasticity Estimates		
3.	0.896 (0.119)		0.55
4.	0.981 (0.179)	-0.098 (0.154)	0.56
Mean	2.06	37.3	
Standard Deviation	0.42	9.0	

Source: U. S. Bureau of the Census, Statistical Abstract of the United States, 1963 (84th ed. Washington: U. S. Government Printing Office, 1963), pp. 10, 20 and 329, and Health Insurance Institute, Source Book of Health Insurance Data, 1962 (New York: Health Insurance Institute, 1962), pp. 42-43.

health insurance premiums paid by third parties. When these contributions were included, the 1960 estimate was increased to about 0.883. 21/

V. Conclusion

A. Estimate of the Income-Expenditure Relationship

We began this analysis by noting the relationships which have been determined between medical care spending and income from unadjusted cross-section data. In 1950 these indicated an elasticity of about 0.7. 22/ Allowance for the transitory component of income raised the elasticity measure considerably, to over 1.0 in 1950. The elasticity estimates derived from the 1960 city-average cross-sections, however, dropped to about one-half their 1950 levels while the elasticity derived from the cross-section of families classified by income remained nearly constant at 0.7. Evidence suggests that this discrepancy may be explained largely by transitory income in the 1960 city averages and the growth of health insurance enrollment. When these factors were allowed for, the estimated elasticity for 1960 increased to approximately 0.883. Considering that the elasticity measures were all based on lower-bracket estimates, it is our belief that the adjusted income elasticity in 1960 was probably near 1.0. 23/

FOOTNOTES

1/ Data on private consumer medical care expenditures in relation to national disposable personal income may be found in Louis S. Reed and Dorothy P. Rice, "Private Consumer Expenditures for Medical Care and Voluntary Health Insurance, 1948-62," Social Security Bulletin, Vol. 26, No. 12 (December, 1963), p. 7.

The change in private medical care spending was accompanied by a nearly proportionate increase in government provision for medical services. Public expenditures for personal health care amounted to 20.1 per cent of the total in the 1949-1950 fiscal year and 21.2 per cent in 1961-1962. See Ida C. Merriam, "Social Welfare Expenditures, 1962-63," Social Security Bulletin, Vol. 26, No. 11 (November, 1963), p. 10.

2/ I. S. Falk, Margaret C. Klem, and Nathan Sinai, The Incidence of Illness and the Receipt and Costs of Medical Care Among Representative Families: Experiences in Twelve Consecutive Months During 1928-31. (Publication of the Committee on the Costs of Medical Care, No. 26, Chicago: University of Chicago Press, 1933.)

3/ Odin W. Anderson and Jacob J. Feldman, Family Medical Costs and Voluntary Health Insurance: A Nation-Wide Survey (New York: McGraw-Hill, 1956).

4/ Odin W. Anderson, Patricia Collette, and Jacob J. Feldman, Changes in Family Medical Care Expenditures and Voluntary Health Insurance: A Five-Year Resurvey (Cambridge, Mass.: Harvard University Press, 1963).

5/ Data from another study of medical care spending in relation to family income are contained in U. S. National Center for Health Statistics, Medical Care, Health Status, and Family Income, United States (Public Health Service Publication No. 1000, Series 10, No. 9. Washington: U. S. Government Printing Office, 1964). In this report, however, expenditures are presented on a per-person rather than a per-family basis.

In addition to the national surveys, there have been numerous local and specialized studies of health care expenditures and the utilization of medical services. References may be found in footnote three of Odin W. Anderson and Jacob J. Feldman, op. cit., pp. 1-2, and footnote six of Odin W. Anderson, Patricia Collette, and Jacob J. Feldman, op. cit., p. 3.

6/ In each of the three studies, charges were also related to other characteristics of families and their members, such as age, family size, and area of residence.

7/ Cross-sections in addition to the ten presented here may be found in Helen Hollingsworth, Margaret C. Klem, and Anna Mae Baney, Medical Care and Costs in Relation to Family Income: A Statistical Source Book (U. S. Social Security Administration, Bureau of Research and Statistics, Memorandum No. 51, 2nd ed. Washington: U. S. Government Printing Office, 1947); U. S. Bureau of the Census, Historical Statistics of the United States: Colonial Times to 1957 (with the cooperation of the Social Science Research Council. Washington: U. S. Government Printing Office, 1960); Life Study of Consumer Expenditures: A Background for Marketing Decisions, Vol. 1 (conducted for Life by Alfred Politz Research, Inc. New York: Time, Inc., 1957); and George Katona, Charles A. Lininger, and Richard F. Kosobud, 1962 Survey of Consumer Finances (The University of Michigan, Institute for Social Research, Survey Research Center, Monograph No. 32. Ann Arbor, Michigan: The University of Michigan, 1963). These cross-sections were not used because they relate to nearly the same population as those discussed or to specialized areas, or because averages were not provided for both income and medical care spending.

8/ The income elasticities were calculated by transforming the income and expenditure data to logarithmic form and fitting a regression line to the logarithms, the elasticity being the coefficient of the slope of this line. This regression model differs from the one used previously in that it assumes a constant percentage change in medical care spending per per cent change in income rather than a constant dollar change in expenditure per dollar change in income.

For reasons to be discussed below, the calculated elasticities are not necessarily the "true," or structural, elasticities of demand because they do not measure the effect of income on medical care spending exclusive of the other factors affecting the level of expenditure.

9/ For a contrary view, see Reuben Kessel, "Price Discrimination in Medicine," Journal of Law and Economics, I (October, 1958), 20-53.

10/ It is difficult to think of medical care as being a service of homogeneous quality. Therefore, even with similar "quantities" of service, a person with a relatively high income paying a higher price may be receiving a somewhat different product. Because of this element of "trading-up," income elasticities calculated from expenditure data will result in higher estimates than elasticities based on quantities as they are usually measured, e.g., dental visits.

11/ The distinction between permanent and transitory components of income and their relationship to consumption is set out in Friedman's permanent-income theory of consumption. See Milton Friedman, A Theory of the Consumption Function (Princeton, N. J.: Princeton University Press, 1957).

12/ Transitory income may have an important effect on expenditures involving investment, such as the purchase of durable goods, even though the consumption of services provided by these items is not significantly affected by temporary income variations. This phenomenon would apply to medical care to the extent that it may be considered as an investment and that the timing of expenditures is discretionary.

13/ Random errors in the amount of income reported will bias the income-expenditure regression in the same direction as differences in transitory income because they affect the level of income reported but not the level of expenditure.

14/ It is possible that the increase in the estimated effect of income could have resulted from the averaging of other factors affecting medical care expenditure. However, a multiple regression analysis based on family data using the same variables as those in Table 3 resulted in a lower income elasticity estimate than one based on family income alone.

15/ Those who find difficulty in believing that failure to take the transitory component of income into account can lead to such a large downward bias in the measured income-expenditure relationship should consult Margaret G. Reid, Housing and Income, (Chicago: University of Chicago Press, 1962). The thorough and painstaking analysis in this monograph reveals that the elasticity of housing expenditure with respect to normal income is in the range of 1.5 to 2.0, whereas other cross-section estimates, based on reported income, had placed it below 0.5. Most of the techniques of analysis applied to medical care here were suggested by Miss Reid's work on housing. Those interested in the application of the concept of permanent income to consumer demand studies will find worthwhile reading in Chapter 2 of Housing and Income, in which the theory is set forth and methods of deriving estimates are summarized.

16/ Urban area estimates were also calculated for 1950 by combining the city average data but did not raise the estimated effect of income.

17/ In considering the relation of third-party payments to health insurance plans, we have been aided considerably by discussion with Robert G. Rice of the University of Chicago, who is carrying out an extensive study of factors affecting employer expenditures for private wage supplements.

18/ See Table 2 of Louis S. Reed and Dorothy P. Rice, op. cit., p. 4.

19/ Degree of urbanization was included as a variable because it was thought that there was a greater likelihood of enrollment in group plans in the more highly urbanized and industrialized states. The fact that imputed income from farm products is not reflected in the income variable may have been a factor in lowering the measured effect of urbanization. This is the case because the level of insurance premiums predicted from income alone would be too low in the less urbanized states due to the understatement of income.

20/ Robert Rice has estimated that roughly one-half of total health insurance premiums are paid by third parties. Thus, the fact that the elasticity of family health insurance expenditures in 1960 was nearly zero (Table 4, equation 8) implies that the income elasticity of third-party payments must be much greater than 1.0.

21/ Using the standard formula for point elasticity:

$$\eta = \frac{dM_1 + dM_2}{dY_1} \cdot \frac{Y + M_2}{M_1 + M_2} =$$

$$= \frac{\$42.8 + \$24.0}{\$1,000} \cdot \frac{\$5,550 + \$86.6}{\$335.8 + \$86.6}$$

$$= 0.883$$

where M_1 = average family medical care expenditures = \$335.8, M_2 = average third-party payment of health insurance premiums (per family) = \$86.6, Y = average family income (Table 5) = \$5,550, dM_1 = regression coefficient of family medical care expenditures (Table 5, equation 1) = \$42.8, dM_2 = estimated regression coefficient of third-party payment of health insurance premiums (Table 6, equation 2, minus the value in Table 4, equation 6) = \$24.0, dY = income unit used in calculating regression estimates = \$1,000. Average third-party payments of health insurance premiums were assumed equal to family expenditures for health insurance. Although the data upon which this estimate of the elasticity was based were obtained from different sources, it is believed that the derived estimate is a reasonable approximation.

22/ The real elasticity of medical care spending may have been more stable than the other cross-section estimates in Table 2 appear to indicate. Much of the variation in measured elasticity may be accounted for by differences in the degree of transitory income or by variations in survey eligibility requirements. For a discussion of these considerations, see Milton Friedman, op. cit., Chapter 4, pp. 38-114.

23/ The estimate of income elasticity derived in this paper is higher than that found in two previous studies: Grover Wirick and Robin Barlow, "The Economic and Social Determinants of the Demand for Health Services," in The Economics of Health and Medical Care (Ann Arbor, Michigan: The University of Michigan, 1964), pp. 95-127, and Paul J. Feldstein, "The Demand for Medical Care," in Report of the Commission on the Cost of Medical Care, Vol. I, General Report (Chicago: American Medical Association, 1964), pp. 57-76. It is possible that these differing estimates may be reconciled, however. In the former study, for example, data relating to individuals were used. In the latter study, employer contributions to health insurance were not added to the family data.

DISCUSSION

Herbert E. Klarman, The Johns Hopkins University

Although the Feldstein and Carr paper and Rosenthal's paper share a common tradition, that of applying advanced statistical techniques to the study of economic problems, they are sufficiently different to warrant separate discussion.

FELDSTEIN-CARR PAPER

This paper is instructive, ingenious in design, and lucid in presentation. It is lucid, in that complicated ideas are presented simply. It is ingenious, in that diverse techniques are brought to bear on the question at hand. It is instructive, in that existing data which have lain dormant in many reports and repositories receive a fresh interpretation.

Technical Comments

The authors aver two purposes: (1) to ascertain the net effect of a change in income on (private) medical care spending; and (2) to appraise the stability of the relationship over time. Their major conclusion is that the income elasticity for medical care spending was more than one in 1950 and probably near one in 1960. The calculated figure is often smaller because the reported income of consumers includes transitory income and because employer contributions to health insurance premiums are excluded. The downward effects are aggravated when simple correlation is replaced by multiple correlation, and income becomes only one of several independent variables.

By following the authors' formula it is possible to raise the 1960 estimate of income elasticity from .883 to .930, if the proportion of employer contributions to total health insurance premiums is reduced. This is warranted because the figure of one-half employed in the paper is too high. For the year 1960 two-fifths appears to be more nearly correct.

In order to gain additional insight the authors go beyond correlation analysis of cross-section data for households. In this discussant's opinion the analysis of time series does not enhance our understanding. As Louis Paradiso has pointed out, in the post-war period annual changes in consumer spending have reflected growth relationships almost exclusively, rather than sensitivity to changes in income. Nor should one regard D. S. Lees' incidental observations on this score as considered comments.

To deal with the problem of transitory income the authors resort to an analysis of data for 91 cities. The question arises whether equally good, or better, results would not be obtained in the preceding analyses by substituting family expenditures for reported family income. The sacrifice of information concerning the characteristics of individual families would thereby be avoided.

It is appropriate to note that Feldstein and Carr are not measuring income elasticity in a strict sense. Given a one percent change in income, they aim to measure the associated percentage change in medical care expenditures, rather than in the quantity of services taken. As we shall shortly see, this distinction is not unimportant when changes take place in the volume of free care rendered or in the application of the sliding (variable) scale of fees.

Support for Hypothesis that Income Elasticity Has Declined

The authors may wish to consider the possibility that the income elasticity of medical care spending has declined. Their data in Table 3 show lower figures for 1960 than for the earlier years. Certain institutional changes appear to lend support to the hypothesis that such a shift has taken place.

1. The volume of free care has declined. This would serve to raise the proportion of medical care spending to income for families with low income, thereby reducing the income elasticity for all families. (It makes no difference whether free care is rendered without charge or is paid by government, as long as it is not reflected in private expenditures.)

2. The sliding scale of fees is less pervasive than formerly and its range is probably narrower. The effect is to reduce the income elasticity of medical care spending.

3. The proportion of total health insurance premiums paid by employers has increased. A constant sum is, in effect, added to the income and expenditure distributions for employed persons, and the effect is the same as above.

Implications for Policy

If the income elasticity of private medical care spending has in fact declined so that it now falls below one, what are the implications for the volume of medical care spending in the future? At first impression the answer is that the proportion of aggregate income devoted to medical care will decline.

There are certain offsets, however, which should be recognized. Health insurance is likely to continue to increase medical care spending in two ways. The two-price system (a lower one for beneficiaries of insurance) leads to increased utilization by lowering the price of insured items relative to other objects of expenditure. In addition, persons with health insurance seem to spend more than uninsured persons on all types of health and medical service, including uninsured items.

Also to be taken into account is the prospect that the unit cost of hospital service will continue to rise at a high rate, owing to the hospital lag in productivity gains behind other industries. This factor is independent of any increase in the general price level, increased costs of educational programs, or extraordinary improvements in the technology and efficacy of medical care.

The discussant concludes, therefore, that there is little reason to expect that an income elasticity of less than one at a given time will necessarily signify a decline over time in the proportion of aggregate medical care spending to aggregate income. The simultaneous occurrence of other changes may offset any such tendency.

ROSENTHAL PAPER

This paper offers a principal components analysis of the demand for general hospital facilities. It is a progress report on a continuation of the author's original work on the same problem. The findings are preliminary and their interpretation is boldly imaginative.

Conceptual and Technical Comments

The objective of the new work is to simplify and improve the original 12 variable model, by reducing the number of independent variables. Five principal components explain 92-93 percent of the variation in patient days. Since the components are statistical creations, the author seeks to represent each by one of the original variables. Initially this is done on the basis of factor loadings; however, the procedure for selecting representative variables is not mechanical.

For the first component the proportion of the population with low income is preferred to the proportion under age 15 because the former is more meaningful on a priori grounds. Marital status is chosen to represent the second component on a priori grounds plus statistical stability. The urban-rural variable is chosen for the fourth component on the ground of statistical stability. Only for the third and fifth components are age and level of education, respectively, chosen in accordance with the factor loadings.

What is an a priori ground? The dictionary states: "that which can be known by reason alone, not by experience." Rosenthal did, of course, look at experience, for he refers to what other investigators have found and reported. By a priori he must mean that which is plausible or makes sense. How does one deal with mutually contradictory explanations that appear to be equally tenable? (Consider the change in sign for the factor loading of the low income variable between 1950 and 1960.)

The new model is evaluated in part by comparing it with the original model for efficiency of prediction. The conclusion is that the new model is a statistically satisfactory substitute. Yet for the year 1960 a loss of 19-24 percentage points is reported in the value of R^2. How large a loss does it take to render the new model unsatisfactory?

The stated purpose of the analysis is to understand the factors that influence hospital use and to apply this knowledge to predicting use. Since variables pertaining to the organizational structure of medical care are not now contained in the model, the interpretation of findings to-date may attempt to explain too much.

The organizational structure is not clearly set forth. Apparently, it is intended to represent the forces on the supply side. Subsumed are not only the forms of physicians' practice but also the existence of substitute medical care facilities. No reference is made to the possible effect on use of the supply of general hospital beds. Rosenthal dealt with this subject in his monograph, but not conclusively in this discussant's opinion. D. J. Newell's data were not then available; and the factors that impinge on the demand for and supply of a service are by no means identical.

It is doubtful that the multiple regression approach can handle special institutional controls over hospital use. An example is the relatively low use attained by a self-insured fund in whose behalf a labor union has enlisted the protective concern of its members.

The separate analyses of the admission rate and duration of stay are welcome, and in line with steps recently taken by investigators in England, including Norman Bailey, Martin Feldstein, and G. Forsyth.

A comment may be in order on the superior performance of the equations for duration of stay. It may reflect the greater ability of this element of hospital use to adjust to various influences and pressures. To Norman Bailey such adaptability is a focal point for hospital planning.

Findings on the Independent Variables

Perhaps this discussant has unreasonable expectations of a new, exploratory approach. His difficulty is that sometimes he cannot see in the data what the author sees.

It is reported that the low income variable performs well in a statistical sense, being significant in almost all cases. Although Table 3 shows it to be significant for patient days, it shows lack of significance for the admission rate and duration of stay separately.

It is recognized that the contribution of the first component, ability to pay, is less important in 1960 than in 1950. The change is attributed to the enhanced importance of either the organizational structure of medical care or health insurance. Further analysis is indicated.

It is reported that the second component, represented by marital status, has a positive association with the admission rate and a negative one with duration of stay. Table 3 shows the latter, but not the former for the year 1960. Moreover, none of the coefficients is statistically significant.

Marital status has a negative relationship with patient days. The interpretation offered is that the home serves as a substitute convalescent facility. This is consistent with the findings of Brian Abel-Smith and R. M. Titmuss. It may be, however, that the true relationship is more complicated, with married persons having the higher use among the young and the lower use among the aged. It might be helpful to exclude obstetrical use from the data.

Unlike the original model, the new one shows a positive association of use with age (which represents the third component.) It is found that the association is operative more strongly through the admission rate than through duration of stay.

This discussant would have expected the exact opposite. It is not so much that "the disease mix is more heavily chronic" for the aged, but rather that they have a longer stay for almost every type of admission.

A greater impact of the proportion of aged is reported in 1960 than in 1950. This is not evident from Table 3. Notwithstanding, the conclusion may be correct. The interpretation offered is that with the dilution of the economic factor, physiological factors have come into play. It will be recalled, however, that health insurance is not dealt with. Another possibility is that the older aged (75 years and over), who use considerably more services than the younger aged (65-74), are becoming more prominent in the population.

None of the coefficients for the urban-rural variable is statistically significant. As for the interpretation of findings, it is doubtful that persons hospitalized in cities are sicker, except insofar as large cities may attract patients from more distant parts. Even so, such patients must be sufficiently hardy to survive travel. Other factors may be operative, such as the lower proportion of obstetrical (short-stay) patients in cities; the presence of major teaching hospitals, with attendant prolongation of stay; and the presence of long-term units in municipal general hospitals, designated or otherwise.

The findings concerning education and hospital use are tenuous. Their interpretation is necessarily speculative.

Summary

In this paper Rosenthal has submitted early results of an improved statistical model of general hospital use. Since important variables are still not contained in the present model, reported findings are likely to be modified.

Some of the findings appear to be tenuous, and not able to support an elaborate structure of interpretation.

The implications drawn by the author are premature. An appraisal of the implications awaits the findings of more complete analyses.

COMMENTS ON TWO PAPERS ON THE STATISTICAL ANALYSIS OF THE DEMAND FOR MEDICAL CARE:

Paul J. Feldstein and W. John Carr, "The Effect of Income on Medical Care Spending"

Gerald Rosenthal, "The Demand for General Hospital Facilities:
A Principal Components Analysis"

Jerome Rothenberg, Northwestern University

A. Feldstein-Carr Study

I shall discuss the Feldstein-Carr paper first. This is a deep, well-thought out study. The field of medical economics could use many more like it. I have three comments to make, two on the permanent income adjustment, the other on the interpretation of results.

1. Appropriateness of Permanent Income Adjustment

A central part of the Feldstein-Carr approach is to regress demand for medical care on an income variable (among the several variables) which represents permanent income - i.e. a variable from which as much of the transitory component of income has been removed as possible. The removal of the transitory component from income has a most significant upward effect on the estimated income elasticity of demand. Yet one can ask whether this adjustment is warranted after all.

The rationale of the adjustment is the Permanent Income Hypothesis, which holds that consumer expenditures are more closely associated with permanent income than with current measured income. Current measured income is deemed to contain both a permanent component and a transitory component; and the correlation between consumption and transitory income is assumed to be zero. So the measured association between consumption and current measured income is an impure reflection of the primary relationship - that between consumption and permanent income.

But in this formulation, consumer purchases of durables - i.e., consumer investments - are not considered a part of consumer expenditures. They are considered rather a portfolio management decision, affecting the form in which consumers hold their wealth, and only the flow of services from these durables during the accounting period is treated as current consumption. The portfolio decision is deemed to be significantly affected by transitory income. Consumer purchases to develop productive skills (as, for example, through education) are generally considered to represent one form of consumer investment - an investment in human capital - and would therefore warrant treatment as portfolio decisions. But so too is at least part of the investment in health. All medical care is directed toward future well-being. Of course, most of the consumption of so-called "necessities" have a future orientation too, in the sense that grossly inadequate provision for food, clothing or shelter will prevent a consumer from having any future. But besides this sense, some medical care is so clearly predicated toward future productivity and well-being, with long-lasting effects (as, for example, correction of defects, preventive care, direct life-saving services), as truly to constitute another form of consumer investment in human capital. This means that purchase of such care should be sensitive to transitory components of income.

If this is so, then it is indeed total measured income, with both permanent and transitory components, that is more closely determinative of the demand for medical care than is permanent income alone. Thus, for at least some portion of total expenditure on medical care, the Feldstein-Carr adjustment to permanent income is unwarranted. The appropriate income variable may well be current measured income after all; and with it, the substantially lower income elasticity of demand.

2. Approximation of Permanent Income

If, notwithstanding my first comment, the authors wish to use permanent income as their income variable, this can be accomplished somewhat more directly and with more flexible results (since, unlike their present methods, this can be applied to individual families). Instead of, or in addition to, the averaging techniques used in the paper to eliminate transitory elements in income, permanent income can be approximated by using total per family consumer expenditures (after subtracting out consumer investments and adding in the value of imputed consumer services from the stock of consumer durables). Under the Permanent Income Hypothesis, total annual consumption is asserted to be proportional to permanent income.

3. Interpretation of the Income Elasticity of Demand

In order to interpret the paper's results, the very concept of income elasticity of demand for medical care must be given deeper consideration than is accorded it in the paper. The incidence of illness - which generates the "need" for medical care - is generally highly irregular among families at every income level, showing more variation within income levels than between income levels; and introducing significant slippage between income and demand despite any tendency for the perception of "need" to be related to income level. Moreover, in no other consumer sector is consumer sovereignty on how to meet the "need" so substantially delegated by the consumer to someone else, nor is the resultant provision of care so apparently income-neutral. Income-tailored features of medical care seem to be relegated to peripheral frills rather than to the essentials of treatment. Further, serious illnesses seem generally to receive treatment, regardless of the income level of the patient, through an important tradition of charity, sliding scales of payment or public provision of care.

Given this background, to be told that the income elasticity of demand for medical care is unity is surprising. It implies that a constant proportion of (permanent) income is spent on medical care. While this is not inconsistent with the background characteristics cited, it represents a knife-edge regularity that is not obvious, to say the least. To know the significance of these results for planning medical needs, or even just to understand consumer behavior, we must know much more about the *composition* of the demand for medical care by different income groups. (The income elasticity here is a cross-sectional one referring to income *differences* among different families, not a temporal one referring to income *changes* for the same families. It may be quite amiss to predict temporal elasticity from the cross-section elasticities, because of some of the issues to be mentioned now.) We must know, for example, some of the following:

a. To what extent does the estimated elasticity reflect simply price discrimination, whereby prices charged differ for recipients of different income levels, so that the total *cost* of services varies by income level far more than does the quantity or even quality of services? To what extent are actual quantities or qualities of service, involved? To what extent do commodity packages going to different income groups differ chiefly in quasi-medical adjuncts of care - such as larger or more sumptuous hospital rooms or more privacy of care - rather than in the more directly medical attributes of the package?

b. To what extent does estimated elasticity reflect different compositions of care - as for example, a larger share of patent medicines for lower income groups, a larger share of physician and hospital care for higher income groups? The relative effectiveness of different bundles - that is, the "real" amount of care received - may well be related to these different configurations.

c. To what extent does the elasticity reflect income-induced differences for some kinds of illnesses, but not for others? For example, do patients at different income levels get much the same care for "serious", non-discretionary ailments, with significant differences showing up chiefly for "non-serious", or discretionary problems (cosmetic operations, corrective as opposed to ameliorative, dental care - orthodontia, etc.)?

d. Is the estimation subject to measurement bias? For example, higher income groups typically obtain psychiatric care in the form of private financing of psychiatrists and private hospitals. These expenditures would be included in the Feldstein and Carr figures. Lower income groups typically obtain psychiatric care (admittedly not the same kinds of treatment, nor for the same composite of illnesses) in publicly supported mental institutions. The cost of these services would be excluded from figures on private medical care expenditures. This kind of systematic difference would produce an upward bias in the computed income elasticity.

B. Rosenthal Study

This is an imaginative, thought-provoking study. It attempts to reduce the dimensionality of an explanatory cross-section schema in order to perform time series analysis, where data availability is a critical barrier. An important collateral goal is to reduce the possibly substantial multicollinearity of the original schema. While the method and some of the findings are suggestive, some of the procedures used raise serious questions.

1. Dimensionality and Multicollinearity

The dimensionality of the model is reduced by finding five orthogonal principal components. Yet these components are not themselves used as explanatory variables in the reduced model: only one original characteristic with high factor loading (out of the twelve constituents) is chosen as a proxy for each component. One can raise the following questions:

a. The factor loadings for each principal component are not so *uneven* that any *one* factor clearly and exclusively represents it. Moreover, no single factor captures what appears to be the broader sense of any of the principal components.

b. While the system composed of the five principal components as explanatory arguments minimizes multicollinearity, since the components are orthogonal, the set of proxy variables selected from the principal components is not necessarily orthogonal; so the system employing these as arguments does not necessarily exclude multicollinearity. Specifically, the loadings on each principal component are generally quite low and diffuse; and a factor selected to proxy for one component will sometimes have significant loadings on other components (especially urbanization and education). Very likely the education, low income and urbanization variables are related; likewise, old age and marital status. Thus, in fact the reduced model may well have retained a significant degree of multicollinearity. This is an empirical question. It would be very helpful to have a correlation matrix for the explanatory variables in order to evaluate the seriousness of this problem.

c. There are devices open to Rosenthal that might decrease multicollinearity and give better explanatory power. Contrary to Rosenthal's statement about the principal components ("Since these influences are statistical creations, however, they can not be measured directly." Page 8.), they can indeed serve as arguments in a reduced model. Each is a specific linear combination of the twelve original factors and can be measured as a separate variable by measuring its constituents. Their use by Rosenthal is, however, precluded for time series analysis because of the unavailability of data. To bypass this barrier, however, a modification of the procedure can be employed: linear combinations of various representative clusters of three or four constituents from each principal component can be used, selected for data availability. This could give much better coverage to the sense of each principal component.

2. Choice of Variables

Within Rosenthal's method, his choices of proxy variables are frequently questionable. Consider some details.

a. Principal Component 1 (C_1): Low income was chosen. It had high loadings in both years but - and this is true of almost every other factor in this component - these loadings have different signs in 1950 and 1960! To interpret "low income" as representative of something that is claimed to have the sense of "economic prosperity" is most peculiar when this proxy has a high positive association with C_1 in 1950 and just as high a negative association with it in 1960.

The characterization of C_1 as an "economic prosperity" variable is not apt. Price, high income, young age and urbanization also have high loadings (with opposite signs between the two years). Yet, by a priori reasoning, price should have a different kind of marginal influence on hospital demand than the two income variables; and these latter two are themselves not nearly perfectly correlated: it is perfectly conceivable to have both a high percentage of the population with incomes above $5,999 and with incomes below $2,000. Further, the very use of these absolute figures to represent distinctive situations is itself suspect, since these absolute income levels have a very different "real income" significance for hospital demand between 1950 and 1960, because income and hospital costs changed appreciably during the period. If anything, these two income variables characterize not the degree of prosperity but the population structure within each state. This is an example of my earlier argument about the desirability of using clusters of constituents as individual variables in order to get better representation.

As a compromise here, however, one could use per family income as the proxy variable. This would incorporate information from both income variables used by Rosenthal.

b. Principal Component 4 (C_4): The urban-rural variable was chosen. This is referred to by Rosenthal as reflecting the degree of crowdedness in housing. Such a characterization is misplaced. The actual association is poor: rural slums have as much crowding per dwelling unit as urban slums; perhaps more, since rural family size is probably larger. Indeed, the family size variable seems the better representative of crowdedness, if crowdedness is what is wanted.

Yet crowdedness is probably not what Rosenthal wants after all. In his analysis he refers to this variable as indicating the existence of alternatives to hospital care - but not so much for convalescence, where crowdedness is relevant, but for the availability of treatment services. Degree of urbanness is relevant to the availability (or accessibility) of hospital and other medical services. Yet this fact leads to a more serious problem. The supply of hospital accommodations is related to the urban-rural ratio. Moreover, there is good reason to believe that the supply of hospital services influences hospital use. Thus, inclusion of this supply side of the market begs the whole identification question. Introducing a variable associated with supply into a demand function, without also explicitly introducing a separate function explaining supply, makes it extremely difficult to interpret the estimated function as either a demand or a supply function. It is in fact neither, but only a composite.

Thus, the urban-rural factor - which shows no significant relationship in the regressions anyway - should be excluded to preserve identification of the estimated function as demand; its place should be taken by either the family size or race variable, both probably a better proxy for crowdedness anyway (and crowdedness is a relevant argument in a demand function).

3. Interpretation of the Results

Generally, only the coefficients explaining total patient days utilization are statistically significant. With two exceptions, no coefficient referring to admissions or length of stay is significant. (The exceptions are the old age and marital status variables for admissions in 1950.) Yet non-significant coefficients are subjected to heavy interpretation as though their signs were meaningful. The urban-rural variable is so used for all three dependent variables; most of the others are used for admissions and average length of stay. Rosenthal's interesting - and by no means far-fetched - conjectures are nonetheless based on non-significant findings. They are pushed much too hard.

On the other hand, no attention is paid to the extreme instability of the otherwise splendidly significant partial coefficient for the effect of education on total patient days between 1950 and 1960. It changes from 23.13 to 2.15! Some thought to the interpretation of this finding is called for.

4. Effect of Economic Constraints

Rosenthal conjectures that "economic constraints" and the availability of substitutes affects hospital utilization via the average length of stay rather than via number of admissions. This throws some light on my discussion of the interpretation of the results in the Feldstein-Carr paper. While the conjecture is, as I have argued, not backed up by statistically solid evidence, it is consistent with my comments on the background of the demand for medical care. It suggests that the extra medical care spending by higher income groups is not due to a greater "need" (incidence) but to greater "quality" or quasi-medical concomitants of care for each medical condition. This suggestion must of course be qualified, not only by the lack of statistical significance of the Rosenthal evidence, but also by the fact that the income variables used in the two studies are quite different. Nevertheless, a sign that these studies are useful, whatever their difficulties, is that the reader is sorely tempted to confront each with the other to generate a dialogue. So progress may be generated.

VI

STATISTICAL DEVELOPMENTS IN EDUCATION RESEARCH

Chairman, Herbert Solomon, Stanford University

Page

Development of a Data System for Longitudinal Research - Robert W. Marker, University of Iowa. 114

Statistics, Measurement, and Educational Research - Henry F. Kaiser, University of Wisconsin.. 117

The Office of Education Cooperative Research Program and the Statistician - Howard F. Hjelm, U. S. Office of Education...................... 120

DEVELOPMENT OF A DATA SYSTEM FOR LONGITUDINAL RESEARCH

Robert W. Marker, University of Iowa

Professor E. F. Lindquist, originally scheduled for participation on this panel is unable to be with us. He has asked that I express to you his sincere regret in missing this opportunity. I shall attempt to treat the subject with what abilities I can muster but I cannot help but share with you a sense of loss brought about by this program change. I would count it a privilege if I could join your ranks and listen to his views on longitudinal research.

Fortunately, my task is relatively simple. I intend to outline our approach to the development of a data collection and retrieval system in Iowa. We are currently engaged in the design of a statewide data system which, if successful, should prove valuable for purposes of longitudinal research. I shall attempt today to describe;

1) something of the growing need for information,
2) an agency created to design and implement a statewide data system,
3) the potential of such a system for longitudinal research,
4) our base of service activities and how these relate to research information, and
5) the nature and extent of the support necessitated by this approach to data collection.

THE GROWING NEED FOR INFORMATION

We are witnessing dynamic changes in education. Changes in quality and quantity. Permit me to crystalize this point by citing some obvious examples. We continue to see gross increases in the number of persons educated as well as the duration of their formal schooling. Changes are occurring in the internal and external structure of educational organizations. Experimental programs in elementary and secondary curricula are being initiated in the social sciences as well as in the physical and biological sciences, mathematics and foreign languages. A mounting concern is evidenced for educational excellence at all levels, kindergarten to graduate school. The American urge to extend our systems of general, liberal, and vocational education is finding expression in legislation and practice. This mounting concern is complemented by another, no less intense, for intelligent and prudent investment of the nations resources allocated to educational functions. The changes are creating an unprecedented need for educational information. We need to identify trends, examine new programs, evaluate outcomes, communicate desirable adaptations, and develop appropriate operational educational policies at federal, state and local levels.

Change is not the only agent precipitating the need for more and better information. Education has a poor record in the matter of acquiring timely and accurate information for operational purposes. Broad scale research efforts in education can best be implemented when the required data are "creamed" from day-to-day operational information processes. Both researchers and administrative personnel are now demanding that the "information lag" be closed and that systems be designed to provide for operational needs as well as research purposes.

School teachers, supervisors, administrators, board members, parents, professional organizations, and the general public are asking for more information about our schools; universities, state education departments, accreditation bodies, and the United States Office of Education also require more and more data for policy and operational purposes. Add to this the increasingly important role of the researcher within all these groups and a case has been stated concerning the growing need for processed data.

Schools and various agencies collect data of many types and in many ways, but there is serious inefficiency and lack of coordination, considerable overlap and duplication, and frequent repetition of tasks required of persons supplying this information. Teachers and administrators, filling out numerous questionnaires and report forms for different agencies and for different purposes, find that they are supplying exactly the same detailed items of information over and over again, from agency to agency, and from year to year. Local school systems also find many inadequacies and imperfections in their systems of collecting data for their own use as well as in the systems of maintaining and utilizing their own records.

Although it is impossible for the schools and agencies to be too well informed about the many facets of their operations, it is possible to collect and have available more detailed information than can be digested and assimilated, or interpreted and used effectively. The needs of school administrators for better methods of data collection and for more information are surpassed only by their needs for help in organizing, integrating, interpreting, and using the information that is being and may be collected.

A CENTER IS CREATED

We have created an organization to search for and implement answers to this informational problem. This agency is called the Iowa Educational Information Center. It is closely associated with the policies, personnel, and equipment of the State Department of Public Instruction, the University, and the Measurement Research Center notable for its record in test development and test scoring services. This new Center is organized to establish and maintain a comprehensive system for the continuous, systematic, and routine collection, storage, and retrieval of information about all phases of the educational enterprise - about the professional staff, students, curriculum, school facilities, and financial activities. Data thus collected, are being transcribed to a magnetic tape file, merged, integrated, organized, and stored in a form for most convenient access to large scale computers. This file will be kept continually up to date. From this file, schools and educational agencies can draw the information needed to perform their many functions most intelligently. From it can be drawn, also, the information needed for a great variety of research purposes.

LONGITUDINAL RESEARCH THROUGH CardPac

Longitudinal research is dependent upon the availability of this kind of systematic data collection. Information collected and stored in the central file will be made readily available to research workers everywhere. Much work is needed in educational and psychological research to determine the antecedents, the earliest observable symptoms, and the basic causes of various types of human behavior not easily recognized and identified until relatively late in the life of the individual. The school "dropout," the juvenile delinquent, the underachiever, the educationally and vocationally dis-oriented or non-oriented student, the leader, the specially talented individual, the creative individual, all evidence types of behavior that have their roots in the earlier experiences, environment, and inheritances of the individual. These should be recognized, precluded, or strengthened as early as possible. Heretofore, longitudinal research studies into these problems have necessarily been of the backward-looking type, in which the needed historical data about the individual have been collected only after he has clearly manifested the critical behavior.

The approach now being taken by the Iowa Educational Information Center makes possible a new kind of "Follow-back" study wherein various kinds of data, such as achievement test scores, academic marks, and vocational-educational aspirations may be integrated for a given student and also traced longitudinally year by year. Given a data file on a complete population several other possibilities present themselves to researchers. It will be relatively simple to select specific samples such as all identical twins within the state. Random samples can be generated automatically with increased potential for meaningful stratification. One could even use this data file as painless instrument for testing sampling theory.

In the Spring of 1965, the Information Center is initiating a statewide system known as CardPac for the collection of information from every secondary school in Iowa. Empasis this year is to identify data from every teacher and every student. The system employs the principle of batching information about each of the courses taken by a given student with questionnaire information supplied by the student himself into a pack of tabulator cards. The teacher places identifying information about herself on top of this pack of cards and this completes the CardPac. Each teacher, course, and student is identified by a continuing number. Thus, the several sources of information can be connected and integrated for a given year and in the future longitudinal information can be supplied for as many years as the system has been functioning. The population supplying information includes about 30, 000 professional educators and 250, 000 students.

STUDENT ACTIVITIES ARE BASIC

Service to local schools make possible the kind of approach we are taking in Iowa. A data collection project of the magnitude just described can only hope to succeed if it is tied to some useful purpose as perceived by the persons supplying the information. For this reason a great deal of our effort centers on supplying local school personnel with timely and useful information selected from the CardPac system and other sources. This "feedback" service provides some insurance of intimate support by those providing the data. Indeed the early activities of the Iowa Educational Information Center have been characterized by the service activities such as secondary student scheduling, grade reporting and attendance accounting. These services and other being developed are provided through our "UPDATE" program of local district computer services. We now schedule about 50, 000 students in some 50 schools annually. We will continue to provide these and similar services while awaiting full scale action by state and local authorities in organizing to take advantage of automation in the schools.

RESEARCH SUPPORT REQUIRES MACHINES, MONEY AND MEN

One reason which makes it possible for us to handle this approach to the collection of research data is the availability of large-scale and specialized computer hardware. The dramatic growth of the computer industry and the methodologies of processing data are making possible an answer to the total information problem. Output speeds have consistently increased for tape processors as well as printers. We can now print at a rate of 1100 complete lines per minute. Processing speeds have increased even more markedly to the point where some individual processes are measured in nano-seconds or billionths of a second. In addition, parallel and multi-processing techniques have improved the data handling characteristics of computers. The near future will see similarly startling modifications and extensions of the necessary hardware. Improvements over the key-punch method of in-put preparation are much needed. Optical scanners which read characters or positions have proved highly beneficial in the data collection process. The Measurement Research Center is currently designing and manufacturing an optical scanner which will read both punched information and scannable marks at the rate of 35, 000 cards per hour. This scanner will be used to "read" the millions of cards associated with CardPac.

The Iowa Educational Information Center's initial pilot program began under a grant from the Measurement Research Center. MRC is continuing to develop and improve its automatic electronic scanners which can now transcribe information from an original data collection document or report form directly to magnetic tape. MRC provides to the Information Center, on a strictly at or below cost basis, all required data processing services on document scanners and IBM 1460 computer equipment, as well as the services of its staff.

The College of Education conducts the Iowa Testing Programs in which the 5 to 7 hour comprehensive Iowa Tests of Basic Skills and the Iowa Tests of Educational Development are administered annually under closely controlled conditions to almost the entire school population of the state in grades three to twelve. Over 400, 000 Iowa pupils took these tests in 1963-64. The tests are centrally scored by MRC, the scores permanently recorded on magnetic tape, and the information made immediately available to the Information Center without imposing any additional reporting burden on the schools. The College of Education provides personnel and office space for the Information Center staff, and members of the faculty contribute professional assistance in specifying the information to be collected. Facilities of the University Computer Center are also available to the Information Center, including a large scale, IBM 7044 processing system.

The State Department of Public Instruction provides personnel and access to its facilities as well as financial support for the Information Center. The Department has traditionally enjoyed unusually good relationships with the public schools of the state, and its support of the data collection activities insures that the plan as a whole will be an integral part of the legally constituted educational structure of the state. As the Center grows the Department will depend increasingly upon it for collection of data.

Funds in support of the agency and its development of data systems have been granted by the U.S. Office of Education and the Ford Foundation. It is expected, if successful, that the Iowa Educational Information Center will serve as a model for the development of information systems in other states. Iowa possesses a community of dynamic and dedicated educators. It has a rich heritage in data processing and educational innovation. A good beginning has been made in securing the necessary machines, money, and men. Given a period of years for implementation, evaluation, and revisions of the data system being developed we should be able to fully explore the potentially fruitful realm of longitudinal research.

STATISTICS, MEASUREMENT, AND EDUCATIONAL RESEARCH

Henry F. Kaiser, University of Wisconsin

This afternoon I should like to make a short survey of some random ruminations and polemical personal pontifications regarding educational measurement and statistics, and point out possible implications of my remarks with respect to the scientific study of education.

Preliminary Thoughts

Because of a failure to distinguish measurement problems from statistical problems, traditional training has often confused and thus confounded these two distinct aspects of quantitative methodology; witness the unfortunate synonymous use of "reliability" and "significance." Current thought attempts to react against the earlier confusion to separate, probably to too large an extent, measurement from statistics. For in one sense such a distinction is a little artificial: almost invariably in educational research the problem of devising measurement instruments on one hand and the statistical problems of sampling and design cannot each be considered in isolation.

I think it could be said that the problem of measurement is closer to the concerns of the educational researcher than is statistics. Statistics, as an independent discipline, is only a formal set of procedures for analyzing data, while the scientist must take upon himself the principal responsibility of devising his own measurement tools--for only he knows, however vaguely, the concepts with which he is concerned.

In educational research, perhaps the largest methodological difficulties stem from a failure to plan ahead with sufficient care. Once we perceive a problem, we are tempted to blast forward in an ill-conceived fashion to attempt to solve it. Oftentimes, technical problems of measurement and statistics are only vaguely conceived of *a priori*, being dismissed with the thought, "We can cross that bridge when we come to it." It is found only later that data so enthusiastically gathered cannot be analyzed in a systematic fashion. Some people who have been denied this allegation, suggesting that if enough data is gathered with enough enthusiasm, solutions to problems will surely come forth like a bolt from the blue. Perhaps so; but quite likely these solutions will be to the wrong problems. Rather than "cross the bridge when you come to it," a better maxim would be, "Look before you leap." Careful thought on what to measure and how to measure it, considered simultaneously with appropriate methods of statistical analysis is the sound way to do business. Yet also I would never suggest that such planning should inhibit subsequent effort; one should not be stunned into silence because of difficulties in planning. A struggling start is certainly better than no start at all.

Let us think about statistics in a little more detail. As I suggested before, statistics is an independent discipline, having nothing necessarily to do with any science. From this viewpoint, statistical methods are capable only of providing us with decisions about the probability distributions of random variables. It is the responsibility of the scientist, as a scientist and not as a statistician, to consider the relationship of statistical decisions to scientific decisions in education. In playing the role of statistical consultant, there is nothing more distressing to me than having a purported scientist ask me to state his problems; who am I to speak with authority about the problems of the administration of secondary school guidance for curriculum workers in a laboratory school setting?

Scales of Measurement

The relationship of kind or level of scale of measurement of the educator's data to statistical procedures appropriate for working with these data persists as a topic of great controversy. In one school statistics is considered as above: formal discipline which bears no necessary relationship to the real world. This school of thought--to which I must admit for the most part I adhere--asserts that considerations of scales of measurement are irrelevant to statistical procedures. Actually, this is more than an assertion: it is a fact. Statistically, we can do anything we please perfectly "legally" --so long as the formal statistical assumptions are more or less met. But, whether the statistical results have any scientific meaning is an entirely different question, and should be thought of as such. *For example*, we have often heard the statement that a variable must be measured on an interval scale in order to compute a mean; really, this is hogwash. However, for our scientific *interpretation* of a mean, such considerations *may* be of importance. Pertinent here is the distinction between a scientific and a statistical hypothesis. While statistical hypothesis is nothing more than a statement about the probability distribution of a random variable, a scientific hypothesis is a statement about something in the real world. For example, the question, "Are boys smarter than girls?" is not a statistical hypothesis. However, corresponding to this, as scientific hypothesis, there may be a reasonable isomorphism to a statistical hypothesis; in this case, it could be the assertion that the population mean IQ of boys is greater than that of girls, given that IQ both for boys and girls is normally distributed and that each of these distributions has the same variance. This statistical statement obviously leads to a traditional *t*-test. What it appears that we do in practice, then, as scientists using statistics, is first to state a scientific hypothesis, then translate this to a seemingly reasonable statistical hypothesis, formally test this statistical hypothesis, make a statistical decision, and finally make a corresponding equally reasonable or useful scientific decision. In making these reasonable translations, it would appear that in some not-well-defined way measurement considerations are of importance.

The other school of thought on the question on the relationship of scales of measurement to statistical procedures is due to S. S. Stevens, who is responsible for the insightful taxonomy of scales of measurement into nominal, ordinal, interval, and ratio scales. Although after the fact, this taxonomy seems particularly obvious, it was not fully articulated until the late '40s. It may surely be considered one of the major landmarks in the theory of measurement. However, it would seem that Stevens has gone a wee bit too far. This is represented by his pontifications on what you can't do. For example, you can't do a t-test unless the variables are measured on an interval scale. As suggested above, of course you can; what I think Stevens means to say is if you want to make scientific sense out of the results of your t-test it is really sufficient that the variables be measured on an interval scale--and it may be necessary.

Perhaps my disenchantment with these prescriptions stems from the experience of dealing with educationists who have taken Stevens too literally, i.e., I have often been confronted with insecure and terrorized consultees with hollow and haunted eyes, so frightened that they will do something "wrong" that they are inclined catatonically to do nothing. After all, many of us are scientists first and statisticians second. To allow statistics to repress our ideas is an anathema of the worst kind. The tail should never wag the dog.

To the extent that research conclusions hold up, one has pragmatic evidence of the efficacy of his scaling assumptions. For example, the massive weight of evidence would indicate that most intelligence tests are essentially measured on an interval scale. As an educational psychologist I would say with a high degree of confidence, that a difference in IQ of 75 and 85 may be considered the "same" as the difference between 105 and 110. To have established this statement *a priori* is like proving the existence of God. On the other hand, if one were newly to devise a rating scale, say, and handle the data as if it were on an interval scale, the generalities of the scientific translations could well be suspect--although the statistics was perfectly dandy.

Thus it would seem that the critical issue in scaling is the scientific generality of the resulting conclusions. The more general the scaling, the more general the scientific (as opposed to statistical) results will be. A careful consideration of levels of measurement would then seem essential to scientific conclusions, although such cogitations are irrelevant to statistical procedures.

Significance Testing

Let me turn now to some comments on the applications of statistics. In reading the research literature in education, I am impressed --I might say appalled--by the relative frequency with which tests of significance are performed as a matter of thoughtless ritual. Although I have strained for years to understand the meaningfulness of the seemingly ever-popular significance test, I remain convinced that there is little relationship in this ritualistic procedure to the scientific thinking of educational investigators. First, typically one tests a null hypothesis against all possible alternatives. Appealing to subjective probability, such hypotheses are simply absurd. What scientist, on this green earth, would ever state that girls and boys are exactly equally bright? Or that the Dandy-Dan method of teaching arithmetic is exactly as effective as the Johnson-Kleinsohn procedure? Testing such closely specified null hypotheses against omnibus alternatives simply doesn't make sense, for such null hypotheses will be rejected, or not, simply as a function of the sample size and the power of the test used. Even were boys much, much brighter than girls, a sample of size 2 would rarely show significance; or, if boys and girls were essentially equally bright--but not quite--a sample of size 4,000,000 would almost certainly pick up the negligible difference. Significance testing is a myopic way of doing business.

It seems to me that the only time when testing procedures in statistics are valid is for the purpose of final adjudication between two or more equally specific theories, where each can be translated into statistical hypotheses of the same dimensionality in the parameter space. Thus, rather than having "everything-else" alternatives, the scientist should state a particular difference which he, as a scientist, considers to be educationally significant. Once this is done, he should assert precisely what sort of risks he is willing to take for all possible errors. Then, the standard application of statistics (à la Neyman-Pearson) will do the deed. But in areas such as education and psychology--the behavioral sciences generally--studies which are concerned with such final adjudication would seem rare indeed.

It's just that in the vast majority of educational research, theory has not reached a level of sophistication which allows scientists to make precise quantitative predictions for alternative hypotheses. For these studies, a more appropriate statistical procedure would be to estimate the differences of interest or the degree of relationship rather than dichotomously succeeding, or failing to succeed, to see "truth." Thus, the first concern would be a point estimate of the parameter for the problem in question. (I must admit that this major interest would seem implicit in the somewhat irrational defenses of so-called "descriptive" statistics.) After this primarily important point estimate is made, it would seem nice to jazz it up by putting an interval about it and indicate the degree of confidence which we have in the interval. Finally, but least important, we might sneak a peek to see if our interval covers zero. It is most unfortunate that many popular texts emphasize significance testing first --not last...

Nonparametric Methods

At this point, it seems appropriate to comment on the recent rise of nonparametric methods in educational statistics. I consider the stampede to these procedures unfortuante. First, most nonparametric methods emphasize the significance testing viewpoint. Usually, in nonparametric procedures, distributions are computable only under a traditional null hypothesis--have you ever heard of the sampling distribution of the rank-order correlation coefficient for a population rank-order coefficient different from zero?--and thus, to a certain extent, my previous diatribe about the thoughtless use of significance testing applies. A second consideration is related to the question of scales of measurement. For there are many poor souls who are driven into a dark corner by the imprecations of the overly serious scales of measurement boys and are unwilling ever to accept the notion of an interval scale and thus, at best, apply the much less informative nonparametric methods to their safe and sure ordinal data. Third, we often hear the cry that it is so important to meet the formal assumptions in statistical procedures. The question, "Have the assumptions for the *t*-test been met?" is an example of the watchword of these folk. Well, of course, the statistical assumptions have not been met. Nor have the assumptions of the corresponding nonparametric approach been met. For the assumptions in a formal statistical model are abstract assumptions and never can exactly be met in the real world. There is no such thing as normal distribution in Nature. Or, for most nonparametric methods, who ever heard of a continuous distribution existing in Nature? The correct question of "just" meeting assumptions is somewhat more difficult. It is simply a matter of how closely one comes. And to assess how close one must be seems a subjective, almost arbitrary decision. Fortunately, the problem of meeting statistical assumptions--considering that they never can be met exactly--is not really so bad for many traditional procedures using metric data. Empirically it is well known that standard, very useful things are robust, i.e., relatively insensitive to the underlying statistical assumptions, and thus one can blatently go forward with only slight distortion in his probabilistic conclusions.

Multivariate Analysis

In the often encountered situation in the behavioral sciences where there are a number of criterion or dependent variables, there has been built up in recent years a large number of techniques subsumed under the general title of multivariate analysis. Unquestionably the most common multivariate procedure in use today is factor analysis, a technique which, at the exploratory level, has probably done more than anything to bring some sort of preliminary order out of chaos.

I can't resist the opportunity to get in a plug for some recent developments in traditional factor analysis. First, regarding the communality problem, Chester Harris of Wisconsin has recently published some remarkable results linking the important statistical work of Rao with the important psychometric work of Guttman. His paper has clearly demonstrated the crucial notion of *scale-free* solutions--solutions which are metric invariant, so that we are no longer tied to the traditional normalization of observations. With regard to the transformation problem, Harris and I have invented methods which can yield all possible solutions--involving correlated or uncorrelated factors--using orthogonal transformations only. This seems important, for we can now attack the general problem with tractable machinery, for the first time. Finally, a mathematical statistician, Karl-Gustav Jorëskog of Uppsala, has begun to look at the "right" problems in factor analysis (from a psychometrician's viewpoint) and those things that we have been doing with such great gusto for so many years are now being annointed with the propriety of sampling distributions, etc.

But, of course, factor analysis is not the only multivariate technique. Generalizations of the *t*-test and of the analysis of variance have been made to the multivariate case. The ultimate fruitfulness of these approaches is probably yet more or less an unknown quantity. In educational research, undoubtedly the most vigorous activity in the application of the multivariate analysis of variance has been led by Professor Darrel Bock of the University of North Carolina. It will surely be interesting to continue to watch the progress of this provocative area of statistical methodology in education.

THE OFFICE OF EDUCATION COOPERATIVE RESEARCH PROGRAM AND THE STATISTICIAN

Howard F. Hjelm, U.S. Office of Education

The U.S. Office of Education through its Cooperative Research Program provides financial support for investigators in colleges, universities, and State departments of education to conduct research of significance to education. The Program is general in nature in that it is not mission-oriented to any one particular aspect of education such as special education or technical education. It is concerned with all levels of education from preschool through graduate school and including adult education. Support is given to basic and applied research projects having potential significance for education. They can deal with the psychological foundations, the sociological foundations, instructional and curricular foundations, and the administration and organization of education.

Authorization for the Program is in Public Law 531 passed by the 83d Congress in 1954. The first appropriation of one million dollars was granted in fiscal year 1957. The size of the annual appropriations has steadily increased, with the amount for fiscal year 1965 being $15,840,000. With the current emphasis being placed on the role of education in our society, there is every reason to believe that the amounts appropriated for educational research will continue to increase in the coming years.

The U.S. Commissioner of Education employs the use of nongovernmental experts as consultants to assist in the evaluation of proposals submitted to the Program. The primary evaluations of the proposed projects in regard to their scientific merit are made by such consultants. The evaluative criteria fall basically into four categories. They are: (1) significance to education, (2) soundness of the research design, (3) adequacy of the personnel and facilities, and (4) economic efficiency.

Some of the basic and applied research projects, both experimental and descriptive, supported by the Program involve the development of statistical procedures as appropriate ones do not exist or involve the application of existing techniques that have not previously been applied in an educational context. Furthermore, projects whose primary objectives are the development of statistical methodologies can also be funded by the Program. The plans for such projects must be presented in an educational research frame of reference. Such projects are judged of potential significance to education on the basis that they will ultimately have the effect of improving educational research methods. Illustrations of projects supported by the Program are as follows:

1. Tate and Brown (13) at the University of Pennsylvania constructed tables for comparing percentages from small, related samples.

2. Norris and Hjelm (8) at George Peabody College for Teachers conducted an empirical investigation of the effects of non-normality upon sampling distributions of Pearson's r.

3. Solomon (9) at Stanford University received support for attempts to develop new statistical techniques for analysis of items in a test and of tests within a battery.

4. Stake (10) at the University of Nebraska performed a project aimed at generating sampling distributions necessary for discriminating between error and non-error multidimensional scaling factors.

5. Baker (1) at the University of Wisconsin received support for a project to empirically determine sampling distributions of item discrimination indices and the Hoyt Reliability Coefficient.

6. Harshbarger (6) at Virginia Polytechnic Institute received support for a study of statistical models applied to educational criteria.

7. Harris (5) at the University of Wisconsin has received support for a study involving the application of new methods of factor analysis to existing data on intellectual abilities and measures of motor skills and physical fitness.

8. Stanley (12) at the University of Wisconsin conducted a project concerned with the development and analysis of experimental designs for ratings.

A type of project which can be of inestimable value to educational research is one that consists of a critical review of the statistical methodologies which have been employed in research projects in a given educational area. Stanley and Beeman (11) conducted a Cooperative Research project in which they examined the statistical designs that were employed in research projects concerned with the education of the mentally retarded. They found certain serious weaknesses in the designs being used, and these findings together with their recommendations have had a beneficial effect on research in this area. Similar studies could be done in other areas.

The Program has also supported research and development projects dealing with data processing as applied to the field of education. Although such projects typically are situated in local institutional and regional settings and are of direct benefit to the institutions and regions, they must have enough uniqueness, in the sense of adding to our general knowledge in the area of educational data processing, and generalizability that the use of Federal research funds in supporting them can be justified. Some examples of projects supported by the Program are as follows:

1. Lindquist (7) at the State University of Iowa has received support for developing instruments and procedures for collecting and disseminating educational data on students, certified personnel, and school districts.

2. Grossman (4) of the California State Department of Education is receiving support for developing a model educational information system for pupil personnel and curricular services.

3. Carroll and Ellis (2) at Harvard University are investigating the development and use of educational data banks for research purposes.

4. Goodlad and Caffrey (3) at the University of California at Los Angeles received support for a status study of educational data processing.

Proposed research projects such as have been illustrated in this paper are appropriate for consideration in two programs of the Cooperative Research Program. They are the Basic and Applied Research Program and the Small Contract Program. The Basic and Applied Research Program has no dollar or time restrictions. Individual projects in this Program have been funded for less than one thousand dollars and for more than one million dollars. A typical project runs for two years receiving a Federal financial contribution of about $25,000 per year. Deadlines for the submission of proposals to the Basic and Applied Research Program are September 1, December 1, and March 1.

The Small Contract Program entertains proposed projects requesting $7,500 or less in direct costs. There are no deadlines for the submission of proposals to the Small Contract Program.

Instructions for the preparation and submission of proposals to the Basic and Applied Research Program and the Small Contract Program of the Cooperative Research Program may be obtained from the Director, Basic Research Branch, U.S. Office of Education, Washington, D.C. 20202. Inquiries concerning the appropriateness of contemplated projects in areas such as those presented in this paper may also be directed to the Director of the Basic Research Branch.

REFERENCES

1. Baker, Frank B. Empirical Determination of Sampling Distributions of Item Discrimination Indices and the Hoyt Reliability Coefficient, Cooperative Research Project 1299.
2. Carroll, John B., and Ellis, Allan B. Planning and Utilization of a Regional Data Bank for Educational Research Purposes, Cooperative Research Project F-053.
3. Goodlad, John I., and Caffrey, John G. Application of Electronic Data Processing Methods in Education, Cooperative Research Project F-026.
4. Grossman, Alvin. Data Processing for Pupil Personnel and Curricular Services, Cooperative Research Project D-050.
5. Harris, Chester W. New Methods of Statistical Analysis of Tests of Intellect and of Motor Performance, Cooperative Research Project S-094.
6. Harshbarger, Boyd. A Study of Statistical Models for the Evaluation and Interpretation of Educational Criteria, Cooperative Research Project 1132.
7. Lindquist, E. F. Educational Information Project, Cooperative Research Project E-031.
8. Norris, Raymond C., and Hjelm, Howard F. An Empirical Investigation of the Effects of Nonnormality Upon the Sampling Distributions of the Product-moment Correlation Coefficient, Cooperative Research Project 0637.
9. Solomon, Herbert. Item Analysis, Test, Design, and Classification, Cooperative Research Project 1327.
10. Stake, Robert Earl. Sampling Distribution of Error Factors in Multidimensional Scaling, Cooperative Research Project 1253.
11. Stanley, Julian C., and Beeman, Ellen Y. Restricted Generalization, Bias, and Loss of Power That May Result From Matching Groups, Cooperative Research Project 0149.
12. Stanley, Julian C. Development and Analysis of Experimental Designs for Ratings, Cooperative Research Project 0789.
13. Tate, Merle W., and Brown, Sara M. Construction of Tables for Comparing Related-Sample Percentages, Cooperative Research Project S-010.

VII

THE USE OF SOCIAL STATISTICS IN URBAN POLICY RESEARCH

Chairman, Scott Greer, Northwestern University

	Page
Statistics and the Urban Revolution - Albert Mindlin, Government of the District of Columbia.	124
Problems and Prospects of Enumerative Data in Urban Political Analysis - Peter Orleans, University of California, Los Angeles...........	131
Social Areas: A Case Study in the Methodology of Mass Data Analysis - Dennis McElrath, Northwestern University.............................	137

STATISTICS AND THE URBAN REVOLUTION

Albert Mindlin, **Management Office**, Government of the District of Columbia

I

A convenient way to grasp the present state of metropolitan affairs from the viewpoint of the government statistician is to perceive our urban centers and their governments as being in the throes of two major revolutions. The first revolution is the sheer growth in size of urban centers, together with the tremendous increase in a) the Negro population in the central cities and b) the suburban commuting population. The second revolution is that of computer technology, which potentially can fundamentally transform some basic municipal administrative practices. I will discuss each of these revolutions in turn.

Most people by now are fairly familiar in a general way with the vast population shifts that have been going on in our country in recent years -- the movement of most of our people to metropolitan centers, the increasing concentration of Negroes in the central city, and the vast white migration to the suburbs. The actual statistics of these movements may be less known. They are worth calling to your attention. Let us compare the 1950 and 1960 censuses for the 25 largest cities in 1960.

Table 1 -- Median Family Income

central city (cc)		SMSA		urban fringe
1950	1960	1950	1960	1960
$3,526	$5,935	$3,476	$6,545	$7,082

SMSA = Census Bureau's Standard Metropolitan Statistical Area

Median family income between 1950 and 1960 increased substantially in both central city and the entire SMSA. But it increased more in the SMSA than in the central city. Since the SMSA includes the central city, it is clear that the increase in the suburbs was substantially greater than that shown for the SMSA as a whole. This is brought out explicitly in the 1960 comparison between the urban fringe, which is that part of the suburbs to which most people moved, and the central city. The median family income here was nearly 25% higher than in the central city.

Table 2

Total Number of Families, 1960 minus 1950

	cc	suburbs
white	-346,000	2,733,000
non-white	450,000	112,000

Table 2 dramatically shows the racial change in the central cities and the flight to the suburbs of white families. The relatively greater family income of the suburbs reflects the fact that these migrants were mainly middle class white families.

As Table 3, bank 1 shows, short of large-scale migration of non-whites to the suburbs, or massive return of whites to the central cities, the central cities are bound to be increasingly non-white by natural increase alone, as the female fertility ages become increasingly non-white. This shift in child-bearing females is also shown in bank 2. The proportion of childless families remained about the same during the decade for white families in the central city, but decreased substantially for non-whites. In the central cities there has been a substantial increase in the number of children in proportion to the total population (bank 3), and an increasing proportion of these children are non-white. There has also been a relative increase, less dramatic, in the number of aged (bank 4). These are the ages that are the least productive and the most dependent on community services. There has been just as great a proportional increase of children in the suburbs. But the higher incomes of the suburbs make them much better able to pay for the necessary community services than the central cities.

Drastic changes are also occurring in the character of the non-residential sector of many metropolitan areas. In the core city new construction is increasingly in office buildings. There has been a relative decline in retail merchandising in the core city, corresponding to the phenomenal growth of the suburban shopping centers. This has produced a mounting crisis in the classical "downtown" section of the central city.

II

The changes outlined above have created vast new metropolitan problems. Some of these are:

a. The critical commuting problem.

b. The massive problem of urban renewal in slums and deteriorating areas.

Local and Federal government together have

Table 3 -- Median Values, 25 Largest Cities in 1960, and their metro areas

		cc	suburbs
1. Females, ages 15-44, 1960/1950	white	0.79	1.51
	non-white	1.34	1.27
Non-white proportion of total in this age group	1950	0.14	0.05
	1960	0.21	0.04
2. Families without children under 18:			
Proportion of all families			
white	1950	0.53	0.45
	1960	0.51	0.38
non-white	1950	0.56	0.49
	1960	0.44	0.43
Non-white families proportion of all families in this group	1950	0.13	0.04
	1960	0.17	0.03
3. Persons under 18 years of age:			
Non-white proportion of all persons in this age range	1950	0.16	0.05
	1960	0.23	0.04
This age range as proportion of all persons	1950	0.26	0.30
	1960	0.33	0.38
4. Persons 65 years and over, proportion of all persons	1950	0.08	0.07
	1960	0.10	0.07

NOTE: In 16 of the 25 cities there were boundary changes in central city or SMSA or both between 1950 and 1960. Changes in the 9 cities with unchanged boundaries were almost the same as for all 25 cities. Thus geographic changes were not a significant factor in these phenomena. Furthermore the changes for the cities taken altogether, where the bigger cities count more than the smaller cities, are roughly the same as when the cities are treated equally by the use of medians. Thus experiences of the larger cities are similar regardless of size differences.

been wrestling with these problems. There are vast highway and urban renewal programs, jointly financed. Both local and Federal interest in interurban mass rapid transit systems have greatly increased.

But efforts to solve these problems have created new problems. a. The highway program and the rapid transit program compete with each other for attention, priority, and for rights-of-way. This competition has created one of the liveliest, at times bitter, controversies in contemporary urban politics. At stake are competing philosophies about how people should live, competing economic interests, competing views on the tax base of the community, and competing arguments on how to keep the city alive. b. Large urban renewal and highway construction programs have created a sizeable problem of displacees. So the city is faced with a big new headache -- relocation.

c. With an increasing proportion of the core city's population in the young and old dependent age groups, community services in health, education, welfare and recreation are inevitably increasing, with an attendant spiraling of budget and taxes. These increasing taxes rest on a proportionally diminishing base of income-earning producers.

d. This proportionally diminishing base of income earning producers in the central cities is proportionally increasing non-white and poor. Clearly an improvement in the quality of community life depends on an accelerated improvement in the productivity and income of these people. There is now going on a dramatic shift in emphasis in the civil rights movement, a joining hands of that movement with the war on poverty. Only about one fourth of the nation's families with income under $3,000 in 1960 were non-white, but in the large cities poverty has become increasingly a non-white problem. Here the war on poverty and the civil rights movement are merging into a single new frontier of social reform.

e. Many urban problems are metropolitan-wide in character. They cannot be solved without close cooperation of the many independent jurisdictions that compose the modern metropolis. Transportation is only the most visible problem. Another is water -- supply, siltation, pollution, and sewage. A third is the preservation of park land and open spaces. A fourth is the racial distribution of the metropolis. The suburbs so far have largely resisted the pressure of out-migrating middle class Negroes. This resistance cannot be maintained much longer without developing serious friction between core city and suburbs.

III

What is the role of the government statistician in this urban revolution? This paper will not discuss the statistician's role in internal management improvement. The panoply of statistical technics such as quality control, work sampling, operations research, and other statistical tools are every bit as applicable to local and regional management as they are to Federal Government and industry management; and indeed it is by such internal management applications that a statistician can often play his most prominent role. But the focus of this paper is on research design and data needs of community programs -- local, regional and Federal.

There is great overlap in the data needs of urban social and economic programs.

a. A great deal of demographic data are needed -- population by age, sex and color; family composition; education; housing data such as condition and overcrowding; place of work related to place of residence.

b. Economic data are also needed -- personal and family income; housing costs, transportation costs, and medical costs.

c. The geographic distribution of land use is needed -- where different kinds of households live, where different kinds of non-residential uses are located; and how these are moving.

d. Most of these data are needed for small areas -- blocks, census tracts, neighborhoods.

e. Most programs require projections into the future of some sort, so that current and continuing data are needed.

f. For effective regional cooperation, the data are needed with comparable definitions and reconcilable, if not standardized, classification systems.

It is the responsibility of the urban statistician 1)to locate as much of this information as possible; 2)to generate as much of it as possible; and c)to design and carry out statistical research that will provide valid underpinning of community programs -- and equally important altho less recognized, to test proposals and programs for usefulness and validity, to the extent that they are statistically testable.

IV

It is at this point that we turn to the second major revolution that is overtaking the local governments -- computer technology. There is no question that the new technology can revolutionize municipal administration -- and is actually on the threshhold of doing so in many communities. To the statistician this revolution is partly real, partly mythical -- and it is important that the myths be separated from the reality, for there is much wishful thinking on this subject.

When the statistician seeks data to solve community problems, if he does not generate original data where can he turn? Aside from the periodic national censuses, the most obvious information sources are records of operating agencies, both local and national. It is when we examine closely these records in relation to our needs that we perceive the real and the mythical aspects of the computer revolution.

A. The real revolution. An overwhelming barrier to effective statistical use of local records is that the bulk of them are not mechanized. Operating agencies as a general rule mechanize only the items which are critical for the regular operations of their program. Only rarely can such mechanized data fit, even reasonably well, the needs of other agencies and projects for which they were not intended, designed, or gathered.

Much extremely valuable information which is obtained almost incidentally as part of an operating program is left unmechanized. For example, many jurisdictions gather land-use and structure-use information as part of their assessment program -- land-use classification, number of living quarters in the structure, housing condition and other extremely valuable items -- but do not mechanize it. At times, needed information is laboriously copied from these records by hand; at other times, original field surveys are made to obtain information already present in inaccessible assessment records; most often, action programs that would benefit greatly by this information simply do the best they can without it.

The new technology has opened up new horizons for the statistical utilization of administrative records. To continue the example of the last paragraph, for the first time mechanized assessment files can become widely accessible by being taped. A deck of hundreds of thousands of punch cards is not a viable tool to be used by many different agencies or to be statistically manipulated --but three or four reels of tape are. On the computer we can apply statistical methods with mass data; and on tape it becomes feasible to add to the mechanical record many items not previously mechanized. No operating agency will automate primarily to satisfy statistical research or the needs of community programs. It will automate only when that is shown to be a more effective method of conducting operations. But once automated, it is far easier and cheaper than ever before to add in the extra information so important to other community programs. Furthermore it becomes feasible to bring together in a single record information about the same property or person from different agencies. The vision of a "data bank", so dear to the hearts of planners and some statisticians, has become a real possibility.

B. The mythical revolution. The possibilities of the new technology are so exciting and the pressures of salesmanship and "machine politics" so great, that a number of myths have arisen. These must be dissipated.

One myth is that there is great duplication in local government records that will be eliminated by an integrated electronic data processing system. In general, neither part of this assertion is true. There is not great duplication in records, and what duplication there is will not for the most part be eliminated by automated processing. For the most part municipal agencies do not duplicate substantive information. The Fire Dep't., Health Dep't., Assessment Office, License Office, Housing Code Division, and Highway Dep't., do not maintain on their records the same substantive information about a piece of property. They operate different programs and maintain different information. They do maintain the same non-substantive information for identification -- block, lot, census tract, etc. But this duplicated information would have to be maintained even under an automated system in order to query or enter the central system for a unique record.

Another myth much more important to the statistician is that the bulk of our statistical problems can be solved with records now collected by operating agencies. A third myth is that the main methodological problem is simply how to integrate these records in a mechanical system and make them statistically available. Both of these are indeed serious myths that have created much delusion among administrators and non-statistical program planners and researchers. In fact, even if most presently gathered administrative information were mechanized and accessible, this would not, by itself, adequately satisfy many, including some of the most critical, statistical needs. Almost all information gathered by local agencies, either public or private, is designed and tailored to serve primarily operating programs. In spite of the seemingly large amount of information gathered, some of the largest and most pressing needs of non-operating agencies, such as urban renewal offices, planning commissions, and mass transit planning agencies, cannot be met, or can be met only minimally, very inefficiently, and unsatisfactorily, by data regularly gathered by operating agencies. The same statement holds for problems constantly coming before the operating agencies themselves.

Two examples will suffice. 1)Much useful information about housing condition, overcrowding, etc. is often gathered by a municipal housing agency. But aside from licensed premises, such as apartments, and specific programs in designated urban renewal areas, this information is usually generated only by complaints. Hence it is fitful, irregular and spotty, and therefore largely unusable for probability sampling or for estimates. 2)The small area identification of place-of-work in relation to residence, so important to transportation, city planning, and civil defense, simply does not exist because it is not part of any operating program.

A much more central problem than making accessible current records, is the absence of professional statistical competence. Most municipal and regional governmental agencies do not have the statistical know-how to either generate or analyze statistical data in a sophisticated or efficient way. There is a drastic paucity of well-qualified professional statisticians in most metropolitan areas. For example, many statistical needs that cannot be filled by existing records can be filled by properly designed sample surveys. But the professional talent to design sophisticated sample surveys is often lacking. All of the automating in the world will not produce non-existent input data and non-existent statistical talent.

V

In light of all of the above, what are some genuinely realizable elements of a modern local statistical program? The following proposals are advanced, in no special order of importance.

1. There should be created a professional demographic unit to serve the entire governmental apparatus. This unit should make periodic population estimates by age, sex, color and various other groupings. These estimates should be for areas within the city as small as can be professionally supported. This unit should be responsive to requests for demographic assistance from any governmental agency, e.g. to put together all available census material for a small

area and to relate such material meaningfully to local material such as vital statistics and welfare data. This unit must have professional demographers and be centrally located, not dominated by any single Departmental program.

2. There should be developed a regionwide set of comparable subject-matter classification systems -- comparable both for agencies within a jurisdiction and between jurisdictions. For example, three areas where lack of comparable data has hampered not only regional programs but even programs in a single jurisdiction,are land use, juvenile delinquency, and housing condition.

3. There should be developed a master sample survey design of properties that can be used repeatedly and that can form the basis for longitudinal studies. The full design should allow reliable neighborhood estimates.

4. There should be maintained a comprehensive statistical library. This may well be a responsibility of the demographic unit, which would probably use it more than anyone else. This library would maintain the local area output of the Census Bureau, Bureau of Labor Statistics, Internal Revenue Service, and other Federal agencies; would also maintain all other obtainable tabulated data from local public and private sources of more than purely internal workload interest; and would constitute a responsive reference service that would tell an inquirer readily what organization has what data, how complete they are, how current, how available, and how much they cost.

5. There should be developed a comprehensive locational directory. This directory would cross-classify the most important ways that a given location is identified -- for example, address, lot, city block, census block, census tract, census enumeration district, neighborhood, and police precinct. More generally, a feasible master coding scheme might be developed that would identify any property to a very small geographic building block, usable to approximate most geographic classifications. The master code need not be introduced into operating records, but could exist as a tape or deck providing a master switchboard between one classification and another. Thus all available data for an area -- local, state and national -- could be brought together. Indeed research is now going on in several places to locate all properties on a fixed coordinate system that would be independent of any area changes. Property coordinates can be entered into a mechanical data plotter which mechanically constructs data maps.

6. Leadership should be exerted in mechanizing or automating those governmental records that are the most valuable for community research. In carrying forward this task, it is well to perceive clearly two different kinds of records -- property records and social records. It is much easier to automate the first kind than the second, hence most progress has been made in the first area.

a. Property records. Almost every community has one operating file that is geographically universal -- its property assessment file. In most large communities this file is already mechanized to the extent necessary to calculate and distribute property tax bills. However, as stated earlier, the non-mechanized records usually contain a statistical cornucopia of information. The most important of these items should also be mechanized. Such mechanization would serve two fundamental purposes: 1)It would provide critically needed statistical information, and 2) probably of much greater long-range value, it would provide various bases for stratifying and classifying properties in order to draw samples for field surveys. An appropriately mechanized assessment record file would be of inestimable value as a sampling frame for small area samples, or samples of various types of property. Indeed such a mechanized file could virtually eliminate the more cumbersome classical method of area sampling in residential surveys. Furthermore tape reels of this file, which is usually all public information, distributed to a Planning Commission, Urban Renewal Office, Highway Dep't., Mass Transit Agency, and even public utilities and real estate institutes, Board of Trade, would enormously improve community research.

b. Social records. These are records about people and families from the Welfare Dep't., Health Dep't., Police Dep't., Juvenile Court, Unemployment Compensation Board, School Dep't., etc. This is a more difficult problem, but one that is receiving increasing attention. The creation of an integrated file of such records would stimulate effective and intelligent social planning to an extent that cannot be exaggerated. Here the computer technology opens possibilities previously closed. In particular it offers solutions to two of the stickiest problems -- 1) the necessity to develop identification search rules that will assure the proper match of records about the same person or family; 2) the preservation of confidentiality of information, which is entirely feasible on a computer.

7. Probably the greatest single step forward that could be accomplished for urban statistics is the establishment and adequate staffing of a professional statistical unit available to the entire local and regional governmental complex for consultation, design, analysis, and other statistical research.

VI

The Federal Government could exert great leadership in helping solve the statistical problems produced by the urban revolution. Federal assistance to improve local statistical programs will directly and greatly benefit the Federal Government itself, because many of the new programs are largely Federally financed. Indeed, in 1960 - 1961 the Federal Government spent the astounding amount of nearly $7 billion on grants-in-aid programs administered by State and local governments. The current anti-poverty program will add millions more to this. Below is a too-brief, highly incomplete, and superficial discussion of some things the Federal Government could do.

1. A mid-decade population and housing census: It is probably too late to mount a mid-decade census now, but this issue must be kept alive. The national census is by all odds the greatest single source of local information. Our metropolitan growth is simply too dynamic to have to rely on this most fundamental of all statistical tools only once every ten years. There have been hundreds of pages of Congressional testimony on this subject, indeed probably more than on any other Federal statistical subject. There is nothing unprecedented about a shorter census period. Canada, Japan, New Zealand, and Turkey have five-year censuses, and the United Kingdom is about to have its first. The opposition of the Bureau of the Budget must be overcome. The last half of an intercensal decade is the dark ages for urban research. This is the most important single statistical step the Federal Government could take to advance local and regional community planning.

2. Demographic improvements in the census: Experience and new needs have indicated several areas of improvement in the population and housing census data, e.g.:

a. Housing condition: This is one of the most important and least reliable census items. Efforts to improve it must continually be made. In addition to improved condition reporting itself, the information would be more useful if it applied to a structure rather than to a housing unit.

b. Value of structure: Ways should be found to extend this beyond the present highly limited owner-occupied-one-housing-unit detached structure.

c. Block statistics bulletin:This should be greatly expanded.

d. Census blocks: These should be made coterminal with city blocks in order better to integrate census and local data.

If the Census Bureau abandons the enumeration district, as it currently contemplates doing, points c and d become both critical and feasible.

e. There should be a much more refined work-residence item -- if not on a census tract level, at least on some gross neighborhood level.

f. There should be much finer cross-tabulations by color. There is now going on a most interesting reversal of thinking on this subject. For many years pressure mounted to remove color from all kinds of records, including census, because color was not only offensive but was often used for racial discrimination. This campaign was successful in many areas. This trend is now being reversed. The civil rights movement has become so pervasive that it now perceives that statistical information by color is one of its strongest weapons for carrying forward its objectives -- something that most statisticians have long known. For example, it is impossible to make adequate studies of employment discrimination in local areas without detailed color cross-tabulations by occupation, industry, wages, educational attainment, age, and class of worker -- all of which are unobtainable in present census publications.

3. There is a cornucopia of important local information locked in the records of some Federal operating agencies. These agencies must recognize responsibilities beyond their immediate operating programs, and make their data more easily available for community research. Two of the most important agencies to which this applies are the Bureau of Old Age and Survivors Insurance and the Internal Revenue Service. One feasible way to do this is to mechanize one or two more locational codes for ready access to tabulated data on a small area basis. With respect to the two mentioned agencies, another way is a vigorous and imaginative effort to integrate their records. A third is to develop regular series of local statistics from agency data, as the IRS has done in its SMSA "Statistics of Income on Individuals" -- preferably for central city and suburbs separately, since these are often separate political jurisdictions.

4. The Dep't. of Labor, which publishes extremely useful employment and earnings information for SMSA's, should if possible develop separate estimation programs for central cities.

5. The Federal Government could greatly stimulate urban statistical research by an expanded program of planning, demonstration and research grant funds. The variety of possible relevant and important statistical research projects is too numerous to list. Two examples of demonstration projects currently being funded by the Urban Renewal Administration are the study of automation of local records for use by planning commissions centered in Tulsa, Oklahoma, and the effort to develop a reconcilable and operable set of land use and housing classification systems over an entire metropolitan area centered in Washington, D.C.

6. Finally, the Federal Government should explore ways of bringing to local

communities professional statistical assistance which they sorely need. A string of Federally assisted metropolitan statistical centers (not data centers) across the country should go a long way toward channeling our sprawling urban revolution into reasonably ordered and controlled progress.

PROBLEMS AND PROSPECTS OF ENUMERATIVE DATA IN URBAN POLITICAL ANALYSIS

Peter Orleans, University of California, Los Angeles

Introduction

Where electoral behavior has been the focus of attention researchers have favored the case study approach (Lazarsfeld, et. al, 1944) or the use of survey methods (Campbell, et. al, 1960) and either by oversight or by design have tended to ignore the possibilities of comparative analyses of the interunit or intraunit type. The major exceptions are to be found in the early attempts of ecologically oriented political scientists to relate areal characteristics to individual behavior considered in the aggregate (Litchfield, 1941; Gosnell, 1936). The fallacy of the latter type of study, as W. S. Robinson has pointed out, lay in the use of the ecological correlation to make inferences about individual behavior on the basis of aggregate data (Robinson, 1951). More recently Schnore has aptly dubbed such efforts as psychological sociology (Schnore, 1961).

Now, it is increasingly possible for sociologists concerned not with individual aspects of the collective life, but rather with the analysis of structural constraints upon social organization, to revise earlier interpretations and to further develop already available techniques. Administrative units, school districts, police precincts, census tracts, political wards, land-use zones and the like, as sources of enumerative statistics constitute both a help and a hindrance to the researcher involved in such an effort. They are a help inasmuch as they provide the researcher with raw data which ordinarily cannot be obtained directly on an individual basis. Anonymity is preserved in the aggregate. And, administrative units circumscribe spatially based aggregates. They present difficulties, however, in that raw material available on an administrative unit basis carries with it the liabilities of data collected at other times, in other places, by other people, for other purposes. Neither time nor space remain constant. Thus, the boundaries of various kinds of administrative units, which rarely coincide to begin with, are often altered, and the constituents of administrative units may change.

Nevertheless, the analyst of urban political organization, cognizant of the availability of enumerative statistics collected on an administrative unit basis, has at his disposal a tantalizing array and an enormous quantity of raw material. And, what is more important, the collection of data on such a subunit basis allows for the analysis of both interunit, that is between city, as well as intraunit, or within city, variation.

In an attempt to use this material the researcher is confronted, by the nature of the data, with a series of procedural problems which have both pragmatic consequences and theoretical implications. It is the purpose of this paper to consider some of the problems involved in the use of enumerative statistics, and to note some of the assumptions which underly decisions made by the analyst in resolving such problems. This will be done by examining three specific problems encountered by the author in an analysis of the voting of selected subarea populations in two cities in the United States in elections occurring over a four year period (Orleans,1964).

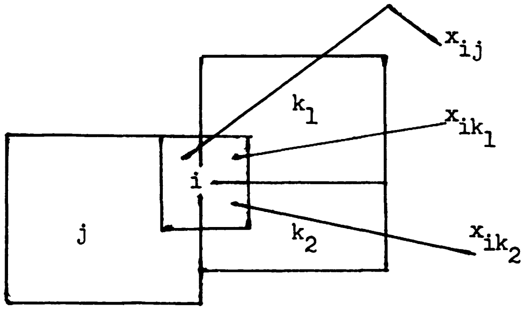
Although the present discussion is limited to problems which occurred in a particular piece of research, the problems encountered are of a general nature. Thus, for instance, a method of estimating election returns for census tract populations will be considered as a specific example of the more general problem of fitting any two non-coterminous administrative units for the purpose of collating diverse sets of data. Geographic mobility and incremental growth of subarea populations as possible sources of error in the measurement of proportional participation will be examined as one example of the more general problem of anticipating and accounting for possible bias in measurement due to changes in the composition and/or constituents of administrative units. And, finally, the general problem of how to standardize measures employed in subarea analyses will be considered by examining the assumptions involved in standardizing social area indices developed by Shevky (Shevky and Bell, 1955; Bell, 1959). In short, in each case, a specific problem is employed to illustrate a more general problem.

Fitting Non-Coterminous Administrative Units

As has already been suggested one of the principal uses of the various administrative units is in the collection and tabulation of enumerative statistics. It is often the case that the researcher wishes to collate data which have been organized in terms of two or more different kinds of administrative units. When the administrative units of concern are not spatially coterminous the researcher must employ some sort of fitting or estimation procedure if he is to be able to relate independent variable data available for one type of administrative unit to dependent variable data available for another. Such a procedure allows the researcher to estimate how much of the information enumerated on the basis of one type of administrative unit is to be allocated to another. The more accurate the procedure, the more reliable the results of the analysis.

Analyses of voting restricted to the use of enumerative statistics provide an appropriate example for a consideration of this problem. Data collected by the Bureau of the Census, and tabulated by census tract, provide the researcher with a wide array of independent variable (or what sociologists often refer to as facesheet) data. Attempts to relate data descriptive of census tract populations to voting data collected and tabulated on an election precinct basis involves fitting these two types of administrative units together. Election precincts and census tracts rarely coincide. Election precincts are smaller (both in terms of geographic space and of population size) than are census tracts, and therefore it is the election precinct which must be fitted to the census tract. In other words, it is the vote summarized by election precincts which the researcher must allocate to census tracts across which election precincts fall.

When it is available, total population (though not voting age population) data enumerated on a census block basis can be used in a direct assessment of the distribution of the population in census tracts, election precincts, and areas of overlap between the two. This is because census blocks are smaller than either census tracts or election precincts, and because they do not cross the boundaries of either. Thus, an increment in accuracy is obtained by using census block data in fitting election precincts to census tracts. By doing so one does not have to make the dubious assumption that the total population of a census tract is evenly distributed in geographic space. The exact distribution of the total population can be determined on the basis of census block data. Once this distribution is determined the process of estimation involved in the fitting procedure becomes a matter of extrapolation. By making the assumption that the distribution of persons of voting age follows the known distribution of the total population, it is possible to estimate, on a nonrandom basis, the number of persons of voting age located in that portion of a given election precinct which overlaps the census tract of concern in the analysis. This information can be used, in the manner indicated below, to develop a weight for use in allocating election precinct data to census tract data.



i = precinct to be allocated
j = tract into which a portion of precinct i is to be allocated
k = other tracts containing a portion of precinct i (k may take values of 1, 2, 3, ... n)

x_{ij} = number of persons in precinct i and tract j

x_{ik_n} = number of persons in precinct i and tract k_n

$Y_{.j}$ = number of persons 21 years of age or older in tract j

$Y'_{.j}$ = total number of persons in tract j

$y_{.j}$ = proportion of tract j population that is 21 years of age or older; i.e., $Y_{.j}$ divided by $Y'_{.j} = y_{.j}$

$Y_{.k_n}$ = number of persons 21 years of age or older in tract k_n

$Y'_{.k_n}$ = total number of persons in tract k_n

$y_{.k_n}$ = proportion of tract k_n population that is 21 years of age or older; i.e., $Y_{.k_n}$ divided by $Y'_{.k_n}$

$t_1 = x_{ij}(y_{.j})$ the number of persons in precinct i and tract j multiplied times the proportion of the population in tract j that is 21 years of age or older; i.e., the number of persons in precinct i and tract j who are 21 years of age or older

$t_2 = \sum_{k=1}^{n} \left[(x_{ik_n})(y_{.k_n}) \right]$ the sum (overall k_n tracts) of the number of persons in precinct i and each k_n tract multiplied times the proportion of the population in each k_n tract; i.e., the number of persons in precinct i and <u>not</u> in tract j who are 21 years of age or older

$T = t_1 + t_2$ number of persons in precinct i who are 21 years of age or older

$M = t_1 / T$ the number of persons in precinct i and tract j who are 21 years of age or older divided by the total number of persons in precinct i (i.e., those in both tract j and the k_n tracts) who are 21 years of age or older; i.e., the proportion of persons in precinct i who are 21 years of age or older who are also in tract j <u>or</u> the weight used to allocate precinct i votes to tract j

x_{ij} = 511 total population in precinct i and tract j <u>determined</u> on the basis of block data

x_{ik_1} = 383 total population in precinct i and tract k_1 <u>determined</u> on the basis of block data

x_{ik_2} = 420 total population in precinct i and tract k_2 determined on the basis of block data

$Y_{.j}$ = 2257 voting age population in tract j determined on the basis of tract data

$Y'_{.j}$ = 2942 total population in tract j determined on the basis of tract data

$y_{.j}$ = .767 proportion of tract j population which is of voting age determined on the basis of computation

$Y_{.k_1}$ = 2068 voting age population in tract k_1 determined on the basis of tract data

$Y'_{.k_1}$ = 3075 total population in tract k_1 determined on the basis of tract data

$y_{.k_1}$ = .673 proportion of tract k_1 population which is of voting age determined on the basis of computation

$Y_{.k_2}$ = 743 voting age population in tract k_2 determined on the basis of tract data

$Y'_{.k_2}$ = 1124 total population in tract k_2 determined on the basis of tract data

$y_{.k_2}$ = .661 proportion of tract k_2 population which is of voting age determined on the basis of computation

t_1 = 511 (.767) = 392 estimated number of persons of voting age in precinct i and tract j

t_2 = 383 (.673) + 420 (.661) - 535 estimated number of persons of voting age in precinct i outside of tract j

T = 392 + 535 = 927 estimated number of persons of voting age in precinct i

M = 392/927 = .423 proportion of estimated voting age population in precinct i which is also estimated to be in tract j or the weight used to allocate precinct i votes to tract j; e.g., 423 of each 1000 precinct i votes are allocated to tract j

Shifts in the Composition and Constituents of Administrative Units

A second problem often encountered by researchers who use enumerative data is that of accounting for possible bias due to changes, through time, in the composition or the constituents of administrative units. Again limitations inherent in enumerative statistics employed in analyses of voting make an apt example for our consideration.

Longitudinal analyses of voting usually involve the collection of election returns at a date considerably later than the time when each election examined was held. As a result it is often the case that voter registration information is not available. This is a consequence of the widespread practice of Boards of Election of up-dating registration rolls by continually revising one list of registered voters instead of periodically developing new lists while retaining and filing old ones. Boards of Election are established to facilitate the electoral process, not research into that process and the updating procedure is most efficient for that purpose. In practice the updating method adopted by most Boards of Election means that longitudinal analyses of voting are restricted to measures of turnout indicated by ratios of persons voting to persons of voting age instead of ratios of persons voting to persons eligible to vote; eligibility simply cannot be determined.

Because the base N involved in such a measure refers to the voting age population and not to the registered population it is especially susceptible to error resulting from the inability to account for disfranchisement due to geographic mobility and the failure to meet residence requirements, as well as error due to shifts in the size of the voting age population prior to and after the date of enumeration.

The first of these sources of error, bias due to disfranchisement resulting from geographic mobility, pertains to changes in the composition (though not necessarily the constituents) of census tracts. The second source of error, shifts in the size of voting age population, is a product of a possible shift in the number of constituents. The first is a qualitative matter, the second a quantitative matter.

It is possible to use a stability coefficient to compensate for error due to differential disfranchisement. Data, available on a census tract basis, indicate (a) the total number of persons 5 years old and older as of the date of enumeration, (b) the number of persons 5 years old and older whose residence is the same at the time of enumeration as it was 5 years prior to that date, and (c) the number of persons 5 years old and older who moved to the enumerated residence from another residence in the central city of the same SMSA within the 5 year period prior to the date of enumeration. This information is sufficient to permit the construction of a gross measure of stability based on a five year interval, referring to the 5 years old and older population, and accounting for movement across SMSA boundaries. By making the assumption that movement into or out of each census tract is constant (i.e., is equivalent in each of the 5 years) a stability coefficient can be constructed, for each census tract, which would express an estimated ratio of persons with a residential tenure of one year or more to the total number of residents in the census tract. Such a ratio may be expressed as follows:

$$\frac{B + C + \{.80 [A - (B + C)]\}}{A}$$

A = total persons 5 years old and older as of 1960

B = number of persons who lived in the same house in 1960 as in 1955

C = number of persons who moved to 1960 residence since 1955 and who lived in different house in the central city of the same SMSA.

Such a coefficient gives an estimate of the proportion of the population aged 5 years or over residing at the same address for one year or more. Assuming that the geographic mobility of the voting age population is roughly equivalent to that of the 5 years old and older population, the coefficient can be used to estimate the eligibility of the voting age population in terms of meeting a one year residence requirement. (On the basis of the one year criterion it was estimated that the ineligible voting age population in selected census tracts in the two cities involved in our analysis ranged from 2% to 11%).

To cope with the second limitation mentioned above, that of shifts in the size of the base population, estimates of growth and decline in the size of census tract populations would have to be made. The problem here is entirely separate from the one involved in the assessment of bias due to disfranchisement resulting from geographic mobility. It is possible to have a complete turnover in the population of a census tract without necessarily having any increment or decrement in the number of persons involved. Thus, the effect of shifts in the size of census tract populations would have to be determined by reference to previous censuses or by extrapolations into the future on the basis of past trends and current data. However, estimates over the decennial census period, for the relatively small number of persons located in each census tract, can be expected to be subject to gross distortion. Therefore, it would appear that all one can do is <u>select elections which are temporally proximate to the date of the census</u> enumeration, note the probable bias due to this source of error, temper the interpretation of his findings, and caution the reader about his results.

Standardization of Measures in Comparative Analyses

In any analysis involving the comparison of two or more sets of data a question is always raised as to the comparability and the generality of the measures employed. The question is inevitable inasmuch as the usual purpose of comparative research is to determine the pervasiveness of the phenomenon under consideration. To do this the researcher must be able to attribute his results, not to the procedures employed, but to the observations involved. In this sense the problem of selecting a method for standardizing measures in comparative analyses is a general question (Jaffe, 1951).

The existence of discrepancies in the assessment of positions which occur in comparative analyses of social structures first became apparent in research dealing with stratification. This occurred when researchers began asking, for example, whether the middle class in middle sized cities could be considered to be equivalent to the middle class of the metropolis. That the sociologist's concern with conceptual consequences of standardization should have occurred in the realm of stratification research is not surprising inasmuch as class has acquired the rather exalted status of a major independent variable in sociological analysis. As one observer aptly puts it, "a sociologist worth his salt, if given two basic indices of class such as income and occupation, can make a long list of predictions about the individual in question even if no further information is given." (Berger, 1963).

The problem of standardization in stratification research came to the fore because of the inability of sociologists to obtain consistent results, across communities, using reputational techniques to assess positional status. Thus, the problem came to be defined as one having to do with the instability of subjective assessments. As a result, a common resolution of the problem has been increased reliance upon objective, rather than subjective, measures of stratification. However, we would contend that even when one works exclusively with objective measures, such as occupation, education, and income, instead of subjective assessments of prestige (honor), the possibility, and therefore the problem, of discrepancies between positions in different settings (structures) remains.

Regardless of the criterion which is used as a basis for differentiating various positions, the location of positions <u>vis a vis</u> one another within a particular setting may depart from the relative location of positions across settings. But, the question with which one is ultimately concerned in the comparative analysis of social structures is whether differences in the relative location of positions across settings makes for differences in the organization of opportunities for various types of behavior. For example, where interest centers upon the relationship between religious affiliation and political behavior, one might be concerned with whether Catholics in a predominantly non-Catholic area are as likely to vote, or to vote Democratic, as Catholics in a predominantly Catholic area. If this were of concern, the distribution of Catholics would have to be accounted for in standardizing measures of the incidence of Catholics in various areas. Similar considerations would obtain whether the independent variable is religious affiliation, level of education, children per household, per capita income, or whatever. In other words, regardless of the form of differentiation that is the basis of classification with which one is concerned, the location of positions <u>vis a vis</u> one another within a given setting would seem to be a critical matter--one to be considered if a structural sociology is to be established within a comparative framework.

To illustrate the problem of standardization I will refer to a specific paradigm; social area analysis. (Shevky and Bell, 1955) I prefer to consider the problem in these terms because this particular paradigm, constructed for the purpose, already has a history of application in comparative urban research involving cities in Europe, Africa, and Latin America, as well as in the United States.

Social area analysis involves the use of indices, for example an index of social rank, which consists of the unweighted average of a number of component ratio scores; education and occupation in the case of the social rank index. In their monograph outlining computational procedures Shevky and Bell suggest that the component ratio scores should be standardized to ranges of scores empirically extant in Los Angeles in 1940 at the time of the first application of this technique of analysis. Thus, for example, the education ratio, a measure of the number of persons per 1000 (in each census tract) having elementary schooling or less, ranged from a value of 130 to one of 900, and this yields a conversion factor of .129. The standardized education ratio score for each census tract population is obtained by multiplying the difference between the score for that census tract and the lowest score, in this case 130, by the conversion factor. General acceptance of the conversion factors developed by Shevky and Bell would permit comparison of two or more census tracts, regardless of where they are located or when they are observed. The index scores can be directly related to one another because the component ratio scores involved can be referred to a common standard--the ranges of component ratio scores extant in Los Angeles in 1940.

Dissatisfaction with the standardization procedure adopted by Shevky and Bell stems from the fact that it arbitrarily anchors all subsequent analyses to the Los Angeles SMSA as defined by the Bureau of the Census in 1940, and from the conviction that there is no reason to assume that it is either theoretically representative or empirically inclusive of other urban areas.

There are several alternatives. The most common departure, in recent research, from the procedure set forth by Shevky and Bell has been the standardization of component ratio scores for a set of administrative units in a given city to their own range as of the date when the data were collected. This procedure frees each analysis from the arbitrary base established by Shevky and Bell and it preserves the autonomy of each urban area considered, but it impedes interunit, between city, comparison which cannot be accomplished without statistical interpolation. Accordingly, this standardization procedure has been adopted most often in research where concern has been with intraunit analysis, the comparison of administrative units within a given city.

A logical extension of this alternative, where interunit or between city comparisons are of concern, would be the standardization of component ratio scores to a range determined by the highest and lowest scores found among all of the administrative units in all of the cities involved in a particular analysis. In this way comparability of the numerical scores across cities would be achieved within a given analysis, but this would be done at the expense of preserving the autonomy of each urban area and therefore the importance of the interrelation of the scores within a given city. Moreover, such a standardization procedure still would not assure the comparability of component ratio scores obtained in a given analysis, with those obtained at other times, in other places, by other researchers.

A final procedure which may be considered is one in which component ratio scores are standardized to the most inclusive arbitrarily defined range of possible scores. Such a procedure would differ only in detail from the one originally proposed by Shevky and Bell. It would substitute an arbitrary, but more inclusive, range of scores for the range of scores obtained in Los Angeles in 1940. In this instance, the previously mentioned education ratio would range from 0 to 1000 and the conversion factor would be .100. A general standard will have been established permitting both interunit and intraunit comparisons, but the autonomy of specific urban areas will have been sacrificed. The numerical value of a component ratio score will always have the same meaning in absolute terms, but the relation of a given score to others in any particular configuration would remain problematic.

It has not been my intention to propose that one of the aforementioned alternatives be adopted. Rather I only want to indicate that different standardization procedures place different conceptual limitations on the data, and that such limitations and their implications must be recognized and evaluated in the process of selecting an appropriate procedure. Only future empirical research can determine how various kinds of dependent variable data are affected by the various procedures discussed above. And, eventually, when this has been done, one procedure will have to take precedence over the others for only in that way can a generic methodology for comparative urban analysis be achieved.

Conclusions

In this paper I have attempted a preliminary discussion of some of the unexpected pragmatic consequences and unrecognized theoretical implications which often result from procedural decisions in urban political research. It was suggested that the researcher should be especially cognizant of three types of problems, problems associated with the fitting of non-coterminous administrative units, problems resulting from shifts in the composition and constituents of administrative units, and problems involved in the standardization of administrative unit measures employed in comparative analyses. All of these problems require procedural decisions which must be made by the researcher who uses enumerative statistics in comparative analyses of aggregate phenomena, distributed through time and over space. As electronic data processing technology and data collection methodology become more refined the prospects for using enumerative statistics will be enhanced further, thus permitting the development of a structural sociology within a comparative framework. But, more problems of the type discussed here can be expected as well. We would suggest that such problems increasingly will make appropriate grist for the mill of statisticians interested in social science research.

References

Bell, Wendell, "Social Areas: Typology of Neighborhoods," in Marvin Sussman (ed.), Community Structure and Analysis, New York: Thomas Y. Crowell Company, 1959; pp. 61-92.

Berger, Peter, Invitation to Sociology: A Humanistic Perspective, Garden City: Doubleay Anchor Books, 1963; p. 80.

Campbell, Angus, et. al., The American Voter, New York: John Wiley and Sons, Inc., 1960.

Gosnell, Harold, Getting Out the Vote, Chicago: University of Chicago Press, 1936.

Jaffe, A. J., Handbook of Statistical Methods for Demographers, U. S. Bureau of the Census, Washington; U. S. Government Printing Office, 1951.

Litchfield, E. H., Voting Behavior in a Metropolitan Area, Ann Arbor: University of Michigan Press, 1941.

Orleans, Peter, "Partisan and Nonpartison Politics in the Metropolis: An Analysis of the Social Contexts of Political Participation," unpublished doctoral dissertation, Northwestern University, 1964.

Robinson, W. S., "Ecological Correlations and the Behavior of Individuals," American Sociological Review, 1950, pp. 315-357.

Schnore, Leo, "The Myth of Human Ecology," Sociological Inquiry, 1961, 128-139.

Shevky, Eshref, and Wendell Bell, Social Area Analysis: Theory, Illustrative Application, and Computational Procedures, Stanford: Stanford University Press: 1955.

U. S. Bureau of the Census, U. S. Censuses of Population and Housing: 1960. Census Tracts. Final Report PHC(1) Series.

SOCIAL AREAS: A CASE STUDY IN THE METHODOLOGY OF MASS DATA ANALYSIS*

Dennis McElrath
Department of Sociology and Center for Metropolitan Studies, Northwestern University

Today it is not at all uncommon to find the son of a cattle herder or a peasant seated at a key punch, busily recording population changes in the middle of a government compound in Africa or Asia. As a result, the volume of raw facts on the desk of a social scientist has increased enormously in the last decade. But this quantitative growth in information is the product not only of advances in data-accumulation and processing techniques; it also stems from the increased bureaucratization of the world. Not just technological innovations, but changes in the organization of society as well sponsor the diffusion of social bookkeeping.

The enlarged scale of government and of business enterprise has increased the demand for detailed information. Much of this information is now gathered on a regular basis: censuses, vital statistics and labor force data are prepared every decade, or yearly, or quarterly, or monthly. Thus, (1) a greater volume of mass data now is accumulated from all over the world (2) on a greater variety of topics (3) at regular intervals. The advantage of mass data analysis, for both the social scientist and the policy maker, lie in these three characteristics of mass data: volume, variety and regularity.

Social bookkeeping records are not ordinarily maintained to answer specific policy issues. The mass data analyst, when facing a particular question, must therefore select relevant data out of a plethora of periodic observations and somehow reduce them to manageable proportions. Currently three major approaches are used to handle this problem of selection and data reduction. These procedures are: correlation analysis (such as multiple regression analysis and multiple factor analysis); classification; and construction of typologies.

The correlation approach frequently is used here where (1) little is known about the phenomena under investigation, or (2) where a strong inductive tradition exists in a discipline, such as in engineering or medicine. Correlation techniques attempt to reduce a variety of information to a few critical variables by observing the way in which a large number of variables are interrelated. By examining their inter-correlations, it is possible, with present computer techniques, to reduce these many measures to a few sets of variables which are closely related to one another. These few combinations of sets are then analyzed, treating each set as if it were a single variable. A mass of information is thus reduced to manageable dimensions.

The relevance of a set of variables for the analysis of a specific problem is determined by following similar procedures. Strategic variables are identified through a correlation analysis which determines how well a particular measure or combination of variables predicts any specified "dependent" variable. Thus, correlation is used both to reduce information and to determine its relevance.

The problems which plague this "shotgun" approach, of course, are that (1) relevance is established by correlation but correlation is not a cause and does not necessarily lead to understanding; and (2) that there is no way of determining whether or not reduction by correlation has obscured the critical variables by mixing them with related but unimportant measures. Thus, while data are reduced and predictions calculated understanding is not always advanced by this approach.

Development of classification schemes often attempts to instill explanation and understanding into mass data reduction operations. Here, a large number of observations are reduced by selecting from and combining original categories guided by some frame of reference. The approach is deductive in that the analyst picks a few observations or sets of observations out of many because his theory or explanatory frame of reference suggests that these are critical. His bent is toward explanation and to variables with strong interpretive power rather than to the empirical regularities guiding correlation analysis. Thus, a thick volume of occupational titles may be re-classified to a simple dichotomous classification of manual and non-manual occupations based on a theory indicating that this split in the skill hierarchy is critical in the development of nations.

Construction of typologies usually follows the same theoretical bent as classification. However, reduction of mass data to typologies has one advantage over classification in that data are reduced to several independent dimensions whose joint or combined variation is used to interpret a policy question. Because of this they are suited to the more complex theories now current in the social sciences. Their construction, however, is less standardized than the fairly automatic numerical reductions by correlation or factor analysis. All three are designed to handle a large volume and great variety of data. Differences between them lie in their theoretical relevance and in the extent to which a catalog of appropriate techniques is available.

The volume, variety and regularity of mass data have their disadvantages though they do offer opportunities to social scientists and to policy makers. The analyst usually must accept as given the observations, categories and areal units used in social bookkeeping records. These data are highly structured before they reach the analyst and this pre-structuring strongly influences the kinds of understandings or recommendations he can make. First, the observations themselves suffer from a variety of levels of interpretation which occur throughout the data accumulation process. These range from obvious instances of interviewer bias to subtle interpretations used in fitting a perceived world into prearranged categories. Second, often mass data are classified in categories derived from bookkeeping traditions of a data-gathering institution. Continuity or comparability of categories frequently is more important than current relevance. In addition, classifications may be designed to serve diverse interests of various agencies, bureaus or nations and relevance to a particular issue at times may be obscured for the sake of generality. The development of standard classifications and classification procedures which are widely used throughout the world has yielded fairly reliable information of great generality, but sometimes at the cost of some fairly revealing local classification schemes.

Legal restrictions also often constrain data presentation. Most censuses must guarantee the anonymity of respondents in their published records so that categories must be fairly large and cross tabulations limited. Otherwise an individual could be singled out from the crowd.

The third problem with mass data arises from the fact that they are collected on the basis of areal units and are tabulated areally. Nearly all censuses collect observations using a small local area as a basis for assigning individuals or households to an enumerator. Data usually are summarized for this enumeration area or for a set of such territories. If the areal unit is large enough, a great deal of material may be summarized and cross-classified while the anonymity of individual residents and households is preserved. Thus a great deal may be learned from census publications about a particular census tract, but nothing may be learned about a particular individual.

Some students of the city have attempted to talk about individuals from correlations of areal data. They have reasoned (erroneously) that these ecological correlations are closely analogous to individual correlations. On the other hand, there have been serious attempts to develop forms of analysis where the local area is the basic unit of observation <u>and</u> interpretation. This sociology of locality groups attempts to avoid misuse of ecological correlations and still take advantage of the wealth of local area data presented in almost all large scale social bookkeeping accounts.

Today I should like to report on several interesting recent developments in social area analysis.[1] These modifications and changes which I shall describe stem directly from attempts to take advantage of the increased volume, variety and regularity of local area data now available in national censuses of metropolitan areas throughout the world. The history of these changes represents a case study of more than a decade of effort on the part of a number of scholars to develop an approach to mass data analysis which takes advantage of these prevalent characteristics and also provides a theoretically meaningful and empirically grounded frame of reference for interpreting these data. Today I can only highlight some of these developments.

Most of you are familiar with social area analysis in the form originally applied to Los Angeles in 1949 and elaborated by Shevky and Bell in 1955.[2] In this second monograph, which included a number of revisions of the earlier work, the authors rigorously describe how census tract populations may be located in a "social space" or three-dimensional typology. This space is defined by three axes along which resources and opportunities are distributed in modern society. The location of tract populations along each axis is determined by combining several standardized ratios computed for each tract. "Urbanization", the first dimension of the typology arrays subareas according to prevalent alternative styles of family life ranging from "familism" on the one hand to "urbanism" on the other. It is measured by subarea distributions of fertility, women at work and house type. Thus, a familial area is characterized by a low proportion of women at work, a high fertility ratio and a high proportion of single family dwellings. "Social rank", the second dimension, arrays subpopulations by distributions of literate and non-manual skills. It is measured by combining education and occupation ratios. Finally, "Segregation" arrays subareas by using tract measures of the distribution of racial and nationality groups living in relative isolation.

From its inception, then, social area analysis (1) uses a typology to compress a large variety of widely recorded population characteristics; (2) forms this typology by selecting from and combining these characteristics guided by a general theory of social differentiation which may be broadly applied; (3) views local area populations as fundamental units of observation, and (4) interprets variations in local area populations in terms of combined importance of their location along all three axes of social differentiation. These assets of the original formulation make the approach highly generalizable and suggest that its development may have broad implications for analyses of mass data in a variety of quite different settings. Subsequent modifications represent then, an important case study in the methodology of mass data analysis.

A survey of revisions and modifications in social area analysis reveals that significant changes have occurred in five major areas. First, there have been important changes in the dimensions themselves. These occurred in response to observed empirical regularities and, more importantly, in several attempts to achieve greater theoretical clarification than existed in the original formulation. The segregation dimension is the major point where both of these considerations are operative.

In the original Shevky formulation, tract populations were arrayed according to the distribution of "subordinate" groups. The general view was that a large scale society included within its social boundaries a variety of populations of varying ethnic, racial and national backgrounds. Subordination was the outcome of their differential date of entry into the social system as well as shared physical and cultural visibility. Spatial isolation is viewed as a consequence and perpetuator of subordination. An empirical observation occurred in a social area analysis of Accra, Ghana, which stimulated further exploration of this dimension.[3] It became clear that here segregation involved two quite different conceptual dimensions which are usually compounded in the American situation.

Urban subpopulations in Accra were differentiated by tribal origin, on the one hand, and by migration experience on the other. While these frequently are empirically closely intertwined, they would seem to have quite different theoretical implications -- for example, assimilation of migrants involves socialization of the urban scene while ethnic assimilation involves the erasure of social stigmata. Yet both of these differentiators are involved in the single dimension of segregation. These empirical observations coupled with a series of comparisons conducted by John Barkey relating segregation to the assimilation process in Chicago led to an exploration of the theoretical import and analytic utility of separating ethnic status -- based on physical and cultural visibilities -- from migration status -- based on such shared migration experiences as the volume, variety and intensity of migration, the steepness of social boundaries crossed in the process of migration; and structural differentials in time of arrival.[4] By treating migration status separately from ethnic status, it is possible to examine situations where the statuses are compounded (e.g., recent Negro arrivals in Chicago), as well as their separate occurrence among subpopulations (e.g., long term resident Negroes in Chicago or recent rural-to-urban migrants in Mexico City). Thus empirical observations stimulated examination of these conceptually distinct dimensions. This separation modifies somewhat the underlying theory of social differentiation and suggests that the original three-dimensional typology be supplanted by dividing segregation into ethnic status and migration status. Work along these lines is continuing.[5]

A second modification of social area analysis concerns index construction. A basic requirement of the indexes used to measure each dimension of the typology is that they range evenly across the variety of situations studied. This means that each index should be "univocal" (i.e., measure the same presumed reality in each test situation) and that the value of the index (index score) be comparable across all urban areas. When the approach was applied across a variety of situations, several modifications of the indexes were required. One such modification occurs when a component is inoperative in a particular situation. The application of the urbanization dimension to Rome is an example of this.[6] One of the three components of this axis is house type. In the United States, the proportion of all housing units which are single family units is used as a partial indicator of life style. Obviously this index is not useful where the opportunity for subpopulations to be distributed in single units is non-existant. In this instance, between tract variation in house type is negligible. In fact, the slight variation which is observed is probably not indicative of life style at all but rather of the distribution of a few shacks (barrachi) occupied by poor migrants. That is, it probably measures migration status and social rank rather than urbanization. In this situation all that can be done is to throw out this component and perhaps substitute some other measure of this aspect of urbanization. I, personally, doubt that this measure is essential to the underlying concept of urbanization and probably could be discarded in all situations. Some of the recent work of Theodore Andersen suggests that this may indeed be the case.

Somewhat similar index problems occur in applying the education ratio (a component of social rank) in cross-societal studies. However, in this instance, functional literacy seems to be the operative concept and may be handily measured from most census sources.[7]

There is an additional index problem which may be partly involved in the above difficulties. This concerns the extent to which any index of social structure may be "culture free". This is, of course, a question which should be examined empirically whenever an index is applied across societies. The relevant proposition guiding such research using the Shevky frame of reference is that all large scale urban societies share fundamental social structures and that these in turn are subject to common measurement. A partial substantiation of this position is observed when all components of each axis vary together in an expected pattern across a set of subareas. When they do not, one should examine both the scale of the society as well as a cultural contamination of indexes. It may well be that in a small scale society, a particular form of differentiation is inoperative. In this situation, it may also occur that a particular index is inoperative or operates at variance with expectations because of the culture-bound social meaning of the index component. This latter difficulty is the case with house type in Rome, while the former occurs with the use of women in the labor force in Ghana.[8]

A third way in which the Shevky approach has been modified involves methods of standardizing components of each axis. Since Professor Orleans is going to discuss this subject in some detail, I shall merely mention that the question of standardization has not been resolved and we have therefore built a good deal of flexibility in this area into our computer programs.

Dr. Orleans will also discuss several solutions to problems encountered in shifting from one areal unit of analysis to another -- the problem of fit, for example. This is a fourth way in which the social area approach has been modified. I should like to preface his thoughtful paper with the single observation that, unlike the traditional "natural area" concept of classical ecology, social area analysis does not assume homogenous local area populations. Rather, the assumption is that each local area, census tract, enumeration area, gruppi di sezioni, precinct or what have you, is characterized by a distribution of attributes which may or may not vary systematically within a local area. It is this distribution which is either the object of interpretation or defines the context within which other dependent variables are interpreted.

Finally, a major area of change in social area analysis has been the development of a tool kit of supplementary techniques which facilitate application of this approach. I shall merely list them.

CENSAN - a flexible program adapted to the 709 which yields social area distributions from a variety of local area data. Options for standardization of components and the segregation dimension are built into this program.[9] Figure 1 presents sample output of this program based on tract statistics from ten metropolitan areas.

RATIO - a flexible program which yields a variety of local area ratios which may be used in conjuntion with CENSAN.

SYMAP - a computer graphic technique developed by Professor Howard Fisher which supplements CENSAN and yields highly legible social area topographies indicating the relative intensity of local areas within the social area grid. Figure 2 presents sample output from this program.

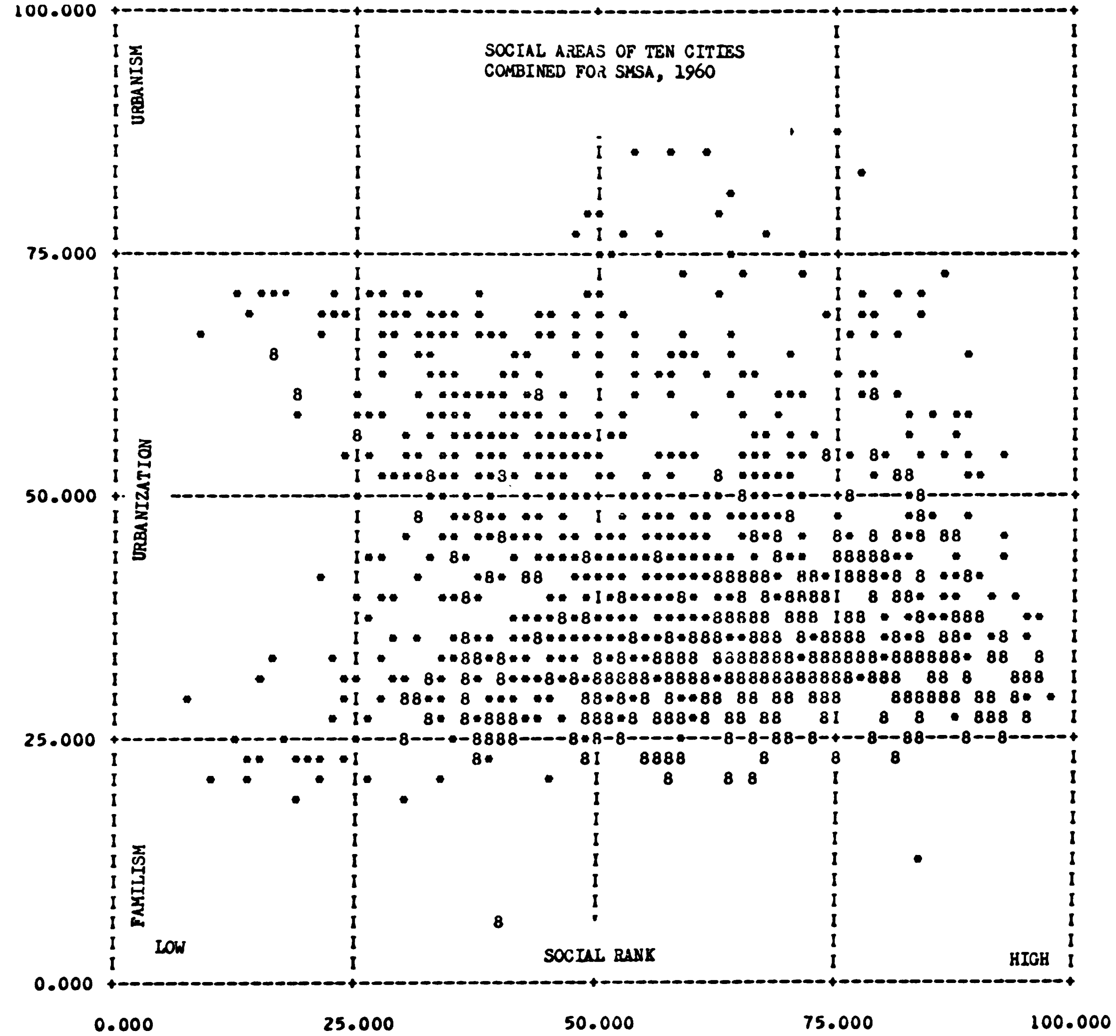


Figure 1

8=80% or more automobile commuters to the central city

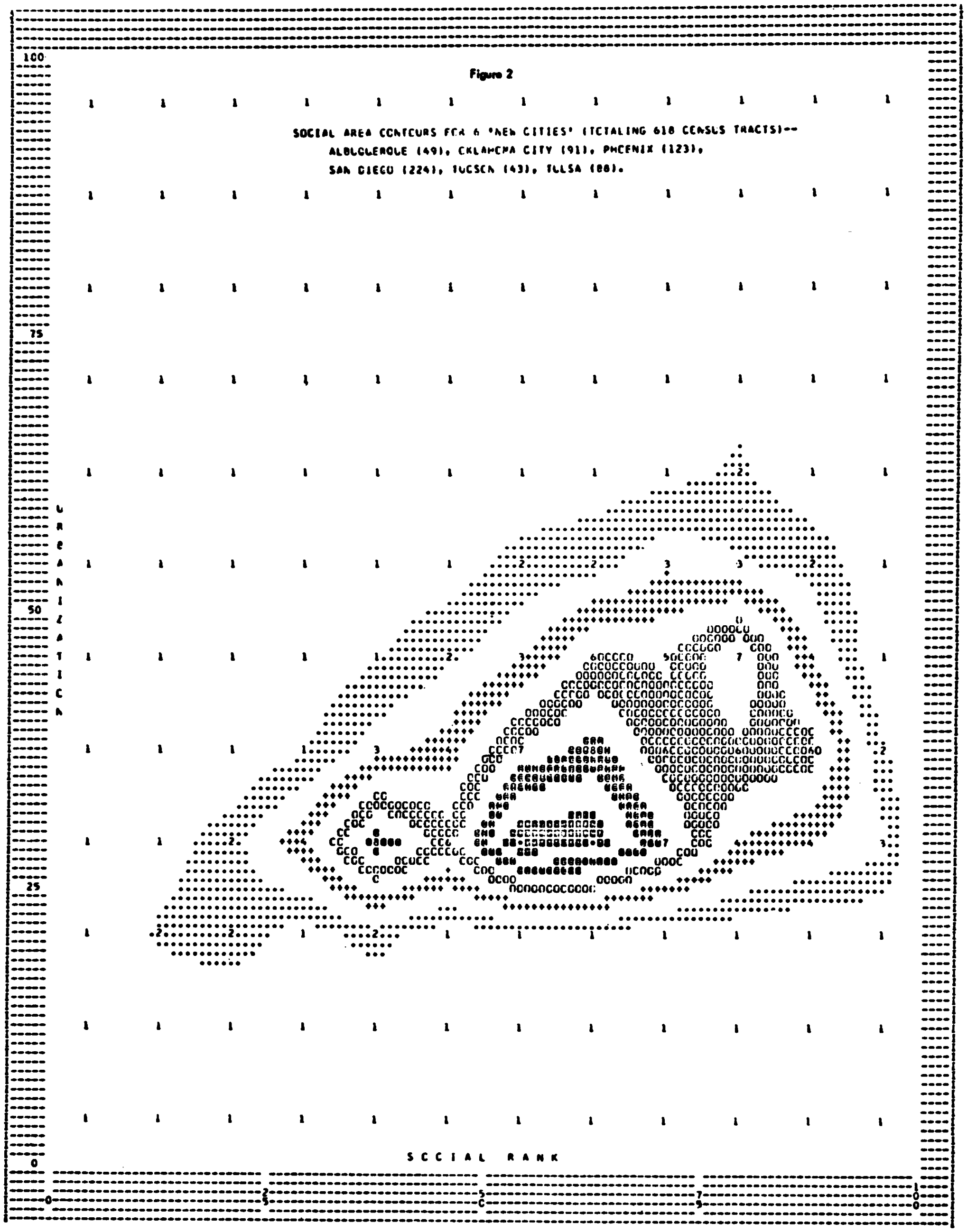
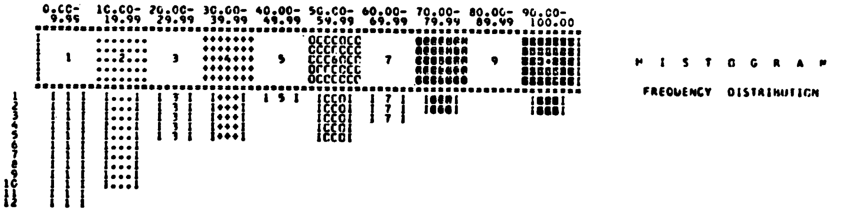


Figure 2

SOCIAL AREA CONTOURS FOR 6 'NEW CITIES' (TOTALING 618 CENSUS TRACTS)--
ALBUQUERQUE (49), OKLAHOMA CITY (91), PHOENIX (123),
SAN DIEGO (224), TUCSON (43), TULSA (88).

BASED ON STANDARD METROPOLITAN STATISTICAL AREAS, 1960 CENSUS.
SCALE IS IN PERCENTAGE DECILES-- ABSOLUTE RANGE 0 TO 4.
'SOCIAL RANK' = A STANDARDIZED INDEX BASED ON EDUCATION AND OCCUPATION.
'URBANIZATION' = A STANDARDIZED INDEX BASED ON EMPLOYMENT OF WOMEN,
MULTIFAMILY HOUSING, AND INFERTILITY. (HOWARD T. FISHER)



In addition to several special programs, several special programs, several non-computerized techniques have been developed as well. These include a technique for sampling local areas by distance and direction weighted for the typical density gradient of single centered metropolitan areas. This facilitates initial exploration of relations between geographical distributions of local areas and their configurations in social space. In all of these efforts we have tried to develop tools which are sufficiently flexible to take advantage of the increased volume and variety of local area data and which will be ready for the 1970 round of censuses. I should greatly appreciate hearing your comments on these revisions before that date.

---

[*] Paper prepared for delivery at the annual meeting of the American Statistical Association in Chicago, December 29, 1964. The first section was prepared with Raymond W. Mack.

1. An annotated bibliography of work using this approach is available from the author.
2. Eshref Shevky and Marilyn Williams, The Social Areas of Los Angeles: Analysis and Typology, Berkeley and Los Angeles: University of California Press, 1949; and Eshref Shevky and Wendell Bell, Social Area Analysis, Stanford University Press, 1955.
3. Dennis C. McElrath, "Social Change and Urban Social Differentiation: Accra, Ghana", (Mimeographed).
4. John W. Barkey, "Selected Aspects of Change in the Social Areas of Chicago, 1930-1960: A Research Proposal". Unpublished Master's Thesis, Department of Sociology, Northwestern University, Evanston, Illinois, 1963.
5. Dennis C. McElrath, "Migration Status in Accra, Ghana", (Mimeographed).
6. Dennis C. McElrath, "Social Areas of Rome: A Comparative Analysis", American Sociological Review, Vol. 27, No. 3. June, 1962, pp. 376-391.
7. Loc. cit.
8. Dennis C. McElrath, op. cit., "Social Change and Urban Differentiation: Accra, Ghana".
9. Program available on request from the author.

VIII

RETRAINING AND RESEARCH FOR AMERICA'S MANPOWER NEEDS

Chairman, Margaret E. Martin, U. S. Office
of Statistical Standards

Page

Congressional Expectations for Manpower Research Under the Manpower Development and Training Act - Curtis Aller, San Francisco State College. 144

Administrative Statistics on Retraining - Frederick W. Suffa, Office of Manpower, Automation and Training............................ 150

Blue to White Collar Job Mobility: A Preliminary Report - James L. Stern and David B. Johnson, University of Wisconsin........................ 166

Discussion - Marvin Friedman, AFL-CIO............. 187

Discussion - Richard A. Lester, Princeton University.......................... 189

CONGRESSIONAL EXPECTATIONS FOR MANPOWER RESEARCH UNDER THE MANPOWER DEVELOPMENT AND TRAINING ACT

Curtis Aller, San Francisco State College

I. Research and the Manpower Act

There is, I submit, very little that can be said directly on this subject. The formal action of Congress as reflected in Title I (see appendix A) is certainly impressive and suitably broad. Title I is permanent and represents a significant extension of the underlying commitment of the Employment Act of 1946 to a full employment economy. In the 1962 Act manpower policy was elevated to the level of aggregative economic policy. As one vivid indication of this new status Congress required the President to furnish an annual manpower report that is comparable in scope and importance to the older annual economic report of the President. Moreover, the record is clear that this was a congressional innovation since the Senate Committee on Labor and Public Welfare featured this as one of the changes it had made in the bill as introduced.

In reviewing the record I could find no better summary of the balance of Title I than that contained in the statement by Seymour Wolfbein, Director, Office of Manpower, Automation, and Training, Department of Labor before the House Appropriations Subcommittee on February 6, 1963. I quote:

"The act imposes the obligation on the Department of Labor to carry out one of the most extensive and meaningful research programs ever required by law. Specifically, the Secretary is directed to: (1) Evaluate the various factors of the problems created by automation, technological progress, and other changes in the structure of production; establish techniques and methods for detecting in advance their potential impact; develop solutions to these problems and publish relevant findings. (2) Establish a program of studies of practice of employers and unions which affect the mobility of workers and to report on the results of these studies. (3) Appraise the adequacy of the Nation's manpower development efforts to meet foreseeable manpower needs and recommend necessary adjustments for the most effective utilization of its manpower. (4) Promote, encourage, and engage in programs of information and utilization, and the amelioration of undesirable manpower effects. (5) Develop, compile, and make available manpower information regarding skill requirements, occupational outlook, job opportunities, labor supply, and employment trends. In addition, the statute requires the Secretary to 'arrange for the conduct of such research and investigation as give promise of furthering the objectives of this act.' The problems to be investigated under this research program are huge in scope with constantly shifting dimensions. During a period in which an industrial revolution of broad magnitude and significant implications is taking place it becomes necessary to obtain as much manpower and related information as possible so that both immediate and ameliorative measures and long-range solutions can be developed without creating additional problems and dislocations during this period."[1]

It seems remarkable that a program requiring "one of the most extensive and meaningful research programs ever required by law" could have been adopted by Congress with so little apparent discussion. The record is exceedingly sparse and I have discovered no source that would indicate any more extensive background consultation or debate. The Senate held only four days of hearings on the bill and the House three. It should be remembered, however, that in the immediately preceeding years, Senator Eugene McCarthy and Representative Elmer Holland had headed special committees which had been probing deeply into our manpower problems. The stage had been set by these investigations and presumably the time for action had arrived. Therefore the fact that the concern of both the Senate and House at the Committee stage and during subsequent floor debate seemed to be focused more on the operational (Title II) apsects of the program, becomes more understandable. Nevertheless it remains surprising that the totality of Congressional discussion on manpower research could be reproduced in a few pages. I will sketch the highlights even more quickly.

Senator Clark in his committee report on the bill almost seemed to be taking pains to deemphasize Title I when he wrote:

"To a considerable degree, title I of the bill is a restatement of existing responsibilities of the Department of Labor. The Secretary of Labor now possesses the authority to evaluate the impact of automation, the mobility of labor, and conduct research and information activities in the manpower field. What is added is a specific directive 'to appraise the adequacy of the Nation's manpower development effort' as a while and analyze manpower requirements, resources, and use to provide a sound basis for public and private training efforts throughout the country. That the Secretary undertake these tasks is in the interest of avoiding waste, providing a focus for the coordination

---

[1] Seymour Wolfbein, Hearings Before a Subcommittee of the Committee on Appropriations, House of Representatives, 88th Congress, 1st Session, Subcommittee on Departments of Labor and Health, Education and Welfare and related agencies appropriations, p. 168.

of Government activities affecting manpower requirements, development and utilization, and making it possible for the Nation to meet the staffing requirements of the struggle for freedom."[2]

In later floor discussion Senator Clark became somewhat more expansive in his description of Title I as follows:

"First, the Secretary of Labor is directed to find out where job opportunities are. This is a most important part of the proposed legislation and is set forth in full in title I. We do not really know today what our manpower requirements are. We do not know really what skills are in short supply. We do not know what the requirements for everything from ditchdigger to nuclear physicist are likely to be in the years ahead. In short, we do not know how to staff freedom, man our economy to meet the worldwide challenge it faces. Title I of the bill directs the Secretary of Labor to find out how to staff freedom."[3]

The only full explanation of Title I was given in a speech by Senator McCarthy. I will quote one section that caught my fancy.

"For a long time there has been a serious gap in our knowledge about our working population. One can open the Statistical Abstract of the United States and learn, for example, that on January 1, 1961 there were 28,688,000 stock sheep and 55,305 000 hogs in the country; yet, we do not know with the same exactitude how many tool and die makers, electricians or physicists we have.

"Too few people realize that the only comprehensive detailed occupational statistics in this country are those collected every 10 years in the decennial census of population. The data from the decennial census, even when first available, are about 2 or 3 years old and are about 12 or 13 years old before the next set of data is published."[4]

Still later in his speech supporting the conference version of the bill Senator Clark alluded to another possible dimension of Title I in these words:

"I believe this action today to recognize manpower planning as an element of national policy is one to which we shall look back upon in the years to come with merited pride.

"I believe this function of manpower planning, which appropriately belongs in the Department of Labor, will make an enormous contribution toward overcoming unemployment in the days ahead and will better enable us to staff freedom in the constant cold-war struggle with our Communist opponents."[5]

Representative Powell for the House Committee on Labor and Education in his report described Title I somewhat more fully and I quote:

"One of the important elements of this proposal is the assignment to the Secretary of Labor, in order to further the broad training purposes of the bill, of additional responsibilities in the overall manpower field.

"Title I of the bill will enable the Secretary of Labor to establish a continuing review of the national skill development effort and to recommend actions needed to achieve improved balance between occupational resources and requirements. Combining these manpower functions in one agency will give much needed overall unity to the Federal Government's responsibility for leadership in the field of skill development. It will also more effectively relate the separate activities of the various agencies in this field to an overall program of optimum development and employment of manpower resources.

"To assist the Nation in accomplishing the objectives of technological programs, while avoiding or minimizing the harsh and tragic consequences of labor displacement, title I also requires the Secretary of Labor to evaluate the impact of automation on the utilization of the Nation's labor force, to appraise the adequacy of the Nation's manpower development efforts to meet foreseeable manpower needs, and to arrange for the conduct of such research investigations as give promise of furthering the purposes of this proposal.

---

2. Senator Joseph S. Clark, Manpower Development and Training Act of 1961, Report No. 651, Committee on Labor and Public Welfare, 87th Congress, 1st Session, pages 8-9.

3. Senator Joseph S. Clark, Congressional Record, August 23, 1961, p. 15688.

4. Senator Eugene McCarthy, Congressional Record, August 23, 1961, p. 15696.

5. Senator Joseph S. Clark, Congressional Record, March 8, 1962, p. 3353.

"Many of the beneficial practices that have evolved as a byproduct of labor-management relations (e.g., pension plans and other fringe benefits) have introduced rigidities that impeded labor force adjustments and mobility, thus contributing to unnecessary unemployment. Title I, therefore, directs the Secretary of Labor to make intensive factual studies of what causes lack of occupational mobility and to encourage the voluntary adoption of equitable means by which these impediments might be removed. It also directs the Secretary of Labor to study and report on how the gradual retirement of long-service workers, the vesting of pension rights, and the development of other devices freeing the laid-off workers from equity losses incurred by moving might be encouraged by Government and private actions.

"Title I authorizes the Secretary of Labor to develop, compile, and make available information regarding skill requirements, occupational outlook, job opportunities, the labor supply in various skills, and employment trends on a National, State, or other area or appropriate basis. This is in effect an inventory of the occupational resources and needs of the Nation which will be used in the educational, training, counseling, and placement activities performed under other provisions of this act.

"Finally, title I will require the Secretary of Labor to report to the President on manpower matters, and the President to transmit an annual manpower report to the Congress."[6]

Title I remains, therefore, something of a mystery. The Department of Labor appears to have played the major role in developing the concepts and the language. Yet no historian then or later has reconstructed the sequence of events. I am informed the process was chaotic and that the structure was built in a series of meetings. Much of the language reflected an awareness of some of the Congressional concerns. Thus the references to automation and the mobility consequences of private pensions are two such examples. Aside from these attention getters it would appear that Congressional concerns lay elsewhere. After all three million dollars for research and related activities must have seemed small alongside a total of 161 million dollars being requested for the operational program. Not only was the money request here a large one in the eyes of the economy bloc but there were also doubters who questioned the wisdom or the practicality of the program. Research, it would seem, simply slid through under the umbrella of larger controversies.

More, I suspect, could have been accomplished if it had not been for the timidity of the Labor Department. This department has not fared well historically in Congress when it comes to research and I'm sure this has contributed to its low expectations. I can cite two bits of evidence for what may be the larger pattern in this respect. During the House consideration of the Manpower Act in 1963 several efforts were made to interest the Labor Department in opening up Title I for expansion in the budget and possibly an enlargement of functions. Every time the suggestion was made great fears were expressed that Congress if it really considered research questions on their merits would end by reducing the level of activity. Similarly when the proposed Automation Commission was before the House, labor department representatives informally suggested that only a modest amount of money, possibly a quarter of a million, would be needed. The House committee settled on two million dollars as being more appropriate and this figure was not questioned on the House floor. Nonetheless a key representative of the Department expressed a view at Senate hearings on the bill that such an astronomical sum of money could not be used. Such an admission, though recanted, proved to be fatal.

One consequence of the department's exceedingly modest view of the resources required for manpower research is now becoming evident. Congress in the Vocational Education Act of 1963 provided for an allocation of ten per cent of the funds off the top for research and demonstration projects with additional amounts to be set aside from state allocations. For the current year the amount is 11.8 million and this sum could grow to 22.5 million. Much of this will go for manpower research and since the Labor Department is the repository of trained personnel in this area one can predict an inevitable shift in personnel. The Labor Department will continue to play a role, no doubt, through coordination, but it will now be the tail on a much larger dog.

II. Congressional Needs

If Congress bought manpower research largely on faith and without any constructive discussion, as I believe to be the case, then we need to ask how manpower research can meet Congressional needs. Without attempting to be exhaustive let me suggest some of the elements of an answer.

First, we need to recognize that Congress stands at the apex of our policy making institutions. Yet there are few in Congress who have the time or the interest to become manpower experts. Even those who serve on the appropriate committees are pulled in dozens of different directions and can devote little time to the area. They, and ultimately the balance of the members, are heavily dependent upon the flow of ideas and information from the administration. To some extent this channel may be supplemented by outsiders but even here a pro-

---

[6] Representative Adam Clayton Powell, Manpower Development and Training Act of 1961, Report No. 879, Committee on Education and Labor, 87th Congress, 1st Session, pp.8-9.

cess of selection may often be at work ensuring a relatively narrow range of information. At times the process may be leavened by the turmoil and eddies of conflicting political pressures. But in that event the demands for information from research sources becomes even more pressing.

Second, individual congressmen are keenly aware of the unmet social and economic needs of our system. The effects of racial discrimination, the problems of older workers, the forthcoming surge of unemployed youth, to mention a few, are all aggravated by the persistent high levels of unemployment we have been experiencing. Letters and other contacts of constituents personify these problems and lend urgency to the desires of congressmen to find solutions. Quick, easy, and complete answers are sought and manpower experts are expected to provide guidance.

Third, this means that research efforts are welcomed to the extent that they provide useful guides to policy or suggest ways for improving the operations of our institutions. Congressmen recognize the problems are complex but they are impatient with elucidations of these facts and want to jump beyond to the answers. Here again I can cite two examples.

Economists have certain concepts about mobility of resources and many manpower experts have been contending that labor mobility should be facilitated through subsidies. Other voices have strongly condemned these ideas. Faced with this welter of contending ideas Congressmen become uncertain. They welcomed the 1963 Manpower amendment which provided up to 4 million dollars for experiments with labor mobility demonstration projects. It removed the idea from the arena of controversy for awhile at least. More important it offered the opportunity to find out if subsidies or other assistance could be made to work and whether these would be useful. With information in hand the Congress would thereafter feel more secure in tackling the difficult task of writing new law in this area.

The proposal for an automation commission provides another example. There were heavy political pressures for this proposal and strong, although unexpressed, doubts. Key members of the House committee recognized these political overtones and yet sought to convert the proposal into a mechanism that would, as a byproduct, produce some solid new knowledge in the area. This was basically the reason for the high authorization on the House side. In the end, however, I am sorry to say the concept of the commission as a political maneauver appears to have won out.

Fourth, I think it important to note again that the President's Manpower Report is a creation of Congress. Nevertheless, Congress has done little with the report so far. What is needed is a vehicle for Congressional hearings on this report each year along the lines of the reception Congress gives the annual economic report. This could be of enormous value in stimulating a dialogue at the Congressional level of manpower problems. From this we might begin to get more indications of Congressional needs and expectations. There is good reason for believing that such a pattern may be followed and we have been close to it this past year. However, a new vehicle in the form of a Joint Committee may ultimately prove necessary. For this potential to be realized, however, leading members of the profession would have to persuade Congress by attesting as to its value and by signifying their willingness to participate.

III. Contributions of the Scholarly Community

With the cooperation, perhaps acquiesence would be better, of Congress we are now committed to an active manpower policy. The exact dimensions of this policy have not been prescribed by Congress. We are only at the threshold of an uncharted future. In a very real sense, I would argue, Congress by necessity must rely heavily upon the scholarly community for the further evolution of this policy. The bureaucracy will be of assistance but it is well to recognize its contributions over time will tend to be limited by built-in caution, concern with jurisdiction and absorption of time in operations. The strategic contributions will, I believe, have to come from outside government although in most instances fruitful vehicles for inter-meshing the two can be used.

Because time is limited I will simply sketch come possibilities here. First, I think caution is needed in seeking to exercise our proclivity for basic research. The definition of this term is, I know, partly a matter of taste and yet much of what many would include in this area falls within the province of what is already the obligation and commitment of academia. At the other extreme we have government agencies geared to undertake the massive and routine data collection activities our society requires. There is an intermediate area where in view of my remarks on the policy needs of Congress major contributions can be made by the scholarly community. That is to take the emerging ideas stemming from basic research that now exist and will continue to develop and translate these into useful hypotheses for applied research.

Second, I have noted that present resources for manpower research are relatively meager. Yet we are fortunate in having in being an infrastructure for research that can be adapted to our needs with only modest effort. These are the industrial relations centers that were established after the war. Here we have an opportunity for labor economists to once again use the second half of their title.

Third, beyond these centers now in being there may be justification for the creation of some new kinds of research centers that will utilize different sources of research talent and knowledge. Title II of the manpower act brings together for the first time in operational harness education and labor agencies. It might be useful to consider new centers in those colleges that now specialize heavily in research and training for our educational complex so that at this level both education and labor market analysis are tied together.

Research centers, however organized or located, will as a by-product be producing a new supply of trained personnel to staff the emerging manpower programs and to undertake continuing research. In both respects we are finding that our pool of talent is extraordinarily thin given the new demands we are placing on it. This is even more remarkable considering the rather meager 2.1 million dollars a year so far allocated to outside research under the manpower act. And when the new vocational act research funds are added the picture becomes extremely bleak.

These talent shortages are more critical than is generally realized.[7] New programs have followed one another with dazzling speed. Each competes for essentially the same basic pool of trained manpower and the resultant shortages have been predictable. The need for trained employment counselors, for example, now appears to be double the supply. Similar, though less dramatic, shortages exist for other skill categories. The scarcity of these resources may spell the difference between success or failure for an active manpower policy and yet awareness lags as political sex appeal lies with the new programs and not the operational necessities. For these to secure the attention they deserve may require persistent pressure from the scholarly community.

Fourth, the experimental and demonstration projects under the manpower act have proven to be a most promising vehicle. Outsiders can play an important role by proposing new projects, developing concepts to be tested and otherwise assist in maximizing the research potential of these activities. Aside from their obvious advantages, I have argued elsewhere that these projects can prove to be a strategic vehicle for stimulating change in established bureaucracies because to a limited extent they provide a competitive challenge.[8]

Fifth, I would commend for your attention the suggestion of Frederick Harbison that we need to constitute a group of manpower planners for the purpose of developing unifying ideas that can begin to tie together in some cohesive form the bits and pieces that will emerge from discrete research.[9] If this proves too ambitious it might still be possible for groups of manpower planners to begin the process of selecting some key areas where current research permits a sharp focus on policy possibilities but hasn't yet done so.

Alternatively, each of us might consider devoting some time to one of our existing manpower institutions where conditions are ripe for reform. We could take our existing knowledge, add some dashes of healthy revolutionary zeal and go to work. We might be surprised by the ease and success of these endeavors. My particular target at the moment is the Employment Service. The woods are full of others.

Finally, let me close with a modest note of optimism. I have implied that present research funds may be more meager than perhaps they could have been. Yet I would suggest that the real need is for truly creative ideas. These, while rare, may not require much research hardware. As a case in point let me remind you of the study One Third of a Nation. This is one of the most profound studies of the recent past. Its cost was only $30,000.

---

[7] See for another version of the needs, Margaret Gordon, "Training Programs at Home and Abroad", Proceedings of The Annual Meeting Of The Industrial and Labor Relations Association, December 1964.

[8] See Summary Report of the OMAT Training Conference, Manpower Administration, August 24, 1964.

[9] Frederick Harbison, "Problems in American Manpower Policy and Practice", Proceedings of The Annual Meeting of the Industrial and Labor Relations Association, December 1964.

APPENDIX A

Excerpt From Title I, Manpower Development and Training Ace of 1962

Sec. 101. . .It is therefore the purpose of this Act to require the Federal Government to appraise the manpower requirements and resources of the Nation, and to develop and apply the information and methods needed to deal with the problems of unemployment resulting from automation and technological changes and other types of persistent unemployment.

Evaluation, Information, and Research:

Sec. 102. . .To assist the Nation in accomplishing the objectives of technological progress while avoiding or minimizing individual hardship and widespread unemployment, the Secretary of Labor shall:

(1) evaluate the impact of, and benefits and problems created by automation, technological progress, and other changes in the structure of production and demand on the use of the Nation's human resources; establish techniques and methods for detecting in advance the potential impact of such developments; develop solutions to these problems, and publish findings pertaining thereto;

(2) establish a program of factual studies of practices of employers and unions which tend to impede the mobility of workers or which facilitate mobility, including but not limited to early retirement and vesting provisions and practices under private compensation plans; the extension of health, welfare, and insurance benefits to laid off workers; the operation of severance pay plans; and the use of extended leave plans for education and training purposes. A report on these studies shall be included as a part of the Secretary's report required under section 104.

(3) appraise the adequacy of the Nation's manpower development efforts to meet foreseeable manpower needs and recommend needed adjustments including methods for promoting the most effective occupational utilization of and providing useful work experience and training opportunities for untrained and inexperienced youth;

(4) promote, encourage, or directly engage in programs of information and communication concerning manpower requirements, development, and utilization, including prevention and amelioration of undesirable manpower effects from automation and other technological developments and improvement of the mobility of workers; and

(5) arrange for the conduct of such research and investigations as give promise of furthering the objectives of this Act.

Skill and Training Requirements:

Sec. 103. . .The Secretary of Labor shall develop, compile, and make available, in such manner as he deems appropriate, information regarding skill requirements, occupational outlook, job opportunities, labor supply in various skills, and employment trends on a National, State, area, or other appropriate basis which shall be used in the educational, training, counseling, and placement activities performed under this Act.

Manpower Report:

Sec. 104. . .The Secretary of Labor shall make such reports and recommendations to the President as he deems appropriate pertaining to manpower requirements, resources, use, and training; and the President shall transmit to the Congress within sixty days after the beginning of each regular session (commencing with the year 1963) a report pertaining to manpower requirements, resources, utilization, and training.

ADMINISTRATIVE STATISTICS ON RETRAINING

Frederick W. Suffa, Office of Manpower, Automation and Training

The purpose of this paper is to describe the types of data gathered as a part of the administration of retraining projects authorized under the Manpower Development and Training Act, the system developed to collect and process the data, and to relate these statistics to the needs of the administrator and researcher. No attempt has been made to develop a statistical description of program activity.

The basis for the administrative statistics related to the retraining program is Section 309 (a) of the MDTA, which directs the Secretary of Labor to make a report to the Congress by April 1st of each year containing an evaluation of the retraining programs conducted with funds appropriated in accordance with the provisions of the Act. Section 309 (a) specifies that the report shall include "...the number of persons trained and the numbers and types of training activities under this Act, the number of unemployed or underemployed who have secured full-time employment as a result of such training and the nature of such employment...". Throughout Title II of the Act, the Congress specified various categories of workers who are to receive priority in referral to training, payment of allowances while in a training status, or who are eligible for the various services that can be furnished in an effort to increase their current employability. In addition to describing the manner in which these groups are served, proper evaluation of program achievement must be related to the findings of Congress and purpose of the Act which are set forth in Title I.

Therefore, reporting and evaluation of the conduct of the training program must take into account many factors. For convenience, I will discuss them from three main points of view, the administrator, the evaluator, and the researcher. While collection of information based upon the total number of trainees and gross amount of funds obligated or expended can be used to describe the broad general scope of the program, these data alone are insufficient for the needs of the administrator and the evaluator. In addition to these aggregates, the administrator needs to know the what, where, when, and how of the program as well as the number of persons that are expected to be trained, and the cost in terms of allowance payments and education expense. He needs information on the expectation of future employment for persons in the occupation for which training will be given. For on-going projects, he requires data on the number that complete training and the number that drop out prior to completion. For the latter group, he also needs to know the reasons that trainees find it necessary to drop out. The administrator is also interested in determining the number of completers who obtain employment as a result of the training, as well as the number that obtain full-time employment.

The evaluator and the researcher are interested in going behind the administrative data requirements in order to present a comprehensive picture of program fulfillment, shortcomings, or problem areas. For these purposes, summaries of individual trainee data which can be related to specific projects, occupations, and geographic areas become essential. This information includes the basic demographic descriptors such as age, sex, race, marital status, education, physical condition (whether or not handicapped) and the number of dependents. Economic data, such as the individual's former industrial attachment and earnings, his labor force status at the time the training was offered, his experience in the labor force (gainful employment), his status with respect to receiving unemployment compensation and public assistance, are important complementary items to the basic enrollee demographic data previously mentioned. Then, for the post-training period there is a need for information concerning the total number of weeks that the trainee worked after completing his scheduled course of instruction, the number of jobs he has held, and the number on which he was able to utilize his training, his reasons for leaving a permanent job, and his current labor force status. For those who are employed, data that will identify the industry in which the job is situated, the number of hours he worked on that job, his earnings, the relationship of the current job to his training occupation, and the likelihood that the job is permanent (over 30 days duration), are also important.

Social data important to the researcher and evaluator include such items as the trainee's primary occupation as well as the occupation in which he last found regular employment, the length of time that he worked in his primary occupation, and the time that has passed since his last employment in that line of work, his post-training occupation, his willingness to accept a job outside his immediate commuting area, and the location of his post-training employment.

The categories mentioned represent the most important elements of data concerning trainees that lend themselves to collection and analysis on a recurring basis. Other types of information concerning program conduct and accomplishment require different approaches; generally that of a case study. At the present time, our collection of information concerning program operations and results are carried out through one of the three following methods:

1. Routine collection of data from program operations through use of specially designed reporting forms;

2. Special investigations by the evaluation staff; and

3. Contractual studies by outside investigators.

I do not intend to discuss the latter two programs at any length in this paper. However, I will mention them later to illustrate how they fit into the overall scheme of program evaluation and information.

Routine Data From Program Operations

Source

Perhaps the best way to introduce a discussion of the system for collecting and processing the program data which arise from operations would be to outline the steps that are involved in the approval and conduct of the retraining projects. As the major source of training to date has been the institutional training program, that is the one I will discuss. However, the on-the-job training project development follows a generally similar path. The major difference being that of involvement of the local Bureau of Apprenticeship and Training representative in the initial and approval stages rather than the local employment service office. A flow-chart relating report forms to the steps involved in the approval and subsequent conduct of retraining projects for training under public education or other institutional agencies is given in Chart 1. Copies of the forms are attached as an appendix.

Retraining projects can be undertaken only in those instances where a need for workers has been demonstrated and there is a reasonable expectation of employment for the enrollees. For the purpose of obtaining information concerning the need for training, the BES has developed procedures for the local offices of the various State Employment Services to follow in conducting surveys for the purpose of establishing the need for training courses. These procedures are designed only for the purpose of formulating training projects and do not follow scientific sampling procedures. Thus, the data developed by the surveys do not describe the total needs for the occupation in the area surveyed; neither do they provide suitable estimates of shortages in specific occupations. For this reason, the State agencies have been cautioned to use the data only in a form that would indicate the survey showed an unmet need for workers in the occupations studied, or that a shortage of qualified workers existed.

Following establishment by the local office that there is a need for a training course in order to meet the demand for specific occupations, that there is a trainable supply of workers available, and that there is a reasonable expectation that the trainees will find subsequent employment, the local office completes a Notification of Occupational Training Needs, Form MT-1. The Notification is sent to the State Employment Service Office and to the State Vocational Education Agency. If the State office approves the local office's recommendation for establishment of a training project, the Notification is used by the vocational officials as a basis for development of a training plan. Additional copies of the Notification are furnished the members of the local area Coordinating Committee and to members of the local advisory committee for advice and assistance in the development of an appropriate training proposal. The Notification and its statistical supplement contain data on such items as the number of workers required in the occupation, the number of individuals available for training, the wage rates or range of rates at which trainees are expected to be employed, and a certification that the wage rates, hours and conditions of work in the jobs which the trainees are expected to fill, will conform with existing standards found in area employment. Because of its nature as a planning document, the Notification is not used as a source document for entry into the basic data system. However, a copy of the form is included in the project file maintained in the National offices of the Manpower Administration. For various internal evaluations and for limited research by other Bureaus, the information has some value. (For example, the BLS has found the entry wage rate information useful in preparing their occupational outlook publications.)

Following establishment of the need for a training project and the concurrence of the State officials that the need is one that should be met through MDTA funds, a formal application for project approval is originated jointly by the local employment service office and the local vocational education facility. The Application Form. Form MT-2, identifies the area from which trainees are to be selected for training, the occupation for which training will be given, the number of trainees and estimates as to the allowance payments and training costs that will be incurred. This form is the basis for project approval. Attached to the Application Form is a justification for training - the previously described Notification and its supporting documentation, a detailed training budget, Form OE-4000, which is prepared by the local training agency and indicates the detailed cost estimates for classroom instruction; the training plan, (Form OE-4014) which is submitted by the training facility to outline the enrollment requirements (educational and experience), curricula, instructor qualifications, equipment availability, and finally a summary of the State agency administrative costs which will be incurred during the selection and allowance payment process.

Upon approval of the project proposal, selection and referral of trainees starts in the local office. Local office files of applicants are searched to identify likely candidates for the training course. Notice is given through the local communication media in order to attract other eligible candidates. Following a review of the file of job applicants in the local office including, where appropriate, specific aptitude tests, selection of potential trainees is made. At the time of selection, a report on the Characteristics of Trainees, Form MT-101, is completed. 1/ This form is the basic report used to obtain information concerning individual trainees. When training has started, trainees who are eligible for allowances file a weekly claim that is endorsed by the training facility to indicate satisfactory attendance and progress. This latter report is purely an operational form that does not enter into the basic national statistical system. However, the State agencies summarize the allowance payments for each project on a monthly basis and submit the project summaries to Washington on Forms ES-149.

Another report which is made during the conduct of the training course is the Trainee Termination Report, Form MT-102. This report is completed by the training facility whenever an enrollee is terminated. For our purpose here, the word termination includes graduation or completion, reaching a level of competence acceptable to an employer, or other terminations prior to the end of the course.

Following the conclusion of the course, a series of Post-Training Reports, Form MT-103, are received. These reports contain information concerning the employment status and experience of completers during three reference weeks at 3, 6 and 12 month periods following the completion of the project. This information is obtained either from a personal contact with the completer by a representative of the selection and referral office, or through a mailed questionnaire (Form MT-103a). Also, at the end of the training project, a final training cost report is submitted to the Office of Education on the Form OE-4000. All of the forms cited above except for the Notification (MT-1) and Training Plan Outline (OE-4014) provide data for subsequent compilation.

Compiling System

In the preceding discussion, we have presented an outline of the types of information concerning trainees necessary to describe and evaluate the retraining program as well as the sources for the data. With this background, we will now turn to a discussion of the system that has been developed for tying these data together and summarizing them into meaningful statistics and information. Our system is really two major and nearly independent systems bound together with a common thread. The two major systems are: First, one which has been developed for processing the financial data relating to program approvals and expenditures; and second, the system which has been developed for processing the data related to individual trainees. The data relating to both financial and trainee information are recorded on magnetic tape, identified by a standard identification code and processed through electronic computers. Thus, an analyst is able to call upon the data stored in either or both systems to aid his analysis.

Upon approval of a training project, a copy of the approval document, the Application Form (MT-2) is transmitted to the Office of Financial and Management Services of the Manpower Administration. Codes to identify the areas from which trainees will be drawn, the county in which the training facility is located and its congressional districts, the nature of the training to be given (OJT, Institutional, Prevocational, Basic Education, Experimental or Demonstration, etc.), and occupation for which the trainees are being prepared are entered on the form, and the form is sent to the Data Processing Office for translation into punch cards and magnetic tape. The financial record in machine form has been established.

Just as the Labor Department's Washington offices receive copies of the MT-2 Form, the Office of Education's National office receives copies of the training budget, Form OE-4000. Upon receipt, an office record is established on a Typetronic machine that creates a paper tape containing the project identification information and narrative description of the training facility along with the detailed budget data. At the end of each month, the paper tape records are used to create input to the Manpower Administration's magnetic tape file of financial records. The tape record is up-dated with the budget information from the Office of Education, so that shortly after project approval information has reached Washington, a record in machine language has been created which captures all estimated budget amounts.

During the course of the project, expenditure data relating to the payment of allowances are received by the Bureau of Employment Security. These data are reported to the Manpower Administration through the medium of punch cards and are entered on the magnetic tape record for the project. Subsequent revisions to the original estimates are also received and used to up-date the financial records. Upon completion of a project, the actual training costs are received from the Office of Education and these figures are also entered into the project record.

While these financial reports are being processed, the local offices of the Employment Service are selecting and referring unemployed persons for enrollment in the training course. For each individual offered training, an MT-101 Characteristics Report is completed. A copy of the form is sent to the training facility in order to eliminate as much duplication of reporting as possible. The original is sent to the Bureau of Labor Statistics' Office of Data Processing which serves as a data-processing service Bureau for the Department of Labor. A clerical review is made of the reports and where necessary appropriate editing is done to facilitate the key-punching. We are currently exploring the possibility of having the key-punching done in the States and receiving duplicate sets of cards or magnetic tape in lieu of the report forms.

After key-punching, the reports are converted to magnetic tape and processed through an editing and screening program to check upon internal consistency and usability. The screening program utilizes the financial records which have previously been established to eliminate the need for reporting identical data. Information such as duration and geographic coding which appear on the financial records are transferred to the trainee records thus establishing a uniformity of common data between systems. The screening operations utilize the financial data by comparing the training occupation reported for a trainee against the occupation authorized in the approval document for the designated project. For this match, the occupation is considered as the same if it has been coded with the same three-digit DOT code. Reports which do not agree with the information on the approval tape, or which appear to have internal consistency errors, are removed from the main stream of reports for review and decision as to their validity. Items of internal inconsistency which will result in removal from the main stream are non-valid social security numbers and dates of birth which appear questionable (enrollees who have been reported as born in years whose last two digits fall between 48 and 99; or in more easily understood terms, those less than 16 or more than 64 years of age). Because of the small number (less than 5% of the total reports) that have been screened out by this check, we have not referred them back to the State agencies for verification, but rather have awaited receipt of the termination notice and post-training report which contain the social security numbers and dates of birth in an effort to utilize the electronic computer to the greatest degree possible in eliminating tiresome clerical operations. The termination and post-training reports will be matched against the file of screened reports on a basis of project and name. Where a match is made, the social security number or date of birth will be taken from the latest report. Another screening of the "corrected" reports will be made to insure that the latest data do not themselves contain questionable items.

One of the important demographic characteristics needed to describe program achievement in reaching the hard-core unemployed is data on the racial composition of those trained. Because of various State laws and the regulations of the Department with respect to the maintenance of such records in the local offices, it was not deemed feasible to request this information on the characteristics report form. However, through use of the computer, we are able to communicate effectively with the Social Security Administration for the interchange of information in machine language. The confidentiality of this information is maintained and only statistical summaries are available in connection with any of the data received from the SSA. (In fulfilling requests from State agencies for magnetic tape copies of the trainee records, the information acquired from SSA is deleted).

When termination and post-training reports (Forms MT-102 and 103) are received, they are key-punched and prepared for entry into the system. The narrative reasons for termination are classified into meaningful categories and an appropriate grouping code is used for summarization purposes. Using the State identification, trainee's name and social security number as the means for identification, the termination and post-training follow-up data are matched against the file of enrollees and a new tape record containing the additional information is created. Thus, for those reports for which the identification on the MT-101, 102 and 103 forms is identical, we have captured on magnetic tape the trainee's personal characteristics, his reason for leaving the program and for completers who were located by the selection and referral office at the time of the three follow-ups, his labor force status and employment experience at three, six and twelve-month intervals after the training course is completed.

This, then, is the theoretical design of the system. How do we actually stand today? A relatively small number of enrollee reports were lost during the first six months of operation of retraining projects while the details of the reporting program were worked out. Summary data for these projects are available through the means of an interim reporting program operated by the BES. For the balance of the institutional training program, enrollee characteristics reporting appears to be quite good. Using administrative data as a baseline, the receipt of MT-101 forms appears to be within the range expected when the time lag for reporting and key-punching is taken into consideration. We are currently working on a control program that will utilize OE reports on starting enrollment to alert us to differences between enrollment reports from the selection and referral offices and those from the training facilities.

We have a slightly different picture with respect to enrollee reporting in the OJT program. Reports that have passed through our screening program are considerably under the number that administrative data indicate have enrolled. However, our latest listing of screened reports indicates a substantial number of individuals with occupation codes that did not agree with those indicated on the approval record for the designated project. Investigation has revealed that many of these reports are for trainees in OJT projects which were incorrectly transcribed as institutional. Clerical review of these screening tapes is currently underway. Some slippage is built into the OJT program, for although the selection and referral of trainees for on-the-job training is generally done by the local employment service office, many of the trainees are already on the folls of the establishment and the purpose of the training is to upgrade the skills of the present work force.

Reporting on the termination report has left room for improvement also. Because of the need for translating the termination reason into a machinable classification system, we have had some processing delays in Washington. In addition, we have encountered some communication difficulties. In checking out one serious case of delinquent reports, we learned that the training facilities in the State has been completing the reports and sending the copy that was ultimately expected to end up in our data-processing office to the local employment service office for transmittal to Washington instead of using the Office of Education channels. This would have been fine as it is being done in several other States, however, in this instance the word never got to the local offices that the schools were expecting them to send the reports to Washington. Result - no reports in Washington.

Three steps are being taken to remedy the problems with respect to the termination notice. First, the report form is being revised to eliminate the need for narrative reporting of the reason for termination by placing the groupings on the form and asking the instructor to check those that are applicable and to independently indicate the major reason. Secondly, with the change in program that was authorized in the last Congress, a trainee may be given several different services - such as basic educational training, pre-vocational counseling and testing, institutional or on-the-job training, etc. These services may be given simultaneously and they may be consecutive. If they are consecutive, a report that the trainee has completed one phase of his training will result in the machine record indicating that he has completed training even though he may still be enrolled in another aspect of the training program which is being given by a different training facility. As the selection and referral office is the only source of information as to the status of a trainee with respect to the total program, a revision in the form that will enable the local office to indicate the trainee's current status with respect to the overall program is also in the works. And, as a third step, a computer program has been developed to screen the enrollment reports for delinquent termination reports based on reported starting dates and duration of training.

Finally, we have the post-training employment report. This report did not become operational until September 1964. Prior to its institution, the BES received summary reports for a project and post-training employment status is available from that report according to a limited number of characteristics' summarizations. No earnings or hours worked data are available from the summary report. Because of the lateness in getting the follow-up report into operation, we were faced with the possibility of having serious voids in our data. In an attempt to limit the void as much as possible, we have requested follow-up reports beyond the 12-month period for trainees who completed during the period March 1963 - August 1963. This has meant a great deal of work for many of the State agencies, but the reports are coming in from most of them and we hope that the first tabulations based on MT-103 reports will be off the computer next month.

Other Data

In addition to the data from program operations, there are two other means of obtaining information concerning program activity. The first is information obtained by our program evaluation staff through interviews with individuals who have been either directly or indirectly connected with the training program. During these discussions, the program staff obtain narrative information and opinions with respect to: (1) various selection and referral criteria for evaluation of the effectiveness of the selection process and the realism present in the development of training prerequisites; (2) identification of problem areas such as time lag, lack of training facilities, etc.; (3) effectiveness in meeting employer's needs both on an overall basis and in relation to specific occupations; and (4) similar types of investigations. The results of these inquiries require evaluation of the opinions expressed and background knowledge picked up in the field. Where interviews with trainees are involved, the evaluation staff uses randomly-selected samples from the trainee universe in the data file.

We also have had contractual relations with outside research organizations to study specific areas and develop evaluations based on their findings. Michigan State University has been studying the economic effects of retraining and hopes to come up with an answer to the frequently asked question - What is the net cost or benefit derived from retraining? Another university is studying the changes in attitudes engendered by the interest shown in trainees and upgrading of their skills.

The data collected in our administrative statistical program describes only those individuals who were selected for training. Mass or even sample reporting of those not selected for training presents nearly impossible problems of definition. If a card in the applicant file was turned over, was the individual considered or not? Would your answer be different if the next card in the file represented an applicant who was selected? We reached a decision early in our program that examination of the selection process with respect to those who were not selected for training would be the type of information best gathered through depth studies. A research project in New Jersey has been established to determine what happens to the non-selectee in specified local offices.

Analysis of the administrative data described in this report can at best result only in a superficial, broad, brush examination of the program. These data can, however, point out areas in which more detailed examinations would be fruitful and also can serve as the sampling universe for these depth studies. Two studies are currently being made by contractors to obtain in depth information concerning two problems which our administrative data can only touch lightly. The first is an analysis of the cause for leaving training prior to completion, and the second is a study of the effectiveness of MDTA training from the point of view of the trainee and also his employer.

Utilization of Data

Through use of computer programming, characteristic summaries are prepared in various arrays--State, region, occupation, etc., and cross-tabulation of various types, such as last regular occupation by the training occupation, are obtained. With the existence of a complete record for each trainee on magnetic tape, the sole restraints upon the variety of sequences and combinations of cross-tabulations are those imposed by imagination and the computer-programming time.

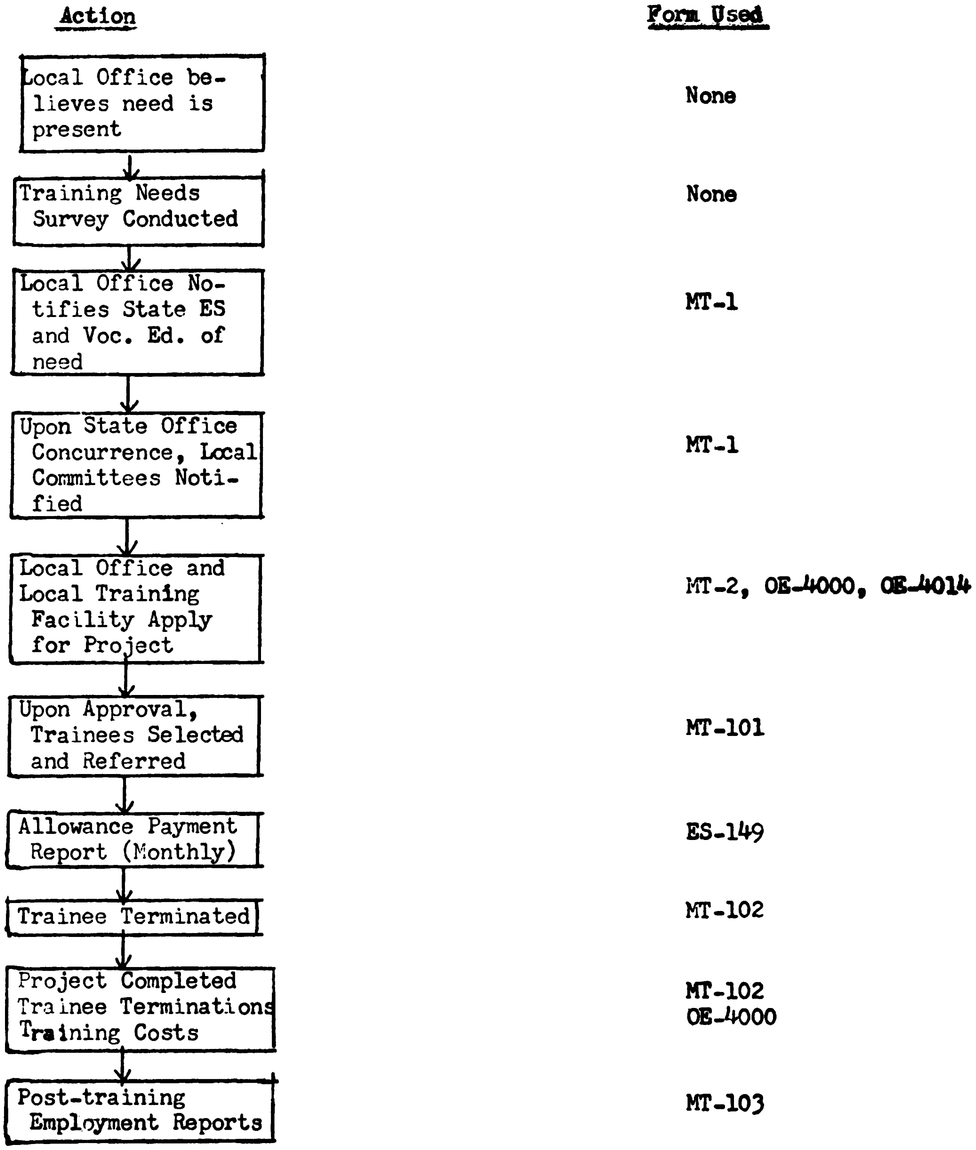
The system lends itself to a form of informational release, too. For example, we are currently establishing two indices which will convert numeric codes into narratives of location and standardized, short, occupational titles. When these indices are established, we expect to be able to produce machine listings of project approvals by State and area on a monthly basis. These listings will be suitable for photo-copy reproduction. Based on actual financial-data records, accuracy is assured.

Depending upon time and staff availability, the reporting system should be capable of meeting the needs of the administrator, evaluator, and researcher. Typical investigations that can be made by the administrator are cost and duration differences involved in identical training between areas; and the effectiveness of local office procedure in the selection process based on analysis of non-completion rates and ability to get training-related jobs. The economist should be able to trace the extent to which the program is providing the means for meeting the changes taking place in America's industrial profile by studying the movement between employing industries and changes in worker's occupations; and he should also be able to relate program achievement in the solution of critical worker shortages and reduction of employment among the hard-core unemployed. Sociologists will find fertile ground for examining the short and long-term effect on trainees' ability to attain employment stability or growth; the general improvement in the skill-level of the working population; and the degree to which the benefits of training are reaching the more disadvantaged groups in our society.

1/ The current form is a revision of the original. In the earlier edition, military status was limited to reporting of the administrative classification of veteran - that is, persons who were entitled to preferential treatment because of their prior military service. The revised form has an additional entry under current labor force status to report separately re-entrants. In addition, the revised form considered the effect of amendment by adding a breakout to indicate trainees with two years of gainful employment and to obtain a limited amount of data concerning the family background of the youth enrolled in projects.

Administrative Statistics on Retraining

Chart I



Action	Form Used
Local Office believes need is present	None
Training Needs Survey Conducted	None
Local Office Notifies State ES and Voc. Ed. of need	MT-1
Upon State Office Concurrence, Local Committees Notified	MT-1
Local Office and Local Training Facility Apply for Project	MT-2, OE-4000, OE-4014
Upon Approval, Trainees Selected and Referred	MT-101
Allowance Payment Report (Monthly)	ES-149
Trainee Terminated	MT-102
Project Completed Trainee Terminations Training Costs	MT-102 OE-4000
Post-training Employment Reports	MT-103

State of ______________________

U. S. DEPARTMENT OF LABOR
FORM MT-1
Bureau of the Budget No. 44-R-1201

NOTIFICATION OF OCCUPATIONAL TRAINING NEEDS UNDER THE MANPOWER DEVELOPMENT AND TRAINING ACT OF 1962

Date: ______________
MT-1 Number: ______________
MT-2 Number: ______________

Labor Market Area: ____
(Name) (Number) (County or Counties)

Local Office: ____
(Name) (Number)

1. Occupation: ____
(Title) (DOT Code)

   a. Description of occupation:

   b. Training objective:

   c. Performance requirements:

2. Estimated number of workers needed over the 12-month period: ______________

3. Estimated potential number of workers available for training: ______________

   a. Characteristics of applicant supply of potential trainees.

4. Wage rate or range of rates at which trainees are expected to be employed ______________________ per ______________________

The undersigned has determined that (1) wage rates, hours, and conditions of work in the jobs for which training is needed conform with standards in the area of employment; (2) the occupational shortage is not due to restrictive hiring practices; and (3) there is a reasonable expectation of suitable employment for ____________ trained persons using skills acquired through training.

Recommended for Training Proposal: ______________________________
Signature, Local Office Manager

Approved: ______________________________
Signature, State Employment Security Agency Administrator

APPLICATION FOR OCCUPATIONAL TRAINING
PROGRAM UNDER THE MANPOWER DEVELOPMENT
AND TRAINING ACT: INSTITUTIONAL

USDL/USDHEW
Form MT-2 (November 1962)

OMAT Project No. ____________

Labor Market Area: Name ______________________________ No. _____ Date ______

Local ES Office: Name ______________________________ No. _____ State ______

State ES Agency: Name ______________________________

Training Agency: ______________________________
Name Address State

Training Facility: ______________________________
Name County Cong. Dist. State

---

1. A. Occupation ______________________________ D.O.T. Code ______

   B. There is reasonable expectation of employment for ______ trainees, of whom ______ are unemployed.

2. Training project information - Summary

A. Length of Course	Proposed	Approved
Total Number Hours		
Hours per Week		
Total Number Weeks		

B. Training Schedule						
Section Number	1	2	3	4	5	6
Beginning Date						
Completion Date						
Number trainees						

3. Estimated MDTA Share of Project Costs -- Summary

Item	Proposed			Project Approval			
	Number of Trnees 1	Cost		Number of Trnees 4	Total Cost 5	Authorization FY 19-- 6	Contingent Next Fiscal Year 7
		Avg. per Trainee 2	Total 3				
Total Estimated Cost	xxx	xxx		xxx			
A. Training (HEW Incl. A.I.)							
1. Local Supervision	xxx	xxx		xxx			
B. Allowances--Total							
1. Training							
a. Family Head							
b. Youth							
2. Subsistence							
3. Transportation							
C. ES Project Admin.	xxx	xxx		xxx			
1. Selection and Placement							
2. Allowance Payment							

4. Attachments (4) - The following attachments are part of the form.
   A. Justification of Training - Form MT-1 and supporting documentation.
   B. Training Budget - Form OE-4000.
   C. Training Plan - See Form OE-4014
   D. State ES agency Project Administrative Costs: Form ES-152

OE-4000 (REV. 7-64)

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
OFFICE OF EDUCATION
WASHINGTON, D. C. 20202

BUDGET BUREAU NO. 51-R400.1
APPROVAL EXPIRES: 8/15/66

COST OF OCCUPATIONAL TRAINING

PROJECT NO.

STATE

(X one) ☐ **MANPOWER DEVELOPMENT AND TRAINING ACT, P. L. 87-415** ☐ **AREA REDEVELOPMENT ACT P. L. 87-27**

NAME AND ADDRESS OF LOCAL PUBLIC TRAINING AGENCY RESPONSIBLE FOR THIS COURSE

NAME AND ADDRESS OF INSTITUTION OR AGENCY (SCHOOL) THAT WILL CARRY OUT THIS TRAINING COURSE

OCCUPATIONAL TITLE

DOT CODE NO.

(X one) ☐ **BUDGET ESTIMATE** ☐ **BUDGET REVISION** ☐ **ACTUAL COST**

(X one) ☐ **INITIAL PROJECT** ☐ **REPEAT PROJECT**
Previous project No. ____________

			AMOUNT REQUESTED OR EXPENDED	*(For State and Federal Use Only)* AMOUNT APPROVED
1.	INSTRUCTIONAL SERVICES	TOTAL →	$	$
	A.	Instructional salaries only		
	B.	Instructional supplies, including shipping costs		
	C.	Rental of instructional equipment		
	D.	Local supervision (where applicable)		
	E.	Guidance and counseling salaries		
	F.	Other allowable items		
		Salaries included in item F above ($)		
2.	FIXED CHARGES	TOTAL →	$	$
	A.	Rental of nonpublic space		
	B.	Employer share of employee benefits		
	C.	Other fixed charges		
3.	EQUIPMENT MAINTENANCE AND REPAIR	TOTAL →	$	$
	A.	Repair and servicing of equipment		
	B.	Other maintenance and repairs		
4.	EQUIPMENT PURCHASE (PUBLIC FACILITIES ONLY)	TOTAL →	$	$
	A.	Initial purchase of instructional equipment		
	B.	Minor equipment, tools, and reference books		
	C.	Minor remodeling of school plant		
	D.	Other capital expenditures		
5.	OTHER COSTS NOT ELSEWHERE CLASSIFIED	TOTAL →	$	$
	A.	Utilities (including telephone)		
	B.	Custodial or janitorial salaries		
	C.	Tuition, fees, or other incidental student charges		
	D.	Trainee transportation		
	E.	Other miscellaneous costs		
6.	**COST OF THIS COURSE** **(Lines 1 through 5)**	TOTAL →	$	$
7.	AMOUNT OF FEDERAL FUNDS INCLUDED IN LINE 6		$	$
8.	AMOUNT OF STATE FUNDS INCLUDED IN LINE 6		$	$
9.	AMOUNT OF LOCAL FUNDS INCLUDED IN LINE 6		$	$
10.	ESTIMATED COST PER TRAINEE HOUR	$		
11.	TOTAL TRAINEES COMPLETING TRAINING	No.		

NAME AND ADDRESS OF LOCAL PUBLIC OFFICIAL RESPONSIBLE FOR ABOVE FUNDS

TITLE

SIGNATURE

COMMISSIONER'S REPRESENTATIVE	DATE	STATE DIRECTOR, VOC. ED.	DATE

OE-4014 (REV. 7-64)

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
OFFICE OF EDUCATION
WASHINGTON, D.C. 20202

BUDGET BUREAU NO. 51-R400.1
APPROVAL EXPIRES 8-15-66

TRAINING PLAN GUIDE

Manpower Development and Training Act (P.L. 87-415)
Area Redevelopment Act (Sec. 16, P.L. 87-27)

The following training plan outline constitutes the minimum information necessary to review the appropriateness of training proposed under the Manpower Development and Training Act or the Area Redevelopment Act. Use code below in preparing the training plan.

1 GENERAL INFORMATION

1.1 Occupational title(s) and DOT Code(s).
1.2 Identify trainees as unemployed or other and ratio, if mixed.
1.3 Total clock hours of training per week, section and project
1.4 Total number of sections.
1.5 Schedule of sections; beginning and ending dates of each; dates of school closures.
1.6 Name of agency to give immediate supervision to this project.
1.7 Name, location, and description of public space to be utilized for training. (When nonpublic space must be leased or rented, describe in budget explanation.)
1.8 List major equipment currently available; identify items previously purchased with MDTA or ARA funds.

2 COURSE INFORMATION

2.1 Topical outline of major units or divisions in course(s); clock hours devoted to each. (Include planned basic education and orientation).
2.2 Educational guidance and counseling services to be provided.
2.3 Instructional material, major texts and references to be used and currently available. (Materials to be purchased should be listed in the budget explanation only.)
2.4 Standards of performance or level of proficency expected at course completion.
2.5 Records to be kept in evaluating trainee progress and achievement.

3 INSTRUCTOR INFORMATION

3.1 Number and titles of instructional personnel, including counselors and full-time equivalency of each.
3.2 Minimum instructor qualifications.
3.3 Hours of employment per week for each instructional staff member.

4 AGREEMENTS

4.1 If articles are constructed during this project explain their disposition. (Such items become the property of the training agency and may be donated to publicly supported institutions but must not enter the commercial market.)
4.2 Verify that employment of personnel and purchase of goods and services will be conducted in accordance with all local, State and Federal laws, requirements, regulations, and policies.

5 BUDGET

5.1 Estimate minimum costs to be incurred in this project on Form OE-4000.

6 BUDGET EXPLANATION

6.1 Estimates of anticipated expenditures in budget are to be explained in detail.

PROGRAM:
MDTA — 1
ARA — 2
Other — 4

TYPE OF ACTIVITY:
Instit. — 1
OJT — 2
E&D — 4
Basic Ed. — 8
Pre-Voc. — 16
Other — 32

CHARACTERISTICS OF TRAINEES
Under the MDTA and the ARA

D/L–D/HEW MT-101 (Rev. 7-64)
Form approved.
Budget Bureau No. 44–R1202.1.

A. 1. State ________________ (Code) 2. L.O. ____________ (No.) 3. Project No. ____________ Section No. ____________

4. Occupation ________________________________ D.O.T. Code ____________

5. Name ____________ (Last) ____________ (First) ____________ (Initial) 6. SSA No. ____________

7. County of residence ________________ ____________ (Code)

8. Date of birth: ____________ (Mo. and year)

9. Sex: Male — 1 Female — 2

10. Handicapped: Yes — 1 No — 2

11. Prior military status:
Veteran — 1
Peacetime service — 2
Rejectee — 3
Other nonvet — 4
Not known — 5

12. Marital status:
Single — 1
Married — 2
Other — 4

13. Primary wage earner:
Yes — 1 No — 2

14. Family status:
Head of family *or*
Head of household:
Yes — 1 No — 2

15. Number of dependents:
— 0 — 2 — 4
— 1 — 3 — 5 and over

B. 1. Highest grade completed:

Code	0	1	2	3	4	5	6
Grade	0	1 2 3 4	5	6 7	8	9 10 11	12

Code College:	7	8	9
Year	1 2 3	4	4+

2. Primary occupation ________________________________ D.O.T. Code ____________
How long worked in *(Months)* ____________ When last worked in *(Month and year)* ____________

C. 1. Did applicant express willingness to accept job out of area? Yes — 1 No — 2

2. Referral to training or services:
Accepted — 1 Refused — 2

3. Enrolled: Yes — 1 No — 2

4. Reason for refusal of referral or failure to enroll:
Obtained employment — 1
Moved from area — 2
Illness (include preg.) — 3
Not interested — 4
Poor location or hours of training — 5
Insufficient allowance for training — 6
Not available (in school, Armed Forces) — 7
No one to look after family — 9
Reason not known — 0
Other *(Specify)* ____________ — 8

5. Was reason considered:
For good cause — 1 Not for good cause — 2

D. 1. At time training offered, applicant was:
a. Underemployed — 0
35–39 hours per week and less than full time — 1
Less than 35 hours per week — 2
Under skill level — 4
Impending technological layoff — 8
b. Reentrant to labor force — 64
c. Unemployed — 16
Weeks unemployed:
Less than 5 — 1
5–14 — 2
15–26 — 3
27–52 — 4
Over 52 — 5
d. Farm worker — 32

2. Years of gainful employment:
Under 2 — 0 2 — 1 3–9 — 2 10 or more — 3

3. Unemployment insurance status: Claimant — 1 Nonclaimant — 2

4. Public assistance status: Recipient — 1 Nonrecipient — 2

5. As defined for reimbursement of training costs, applicant is:
Unemployed — 1 Other — 2

6. Last regular employment:
a. Occupation ________________ D.O.T. Code ________
b. Industry ________________ S.I.C. ________
c. Straight-time average hourly earnings $________

E. Eligible for allowance *(Not applicable for ARA)*:
1. Regular training:
Yes — 1 Augmented — 3 No — 2
2. Youth training:
Yes — 1 No — 2
3. Subsistence-transportation:
Yes — 1 No — 2

F. For youth:
1. What was the most important reason for your leaving school? *(Check only one)*
Graduated from 12th grade — 0
Illness — 1
Had to support self — 2
Had to support family — 3
Preferred work to school — 4
Because of low marks in school — 5
Had to work on family farm or in family business — 6
Trouble with teachers or school authorities — 7
Marriage or pregnancy — 8
Other — 9
(Specify) ________________

2. Living with parents (either own or spouse's)? Yes — 1 No — 2

3. Highest grade of regular school father ever completed?

Code	0	1	2	3	4	5	6
Grade	0	1 2 3 4	5	6 7	8	9 10 11	12

Code College:	7	8	9
Year	1 2 3	4	4+

(Data Processing Office Copy)

☆ U.S. GOVERNMENT PRINTING OFFICE . 1964—O-739-998

U. S. Department of Labor
Bureau of Employment Security
Form ES-149 (Rev. 4/3/64)

MANPOWER DEVELOPMENT AND TRAINING ACT OF 1962
REPORT OF TRAINING, SUBSISTENCE AND TRANSPORTATION ALLOWANCE, OBLIGATIONS AND EXPENDITURES BY TRAINING PROJECT

Fiscal Year Appropriation
During Month of ____________

MDTA Project No. 1/	No. of Paid Trainees	Number of Payments					Amount Paid as Youth 2/ Allow.	Amounts Paid					Project Ceiling Adjustment		
		No. of Paid Youth	Reimb. Payments	Direct Train. Allow.	Subsistence	Transportation		Direct Training	Reimb. State UI	Subsistence	Transportation	Total	Current Project Ceiling		Revised Proj
													Current FY	Subsequent FY	Current FY
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Sub-Tot.															
Reimbursable Interstate Payments															
TOTALS							$	$	$	$	$	$	$	$	$

1/ Indicate by an asterisk (*) any project completed during the month (not section of projects)

2/ Memorandum entry only. Include in Columns (9) or (10) and (13).

Date ____________, 19 __

____
(Signed)

____
(Title)

Facsimile

MDTA ☐ 1 Instit. ☐ 1
ARA ☐ 2 OJT ☐ 2
Other ☐ 3 Other ☐ 3

INDIVIDUAL TRAINEE
TERMINATION OF TRAINING

DL/DHEW—MT-102
Bud. Bur. No. 44-R1204
Expires: 9/30/63

A. 1. Name: ______ *(Last)* *(First)* *(Initial)* 3. SSA No. ______

Date of Birth ______ *(Month)* *(Day)* *(Year)*

2. Address ______ *(Number and Street)* *(City)* *(Zone)* *(State)*

B. 1. State ______ *(Name and Code)* 5. Date section began ______ 7. No. days attended ______

2. Project No. ______ 6. Date trainee terminated ______ 8. No. days absent ______

3. Section No. ______

4. Occupation for which training conducted ______ *(D.O.T. Title)* D.O.T. Code ______

C. Nature of Termination:

1. Completed training ☐ 0

2. Did not complete training:

a. Involuntary termination
- For poor attendance ☐ 1
- For lack of progress ☐ 2
- Other *(Specify)* ______ ☐ 4

b. Voluntary termination
- For training—related job ☐ 7
- For nontraining—related job ☐ 8
- Other *(Specify)* ______ ☐ 9

D. Background information concerning termination *(Do not complete if item C-1 is checked):* ______

E. For the training facility: Date ______

This is to certify that the circumstances of termination for the trainee to whom this report refers have been determined to be:

for good cause ☐ 1. not for good cause ☐ 2.

Name *(Signature)* ______

(Typed or printed) ______

Title *(Instructor or supervisor)* ______

Facility Name ______

Address ______ *(Number and Street)* *(City)* *(Zone)* *(State)*

USED ONLY WHERE GOOD CAUSE IS NOT SHOWN

F. For agency or organization responsible for training: Date ______

I have reviewed the circumstances surrounding the termination of the trainee to which this report refers and have found them to be accurately described.

Name *(Signature)* ______

(Typed or printed) ______

Title *(Agency head)* ______

Agency Name ______

Address ______

DL/HEW FORM MT-103
MDTA ...... 1 Instit. ...... 1 Basic Ed....... 8
ARA ...... 2 OJT ...... 2 Pre.Voc.......16
Other ...... 3 E and D ...... 4 Other 32

POST TRAINING REPORT

Bud. Bur. No. 44-R1246
Expires 6/30/65

State ______________ (Name) ______________ (Code)
Project Number ______________
Section Number ______________

A. IDENTIFICATION

1. Name ______________ (Last) ______________ (First) ______________ (Initial) 2. S.S. No. ______________

3. Date of Birth ______________ (Month/Year) 4. Occupation For Which Trained ______________ D.O.T. Code ______________

5. Report Number ______ 1 ______ 2 ______ 3 For reference week ending ______________ (Month/Day/Year)

6. Source of Data: Trainee or Trainee's Family ____ (1); MT-103a ____ (2); L.O. Records ____ (3); Could Not Locate Trainee ____ (4)

B. STATUS

1. Work History Since Training
   - a. Total Weeks Since Training ______________
     - (1) Weeks Totally Unemployed ______________
     - (2) Weeks In Which Employed ______________
     - (1) and (2) should add to total
   - b. Number of Jobs Lasting 30 Days or More Since Training ______________
     - (1) Training Related Jobs ______________
     - (2) Non-Training Related Jobs ______________
     - (3) Information Not Known ______________
   - c. Number of Placements Through ES ______________

2. Current Labor Force Status
   - a. Employed ______________ 1
   - b. Unemployed ______________ 2
   - c. Not in Labor Force:
     - Keeping House ____3 Illness ____5
     - In School ____4 Other ____6
     - (Explain other) ______________

3. If Employed In Reference Week (if B2a is Checked)
   - a. Employer's Industry ______________ SIC Code ______________
   - b. Trainee's Occupation ______________
     ______________ D.O.T. Code ______________
   - c. Hours Worked In Reference Week
     - Less than 15 ____1
     - 15-34 ____2
     - 35 or more ____3
   - d. Straight-Time Average Hourly Earnings $ ________ per hr. (Excluding Overtime)
   - e. Is Job Training Related? Yes__1 No__2
     - (1) Was Training Useful in Obtaining Job Yes__1 No__2
   - f. Job Is Expected to Last 30 Days or More Yes__1 No__2
   - g. Trainee Had to Move 50 Miles or More to Take Job Yes__1 No__2

During Last Month

If Other than Zero

(a) Has Individual Turned Down Offer of a Training Related Job
Yes __1 No __0

(b) If Yes, Check 1 or More Reasons
Hours Undesirable __1
Pay Below Normal for Occupation __2
Couldn't Afford to Move __4
Unwilling to Move __8
Other ______________ __16
(Explain) ______________

If Other than Zero

(a) Reason for Leaving Last Job Lasting 30 Days or More
Did Not Leave a Job __0
Slack Work __1
Plant Shut Down __2
Illness __3
Other __4
Unknown __9

If Checked

(a) Job Obtained Through:
ES Office __1
Establishment Where Trained __2
School __3
Other __4
Not Known __9

If Checked

(1) Is Individual Waiting to Report to a Job in the Next 30 Days
Yes __1
No __0
If Yes, Is It Training Related
Yes __2
No __3
Not Known __4

If Checked

(1) Hours Worked Were
Normal For Industry, Area or Season __1
Part-Time
Employee's Choice __2
For Economic Reasons __3
Other ______________ (Explain) __4

Not Known __9

GPO 878-361

D/L FORM MT—103a

Budget Bureau No. 44-R-1246
Expires 6/30/65

WORK EXPERIENCE SINCE TRAINING

REFERENCE WEEK ____________ TO ____________

Trainee's name ______________________________

and address ______________________________

____
(please correct if wrong address)

Please complete form by ______________________________

and return to:
(agency name
and address)

Please enter a telephone number where you can be reached ____________

Please read carefully. If you check an answer with an arrow, please follow arrow before answering next question.

1. a. About how many weeks is it since you completed training? ____________
   b. In how many of these weeks did you do no work for pay? ____________
   c. In how many weeks have you worked since training? ____________

2. Since training, have you had any one job which lasted 30 days or more? Yes ☐ No Skip to ☐ question 3
   a. How many jobs lasted 30 days or more? ____________
   b. In how many jobs did you use your training? ____________
   c. Please check or explain reason for leaving last job that lasted 30 days or more
      Did not leave a job ☐ Illness ☐
      Slack work ☐ Didn't like job ☐
      Plant shutdown ☐ Other (explain) ☐
      ____

3. Did you work at any time during reference week specified above? Yes ☐ No ☐ Please answer questions in next column

4. During most of the week were you looking for a job? ☐
   Not looking for a job because you
   Were keeping house ☐
   Were in school ☐
   Were ill ☐
   Had a promise of a job ☐
   Other ____________ (Please specify) ☐
   a. For whom do you expect to work? ____________ (company name)
   ____
   b. Do you expect to start within 30 days? Yes ☐ No ☐
   c. Do you expect to use some of your training? Yes ☐ No ☐ Don't know ☐

GPO 878-360

5. For whom did you work?
   ____
   (name of company)
   ____
   (address)
   ____
   (plant, division or department)

6. What did you mainly do on the job?
   ____
   ____

7. How many hours did you work?
   less than 15 ☐ 15-34 ☐ 35 or more ☐ (skip to question 8)
   Were these hours:
   usual for this kind of work ☐
   all you wanted ☐
   short-term because of slack work ☐
   other reason ____________ (please explain) ☐
   ____

8. How much are your average hourly earnings? (excluding overtime) $________ per hour

9. Did you use your training at least part of the time? Yes ☐ No ☐
   a. Do you think your training helped you get this job? Yes ☐ No ☐

10. Have you worked or do you expect to work at least 30 days on this job? Yes ☐ No ☐

11. Did you have to move 50 miles or more in order to take this job? Yes ☐ No ☐

12. Please indicate where you heard about this job
    local employment service office ☐
    in the place where you took training on-the-job ☐
    school ☐
    other ____________ ☐
    ____
    (Explain)

BLUE TO WHITE COLLAR JOB MOBILITY: A PRELIMINARY REPORT[1]

James L. Stern and David B. Johnson, University of Wisconsin[2]

Introduction

In the past few years considerable publicity has been focused on the growth of white collar occupations relative to blue collar occupations. Between 1953 and 1963 blue collar employment remained at about the 25 million level. During the same period the number of white collar jobs increased from 23.6 million to 30.2 million, a gain of 28 percent. The largest group of blue collar workers, production workers in manufacturing industries, declined from the 1953 peak of 14.1 million to 12.6 million in 1963. At the same time, however, the civilian labor force increased by over 9 million people.[3]

Reflection upon these aggregate trends raised questions about how many of these new white collar jobs were filled by experienced blue collar workers and how many were filled by new entrants to the labor force. In view of the extensive changes in the composition of our labor force, are the occupational mobility patterns of the 1930's and 1940's still valid? Also, to what degree does the growing participation of women in the labor force, presumably in the traditional female white collar jobs for the most part, color the picture (no pun intended)?

Initial explorations into the literature revealed that little work had been done on the questions of the kind of white collar jobs into which blue collar workers have shifted in recent years and the number of people who have made this type of occupational move. Our particular interest, arising in part from our observations of this phenomenon as it relates to developments in the field of collective bargaining, was in the extent and direction of movement by male blue collar workers into white collar jobs. By extent we mean the numbers of such workers who, for example in manufacturing, leave the ranks of production workers and become technicians or supervisors. By direction we mean the different kinds of white collar occupations into which the blue collar workers shift.

A logical consequence of our interest in the extent and direction of blue to white collar job mobility was our increasing curiosity about the process by which such a shift was accomplished. Was the labor market working in such a way that the invisible hand gathered the required number of the appropriately qualified blue collar workers into the different types of white collar jobs? Or was it a case of a person being in the right place at the right time? Were there particular skills that enable people to make this shift easily; were there job paths that led logically from blue to white collar jobs? Did these shifts take place mainly within a firm or across establishment lines and even on an interindustry basis?

If some firms fill some of the white collar jobs with men having blue collar backgrounds, what educational, personal or skill characteristics do such men have that would explain why they were hired? Or, is the shift one of those that are explained by the cliché--It's not what you know but who you know? It should be noted that the shift may be possible in one firm and impossible in another, without regard to the attributes of the individual job seeker, because of the personnel policies and practices that prevail. Other potentially important considerations include the placement procedures of public and private vocational schools and employment services.

The enactment of the Manpower Development and Training Act provided funds which enabled us to inquire systematically into many of the questions about which we were curious, the paths by which people make such shifts, and the ways in which mobility from declining to expanding occupations could be improved. Our first step in this project was to obtain the views of the Wisconsin State Employment Service[4] and personnel directors of several large firms about the kinds of white collar jobs into which blue collar workers were moving. From our extended conversations with them and with OMAT and other Department of Labor personnel we arrived at a tentative list of eight broad occupational groupings which we believed would cover the major blue to white collar job shifts. Appendix A contains a description of

---

[1] The research reported herein was financed through the research program of the Office of Manpower, Automation and Training (OMAT), United States Department of Labor, and supplemented by support from the Graduate School of the University of Wisconsin. To each and all of the individuals and agencies that have helped in the various stages of this project we wish to express our appreciation.

[2] The authors are Professors of Economics at the University of Wisconsin. They were assisted in this project by Industrial Relations Research Center Research Assistants Jan Staggs, Eric Langbaum, Kurt Schneidmiller, Gary Parson, Lloyd David, and Barbara Klees, by Project Assistant Carol Cary, and by Professors Leonard Berkowitz and Vernon Allen of the Psychology Department, University of Wisconsin.

[3] Data taken from the Manpower Report of the President, March, 1964. Table A-7, p. 199; Table C-2, p. 227; Table A-1, p. 195.

[4] Mr. Edward Kehl, Chief of the Research and Statistics Section of the WSES and members of his staff were particularly helpful throughout the development and execution of this project.

these groupings and a list of common job titles found within each group.

Despite the continued growth in employment of service workers outside of private households, service occupations were consciously omitted from this initial list. We thought that these occupations were not ones for which extended training was necessary. For the most part they are low-paying occupations and many shifts into them probably represent unfortunate examples of downward economic mobility. Also, much of the growth of these occupations is among females rather than males and we had decided to confine our inquiry to patterns among the men.

In compiling our occupational list, however, and in the process of checking these lists against census figures on occupational growth, we became aware of how uncertain we were about probable mobility patterns and realized further that unless we made some systematic inquiry and developed a statistical framework within which to assess these shifts, we would face difficult problems at a later date when we attempted to ascertain the relative importance of varying trends. We might find that we had omitted several important white collar groups into which blue collar workers were shifting in large numbers. Even if this were not the case and we simply selected an equal number of blue collar males who had shifted into each of the eight groups we had chosen, we would have no basis for weighting one group against the other because we would have no idea of the number of blue collar males shifting into each of these occupations. Over and above these considerations was the question of how we would obtain the names of 500 blue collar males who had shifted recently into white collar jobs. Where should we look to find these people? How could we draw a representative sample from this unknown universe? The solution to these questions was apparent but time-consuming. Before we could study the characteristics of blue collar workers who had shifted to white collar jobs we needed first to make a survey to determine the extent and direction of these shifts and the relationship of these shifts to employment activity in the white collar market generally.

A questionnaire, Appendix B, was sent to a 10% sample of employing units in Milwaukee County covered by unemployment compensation. Each tenth firm, starting with a random number, was selected from the March, 1962 unemployment compensation register containing approximately 9500 companies listed by descending order of size within industries as defined by the Standard Industrial Classification system. Essentially each firm was asked to list the titles of male white collar jobs it filled in 1961 and 1962, the number of men hired or promoted into each of these jobs, the number of men who had a blue collar background and the total number of men in the occupation as of the end of 1962. The instructions stated that "White collar jobs include professional, technical, managerial, sales, clerical and kindred service jobs. Generally the person is paid a weekly or monthly salary. Blue collar jobs include unskilled and semi-skilled workers, craftsmen and manual workers who are paid by the hour." We identified men with a blue collar background as men "who subsequent to 1/1/56 have had at least 1 year in blue collar job."

Three mailings were made, followed in turn by several hundred phone calls. The final response rate was approximately 70 percent. Comparison of responses and non-responses showed no statistical bias by size of firm or type of industry. We recognized that absence of statistical bias insofar as size of firm or industrial distribution of respondents in no way guaranteed the accuracy of the figures that were supplied. In some instances our intuition suggested that small firms might have answered the questionnaire informally--that is, mentally counting up that Joe, Jim and Pete were hired as salesmen in the last two years, that Pete had worked as a blue collar worker and that they now had about 14 salesmen. Then, after sending back the reply, the office manager who filled it out might have realized that he had been hired during that period and that his job title and the pertinent figures should have been included in the report. Although this illustrative example is based on an actual event and the error was easily corrected, it brought home to us the difficulties of acquiring data in this fashion. In large firms where compliance with our request necessitated a search of hundreds of personnel jackets, we assume that errors were made and that probably there was some understatement of hiring activity. This was avoided in situations where university personnel were used to make the search under the supervision of the company. In one case we encountered an exaggeration of the extent of the shift that gave us the impression that the employee making the search was on piecework. Actually, he wasn't, but he was being paid overtime to come in on a Saturday to do this task and apparently wanted to give us our money's worth. More seriously though, this method of collecting occupational data faced the perennial question of the meaning of job titles. Each title such as draftsman, lab technician and clerk probably covered a wide range of skills and duties because it had a different meaning in different firms. Despite these considerations, however, we believe that the extent and direction survey was most worthwhile and in our case served its dual purposes well.

But, before going on, I think you might be amused to hear about several of the unusual responses to our extent and direction study. One company president sent us a long letter in which he stated that he was ashamed of the University of Wisconsin for cooperating with the United States Government in this project. He obviously thought this a most outrageous waste of time and money and was good enough to tell us that the road to white collar success was open to the blue collar worker who skipped his coffee break and did not join a union. Although he declined to participate in the survey initially, his firm eventually responded to the persistent and guileful efforts of our research assistants and cooperated in supplying names for the sample. Considerations of confidentiality preclude us from publishing whether the blue to white collar shifters at that firm resemble in any way the stereotype suggested by their company president.

A second company president thought very highly of the survey, sent us many articles on the same subject including several he had written, but forgot to return the questionnaire until gently reminded by a second letter. And finally, there was the owner of a tavern in the Negro section of Milwaukee who said that our blue to white collar job shift questionnaire did not apply to him as he "had no white color employees." We are not sure whether this is a case of misspelling or misunderstanding.

Returning to the question at hand, we should note why we think that the extent and direction survey of employing establishments was the most satisfactory approach for our purposes. First of all, blue to white collar shifters as we have defined them make up an extremely small percentage of the work force. We debated for some time whether it would be feasible to construct a statistical frame from which a random sample could be drawn directly. We recognized that although this could be done and should be done on a national basis at some future date, it would be far too costly for this experimental project. Instead, as already mentioned, we drew our extent and direction sample from a 10% cross section of Milwaukee firms covered by unemployment compensation.

Replies were received from 70% of the establishments employing in excess of 26,000 people. Among this group we were told that 1318, or 5 percent, met our first criterion of being male entries into white collar jobs during 1961 and 1962 and that only 210 of these men, or less than 1 percent of the employees covered by the replies, had the necessary year or more of blue collar experiences since 1955. Keep in mind that it was from this less than 1 percent who met our definition of a blue to white collar job shift we were going to draw the 500 men who would be interviewed in our effort to ascertain the personal, educational and occupational characteristics of men who made this particular type of job shift.

Prior to drawing our sample we made an educated guess with the help of experienced guessers from the employment service and the university about what percent of the Milwaukee County labor force fit our definition of a white collar worker with a blue collar background. Because we thought that this group would constitute a very small portion of the working population and because of the possibility of a high non-response rate, we estimated that we should supplement the 10 percent sample with an additional sample, bringing the total coverage up to over half of the workers covered by unemployment compensation. We did this not for the purpose of refining our measure of the extent and direction of blue to white collar shifts, but to garner enough names to provide us with our desired sample size. That we succeeded in this effort is due more to those fortuitous offsetting estimates that make statistical analysis such a pleasant pastime, rather than to the accuracy of our original guesses.

Extent and Direction of Blue to White Collar Job Shifts by Industry and Occupation

Turning now to the results of our extent and direction survey, a weighted projection of the sample indicates that about 22 percent of the men hired or promoted into white collar jobs in Milwaukee County in 1961 and 1962 had a year or more of blue collar experience since 1955. Subject to the statistical limitations explained in Appendix C,[5] we find that about 4000 of the 17,900 male entries into white collar employment were made by men with blue collar backgrounds.[6]

TABLE I *

ENTRIES INTO WHITE COLLAR JOBS AND PROPORTION OF ENTRIES THAT SHIFTED FROM BLUE COLLAR JOBS

By Major Occupational Group in Milwaukee County

OCCUPATIONAL GROUPS	NO. OF ENTRIES	NO. OF SHIFTS	(SHIFTS/ENTIRES) % of Entries With Blue Collar Background	% DISTRIBUTION of	
				Entries	Shifts
Managerial	3,606	1,556	43.2%	20.2%	39.0%
Clerical	3,104	827	26.6%	17.4%	20.7%
Professional & Technical	6,127	945	15.4%	34.3%	23.7%
Sales	5,037	661	13.1%	28.2%	16.6%
TOTAL	17,874	3,989	22.3%	100.1%	100.0%

*The figures in this table should be regarded as approximations subject to the statistical limitations explained in Appendix C.

---

[5] We wish to thank Miss Estelle Goldberg of the Wisconsin Survey Research Laboratory for preparing this Appendix section for us.

[6] The various measures of extent and direction were prepared by Jan Staggs. Further detailed information is contained in his unpublished thesis, "A Measure of Blue to White Collar Mobility," University of Wisconsin, 1964.

When this 22 percent figure is disaggregated by major occupational groups we see that the percentage of shifts to entries in descending order varies from a high of 43 percent in managerial occupations, to 27 percent in clerical occupations, to 15 percent in professional and technical occupations, and to a low of 13 percent in sales occupations. The impact of the high ratio of shifts to entries in managerial occupations is offset in part, however, by the low growth rate of jobs in this field relative to opportunities in the other three white collar occupational categories. In Table I we see that the distribution of entries into white collar jobs ranges from a high of 34 percent into professional and technical occupations, to 28 percent into sales, to 20 percent into managerial jobs and 17 percent into clerical positions. When these factors are combined and we look at the distribution of the shifts by major occupational groups; that is, when we look at the occupational composition of the 22 percent of the total entries who shifted from blue to white collar jobs, we find that managerial occupations still offer the greatest opportunities, 30 percent, but that the opportunities in professional and technical occupations of 24 percent exceed the 21 percent found in the clerical sector and the 17 percent found in sales.

The breakdown of the entries and shifts by major industrial categories, Table II, shows that the ratio of shifts to entries is high in transportation, communication and public utilities, 59 percent, and construction, 44 percent; medium in wholesale trade, 28 percent, manufacturing, 27 percent, retail trade, 24 percent, and in service industries, 19 percent; and low in government, 8 percent, and finance, insurance and real estate, 4 percent. Unfortunately, as was the case in the breakdown by occupations, the breakdown by industries shows the greatest ratio of shifts to entries in the fields where the growth is the lowest and vice versa.

In construction and transportation, where we found a high ratio of shifts to entries, we found also a low percentage of total entries. The low number of entries reflects the combined effect of growth rates, size and turnover rates of these industrial sectors compared to other

TABLE II*

ENTRIES INTO WHITE COLLAR JOBS AND PROPORTION OF ENTRIES THAT SHIFTED FROM BLUE COLLAR JOBS
By Major Industrial Categories in Milwaukee County

INDUSTRY	NO. OF ENTRIES	NO. OF SHIFTS	(SHIFTS/ENTRIES) % of Entries With Blue Collar Background	% DISTRIBUTION of Entries	Shifts
Transportation, Communications & Public Utilities	775	459	59.3%	4.3%	11.5%
Agriculture, Mining & Constr.	623	274	44.0%	3.5%	6.9%
Wholesale Trade	802	224	27.9%	4.5%	5.6%
Manufacturing	5,134	1,389	27.1%	28.7%	34.8%
Retail Trade	2,306	550	23.9%	12.9%	13.8%
Services	4,763	908	19.0%	26.6%	22.7%
Government	1,027	88	8.3%	5.7%	2.2%
Finance, Insurance & Real Estate	2,444	97	4.0%	13.7%	2.4%
TOTAL	17,874	3,989	22.3%	99.9%	99.9%

* The figures in this table should be regarded as approximations subject to the statistical limitations explained in Appendix C.

sectors. Between 1950 and 1960, Milwaukee County employment in both of these industries suffered declines of about 11 percent. By 1960 these two sectors included only 14 percent of the male employment. The net result of these trends shows up in our sample where despite the 60 percent and 44 percent ratios of shifts to entries in transportation and construction respectively, entries into these fields were only 8 percent of total entries. This reduced the percentage of total shifts in these sectors to 18 percent. Although the negative growth rate of construction employment in Milwaukee during the decade of the 1950's differs somewhat from national trends, the general occupational trends in Milwaukee County during the decade are reasonably representative of changes to be found nationally. This can be seen by reference to the tables of Appendix D showing Milwaukee and United States male employment changes between 1950-1960 by occupation and industry.

A further analysis of the distribution of entries by major industrial sectors in Table II shows that manufacturing accounted for about 1/4 of the entries, the service industries accounted for another 1/4, finance and retail trade made up the third quarter of the entries, and the last quarter was divided almost evenly among the four remaining industrial sectors. The picture was changed, however, by the varying ratios of shifts to entries. We found that manufacturing was the most important sector with 35 percent of the blue to white collar shifts, service industries were second, 23 percent, retail trade and transportation were next with 14 percent and 12 percent respectively, then construction and wholesale trade, 7 percent and 6 percent, and finance and government last with 2 percent apiece.

Comparisons of variations in the volume of blue to white collar shifts in different industrial sectors with variations in the major occupational groups raise the question of the degree to which the variations by industry are attributable to the differences in occupational composition of an industry. For example, we found that very few entries into the elementary and secondary school teaching profession had blue collar backgrounds. Therefore, since the large and expanding occupation of teaching is confined mainly to the government industry, the ratio of shifts to entries in this industry is depressed.

On the other hand, the high ratio of shifts to entries in the transportation industry flows in part from the fact that three out of four entries into the dispatcher occupation had a blue collar background. This occupation, which had the largest percentage growth of any male clerical group in Milwaukee between 1950 and 1960, is heavily concentrated in the transportation industry. Examples such as these confirmed our notion that the basic explanation of the overall blue to white collar mobility pattern was to be found in the factors determining the ratio of shifts to entries in each occupation.

This brings us to another consideration raised in the discussions we had with government and industry leaders prior to doing this survey. It was suggested that determinants of entry to an occupation may vary widely among firms according to the size of the firm and the different customs of each industry. Although the information on these characteristics derived from our extent and direction survey is limited, several interesting points were noted.

Among draftsmen, for example, we find 13 firms hiring 44 men. Six of these 13 firms hired men with blue collar backgrounds, seven did not. Among the six that did, one firm hired four draftsmen, all of whom had blue collar backgrounds. The other five firms hired 29 draftsmen, eight of whom had blue collar backgrounds. Among the seven firms that did not, we find four firms in manufacturing that hired no workers with blue collar backgrounds into drafting or any other white collar jobs along with three firms that had blue collar shifts to other white collar jobs but not to drafting. These seven firms with no shifts in the drafting occupation hired 11 of the 44 draftsmen in the 10 percent sample. Presumably our typical blue collar man who became a draftsman did so in part because of his education, personal characteristics and skill, but also did or did not become a draftsman because of the personnel policies of the particular firm that was hiring draftsmen.[7]

Although small firms presumably differ from large firms in many ways, including how promotions are made, we found little evidence on the whole that differences by size of firms had a pronounced net effect on the possibility of blue to white collar job shifts one way or the other. The ratio of shifts to entries was a little higher both in firms with less than 10 employees and in firms with between 51 and 100 employees than the lower but similar level found in firms with 21-50 employees and very large firms with over 100 employees. Appendix Table E shows the ratio of entries and shifts to number of employees by size of firm.

Two Insights Gained from the Extent and Direction Survey

Although the basic purposes of the extent and direction survey were to obtain measures that would enable us to identify the relative importance of various blue to white collar paths and to obtain the names of people who had made these shifts, we gained insights into two other points

---

[7] During the pre-testing of the schedule used in the interviews of 500 men who had shifted from blue to white collar jobs we encountered a draftsman whose explanation of the shift illustrates one of the many ways in which the labor market works. He said that he went to vocational school at night on his own but was unable to get a drafting position. Instead, he continued his work in the shipping room. However, one day, when his load was light, he filled out tags in the impeccable printing style associated with mechanical drawing. One such tag was noted by a front office executive who went out to the shipping room to investigate the phenomenon. After talking to the potential draftsman, he informed the personnel office tht it should hire this man for the next drafting opening. And so, this is how one man became a draftsman.

which appear important to us. The first of these points is that the occupation of "foreman" should be considered a white collar job rather than a blue collar and that there is considerable mobility of blue collar workers into it. The second is that there is an internal labor market within the firm with standards that differ from those of the external labor market. The internal labor market may operate in a manner that promotes blue to white collar mobility in instances where the external market does not. Let us take the foreman question first.

In our questionnaire to firms we did not mention foremen either as blue or white collar workers. The definition of white collar worker quoted earlier did not include the term foreman. The definition of blue collar worker also did not; in fact, we specifically did not mention foremen along with skilled workers. It is our impression that most of the respondents outside of the construction industry considered foremen to be white collar workers and listed the foreman occupation as a white collar job. Insofar as prevailing sentiment in industry is a relevant guide for classification, the Census Bureau and others concerned with this question should review their longstanding practice of classifying foremen as blue collar workers.

In 1947 the National Labor Relations Act was amended to define foremen as supervisors and to exclude them from coverage under the Act. The small, weak foremen's union was destroyed and at the same time in many large companies steps were taken to encourage foremen to identify themselves more closely with other components of the management group. In some cases, white smocks were given to foremen as badges of identification. Where foremen had punched the same time clock and used the same locker room as production workers, they were taken off the time clock or at least given a separate time clock of their own, as well as a separate locker room. Over the past decade foremen have been covered by the fringe benefit programs designed for salaried workers rather than the alternative programs for hourly production workers.

The monthly Bureau of Labor Statistics releases about employment changes in non-agricultural industries include the two familiar tables, one covering all employees and the other production workers in manufacturing. Employees excluded from the production worker category are usually regarded as white collar workers. Foremen are excluded and as such would fall into the white collar category. The 1960 census data supply another reason why it makes sense to treat foremen as white collar workers. It is apparent that employment trends for foremen are more like those of white collar workers than those of blue collar workers. For example, in manufacturing, production worker employment did not increase between 1950 and 1960 while the number of foremen in manufacturing increased by 45 percent.[8]

The data from our extent and direction survey reinforce our opinion that foremen should be classified as white collar workers. All of the firms that reported hiring foremen also reported that there was at least one shift; that is, at least one of the new entries into the foreman job had a blue collar background. This was the only occupation for which this was true. The ratio of shifts to entries among foremen was about 80 percent. Some details are supplied in Appendix Table F. No mention was made of hiring foremen in the industries of finance, trade and construction. In finance and trade we find that the term is not used but that instead a person with equivalent duties is called a supervisor, and as such is classified as a managerial occupation. In construction, foremen are considered to be blue collar workers and are also so regarded under our definition because they are paid by the hour. Typically the construction foreman is paid about 25¢ an hour more than the journeyman in his trade. Also, his tenure as foreman is usually confined to that one job and on his next assignment for another construction firm he again may be working as a journeyman.

Next we would like to turn briefly to this question of an internal labor market as opposed to the various aspects of the conventional external labor market. We first became aware of the difference between the occupational paths of the internal labor market and those of the external labor market when experienced staff of the Wisconsin State Employment Service (WSES) and experienced personnel men from large manufacturing plants described quite different patterns when speculating about how men made blue to white collar shifts. The WSES noted two types of movement: (1) horizontal: a lateral movement from a given occupation to an entirely different occupation in a new field of work; and (2) vertical: shifts of workers to jobs based on acquisition of new skills and/or knowledge coupled with previous skills and/or knowledge. Shifts from blue to white collar jobs could involve either lateral or vertical shifts or both as men shifted into managerial, clerical, sales, or professional occupations. The WSES believed that there was very little horizontal or vertical movement by blue collar workers into clerical jobs. The industry personnel men said that blue collar workers frequently made the transition into white collar clerical jobs, particularly in the processing occupations such as routing and scheduling.

The employment service noted that shifts from blue collar to technical jobs in the processing field still existed and mentioned, in passing, jobs such as production planning and production routing. The industry personnel men agreed that this was so but added also the clerical jobs of dispatcher, expeditor and production scheduler. In part the difference may arise because a job considered clerical by the personnel man in industry may be classified among the technical

---

[8] Production worker employment in manufacturing in 1950 was 12,523,000 and in 1960 was 12,586,000. The number of foremen in manufacturing increased from 513,473 in 1950 to 744,011 in 1960. Data from the Manpower Report of the President, March 1963, Table C-2, p. 164 and Table G-6, p. 202.

occupations by the employment service. But much more fundamental than this is the fact that the clerical jobs in the processing field are in actuality the entry jobs into the technical positions in this field. If the WSES is correct, part of the reason that there is little vertical movement into technical and higher level clerical jobs is that there is little horizontal movement into the clerical jobs which are in turn the entry jobs to the technical processing occupations.

But if industry personnel men claim that there is considerable occupational movement from blue to white collar clerical jobs of the nature that WSES would classify as horizontal moves, how then is the difference explained? As most of you have surmised already, we believe that the difference arises because the industry personnel men are talking about the internal labor market while the employment service representatives are talking about the external labor market. A limited investigation was made of this possibility and it appears to be the case. A young semi-skilled male with a high school education and a few years in the shop who impresses his supervisor with his diligence and regular attendance frequently is given the opportunity to shift to a clerical position. At the time of a layoff from a production job he may be offered a horizontal move to a clerical position, or when the clerical force is being expanded the factory foreman may be asked to recommend bright young operatives for these vacancies.

Our brief inquiry suggested also that the educational and occupational requirements applied in hiring from the outside were relaxed considerably when potential transfers from the factory were under consideration. This seems sensible, as a recommendation from a company supervisor to the effect that the person involved is bright, eager, and conscientious in his application to his job makes him appear to be as good a risk for the job, if not a better one, as a person from the outside who has a superior education and supposedly relevant occupational background. In fact, it was our feeling that if the employment service could convince a company that its semi-skilled laid-off blue collar worker had these same qualities the man would be hired. Instead, what seems to happen is that new white collar male clerks hired from the outside are those who possess particular skills and abilities acquired either in school or on previous white collar jobs. This barrier to mobility is reinforced apparently by the separation of blue and white collar hiring. In effect, there are two personnel departments operating separately with very little interchange between them.

For the reasons we have sketched, we amended our tentative research design in several respects. First of all we decided to include about 50 foremen in the sample of 500 blue to white collar males who were to be interviewed in depth. Second, we amended our preliminary interview schedule to make sure that we would be able to followly carefully and distinguish between moves in the internal and external labor markets and to see how the procedures for each differed. This brings us now to the second stage of our project, the attempt to gain a greater knowledge of the characteristics of men who made blue to white collar shifts.

Preliminary Results of the Survey of Characteristics of Blue to White Collar Shifters[9]

We should note at the outset that this is a preliminary report. When we say preliminary, we don't mean a rough draft based on complete data; we mean instead an early report based on first returns and issued before the complete data have been processed, not to mention analyzed and digested. The interviews were completed several weeks ago. Coding is underway at present and machine runs should be completed this coming spring. However, for this meeting we did make preliminary runs on one of the 38 IBM decks that are in preparation. These runs consist primarily of occupational histories, that is, the chronological job path of men who have shifted into various white collar job groupings. Before discussing these job patterns, we should note briefly what information we sought from our sample so that you have an idea of how these first returns relate to the remaining data that will be published at a later date.

Respondents gave us a summary occupational history in reverse chronological order for the jobs they have held in the past ten years. The history also included a one phrase description of each job, the name of the company, what it made, the starting and ending dates and wages earned on each job. Military service and full-time schooling were treated as "jobs." After obtaining this skeleton we put flesh on it by asking the respondents for details about each job, this time going in chronological order. We attempted to find out how a person got the job in question, what he did on the job, what skills he used, where he obtained these skills, what outside schooling or on-the-job training he received, where he went to school, why he left this job and how he got his next job.

After completing this portion of the interview, the interviewer then probed in detail about the blue to white collar job shift. We asked why the respondent made the move, what skills and training enabled him to do so, what he thought about it, whether other people could do the same thing, whether he recommended that they try it and

---

[9] For their assistance in this portion of the project, we wish to express our appreciation to the University's Survey Research Laboratory, Professor Harry Sharpe, Director, to the staff who conducted and coded the interviews and to Mina Hockstad, Marge Colson and Johan Mathiesen who supervised these procedures. We also wish to thank the Social Systems Research Institute, Professor Roger Miller, Executive Director, Hilde Roubal who supervised the work, and the staff who performed machine computation in cooperation with the Commerce Department Computation Center.

if so how they should go about it. We then asked a set of questions that attempt to measure on a most experimental basis the assertiveness of the individual, in order to see if we can identify personal characteristics of individuals that help to explain why they rather than others made this job shift. Finally we asked the usual questions about age, family status and education so that we can compare our group to other groups in the population.

Occupational Paths and Characteristics

Table III is a frequency distribution by two-digit occupational groups showing the "entry" white collar clusters within each major white collar category.[10] This table and the others in this section of our paper contain raw data reproduced from the first machine runs that we received a few days ago. They are included here

TABLE III

DISTRIBUTION OF "ENTRY" WHITE COLLAR JOBS BY TWO-DIGIT OCCUPATIONAL GROUPS*

TWO-DIGIT OCCUPATION GROUPS	MAJOR WHITE COLLAR CATEGORIES					
	Prof-Tech (00)	M'grial (20)	Foremen (28)	Clerical (30)	Sales (36)	Totals
00 Accountants	3					3
01 Teachers (exc. Vocational)	4					4
02 Teachers (Vocational)	7					7
03 Professional Engineers	2					2
04 Personnel & Labor Relations	1					1
05 Misc. Professional	2					2
10 Draftsmen	37					37
11 Industrial Engineering Technicians	54					54
12 Mechanical Engineering Technicians	25					25
13 Other Engineering Technicians	10					10
14 Electrical Technicians	27					27
15 Materials Processing Occupations	21					21
16 Misc. Technical	6					6
17 Professional-Technical Trainees	5					5
18 Professional-Technical Group Leaders	1					1
20 Executives & Public Admin. Officials		3				3
21 Admin. Specializations		12				12
22 Managers & Supervisors		54				54
23 Assistant Managers, etc.		10				10
24 Management Trainees		3				3
25 Proprietors		2				2
28 Foremen			34			34
29 Assistant Foremen			10			10
30 Stenos, Typists, etc.				4		4
31 Computing & Acc't Recording Clerks				8		8
32 Production Recording & Routing Clerks				60		60
33 Information & Message Distribution Clerks				6		6
34 Misc. Clerical				15		15
35 Clerical Group Leaders				2		2
36 Salesmen of Services					7	7
37 Salesmen of Commodities					7	7
38 Other Sales					3	3
39 Sales Group Leaders & Trainees					2	2
TOTALS	205	84	44	95	19	447

*Preliminary unweighted raw data

---

[10] The occupational code is an adaptation of the census code. Specific occupations are coded from 000 to 999. The first digit of the code designates the major occupational group, technical, managerial, clerical, etc. The second digit identifies the various occupational sub-groups within each major category. For example "1" is the technical group, "10" is the drafting sub-group, "11" the industrial engineering sub-group, etc. The third digit identifies the specific job in the sub-group; "104" for example is a mechanical draftsman.

only to illustrate how we propose to identify occupational paths. In Table IV we show the distribution of "exit" blue collar job statuses of men who shifted into the technical occupational groups. A job status is usually a specific job but it also may be military service, full time schooling, not in the labor force for some other reason, or unemployed.

We identify the job status held immediately prior to the shift as status -1 and, in reverse chronological order, each successive job status held prior to it as status -2, -3 and so on through -6. The entry white collar job is regarded as status 0 and each successive job held after the shift up through the third job status is identified as status +1, +2 and +3. This means that for each status from -6 through +3 for any occupation by one, two or three digit breakdown, we will have a run comparing that status with status 0. Table IV shows the comparison between status -1 and status 0. Two types of summary runs are being compiled. One will summarize the jobs held prior to the shift; the other will summarize the jobs held after the shift. For any occupational group in which we are interested, these summaries will supplement the occupational time path patterns by providing a summary of all blue collar jobs held prior to the shift and a summary of all white collar jobs held after the shift.

Let us turn now to the data involving technicians in Tables III, IV, and V to illustrate the type of analysis that can be made and some of the characteristics and occupational paths of the technicians. The distribution of men entering the technical occupations, all oc-

TABLE IV

DISTRIBUTION OF "EXIT" BLUE COLLAR JOBS OF THOSE MEN WHO SHIFTED INTO THE VARIOUS TECHNICAL OCCUPATIONS*

PRIOR BLUE COLLAR JOB	TECHNICAL OCCUPATIONS INTO WHICH SHIFT OCCURRED									
	Draftsmen	Industrial Engr. Technicians	Mechanical Engr. Technicians	Other Engineering Technicians	Technicians, Electrical & Electronic	Materials Processing Occupations	Misc. Technical & Kindred	Technical Trainees	Technical Group Leaders	TOTALS
41 Electricians	1	1	-	-	4	2	-	-	-	8
42 Machinists	3	6	-	-	-	1	-	-	-	10
43 Tool & Die Makers	5	3	-	-	-	3	-	2	-	13
44 Mechanics & Millwrights	2	4	1	1	-	2	1	-	-	11
46 Metal Workers	-	4	-	-	-	2	-	-	-	6
48 Plumbers	-	-	-	-	-	1	-	-	-	1
49 Linemen, etc., tel & power	-	1	-	-	14	-	-	-	-	15
50 Misc. Craftsmen	2	-	1	-	1	1	-	-	-	5
53 Working Leaders, Craftsmen, etc.	1	2	-	-	-	2	-	-	-	5
54 Working Leaders, Operatives, etc.	-	-	-	1	1	-	-	-	-	2
60 Apprentices	3	6	1	-	-	-	-	-	-	10
62 Assemblers	2	6	3	-	1	1	-	-	-	13
63 Vehicle Drivers	1	-	-	-	-	-	-	1	-	2
64 Material Handlers	1	1	1	-	-	-	-	1	1	5
65 Misc. Operatives	6	8	14	6	3	5	3	1	-	46
70 Press Operators	1	-	-	-	-	-	1	-	-	2
71 Lathe Operators	1	5	-	-	-	-	-	-	-	6
72 Misc. Machine Operators	1	4	2	1	-	1	-	-	-	9
82 Misc. Service Workers	3	-	-	-	-	-	-	-	-	3
86 Non-Farm Laborers	1	2	-	-	1	-	1	-	-	5
90 In School	1	-	2	1	2	-	-	-	-	6
91 In Armed Forces	2	1	-	-	-	-	-	-	-	3
TOTALS	37	54	25	10	27	21	6	5	1	186

* Preliminary unweighted raw data

cupations that have a "1" as the first digit, is shown in the first column of Table III. Data in Table V-A and Table V-B indicate that most of these men are rather young and that they shifted to white collar work early in their careers. Some ideas about the degree of their mobility can be gained by reference to Table V-D showing the number of statuses, jobs, and companies for which they worked. It is of interest to note that about three-fourths of those men shifted from blue to white collar employment within the same company. Table V-E compares the percent of the job shifts that were internal for each of the moves both before and after the shift. As might be expected, the proportion of internal shifts rises as the men settle down after shopping around. Although the original extent and direction survey asked for the number of men who shifted during 1961 and 1962 we relaxed this constraint when we again approached our enlarged sample to obtain the names of men who had shifted from blue to white collar work. Even so, as can be seen from Table V-B, most of the shifts are recent enough to suggest that these particular shift possibilities are not obsolete. Our sample of blue collar males shifting into technical jobs has less education than the census sample of all such white collar technicians. But when we compare our sample with the portion of the census sample that held a blue collar job at age 24, as we expect to do, we suspect we will find little difference in educational attainment.

The relationship between Table III and IV can be visualized more clearly if you note that the nine major technical occupational groups listed vertically in the first column of Table III become the nine columns of Table IV, and in each group, or column, is found the distribution of blue collar jobs held immediately prior to the shift. It is also possible to trace the path of each individual. Again, primarily for illustrative purposes, and without implying any great significance to the particular paths selected, we have included in Table VI the occupational history of several of the men who shifted into a specific type of drafting. In themselves these patterns indicate only that men with a variety of backgrounds have shifted into this occupation. The explanation for the shift on the part of the first and fourth respondent is to be found in their educational background. The other two men were able to make the shift within the same company because of the experience gained from their blue collar jobs. These job patterns are useful mainly to us in our attempts to identify the paths that blue to white collar shifters follow.

With these remarks we must conclude. We recognize that our report is without a conclusion but as we said before, our research is still in progress and this report can cover only what has been discovered to date. We dare not gaze into the crystal ball to speculate about what our completed study will show. The only speculation we have done so far has been about our scheduled completion date and unfortunately we have missed the mark on this by quite a wide margin. We can only reaffirm the estimate that we have made several times since undertaking this project in 1963 --- we hope that it can be concluded in the coming year.

TABLE V

SELECTED CHARACTERISTICS OF MEN WITH BLUE COLLAR BACKGROUNDS WHO SHIFTED INTO TECHNICAL OCCUPATIONS*

A. Distribution by Age

YEAR OF BIRTH	DISTRIBUTION
1909 & earlier	2%
1910 - 1919	10%
1920 - 1929	26%
1930 - 1939	51%
1940 - 1943	11%

B. Distribution by Date of Shift

YEAR OF SHIFT	DISTRIBUTION
1950 - 1954	5%
1955 - 1959	27%
1960 - 1964	68%

C. Level of Education

	OUR SAMPLE	1960 CENSUS SAMPLE OF EMPLOYED TECHNICIANS**
Less than High School graduation	15%	12%
High School graduation	33%	23%
Some College, but no 4-yr. degree	50%	53%
4-yr. College degree	2%	12%

---

* Preliminary unweighted raw data. N = 186.

** Monthly Labor Review, November 1964, Vol. 87, No. 11, page 1279.

TABLE V (Continued)

D. Distribution of Total Number of Statuses and Jobs Held, and Companies Worked for, since August, 1954

	STATUSES	JOBS	COMPANIES
1	-	3%	46%
2	2%	20%	23%
3	8%	24%	13%
4	19%	22%	9%
5	23%	12%	4%
6	16%	7%	3%
7	7%	6%	1%
8	15%	4%	1%
9	4%	1%	-
10 & over	7%	1%	1%

E. Distribution of Internal and External Shifts

JOB	SAME COMPANY	DIFF. COMPANY	NO JOB OR NO PRIOR JOB TO COMPARE WITH
Prior job to job in Status #3	4%	1%	95%
" " " " " " #2	12%	2%	86%
" " " " " " #1	29%	7%	64%
Blue Collar to White Collar Job	73%	27%	-
Prior Job to job in Status -1	43%	38%	20%
" " " " " " -2	20%	32%	47%
" " " " " " -3	10%	22%	68%
" " " " " " -4	9%	13%	78%
" " " " " " -5	3%	10%	77%
" " " " " " -6	3%	5%	92%

TABLE VI

* Preliminary, unweighted raw data for 4 of the 8 men who shifted into this job.

JOB PATTERNS OF MEN WHO SHIFTED INTO MECHANICAL DRAFTSMAN "104"*

INTERVIEW NUMBER	STATUS +1			WC JOB			STATUS -1			STATUS -2			STATUS -3			STATUS -4			STATUS -5		
	Occupation Code	Duration in Mos.	2-digit SIC Code	Occupation Code	Duration in Mos.	2-digit SIC Code	Occupation Code	Duration in Mos.	2-digit SIC Code	Occupation Code	Duration in Mos.	2-digit SIC Code	Occupation Code	Duration in Mos.	2-digit SIC Code	Occupation Code	Duration in Mos.	2-digit SIC Code	Occupation Code	Duration in Mos.	2-digit SIC Code
175		--		104	62	36	681	09	37	930	10	99	681	24	26	864	06	24	913	48	99
176		--		104	42	36	411	34	36	656	24	36	913	48	99		--			--	
294	180	74	35	104	35	35	656	24	35	673	84	73	914	12	99		--			--	
320	129	11	34	104	18	34	879	09	42	904	21	99		--			--			--	

KEY

OCCUPATIONAL CODES

104 - Mechanical draftsman
129 - Mechanical engr. tech. (n.e.c.)
180 - Group leader, draftsmen
411 - Electrician, electrical service & repair
656 - Checker, examiner or inspector (mfg.)
673 - Photographic process worker
681 - Welder or flame-cutter
864 - Lumberman, raftsman or woodchopper
879 - Non-farm laborer
904 - College
913 - Army
914 - Air Force
930 - Unemployed

INDUSTRY CODES

24 - Lumber & wood products, exc. furniture (mfg.)
26 - Paper & allied products (mfg.)
34 - Fabricated metal products, exc. machinery & transportation equipment (mfg.)
35 - Machinery, except electrical (mfg.)
36 - Electrical machinery, etc. (mfg.)
37 - Transportation equipment (mfg.)
42 - Motor freight transp. & warehousing
73 - Miscellaneous business services
99 - Nonclassifiable establishments

APPENDIX A
LIST OF CLASSIFICATIONS

White Collar Jobs Into which Blue Collar Workers May Have Transferred

Draftsman -- Prepares working plans or detailed drawings to scale from rough or detailed sketches or notes. Inks in lines, letters, dimensions, specifications, and legends on pencil drawings as required. May make engineering computations such as those involved in strength of material, beams, and trusses. Uses triangles, T-square, protractors, compasses, French curves, and other drawing instruments.

Common job titles in this occupation are mechanical draftsman, electrical draftsman, architectural draftsman, checker, layout man, tool designer, die design draftsman.

Electronic & Electrical Technicians -- Assists engineers and scientists in designing, modifying, testing, and assembling devices and equipment. Installs, tests, maintains, troubleshoots, and repairs electrical and electronic equipment where a functional knowledge of electricity and electronics is required. May revise or prepare assembly drawings.

Common titles in this occupation are laboratory technician, power technician, electronic circuit technician, radar technician, engineering aid, systems technician.

Engineering & Physical Science Technician -- Provides semiprofessional technical support for engineers and scientists working in such areas as research, design, development, testing, or production. Work pertains to physical, biological, chemical, or mechanical component or equipment. Has knowledge of science and/or engineering.

Common job titles included in this occupation are test technician, analyst, data processing technician, laboratory technician (other than electrical or electronic), chemical technician, civil engineering technician, x-ray technician, medical technician.

Industrial Engineer -- Plans and develops production methods that make the most efficient use of materials, machinery, and personnel. May supervise the layout of plant machinery, conduct time and motion studies, design industrial safety programs, and plan cost and quality control programs.

Personnel & Labor Relations Worker -- Interviews applicants for employment, assists applicants with application forms, and answers questions about the company. Assists in the administration of tests, prepares information for future training needs, studies and analyzes jobs, determines appropriate wage and salary levels, and keeps records pertaining to the personnel of the company. Administers health and welfare programs, assists in grievance settlements and contract negotiations.

Common job titles included in this occupation are manager, industrial relations, labor relations specialist, employment interviewer, job interviewer, health and welfare specialist, employee benefit specialist, job evaluation man.

Processing Occupations -- Prepares and maintains production schedules and records, compiles statistics, and charts related to actual production and estimates, to man hours to meet required schedules, and related data. --Coordinates flow of materials, parts, and assemblies within or between departments in accordance with production schedules, expedites the delivery of merchandise, parts, and equipment by contacting suppliers, transportation and company officials. --Checks shipping routes to and from customers, plans the most efficient methods of packaging and shipping products, may design and select containers.

Common job titles included in this occupation are production planner, production scheduler, shipping supervisor, traffic analyst, traffic manager, expediter, dispatcher.

Purchasing and Sales Occupations -- Sells industrial and technical equipment and supplies using professional and technical knowledge of the adaptation and installation of equipment and industrial processes. May service new products and train employees to operate and maintain equipment. Buys raw materials, equipment, and supplies necessary for the operation of the enterprise. Keeps records pertaining to costs, inventories, deliveries, and goods purchases, may interview salesmen and draw up contracts.

Common job titles in this occupation are salesmen, manufacturing salesman, technical sales, buyer, spare parts buyer, analyst, material analyst, purchasing agent, procurement man, outside salesman.

Machine Accountants -- Plans, develops, and prepares programs and phases of programs for solution by electronic computer. Corrects program errors by revising or altering sequence of instructions. Performs operations on a variety of data processing machines. May perform the wiring or machine set-up, may monitor and control an electronic computer.

Common job titles included in this occupation are data processing machine operator, office machine operator, console operator, auxiliary machine operator, programmer (if not included in engineering and physical science technician).

APPENDIX B

NAME OF FIRM:________________________

Bureau of the Budget
Clearance No. 44-6342

BLUE TO WHITE COLLAR JOB SHIFTS - EMPLOYER QUESTIONNAIRE

COLUMN ONE: Fill in the job title of each white collar job into which men have been newly hired or promoted in the two year period Jan. 1, 1961 - Dec. 31, 1962.

COLUMN TWO: Show the number of men hired or promoted into each white collar job during this time period. We are asking for all white collar hires or promotions during this period, whether or not they have blue collar backgrounds.

COLUMN THREE: Show the number of men hired or promoted into each white collar job who have blue collar backgrounds, whether or not the blue collar work was during their employment with this company.

COLUMN FOUR: Show the total number of men in the occupation as of December 31, 1962.

White Collar Jobs include professional, technical, managerial, sales, clerical and kindred service jobs. Generally the person is paid a weekly or monthly salary.
Blue Collar Jobs include unskilled and semi-skilled workers, craftsmen and manual workers who are paid by the hour.

If you have not hired or promoted any men into white collar jobs between Jan. 1, 1961 and Dec. 31, 1962, please write "none" in the following box and return the form: /________________/

On completion, return in the enclosed envelope to the Industrial Relations Research Center, University of Wisconsin, Social Science Bldg., Madison, Wis. 53706.

(1) White Collar Job	(2) Number of men hired or promoted into this occupation from 1/1/61 - 12/31/62	(3) Number of these men in Col. 2 who subsequent to 1/1/56 have had at least 1 year in a blue collar job	(4) Total number of men in this occupation on or about 12/31/62
(Examples:)			
1. Draftsman	19	6	37
2. Timestudy man	1	0	1
3. Production clerk	7	2	18
4. Salesman	4	0	9
1.			
2.			
3.			
4.			
5.			
6.			
7.			
8.			

(USE ADDITIONAL SHEETS IF NEEDED)

APPENDIX C*

STATISTICAL METHODOLOGY

It should be noted first that the sample of firms was a systematic sample which is frequently approximated by a simple random sample. It is clear that the sample of employees of these firms cannot accurately be considered a random sample of employees in Milwaukee County, since employees of any chosen firm form a cluster of elements of our sample. Nevertheless, in this preliminary analysis we shall proceed as if those employees in the chosen firms in any industry constitute a random sample of employees in that industry.

Weights were assigned to industries according to the proportion of employees of Milwaukee County which they contained. An estimate of male white collar entries in 1961 and 1962 per industry was made by applying the appropriate expansion factor to the number of male white collar entries in that industry. This is equivalent to applying the weighted proportion to N, for if we let h indicate the strata, n_h the number in the sample in strata h, and N_h the total number of employees in Milwaukee County in strata h, we have:

$$\frac{N_h}{n_h}\left(\sum_i Y_{h_i}\right) = \frac{N_h}{N}\left(\frac{\sum Y_{h_i}}{n_h}\right)N = (W_h)(P_h)N$$

where W_h is the weight of strata h and Y_{h_i} is 1 if the person is an entry and zero otherwise. The variance of the estimated number of entries in each strata can be computed using

$$V(N_hP_h) = N_h^2\,V(P_h) = N_h^2\,\frac{P_hQ_h}{n_h}$$

and 95% confidence intervals can be computed for the number of entries per strata. However, in finding the variance of the total number $N_{P_{st}}$ of entries in the county we must take account of the strata weight since estimate

$$V(N_{P_{st}}) = N^2V(P_{st}) = N^2\sum(W_h)^2\,\frac{P_hQ_h}{n_h}$$

In using the above it was found that

$$\sqrt{\sum(W_h)^2\,\frac{P_hQ_h}{n_h}} \;=\; .0018\,.$$

Thus our 95% confidence interval for number of entries became 17,874 ± 1430.

The application of these same formuli to find a 95% confidence interval for the number of white collar entries in 1961 and 1962 who had blue collar backgrounds gave us 3989 ± 708. Similarly, the 95% confidence intervals were computed for the estimates of the totals of entries and shifts for each occupational group and major industry category (See Tables C-1 and C-2). These intervals were rather, as was expected, large because of the small percentages involved. Table C-3 shows the population of employees in the sample and in Milwaukee County by industry category, the expansion factors, and the breakdown of shifts by major occupational group and industry category.

A reasonable estimate of the ratio r_h of shifters to entries in each industry would be the number of entries with blue collar backgrounds in that industry. A weighted sum of these r_h would then give us $\frac{3989}{17,894}$ or .22 for the ratio of shifters to entries for Milwaukee County in the given two-year period.

---

*Prepared by Estelle Goldberg, Wisconsin Survey Research Laboratory.

APPENDIX TABLE C-1

ESTIMATES AND STANDARD ERRORS OF ESTIMATES FOR ENTRIES AND SHIFTS IN THE MILWAUKEE COUNTY LABOR FORCE

By Major Occupation Group

OCCUPATION	95% CONFIDENCE INTERVAL FOR WHITE COLLAR ENTRIES IN MILWAUKEE COUNTY IN 2-YEAR PERIOD	95% CONFIDENCE INTERVAL FOR SHIFTS
Professional	6,127 ± 951	945 ± 375
Managerial	3,606 ± 708	1,556 ± 469
Clerical	3,104 ± 553	827 ± 270
Sales	5,037 ± 665	661 ± 254
TOTAL	17,874 ± 1,430	3,989 ± 708

APPENDIX TABLE C-2

ESTIMATES AND STANDARD ERRORS OF ESTIMATES FOR ENTRIES AND SHIFTS IN THE MILWAUKEE COUNTY LABOR FORCE

By Major Industrial Category

INDUSTRY	ENTRIES	SHIFTS
Construction	623.5 ± 246.5	273.5 ± 163.5
Manufacturing	5,127.0 ± 775.0	1,391.0 ± 407.0
Transportation	775.0 ± 294.0	459.0 ± 228.5
Wholesale Trade	805.0 ± 190.0	224.5 ± 102.5
Retail Trade	2,303.5 ± 340.5	549.5 ± 166.5
Finance, etc.	2,447.5 ± 405.5	97.0 ± 11.0
Services	4,762.0 ± 811.0	910.0 ± 448.0
Government	1,013.5 ± 77.5	86.5 ± 22.5

APPENDIX TABLE C-3

DATA USED TO DETERMINE DISTRIBUTION OF BLUE TO WHITE COLLAR SHIFTS BY INDUSTRY AND BY OCCUPATION FOR MILWAUKEE COUNTY

	NUMBER OF EMPLOYEES			NUMBER OF SHIFTS BY OCCUPATION (1) In Sample (2) Estimated for Milw. County								ESTIMATED SHIFTS MILWAUKEE COUNTY	
Industry	In Milw. County	In Sample	Expansion Factor	Prof. (1)	Prof. (2)	Managerial (1)	Managerial (2)	Clerical (1)	Clerical (2)	Sales (1)	Sales (2)	Total	% Dist. by Industry
Agriculture, etc.	19,009	762	24.9	4	99.6	5	124.5	2	49.8	0	---	273.9	6.9%
Manufacturing	169,598	5,624	30.2	14	422.8	19	573.8	4	120.8	9	271.8	1,389.2	34.8%
Transport.	26,242	914	28.7	0	---	4	114.8	11	315.7	1	28.7	459.2	11.5%
Wholesale	14,201	1,199	11.8	0	---	3	35.4	3	35.4	13	153.4	224.2	5.6%
Retail	61,713	4,715	13.1	0	---	13	170.3	19	248.9	10	131.0	550.2	13.8%
Finance,etc.	17,950	924	19.4	2	38.8	0	---	2	38.8	1	19.4	97.0	2.4%
Services	71,080	1,254	56.7	6	340.2	9	510.3	0	---	1	56.7	907.2	22.7%
Government	17,366	11,000	1.6	27	43.2	17	27.2	11	17.6	0	---	88.0	2.2%
TOTALS	397,159	26,392		945		1556		827		661		3989	
% DISTRIBUTION OF SHIFTS BY OCCUPATION				23.7%		39.0%		20.7		16.6%			100%

APPENDIX TABLE D-1

OCCUPATION OF THE MALE EMPLOYED FOR THE UNITED STATES AND MILWAUKEE COUNTY: 1950 AND 1960

United States

Occupation	Number in Occupation 1950	Number in Occupation 1960	% Change 1950 to 1960	1960 % Distribution by Occupation
Professional, Technical, etc.	2,970,200	4,479,358	50.8%	10.31%
Managers, Officials & Proprietors	4,356,700	4,629,842	6.3%	10.65%
Clerical Workers	2,646,420	3,015,476	13.9%	6.94%
Sales Workers	2,572,637	2,977,872	15.8%	6.85%
Craftsmen & Foremen	7,584,306	8,488,777	11.9%	19.53%
Operatives	8.154,084	8,641,652	6.0%	19.88%
Laborers	3,308,553	2,997,785	-9.4%	6.90%
Services: Private Household	73,365	61,063	-16.8%	0.14%
Services: Exc. Priv. Household	2,376,749	2,598,673	9.3%	5.98%
Farmers	4,193,986	2,387,584	-43.2%	5.49%
Farm Laborers	1,965,757	1,201,922	-38.9%	2.77%
Occupation Not Reported	459,581	1,986,951	332.3%	4.57%
TOTAL	40,662,374	43,466,955	6.9%	100.00%

Table 202. Detailed Characteristics of the Employed, by Sex, for the United States: 1960 and 1950. U.S. Bureau of the Census. U.S. Census of Population: 1960. Detailed Characteristics. United States Summary. Final Report PC(1)-1D U.S. Government Printing Office, Washington, D.C. 1963.

APPENDIX TABLE D-2

OCCUPATION OF THE MALE EMPLOYED FOR THE UNITED STATES AND MILWAUKEE COUNTY: 1950 AND 1960

Milwaukee County

Occupation	Number in Occupation 1950	1960	% Change 1950 to 1960	1960 % Distribution by Occupation
Professional, Technical, etc.	23,823	31,125	30.7%	11.3%
Managers, Officials & Proprietors	27,868	25,136	-9.8%	9.2%
Clerical Workers	22,337	22,621	1.3%	8.0%
Sales Workers	18,365	19,689	7.2%	7.2%
Craftsmen & Foremen	63,575	61,041	-4.0%	22.2%
Operatives	67,157	69,276	3.2%	25.2%
Laborers	18,889	16,464	-12.8%	6.0%
Services: Private Household	205	153	-25.4%	.06%
Services: Except Priv. Household	14,924	14,851	-4.9%	5.4%
Farmers	950	396	-58.3%	.01%
Farm Laborers	624	416	-33.3%	.02%
Occupation Not Reported	2,230	13,284	495.5%	4.8%
TOTAL	260,947	274,452	5.2%	99.4%

Table 121. Detailed Occupation of the Employed, by Sex, for the State, Urban, Rural and for Standard Metropolitan Statistical Areas of 100,000 or more and Counties of 250,000 or more: 1960. U.S. Bureau of the Census, U.S. Census of Population: 1960. Detailed Characteristics. Wisconsin. Final Report PC(1)-51D. U.S. Government Printing Office, Washington, D.C. 1962.

Table 73. Detailed Occupation of the Experienced Civilian Labor Force and of Employed Persons, by Sex, for the State, and for Standard Metropolitan Statistical Areas and Cities of 100,000 or more: 1950. U.S. Bureau of the Census. U.S. Census of Population: 1950. Vol. II, Characteristics of the Population, Part 49, Wisconsin. U.S. Government Printing Office, Washington, D.C. 1952.

APPENDIX TABLE D-3

INDUSTRY OF THE EMPLOYED FOR THE UNITED STATES AND MILWAUKEE COUNTY: MALES, 1950 AND 1960

United States

	Number In Industry			1960
Industry	1950	1960	% Change 1950 to 1960	% Distribution by Industry
Agriculture, Forestry & Fisheries	6,440,903	3,932,225	-38.9%	9.0%
Mining	907.721	622,146	-31.5%	1.4%
Construction	3,359,007	3,662,393	9.0%	8.4%
Manufacturing	11,030,576	13,111,965	18.9%	30.2%
Transportation	3,750,833	3,687,448	-1.7%	8.5%
Wholesale Trade	1,584,527	1,762,082	6.5%	4.1%
Retail Trade	5,365,247	5,635,988	5.0%	13.0%
Finance, etc.	1,137,615	1,464,283	28.7%	3.4%
Services	4,755,116	5,716,948	20.2%	13.2%
Public Administration	1,856,058	2,288,760	23.3%	5.3%
Industry Not Reported	330,119	1,582,717	208.4%	3.6%
TOTAL	40,662,374	43,466,955	6.9%	100.1%

Table 211. Detailed Industry of the Employed, by Sex, for the United States: 1960 and 1950. U.S. Bureau of the Census. U.S. Census of Population: 1960. Detailed Characteristics. United States Summary. Final Report PC(1)-1D. U.S. Government Printing Office, Washington, D.C. 1963.

APPENDIX TABLE D-4

INDUSTRY OF THE EMPLOYED FOR THE UNITED STATES AND MILWAUKEE COUNTY: MALES, 1950 AND 1960

Milwaukee County

Industry	Number in Industry 1950	1960	% Change 1950 to 1960	1960 % Distribution by Industry
Agriculture, Forestry & Fisheries	2,004	1,493	-25.5%	0.5%
Mining	189	169	-10.6%	0.00%
Construction	18,287	16,200	-11.4%	5.9%
Manufacturing	125,255	132,750	6.0%	48.4%
Transportation	24,185	21,481	-11.2%	7.8%
Wholesale Trade	11,076	10,594	-4.4%	3.9%
Retail Trade	32,167	30,777	-4.3%	11.2%
Finance, etc.	7,370	8,404	14.0%	3.1%
Services	26,846	28,998	8.0%	10.6%
Public Administration	11,062	13,069	18.1%	4.8%
Industry Not Reported	2,506	10,517	319.7%	3.8%
TOTAL	260,947	274,452	5.2%	100.0%

Table 127. Detailed Industry of the Employed, by Sex, for the State, and for Standard Metropolitan Statistical Areas of 100,000 or more and Counties of 250,000 or more: 1960. U.S. Bureau of the Census. U.S. Census of Population: 1960. Detailed Characteristics. Wisconsin. Final Report PC(1)-51D. U.S. Government Printing Office. Washington, D.C. 1962.

Table 79. Detailed Industry of the Experienced Civilian Labor Force and of Employed Persons, by Sex, for the State, and for Standard Metropolitan Statistical Areas and Cities of 100,000 or more: 1950. U.S. Bureau of the Census, U.S. Census of Population: 1950. Vol. II, Characteristics of the Population, Part 49, Wisconsin. U.S. Government Printing Office, Washington, D.C., 1952.

APPENDIX TABLE E

UNWEIGHTED RATIOS OF ENTRIES AND SHIFTS TO EMPLOYEES AND SHIFTS TO ENTRIES BY SIZE OF FIRM

SIZE OF FIRM	$\frac{\text{ENTRIES}}{\text{EMPLOYEES}}$	RATIO	$\frac{\text{SHIFTS}}{\text{EMPLOYEES}}$	RATIO	$\frac{\text{SHIFTS}}{\text{EMPLOYEES}}$	RATIO
1 - 10	$\frac{127}{1,726}$	7.4	$\frac{42}{1,726}$	2.4	$\frac{42}{143}$	29.4
11 - 20	$\frac{98}{1,481}$	6.6	$\frac{15}{1,481}$	1.0	$\frac{15}{82}$	18.3
21 - 50	$\frac{168}{1,973}$	8.5	$\frac{20}{1,973}$	1.0	$\frac{20}{168}$	11.9
51 - 100	$\frac{94}{2,034}$	4.6	$\frac{29}{2,034}$	1.4	$\frac{29}{94}$	30.9
101 & over	$\frac{831}{19,178}$	4.3	$\frac{104}{19,178}$	0.5	$\frac{104}{831}$	12.6

APPENDIX TABLE F

FOREMEN

Number of Entries and Blue to White Collar Shifts by Industry

		TOTAL NUMBER OF ENTRIES		NO. WITH BLUE COLLAR BACKGROUND		
INDUSTRY	INDUSTRY WEIGHTS	In Sample	Weighted	In Sample	Weighted	% WITH BLUE COLLAR BACKGROUND
Agriculture, Mining & Constr.	--	--	--	--	--	--
Manufacturing	30.2	18	544	14	423	77.8%
Transportation, etc.	28.7	3	86	2	57	66.7%
Wholesale Trade	--	--	--	--	--	--
Retail Trade	--	--	--	--	--	--
Finance, etc.	--	--	--	--	--	--
Services	56.7	3	170	3	170	100.0%
Government	1.6	12	19	6	10	50.0%
TOTAL			819		660	80.6%

Discussion

Marvin Friedman, AFL-CIO

I have not had the opportunity to review Jim Stern's paper. However, from listening to his presentation, I think it is obvious that he is engaged in some imaginative research and that his conclusions might have some very interesting policy implications. In fact, his presentation has prompted me to discard a couple of paragraphs of my notes. Following Stern's presentation, they would have been somewhat out of place because they were perhaps overly-critical of our manpower research. While I am toning down my comments, I think it will still be obvious that I am not happy with our research activity in this area.

Dr. Aller's paper bears out a suspicion I have had for some time, and that is that we not only are in the dark as to where we are heading in manpower research, but we do not even know how we got where we are. I found his brief description of the background of Title I of the Manpower Development and Training Act exceedingly illuminating. Furthermore, I think it is important to underscore his emphasis on the need for a greater policy orientation in our research activities.

I am, after all, a heavy user of the results of the research work being carried on, and not a researcher myself in the true sense of the word. I read the research reports with an eye toward what Dr. Aller has described as "useful guides to policy" or "ways for improving the operations of our institutions." Consequently, I find myself quite concerned over the present tendency of our manpower research activities to move in the direction of a battery of individual research projects, gathering facts and data but without any reference to policy implications; or where there are policy implications, they may be so buried that they are observable only by the closest scrutiny and clearly evident only to the individual who performed the research.

I do not wish to be drowned in an ocean of statistics. Moreover, I look askance at researchers and administrators -- whether in or out of the Department of Labor, or other government agencies -- who constantly advise that we really know very little about automation and technological change or their manpower implications. Such statements can contribute to a paralysis in the policy-making arena -- that is, the adoption of sound legislation -- when logic, even without statistics, tells us action and programs are needed. Properly, Dr. Aller touched on this when he suggested that groups of manpower planners might begin to select "some key areas where current research permits a sharp focus on policy possibilities" -- something which is sorely needed but has not been done in any systematic fashion.

I do not want the foregoing remarks to be interpreted as evidence of any opposition to additional money for manpower research, because such is not the case. I think we need more, and I suspect my present state of unhappiness may be tied to the relatively small amount of money being spent in this field. I would, however, like to underscore Dr. Aller's comment that the real need is for truly creative ideas for research undertakings and in this connection I believe we all have a responsibility to do some serious thinking.

Fred Suffa's paper describes what the manpower administration is doing by way of collecting data on the trainees. It is not a criticism of Mr. Suffa, but rather of the Department of Labor, when I suggest that there are serious shortcomings in the data being collected. I think, however, much of my criticism here is related to a difference of opinion over the general shape and direction of our manpower training programs. I am not sure this is the proper forum in which to explore these differences in any detail. Let me simply mention as an example one item in Mr. Suffa's paper: for the Department's analyses of its training activities, a permanent job is one which lasts for more than 30 days. This is a pretty poor standard.

To evaluate properly the results of MDTA training, I think we have to be concerned equally with the trainee-dropout as well as with the trainee who completes his course, and even with those who never enter a training program. Otherwise, we have no way of measuring the true results of our training efforts and, as a result, may be led too readily to conclude that whatever good fortune befalls the trainee is a result of the training program. The fact that the job success of the trainees may be due to an economic upturn -- which may also provide jobs for non-trainees, and for training dropouts as well -- may be overlooked. What I am suggesting is that a sufficient economic upswing can absorb large numbers of unemployed -- including youths, school dropouts and the unskilled -- and I do not want our training statistics to hide that fact when it occurs. Yet, it seems to me that the evidence is there in Mr. Suffa's paper that this is likely to occur.

From Mr. Suffa's paper one gets the very definite impression that the data being collected, and the analysis being performed on attitudes of trainees, deal exclusively with those whose skills are being upgraded. The question must be asked concerning the problems of those who are being downgraded -- not only in skills but in dollars and cents. In other words, to what extent are we providing training which will lead unemployed workers into jobs which require less skill, and pay lower wages, than the jobs which they were performing before they became unemployed. It seems to me that a question of this kind has some very interesting policy implications, for example, in the area of income supports. Might it not lead us to a conclusion that, under some circumstances -- depending perhaps on age, family responsibility, length of time in the workforce, etc. -- we ought to consider income supplements to workers who are victimized by changing technology?

No doubt, one of the sources of our difficulties -- both in the development of our manpower research program and the administrative statistics program -- is the relative newness of the MDTA. It's been with us only a couple of years. We have not, heretofore, been called upon to maintain a total program of research tied, as this one is, to increasing interest in a total manpower program. And the more we move toward an active manpower policy, assuming this is the direction in which we are now heading, the greater will be the need for related research. We can only hope that it will -- in Dr. Aller's words -- be accompanied by some "truly creative ideas."

DISCUSSION

Richard A. Lester, Princeton University

It would have been helpful if Mr. Suffa had spent less time on the mechanics and more on the limits and advantages of the data for research and analysis. One would have preferred more discussion of the important questions that might be answered by the data. Too much attention to the mere collection of statistics and insufficient thought about their usefulness for particular purposes (including evaluation of the training results and efficiency of the program) may mean less valuable data for the money spent on their collection.

Mr. Aller's paper is an excellent report from a person whose inside position gave him special advantages in discussing "Congressional expectations." I would only take issue with Mr. Aller on his caution with respect to "basic research." The labor field has not had enough basic research. For public policy purposes, there is nothing more practical than, say, fundamental analysis of labor mobility. One of the reasons that "present resources for manpower research are relatively meagre" is that the field during the past decade has lacked enough basic study and, therefore, new concepts, techniques, and ideas, to make it an attractive research area for younger scholars.

The Stern-Johnson paper should help to answer some important questions about mobility and the creation of noncompeting groups. I hope they will put their study in a broad framework. On the supply side, it would be interesting to know, with respect to blue-collar workers, what attracts them to or repels them from, white-collar jobs such as foremen or draftsmen. On the demand side, the barriers that employers place upon mobility through personnel policies (as in the draftsman case that Stern-Johnson mention) should be carefully analyzed.

Since employment expansion has been and will be largely in the low-wage, low-benefit service areas, it would be interesting to discover the extent to which the terms of employment and social status serve to retard worker movement from blue-collar to white-collar and other service jobs. The Stern-Johnson statement that employers' educational and experience "standards are different" for upward mobility in the "internal labor market" from those in "external markets" raises interesting questions of fact and public policy.

The Stern-Johnson study appears to be well conceived and designed. There is, however, a question about the validity of job data, including wages, collected by interviews that require recall of such data for the previous 10 years. Some previous studies have checked the validity of such interview material, and the authors should at least consider that limitation to their data.

IX

MEASURING THE IMPACT OF THE WAR ON POVERTY

Chairman, Genevieve Carter,
U. S. Welfare Administration

Page

Conceptual Approaches to Assessing Impacts of Large-Scale Intervention Programs - Howard E. Freeman, Brandeis University 192

Methodological, Measurement and Social Action Considerations Related to the Assessment of Large-Scale Demonstration Programs - Clarence C. Sherwood, Action for Boston Community Development, Inc. 199

CONCEPTUAL APPROACHES TO ASSESSING IMPACTS OF LARGE-SCALE INTERVENTION PROGRAMS

Howard E. Freeman, Brandeis University

Dissatisfaction with the social order and zealous efforts at community change have characterized the personal and academic lives of social scientists since their emergence as an identifiable group on the American scene.[1] In many ways, of course, the several disciplines and the persons that held membership in them have changed markedly over the last several decades: much to the despair of some, the influence of visionary clergymen, guilt-ridden do-gooders, and political radicals--dedicated to projecting their own humanitarian views in the guise of scientific inquiry--has pretty well diminished.[2]

Today the social scientists' role in the modification of community life and the amelioration of social pathologies is a much different one. He puts forth theories on which action programs may be based, he serves as expert and consultant to policy-makers and he uses his research repertoire to guide program development. Admittedly, much of social science activity is directed at understanding "basic" processes, but, whether by intent or unwittingly, he serves as an agent of social change; and, if one is willing to extrapolate from shifts in occupational settings, it appears that there is a growing movement of research persons who know full-well of the social-change potential of their work.[3] Certainly there are outstanding examples of influence: the work of Stouffer and his associates on military problems, the studies of learning psychologists on educational practices, the manifesto of Clark and other social scientists in connection with the Supreme Court's integration decision, and most recently the document of Ohlin and Cloward on delinquency programs.

During the past 15 years, with the increased emphasis--particularly at federal level--on demonstration programs, there has been much concern with the assessment of therapeutic and rehabilitation efforts and thus the developments of a sub-specialty most typically referred to as evaluation research. Virtually all of the demonstration programs supported by federal funds in the health and welfare field and many of the projects sponsored by philanthropic foundations include a requirement that the worth of the effort be assessed.

For the most part, however, the requirement on evaluation has remained a formality; granting agencies have tended to overlook it in their frenzy to implement programs intuitively believed worthwile, statements and often elaborate designs for evaluation in demonstration--research programs have been included in proposals as a ritual with full knowledge that the commitment would not be met, and researchers have, on occasion, found it expeditious to accept evaluation assignments and then redirect the resources to another type of study.

It is only fair to acknowledge the minimal contribution that we have made to program development and social policy through evaluation studies. To some extent the limited impact of previous evaluation research is related to difficulties of successfully implementing and conducting experimental investigations and to barriers put forth by practitioners. There is no need to underscore the difficulties of undertaking research when the cooperation of practitioners and flexibility on their part is necessary for the development and implementation of an adequate design; conflict between clinician and scientist pervades all fields and the difficulties that medical researchers have in undertaking experiments with human subjects are minimal in comparison with evaluation efforts in the community.[4] Also, of course, many social scientists engaged in evaluation studies regard them as a dilittante activity and their interest in such work centers about partially testing a theory that they are concerned with or because it provides publications and sometimes economic affluence. But the major reason I contend that accounts for the minimal impact of evaluation studies on programs and policy is the sheer infrequency that adequately conceived efforts have in fact been undertaken. Whatever be the explanation, certainly it is difficult to point to many instances in which programs actually have been expanded or terminated because of evaluation findings.

---

[1] Howard Odum, American Sociology, New York: Longmans Green, 1951.

[2] Maurice Stein, Sociology on Trial, New York: Prentice-Hall, Inc., 1963.

[3] E. Sibley, Education of Sociologists in the United States, New York: Russell Sage Foundation, 1963.

[4] R. C. Fox, Experiment Perilous: Physicians and Patients Facing the Unknown, New York: The Free Press of Glencoe, Inc., 1959.

We no longer have the opportunity, however, to learn new methodological wrinkles or to develop slowly and carefully a strategy for rendering the results of evaluation studies into a potent force in the determination of action programs and social policy. Suddenly we have a mandate to participate in massive social change, via community-wide efforts projected to restructure health and welfare activities and to reorient the efforts of practitioners. Despite the failure to work out fully methods and most important a strategy to influence policy on small-scale action programs, we now have been thrust into a predominant role in these massive efforts, one designed to have an impact on virtually all community members and indeed on the very social order. It is simply not possible to retreat from this assignment, any more than it is for all physicists to avoid participation in the development and improvement of destructive devices.

But our position is a dangerous one. Although many individuals, for a variety of reasons, have decried so-called centralized programs of planned change and have expressed alarm over their control by public bodies and large foundations, apparently this is the direction that health and welfare activities are going to take; and, if one may regard the recent election in this sense, certainly there is an overwhelming mandate for these efforts to continue.[5] Perhaps those of us located in professional schools or employed directly by community-based programs are most sensitive to the stakes, but it is obvious that the comprehensive and massive character of projects sponsored by the President's Committee on Juvenile Delinquency and Youth Crime, the Ford Foundation, and now the Office of Economic Opportunity are likely to rock the very foundations of our social system.

Unfortunately, although we have improved at least to some extent our technical repertoire, we have not been successful in developing and explicating a set of conditions that must be met in order for our work to have social policy potential. In a previous paper I suggested that because of a concern not to lose our identification with our academic disciplines and not to sacrifice objectivity, we have not considered how studies need to be developed in order that they have an impact on program and policy persons and have failed to recognize the vested interests of the various groups involved in program and policy-making.[6] Although I will touch on a number of issues here, I would like to focus on two points: the need to define our role appropriately given the environment in which we are being called upon to work and the necessity to organize data and findings so that their potential utility in program and social-policy development is maximized.

The Research Environment

Since most of us have at one point or another been involved with large bureaucracies operating on a continual crash basis, certain rather obvious observations can be made most briefly. It is important to point out that dependence upon the legislature branch or the whims of foundations for funds and the necessity to involve and obtain the cooperation of politically and ideologically antagonistic parties in local communities have led up to now and I am sure will continue to evoke a considerable degree of disorder in most of the massive programs. The development of adequate staffs, personnel policies, and long range planning by community-based mass programs is difficult, some maintain almost impossible given the condition of being affluent one minute and poverty-stricken the next and given the fleeting support of the various political forces involved. The shape, size and goals of programs appear to change from day-to-day, and one of the difficulties of evaluation research in these settings stems from the high degree of organizational and interorganizational chaos.

Even in those efforts in which the overall objectives remain relatively stable, the number of specific programs is large and their goals diverse. The evaluation design needs to be developed in terms of a series of staged inputs and outputs, the situation is much too complex to fit the classical independent-dependent variable model. Further the image of the researcher who remains outside the environment and evaluates what others are doing in no way squares with the reality of his engagement in these programs. It is clear that the researcher is involved in a situation in which he must lock himself into the environment, not only because he has a background that can be exploited by persons designing programs, but because otherwise he cannot accomplish his evaluation task. Unless he participates, indeed leads

---

[5] John R. Seeley, "Central Planning: Prologue to a Critique," in Robert Morris (ed.), Centrally Planned Change: Prospects and Concepts, New York: National Association of Social Workers, 1964, pp. 41-68.

[6] Howard E. Freeman, "The Strategy of Social Policy Research," in The Social Welfare Forum 1963, New York: Columbia University Press, 1963, pp. 143-156.

the dialogue and bargaining required for the identification of goals, for description of input-output variables, and for the elaboration of a rationale that specifies the relationship between input variables and goals, these tasks are likely to remain undone. Once formulated he must continue to remain within the environment, like a snarling watchdog ready to fight alterations in program and procedures that could render his evaluation efforts useless.

It is only fair and from my view unfortunate to note that he can expect little help or guidance from the funding groups in these tasks. In part this is related to the lack of structured expectations of outcome on the part of these groups, but also because of an effort to maintain as non-directed a posture as possible in the light of accusations of authoritarian control. The various President's Committee on Delinquency projects illustrate this point well: from city to city, though the legislation directs attention to the reduction of youth crime and the amelioration or related problems, considerable, no great latitude, has been allowed in individual cities not only in program development but in evaluation design. Thus, not only is there variations in whether one is concerned with area-crime rates, the police contacts of individual youths, or the reduction of deviant though not necessarily illegal behavior, but some cities apparently have not felt a need to be very concerned with any measure of this sort. Unless the situation changes, the researcher is naive to expect that sanctions from above are going to provide him with much support in the specification of objectives, the identification of goals of sub-programs or the outlining of the theoretical links between the goals of sub-programs and overall objectives. The researcher has three choices: he can follow Hyman's recommendation and try to guess the intermediate and over-all goals, and later be told that the ones he selected were not relevant at all; he can insist that program persons provide them in which case he should bring lots of novels to the office; or he can participate or rather take a major responsibility for the development of the action framework.[7] In the companion paper, the impact model of a particular program will be discussed in some detail and this point amplified--but I do maintain that there is little likelihood of developing evaluation designs for these massive programs by either second-guessing the action people or by insisting upon their coming up with an appropriate and explicit flow-chart.[8] Indeed, if the researcher is going to act responsibly as an agent of social change through his evaluation research, I contend it is mandatory for him to engage himself in program development.

I do feel, however, that the task would be much easier if the sponsors of these massive programs would establish and enforce a requirement that the necessary specifications be part of any application and renewal of applications and that they exercised sanctions to prevent slippage. Furthermore, that the sponsors provide a minimal set of outcome variables--uniform measurement would be most valuable for long-range program planning. It is most difficult, indeed probably impossible, to compare the various delinquency prevention efforts of the last three years, the various mental health reorganization attempts over the past ten years and, unless there are marked changes in policy, only limited likelihood of making city-to-city comparisons in the economic, educational and occupational rehabilitation programs now underway as part of the poverty package. Given the lack of structured directions by the government and foundation granting programs, and the lack of commitment to evaluation research on the part of many practitioners on the local level, it is not easy to manipulate the environment so the researcher can undertake his task.

I must acknowledge that the researcher has not always participated in these evaluation studies enthusiastically and with a full sense of commitment; to argue that the problems of evaluation research are solely due to the actions of others is as ludicrous as the general who maintained that the high V.D. rate among his troops was due to the promiscuity of the civilian population. Participation within the action environment obligates the researcher to bring to bear his substantive knowledge in the design of programs and to be a positive influence in their development and to recommend and condemn program plans or at least forcefully report and interpret findings from other research that have a bearing on program development. This we often

---

[7] Herbert Hyman, Applications of Methods of Evaluation for Studies of Encampment for Citizenship, California: University of California Press, 1962.

[8] Clarence C. Sherwood, "Methodological, Measurement and Social Action Considerations Related to the Assessment of Large-Scale Demonstration Programs," read at the annual meeting of the American Statistical Association, Chicago, Illinois, December 29, 1964.

fail to do. If we did exercise our responsibility, we probably would have built into these massive efforts more attempts to use physical means such as brighter street lights to prevent delinquency and have exerted more pressure for cohesive programs such as forced literacy training as a condition of probation and parole in contrast with increased numbers of therapeutic communities and the burgeoning of street-worker projects.[9] The researcher if he is to be part of the action environment must share responsibility not only for the niceties of the evaluation design but for the general over-all frame of the agency.

Most to the point, however, is the researcher's reluctance in many cases to adopt a client perception of his work: many of us in our academic endeavors appropriately behave like small businessmen with a monopoly in a particular neighborhood. An entreprenuial style is not consistent with the research needs in these mass programs and our academic colleagues cannot be the only reference groups for our actions.[10] I contend that research in these mass programs must be regarded as client-sponsored investigations, and variables included in these studies need to take into account the client, i.e., the policy-makers who provide the means and support for the program, not the editors of our journals nor--and I am sure some of my colleagues will not like the remark--the recipient populations. The soap manufacturer is interested in how many boxes of soap he sells and not in whether infant children like the color of the box. In the same way, there is no reason for the policy-maker to be concerned with esoteric measures that might occupy some novel place in an academic field and thus, for example, whoever is paying for a reading program needs and deserves to be told whether it improves kids' reading and not whether it reduces hand tremors. The way it seems to me to approach this latter problem is to have available several conceptions of evaluation and it is to this matter that I would like to turn.

Conceptions of Evaluation

Most of our methodology in evaluation studies stems, of course, from much more highly simplified situations; experimental designs that have worked in agriculture, in the life sciences and in rat psychology. We must be concerned with efficacy of outcome and experimental studies are required in order to determine efficacy. Evaluation research efforts must seek to approximate the experimental model as much as possible--we do not do so often enough and some of the so-called evaluation designs of the current mass programs have completely foregone an experimental or quasi-experimental approach. Admittedly, there is a limit to the extent controlled experiments can be conducted in these programs but nevertheless it is possible in most instances to make use of at least rudimentary or quasi-designs. In the companion paper some of the barriers to experimental research in community programs will be discussed but it seems to me obvious that to do our job it is necessary, be it through randomization or statistical procedures, to approximate the conditions of before-after and/or pre- post test designs.

The situation becomes exceedingly complicated, however, given the need to evaluate a linked input-output system rather than being able to examine specific independent-dependent variable situations. But the kinds of massive efforts going on are of a linked input-output type and it is necessary to measure efficacy of each of the specific programs, the interactions between programs, and to be able to tie together by means of relational analysis the impact that changes due to sub-programs have on the over-all program objectives. For example, an educational program may be designed to improve reading and this must be assessed but if the over-all objective of the community project is to reduce school drop-outs, the relationship between reading improvement and drop-outs must also be demonstrated.

Of the many problems we are up against in the utilization of experimental models, the linking issue seems to me to be the most difficult. Among other things, we are too willing to make use of reflectors of program change rather than program change itself, such as employing shifts in attitudes toward Negroes when the program is concerned with reducing discrimination. Most of us are aware of the limited correlations often found between attitudes and behavior, but as a recent paper points out, the situation may be worse than that: reanalysis of several studies suggest that changes in attitudes may be inversely correlated with changes in behavior. Thus, if one may extrapolate, reducing prejudice may indeed lead to increasing discrimination.[11] Use of attitudinal reflectors may therefore render impossible the linking process.

---

[9] Admittedly, the evidence about the latter two approaches is fragmentary but nevertheless hardly in the direction to encourage the current expansion efforts. See Charles Perrow, "Hospitals: Goals, Structure and Technology," in James March (ed.), Handbook of Organizations, New York: Rand McNally, 1964 and Walter B. Miller, "The Impact of a Total-Community Delinquency Control Project," Social Problems, 10 (Fally, 1962), pp. 168-191.

[10] Howard E. Freeman, loc. cit.

[11] Leon Festinger, "Behavioral Support for Opinion Change," Public Opinion Quarterly, 28 (Fall, 1964), pp. 404-417.

Another dilemma comes about because of our reluctance to keep program efforts within the bounds that permit the maintaining of adequate control groups. In many instances we fail to keep the sub-programs at a level that makes possible the establishment of control groups and thus community programs for previously institutionalized delinquents, nursery-school programs for children of working mothers, and home-maker efforts for the aged are often implemented in such ways that though there are administrative opportunities for randomization nevertheless do not occur. I believe Dr. Sherwood will elaborate this point but I must note that if the "problem" group is so small that some subjects cannot be reserved for a control group, one should question its appropriateness in these mass programs.

Given the size of community efforts under this poverty program, assessing the efficacy of each sub-program in every city is pretty well impossible. Even assuming the availability of research funds, the problem of obtaining necessary professional manpower renders this an unworkable task. Consequently it probably is necessary to sample programs in various cities and this raises knotty problems because of the already-made observations of the linked input-output character of these programs. Sampling must be attempted in terms of the selection of linked programs and the sampling unit needs to be a sub-system of linked programs, analytically if not actually distinct. For example, if one of the goals of a day-care program is to free unmarried mothers so they may receive work-skill training so they may be eligible for employment counseling and training, this "sub-system" of programs must constitute the sampling unit. In order to sample such linked programs, however, it is necessary to have explicit statements of the goals and linkages of the various parts of the community-wide efforts and the necessity to sample speaks too for the need of well-formulated conceptual frameworks for the current efforts.

I say, however, that we have no alternative but experimental evaluation. Should we demand less in terms of the treatment of community problems than we call for in the provision of medical care for ourselves or our pets? Despite the problems of limited sampling and of validity and reliability, many of us seek out Consumer Reports before making major purchases and a few of us even query our physicians about the efficacy of his intended therapies. We reject notions of "intuitive reasonableness" and "impressionistic worth" and seek out comparative assessments before purchasing major consumer items and we have the responsibility to stand for such evaluations in these mass programs as well.

At the present time even the most basic aspects of these efforts are open to question. Many of the mass effects, for example, are heavily committed to community organization programs and to the stimulation of expressive actions on the part of the so-called deprived populations. These programs have as in the case of New York City's Mobilization for Youth been a major source of controversy and yet, despite the resources expended and the conflict occasioned by them, at present they cannot be condemned or condoned in terms of objective evidence.[12] It is possible to mass opinions pro and con but such major issues cannot be settled given the current state of things though it is thirty years ago that community experiments were attempted by a social scientist in Syria.[13]

Let me move on to the second requirement of evaluation research, namely, that of accountability. By accountability I mean evidence first that there is indeed a target population that can be dealt with by means of a program; second, that this population is important either because of its size or the intensity of pathology; and third, that the project program for the target population is undertaken actually with them.

It is not enough to evaluate efficacy--the outcomes of programs--the massive efforts now underway need to be evaluated in terms of accountability as well. While one might be accused of being inhuman for saying it, given the needs, there is little excuse for sanctioning action programs that effect insignificant portions of the population. One of the aspects of accountability is the estimation of the incidence and prevalence of problems. Oftentimes, I am afraid, programs are developed to deal with problems that exist in the minds of practitioners or because of stereotypes held by the public.

---

[12] Roland L. Warren, "The Impact of New Designs of Community Organizations," paper presented at the annual meetings of the National Social Welfare Assembly, November 3, 1964, New York City.

[13] Stuart C. Dodd, A Controlled Experiment on Rural Hygiene in Syria, Beirut, Lebanon Republic; American Press, 1934.

If I may cite one illustration, it is the drug-addiction problem: despite newspaper and public alarm, the incidence in many urban centers is so low that on accountability grounds these efforts hardly merit the attention of so many or the utilization of extensive research resources to evaluate them. If small-size programs use up all the potential clients, then it appears to me only under very unusual circumstances may the researcher be justified in collaborating in their evaluation or even attempting to do so. If the programs are of a large-scale type, than the denial of services or the provision of "ordinary" treatment to a few for control purposes and subsequent estimation of worth is entirely necessary.

Accountability, however, has to do with more than the number of clients served and the size of the potential aggregate of them. Evaluation researchers, in addition to a responsibility for determining efficacy, must deal with the implementation of the prescribed process. In many instances we have engaged in outcome studies without having any knowledge of whether or not what program people maintain is going on actually takes place. This point, I am sure, will be expanded on in the next paper, but it is clear that in many of the sub-programs being implemented as part of these massive efforts--even when evaluation studies of the finest design are accompanying them--we are estimating the utility of programs that never get off the ground; evaluating programs in which volunteers do no more than sign up or week-end educational camping programs in which kids have a good time and do nothing more than play ball or eat marshmallows around the fireside. To say a program fails when it is not truly implemented is indeed misguided, and the evaluation researcher's responsibility here is one of providing evidence and information that permits an accounting of what took place as well as what was the result.

Finally, what we hardly ever worry about, to my knowledge, is efficiency.[14] The various specific programs that are linked together in these massive packages differ extensively in target groups, use of scarce resources and duration. At the risk of being ludicrous, suppose <u>neither</u> individual psychotheray <u>nor</u> group psychotherapy has any impact on the lives of persons but the former costs ten times that of the latter, given such a situation I know where I would put my money. In certain fields of medicine and in certain areas of welfare there is literally no way, given the community's ideological outlook, to cease all treatment even if no efforts are efficacious. But without being too cynical, even when we know this is the case, we refuse to employ a concept of efficiency. Suppose short-term treatment institutions for delinquent offenders do no better than long-term ones, if they are more economical is this not something that the evaluation researcher has a responsibility to take into account?[15]

In terms of all programs, the efficient one is that which yields the greatest per unit change not the one that can be run at the least cost per recipient. What costs the most, takes the longest, and involves the greatest amount of manpower in gross terms may have the greatest net efficiency.[16] Decisions on the continuance of various programs beyond trial--demonstration periods require that we think in these terms. In most evaluation efforts, I would argue we fail to make use of per capita costs, per manpower costs or per time measures.

I would contend that concepts of <u>accountability</u> and <u>efficiency</u> as well as <u>efficacy</u> need to be implemented in order for evaluation research to be properly undertaken. Admittedly, we ought to seek out efficacious programs. But these programs are or at least should be accountable in order for policy and program persons to make rational decisions, and we must also concern ourselves with efficiency of operations.

Concluding Comments

I hope my remarks, though not entirely original, of course, may prove relevant for researchers who have occasion to participate in the evaluation of community-wide programs. The need to become engaged in the action environment, to look at a linked input-output system and to insist on experimental designs, and the necessity to assess efficiency and to recognize the accountability function in evaluation are, to my mind, key points and ones not well-documented in our methods books and not always held to by persons participating in the evaluation of these massive efforts.

---

[14] An illustration of a study that does consider this problem is Julius Jahn and Margaret Bleckner, "Serving the Aged," (Methodological Supplement--Part I), New York: Community Service Society of New York, 1964.

[15] Howard E. Freeman and H. Ashley Weeks, "Analysis of a Program of Treatment of Delinquent Boys," <u>American Journal of Sociology</u>, 62 (July, 1956), pp. 56-61.

[16] Clarence C. Sherwood, "Social Research in New Community Planning Organizations," paper presented at the National Conference of Social Welfare, Cleveland, 1963.

But I would like to feel that I have communicated more than some specific observations--that I have conveyed the potentiality of our role as change agents and the sense of conviction, commitment and responsibility required on our parts. At no other point in time have we had so great an opportunity to have an impact on the social order; if we are to realize our potential within our current stance as social scientists, however, we need more than additional technical innovations. It is an outlook, an ideology, almost a morality if you will, that we must develop in order to function appropriately as agents of social change.

METHODOLOGICAL, MEASUREMENT AND SOCIAL ACTION CONSIDERATIONS RELATED TO THE ASSESSMENT OF LARGE-SCALE DEMONSTRATION PROGRAMS

Clarence C. Sherwood, Action for Boston Community Development, Inc.

The roles which the social scientist can and to some extent is playing in influencing social action are, in my opinion, increasing in both variety and in importance. To a considerable extent, the effective performance of those roles will depend upon the availability of valid, applicable knowledge concerning which social action programs, or interventions, work and which do not. I am personally convinced that the most rational and in the long run the most effective way of accumulating such knowledge will be by conducting action programs as controlled experiments in the community.

I am well aware that the day when such findings will play a dominant part in the decisions concerning the modification, expansion or discontinuance of action programs is probably not in the immediate future. But, I also believe that there is a rapidly developing realization within broad sectors of our population that this kind of knowledge is needed and that the basic questions concerning what works and what doesn't work will be raised with increasing frequency. It is generally recognized, however, that at the present time the social sciences do not have a vast reservoir of findings which are clearly applicable for making decisions about how to attack broad scale social problems. What may not be as clearly or easily recognized is that the social sciences have not had a great deal of experience with large scale social-action programs nor have they thought through how to go about conducting such programs in a way that reasonably hard findings concerning their efficacy can be extracted. We are in a relatively infant stage not only with respect to knowledge of the effectiveness of action programs but also with respect to solutions of the conceptual, theoretical and methodological problems pertaining to the acquisition of such knowledge.

The importance and seriousness of this state of affairs has been augmented many fold by the advent of the multi-billion dollar War on Poverty. Although it is not clear at this moment as to how much emphasis will be placed on the evaluation of anti-poverty programs, there are rumblings that some of the resources will be used to try to ascertain the effects of the attack, or at least of parts of it. I think that social scientists and professionals in related fields should press for a maximum effort in this direction and then organize their own resources toward a major contribution in developing methodologies to fulfill the promises and expectations involved. If some serious effort along these lines is not made, literally billions of dollars will have been spent and we will still not know what works or how to go about finding out about it.

However, no real support for the pursuit of this kind of knowledge is going to be forthcoming until the case for it is presented in practical, useful terms. My recommendation is to point out that a major use of findings from social-action experiments is to provide a basis for a more efficient allocation of financial and human resources to the solution of social problems.

It is this notion of the efficient allocation of resources that I believe is the key to the whole problem of planning and choosing among social-action programs. If this is true, we need a basis for making judgments about efficient allocation, and for this purpose I suggest that we examine very carefully our whole concept of social action or social service programs. Traditionally, service has been viewed--and in a vague way measured--in terms of that which is offered--counseling, guidance, therapy, advice, and the like. Good service is therefore that which is offered in a professional manner by a qualified person who in turn is supervised by a qualified supervisor. My contention is that service must be viewed in terms of impact rather than process. Its success must be viewed in terms of outcome rather than in terms of the quality of the procedures used. The implications of this shift in view are, in my opinion, considerable:

(1) It forces those responsible for program design to clearly specify their objectives, to define what it is they are trying to achieve, what specific changes they are trying to effect At the very least, it requires them to co-operate in efforts to operationalize what they have in mind;

(2) It shifts the emphasis from "procedure as an end" to "procedure as a means." Program personnel must then consider the relationship between the procedures it recommends and the defined outcomes that have been chosen;

(3) It leads to a reconsideration of the whole notion of cost of service. Currently, we are in the grip of the proponents of the "per capita cost of service" point of view. Per capita cost of service is defined as the quotient obtained by dividing the total dollar cost of a program by the total number of individuals exposed to it. If, for example, a particular youth employment program involves 1,000 youths and costs a half million dollars a year, traditional calculations would say that the per capita cost of service is $500. But what if, as well might be the case, only 50 more of the 1,000 youths ended up working steadily (if that is the objective of the program) than would have been the case without the intervention. Calculated on the basis of an impact definition, the per capita unit cost would be $10,000--a vastly different amount. The per capita cost view has resulted in a distinct preference for those programs which "serve" the largest numbers for the least amount of money. If programs were to compete on the basis of how much it costs to achieve one unit (however that may be defined) of desired outcome, our ultimate selection of programs would be, I believe, very different;

(4) And, finally this view forces the inclusion of solid, empirical research into the over-all planning and program operation, because the decisions as to the optimum allocation of the resources available can, within this view, only be made on the basis of empirical evidence.

The challenge is then, at least to me, very clear. Can social experiments be conducted in the community in such a way that the findings resulting therefrom deserve the attention of the policy and decision makers in the community? And what are the current and long-run obstacles to the implementation of action research that prevent it from meeting standards of scientific acceptability?

About two years ago I made the plunge and took on the job of attempting to develop an evaluation program for a large scale community action program, called ABCD (Action for Boston Community Development), in which we are attempting to build research into action program designs. To my knowledge, no similar large scale evaluation research effort had been made before, although variations on the same theme were being developed in New York City, Cleveland, Los Angeles, New Haven, Oakland, Philadelphia, St. Louis, Chicago, Detroit, and a number of other cities and communities. The current state of my thinking is largely the product of my participation in this rather ambitious effort, and therefore, in the analysis which follows I will draw heavily upon my experience for illustrative materials as well as for insights and possible lessons to be learned.

Therefore, before discussing the problems of evaluating large scale social action experiments, let me tell you a little bit about ABCD and the way in which we have tackled this task.

ABCD was envisioned by its founders as the human side of urban renewal. Its first staff people were community organizers and one of their major tasks was to aid in explaining renewal projects to the residents of renewal areas in an effort to enlist their support.

A Ford Foundation grant and a grant from the President's Committee on Juvenile Delinquency and Youth Crime gave ABCD an impetus in another direction. Since then, and until recently at least, our focus has been on developing demonstration programs to be carried out by public and private agencies in the three areas of Boston with the lowest income, highest delinquency and the most deteriorated housing.

ABCD is not designed as a permanent agency. Its function is to act as catalyst in the development of a broad attack on the social problems of Boston. Congruent with this design, ABCD does not intend to operate programs itself. Its approach is to aid in designing programs, to provide part of the financial support required to get them underway and to conduct research to determine whether or not they achieved their chosen ends.

The problems which ABCD posed for itself can be briefly summarized as follows: Can a set of interventions be designed and implemented in such a way that they appear on theoretical grounds to have some chance of reducing the delinquent behavior of youth in the community and of producing knowledge as to whether or not the interventions did have their desired effects? The reference term which we have used to identify community efforts to achieve those goals is "action-research demonstrations".

Let me try to define briefly some of the distinguishing features of this notion, an "action-research demonstration." If we leave out the adjective action-research, for the moment, and focus on the term demonstration, this term can be used to refer to a broad category of social effort, the principal theme of which is that the programs are designed and conducted on a trial basis and do not purport to represent total solutions to a social problem. Demonstrations are therefore knowledge seeking efforts. However, they vary from those which define success in an administrative sense--was the program workable administratively?--to those which define success in terms of effect--did the program produce the changes it was designed to produce? Demonstrations also vary from those which define "knowing" in terms of the judgments of "experts" to those which define knowing in terms of the outcome of controlled experimentation. And they vary from those where the knowledge seeking effort comes after the action part of the demonstration has terminated to those where it is built into the demonstration from the start. Action-research demonstrations are those which seek knowledge concerning the effects of the program through use of a controlled experimental design which is built-into the total demonstration effort.

Therefore, it is assumed in the following discussion that what we are after are findings from action-research demonstrations. It is also assumed that the minimum criteria for the acceptability of those findings includes:

(a) sufficient information so that the program component is repeatable;

(b) knowledge concerning whether the program produced the effects it was designed to produce; and,

(c) knowledge concerning whether the specific ingredients of the program were in any way necessary to the production of those effects.

All social experimentation is likely to encounter some difficulty with respect to each of the above criterion areas; these difficulties are, I believe, considerably aggravated in the case of community-based, action-research demonstrations, and even more so when they are conducted on a broad scale.

We conceptualized our action-research demonstration project at ABCD in terms of three sets of variables. One is the dependent variable of the project--in ABCD's case, this variable is juvenile delinquency; more specifically defined as law-violating behavior of 12 through 16 year old males residing in specific areas of Boston. The second set of variables are referred to as the intermediate variables. According to the Project's hypothesis, changes in the intermediate variables should produce desired change in the dependent variable. The third set consists of the program variables--the specific interventions by which it is hoped to produce changes in one or more of the intermediate variables.

Therefore, the project has two fundamental, interrelated tasks. One is to find ways to produce the intermediate changes which the hypothesis asserts will be followed by desired changes in its dependent variable. The second is to determine if, when such intermediate changes occur, they are in fact followed by the desired changes in the dependent variable.

These two basic research questions:--(1) did the intervention produce the desired change in the intermediate variable? and (2) were changes in the intermediate variable, if they occurred, related to changes in the dependent variable?--appear to involve fundamentally different methodological difficulties.

The first involves all the difficulties inherent in efforts to implement an experimental design plus the complexities and difficulties imposed by the fact that in ABCD's case an attempt is being made to implement experimental designs in the community. The second question--whether changes in the intermediate variables are related to changes in the dependent variable--involves a number of additional problems, including all the measurement and statistical problems of attempting to relate changes in variables.

As a backdrop against which to explore some of these difficulties, one of the ABCD programs will be described in some detail--the Week-end Ranger Program. A regular summer camp site and its facilities are being used as the setting for a program for already delinquent boys--they must be on probation to be eligible for the program--with the ultimate aim of reducing their subsequent delinquent behavior. In this program, approximately 30 boys leave the community each Friday afternoon and travel about 60 miles on a bus to the camp site. Between then and Sunday evening when they return, they participate in an organized series of activities, including discussion groups, council meetings, work activities and recreational programs.

Briefly, the over-all design is as follows: arrangements were made with the State Probation Commission whereby the local probation offices in parts of the City of Boston provided lists of names of boys who were eligible by reason of age, residence and other criteria for participation in the program. These boys were asked to come to the Probation offices and participate in a study.

Upon their appearance at the office, the boys were pre-tested on several attitude scales--an anomie, an alienation and a values scale. The theoretical tie-in involves the possible relationship between what might be called a disengagement of delinquent boys from the values and institutional system of the dominant society, on the one hand, and their delinquent behavior, on the other. An attempt has been made to build procedures into the program which appear to have some hope of changing the attitudes of these youth and ultimately, according to the model, their on-the-street behavior as well. After pre-testing, the youth were randomly divided into two groups and the members of one group were invited to participate in the week-end program. The members of the other group were designated as ineligible for the program.

Remembering that our aim is not only to know that certain effects were obtained but is also to know with some degree of probability that the effects were substantively related to a particular set of stimuli, one major problem confronting efforts to evaluate programs of this type is the problem of controlling the stimulus. In my opinion, real strides toward the accumulation of definitive knowledge about the effects of programs will not be made until we are able to think through and develop procedures for handling the whole problem of what constitutes the stimulus. The basic question is: What is it that should be repeated if the program appears to work? There are two related but nevertheless operationally separate issues here. One is the design of the stimulus or intervention. The other--and perhaps the more difficult one--is the problem of monitoring of the intervention.

Compared with classical conditioning experiments or even somewhat more intricate experiments such as those involving exposing populations to movies aimed at changing attitudes toward minority groups, the "stimulus" in a program like the Week-end Ranger Program is exceedingly complex. We started out at ABCD with very definite and clear-cut intentions to conduct and evaluate "repeatable" programs. But we admittedly grossly underestimated the difficulties which are involved in both designing and monitoring programs with the goal of repeatability in mind. It is becoming clear to us that this problem cannot be satisfactorily resolved by simply reducing it to the problem of spelling out procedures in great and specific detail as difficult as even that may be. The direction in which we appear to be heading in our efforts to deal with this problem and (needless to say, we are nowhere near to solving it) is toward the development of principles rather than procedures. What this has pushed us toward is the notion of what I call an "impact model,"--a set of theoretical concepts or ideas which trace the dynamics of how it is expected that the program will have the desired effects; a theory which logically interrelates a set of principles and procedures with desired outcomes. If the impact model is sufficiently worked out, a set of working principles becomes available upon which practitioners can draw not only for the design of programs but also to make practical decisions about day-to-day program situations.

For example, in the Week-end Ranger Program there are different tasks to be performed, the boys must be allocated to work groups in some way, and the problem of the non-worker in the work group must be dealt with. Should the boys choose the task they work at? Should they choose with whom they will work? And, how is the non-worker to be handled? The point I am trying to emphasize is that a satisfactorily developed impact model would logically imply that certain decisions rather than others be made with respect to such problems. Greater adherence to the development of such models should (a) enhance the probability that such programs may have an impact; (b) provide a basis for training program personnel that does not require that a program procedure be specified for every conceivable situation; (c) provide a basis for outside monitoring of the program; (d) provide a rational basis for modifying the program design should it appear that it does not have the desired effects; and (e) provide the basis for a repeatable program which goes well beyond the mere rote repetition of isolated procedures.

On the more positive side, the Week-end Ranger Program illustrates that it is possible to conduct a reasonably well controlled experiment in the community which involves the co-operation of a number of individuals and agencies. We were able to institute even a modified version of randomly allocating subjects to a treatment and a non-treatment group. I say modified because not all of those randomly selected for the experimental group agreed to participate in the program, and therefore the exposed and the unexposed populations do not constitute two truly random samples from the same population. In addition, we were able to obtain the necessary co-operation for rather extensive pre-testing of both experimental and control youth. It is likely that some version of a pre-post test design is going to be necessary in such experiments because of this element of voluntary self-selection to participate on the part of the experimental group. Thus we are eventually going to have to (and because of this co-operation we will be able to) rely on covariance adjustments to bring the experimental and control groups back into line.

It is worth noting, however, that a main reason we were able to get support for the randomization procedures was because of the very limited number of openings in the program. But there is still great public resistance to and considerable lack of understanding about randomization. This problem is likely to be even more serious in the case of really massive programs in which there appears to be room for everybody. This is likely to be particularly true where randomization to non-treatment groups is involved.

Furthermore, in addition to the ever present abhorrence of "denial of service" there is a very strong proclivity on the part of practitioners to believe that they know which type of person will benefit most from a particular program. Therefore, co-operating practitioners designate more people for a program than there are openings only with great reluctance. There is also a related tendency for practitioners to want the most deserving youth to receive the opportunity to participate in special programs. Unfortunately, in the Week-end Ranger Program it is presently impossible to determine the extent to which these two tendencies are operating in the selection of candidates for the program.

There are two basic problems here which relate to potential findings. One is that if only the most deserving are selected--even from among probationers--the possibility of program impact may be lessened because both the experimental and the control subjects may fare very well according to the outcome criterion. The other problem is that when the selection is left to the personal preference of the practitioners the representativeness of the demonstration population relative to some larger population will be unknown.

I think there are several possible lessons here of relevance to the evaluation of anti-poverty programs. One is that it is already clear that the overwhelming pressure is going to be on doing rather than evaluating. This one-sided emphasis is unfortunate. Be that as it may, random allocation to treatment and non-treatment groups is not likely to be frequently possible. But, random allocations to alternative treatments may be. This means, however, if such an approach is to be carried out well, the alternative treatments should be thought through very carefully so that at a minimum they are different and not camouflaged versions of the same basic idea. The impact model--the set of theoretical concepts or ideas which trace the dynamics of how it is expected that the program will have its desired effects--again rears its annoying head, and in turn a hard look at what the goals--the outcome variable --of such programs are and how to measure them will be required. Since even the broad scale anti-poverty programs are not likely to be any better off as regards knowledge of the representativeness of the populations they will be dealing with and they are also going to have to face the problem of self-selection for participation, extensive pre-testing with good instruments is going to be a must if anything resembling definitive findings is to emerge. Not only should there be common use of some of the same instruments across similar programs within communities but also across similar programs between communities. For the first time we might have some cross-community comparative material concerning the populations being reached and the changes being observed.

Just as there are conceptual and methodological problems (some of which have been discussed previously) in attempting to evaluate the effects of programs on their direct outcome variables--what we have called intermediate variables--there are other, perhaps even more perplexing, ones involved in attempting to relate changes in these intermediate variables with the dependent variable of the over-all project.

The ABCD Youth Opportunities Project hypothesis states that certain changes will be followed by certain other changes. The programs are designed to expose members of the target population to procedures which will hopefully produce changes in the individual or his environment. Each of these changes is expected by the hypothesis to produce an increment of improved behavior--less law violation--on the part of the individual. It is the Project's hope that for each program significantly more of the experimentals than the controls will experience the desired change and those experiencing such change, whether they are experimentals or controls, will manifest a reduction in law-violating behavior.

It must be re-emphasized that the

hypothesis asserts a relationship between two sets of changes, not between two static conditions. Using the Weekend Ranger Program as the example again, the hypothesis does not assert that low anomie or alienation scores or high value scores will reduce the law-violating behavior among those manifesting score changes in specified directions.

The problem of obtaining reasonably reliable change measures preceeds the problem of relating change measures, since attempting to relate sets of unreliable change scores does not appear to be too promising a game to play. There has been, of course, a long-standing concern for the problem of the reliability of scores. Interest in the reliability of change scores is somewhat more recent and is receiving increasing attention among statisticians and psychometricians. Problems arising out of the mathematically demonstrated greater unreliability of change scores relative to the reliability of the scores from which they were derived and problems arising out of demonstrated regression to the mean tendencies in test-retest situations are likely to remain central as well as difficult issues for those who are brave or foolish enough to pursue this change problem.

The problem of the relationship between sets of change scores has, to my knowledge, received little consideration in the literature and undoubtedly also involves serious statistical and mathematical difficulties. Measurements of each variable at a minimum of three points in time are required to provide some estimate of the shape of the curves involved. Two of the problems involved are (1) the relationship between the shapes of the curves--the change curve for the intermediate variable and the change curve for the dependent variable--and (2) the question of the time lag throughout the series and between the two sets of changes: When are the presumed effects of the program on the intermediate variable expected to take place? While the program is going on? After participation in the program has terminated? And for how long are the effects supposed to last? How long a time is expected to lapse between the changes in the intermediate variable and their presumed effects on the dependent variable? What are their relative rates of change? These and similar questions are directly related to some very practical issues such as the amount of success a project can possibly have during some specified demonstration period. If there is considerable lag or the rate of change in the dependent variable is relatively low, much of the effects of the demonstration may take place after the cut-off point for the evaluation of the Project. Again the need for a theoretically based impact model is, it seems to me, underscored.

Of the many other problems which beset efforts to conduct and evaluate large-scale action programs, there are two that I would like to bring to your attention in the short time remaining. One is the problem of the meaning of change in the dependent variable--in our case, a reduction in law violating behavior--and the other is the problem which arises from the fact that members of the target population may--in fact, undoubtedly will--get involved with more than one of the programs and that this involvement is non-random.

The first decision we made concerning the definition of change in the dependent variable was that we could not use time comparisons of area rates of delinquency as a basis. ABCD's aims were to change behavior, not to move law-violating people out of an area and non-violating people into it. Therefore an area delinquency rate comparison over time was rejected as a basis for measuring change since wide variations in delinquency rates may occur over time simply because of changes in the constituency of the population. It was decided that a reduction in law violating behavior would have to be measured in terms of the behavior of a specified population--that is, a cohort of individuals.

Another major problem in defining how change in the dependent variable was to be measured is the known relationship between age and delinquency. Beginning around 10 or 11, age-specific delinquency rates increase rather sharply up into and through the late teens. Therefore to simply compare a given individual's behavior at age 15 with his behavior at age 14, 13, 12 and so on would lose sight of the fact that the probability of a delinquent act increases as he gets older. If a cohort of 15 year-olds committed the same number of delinquent acts at age 15 as they did at age 13, for example, this might not look like a reduction--and in terms of absolute numbers it is not--but in terms of what might have been expected of them it is. We arrived at what we felt to be an inescapable conclusion--within the framework of our approach--namely, that a reduction of law violating behavior must be defined in terms of a comparison of an observed measure with an expected measure. That is, a prediction instrument is required to provide an estimate of the law violating behavior which would have occurred had there been no intervention.

I suspect that very similar problems will arise if efforts are made to take a hard look at the possible effects of various components of large scale community efforts to deal with poverty. To take one variable, employability, which is central to most of the poverty proposals I have heard discussed, this dimension or characteristic of individuals is also a function of age. For example, it is quite well known that the great bulk of the very difficult to employ 16 to 21 year-olds begin to disappear into the job market and from the unemployment roles as they approach their middle twenties. Therefore, if evaluations of community programs dealing with this particular segment of the population are based upon observations of their employment history subsequent to exposure to one or more anti-poverty programs, the success observed may be much more apparent than real. What is needed is a measure of their employment status and prospects at some point in time as compared with estimates of what would have probably been the case at that same point in time had there been no intervention.

The last issue that I have time to discuss with you is that of multiple-exposure to programs. This has presented the ABCD Youth Opportunities Project with distinct methodological difficulties. It is likely to be an even greater problem for any effort to evaluate the effects of anti-poverty programs. Two tendencies combine here, I believe, to aggravate the problem. One is the inclination on the part of practitioners to want to shower programs on the members of the target population. The other is the sheer amount of money that is involved and the resulting large number of programs that are likely to be conducted. This is an extremely important issue if we are serious in our desire to ultimately acquire knowledge concerning the most efficient allocation of human and financial resources. For if the members of the target population participate in a number of different programs and even if desired change occurs and is measured, there must be a way devised to sort out the relative contributions of the different programs to the outcome. Otherwise, in order to produce the same results again the whole menagerie of programs would have to be repeated even though only a relatively few of the programs may have actually contributed to the desired outcome. Again, a cohort and a prediction instrument appear to be indispensible to the solution of this problem. Individuals must be grouped according to the programs they have participated in--in our approach, according to the intermediate variable changes they have experienced--and then the groups compared on the differences between observed dependent variable and expected dependent variable behavior.

In summary, I have tried to sketch for you some of what I believe to be the major issues facing action research. Roughly, these issues fall into three categories: (1) general action-research administrative problems; (2) general over-all action-research design problems; and (3) basic methodological problems. I would include in the first category the need for a much more vigorous effort to increase general understanding of the need for this kind of effort and of the function of some of its tools, such as randomization. I would also include here the careful development of a kind of contract format whereby the various parties involved would understand and commit themselves to specified goals, principles and procedures. I would include in the second category the need for the development of impact models and the recognition that action-research is going to require a delicate blending of action ideas, theory and research technique. And, in the third category I would include the need for solutions to the problems of the measurement of and relationships between changes and the development and application of prediction instrument technique.

X

EVALUATION OF PANEL OPERATIONS

Chairman, Leon Pritzker, U. S. Bureau of the Census

Page

Does a Panel Operation Increase the Reliability of Survey Data: The Case of Consumer Savings - Robert Ferber, University of Illinois........... 210

The Current Population Survey: A Case History in Panel Operations - Joseph Waksberg and Robert B. Pearl, U. S. Bureau of the Census.............. 217

Discussion - Philip E. Converse, University of Michigan............................ 229

Does a Panel Operation Increase the Reliability of Survey Data: The Case of Consumer Savings

Robert Ferber, University of Illinois

The effect of a panel operation on the reliability of survey data has long been a subject of conjecture in sampling circles. Virtually all past work on the subject has focused on measuring these effects indirectly, for as a rule it is only such evidence that can be brought to bear on the question. Hence, the approach in these past studies has been to evaluate panel effects in terms of changes in one or more of these basic forces, namely, changes in population composition, panel mortality and panel conditioning effects.

With regard to estimates of population parameters--usually means or aggregates--the effect of the first two of these factors is likely to be in the direction of reducing the reliability of panel data over time. Thus, changes in population composition (which usually are not paralleled by changes in the composition of the panel), and continuing panel mortality, tend to widen any differences that may exist between the composition of the panel and the composition of the population; and this in turn presumably brings about increased differences between the parameter estimate obtained from the sample and the true value of that parameter.

Moreover, panel mortality, if not offset by new families, tends to reduce sample size. For the same sample design, this serves to increase sampling variances and bring about greater instability in the parameter estimates.

On the other hand, it is well recognized that panel conditioning may have positive or negative effects. In some instances, the repeated interviewing of the panel, with repetition of the same questions, serves as a learning process to educate the respondents in better reporting techniques. In other instances, the repeated interviewing may make respondents less interested or perhaps less inclined to give accurate information, with the result that the reliability of the survey data declines over time.

No attempt is made here to evaluate the various studies exploring these effects, once consideration is given to the different circumstances under which the studies were carried out. Rather, the present paper presents the results of a direct analysis of the problem, with reference to the subject of consumer savings, made possible by the availability of data throughout the course of a panel operation which provided, unknown to the field force and the respondents, true values of some of the variables being requested in the interviews.

This panel operation, and several other panels as well, were carried out as part of the Consumer Savings Project of the Inter-University Committee for Research on Consumer Behavior, with financial assistance from the Ford Foundation. The present paper focuses on the effect of one of these panel operations on the accuracy of time deposits reported by panel members over the period of a year.

The panel operation which served as the source of data for this analysis was designed specifically to study response and nonresponse errors in individuals' reports of time deposits. Sample members were selected by random probability chunk selection from the files of savings institutions in a large metropolitan area. All accounts of individuals selected in the sample were listed separately. A stratification procedure by account size was also used, so that at the time of sample selection equal numbers of sample members had accounts under $100, $100-999, $1,000-4,999, $5,000 or more. (For multiple account owners, the largest account size was used in each case.)

Five waves of interviews were scheduled with the sample members, the interviewing interval being roughly three to four months. The questionnaires used in these interviews were general in scope, covering the entire range of consumer assets and debts. In meetings with the interviewers and in contacts with the respondents, the asset-management aspects of the study were stressed; neither the interviewers nor the respondents were informed of the source of sample selection.

The validation of account balances was carried out after each wave of interviews, data being obtained each time for nonrespondents as well as respondents. Because of the panel nature of the operation, various safeguards could be incorporated against mismatching, which is therefore probably negligible in the present study.

As with other operations of this Project, certain controlled experiments were incorporated into this panel operation, mostly in the first wave. Only one such experiment is relevant for present purposes. This experiment consisted of asking half of the panel for dollar holdings as of the date of the interview and the other half of the panel for changes that had taken place in each holding since the time of the last interview. This distinction was maintained only for the first two waves of interviews; after the second wave all panel members were asked for holdings, and on the fourth wave they were asked in addition for changes since the last interview. This modification made possible comparison of accuracy of reported change against the accuracy of change computed as the difference between two holdings reports. It should be noted that dollar balances were requested as of the date of the interview on all waves and the use of records was stressed, so that memory errors would be minimized.

On the basis of cost and other considerations, the desired sample size was 300 completed interviews. Actually, 316 interviews were obtained on the first wave, of a total eligible initial sample of 411. After five waves of interviews, 205 members in the area were still in the panel. In terms of accounts, which constitutes the unit for this analysis, the initial sample totaled 462. The respondents on the first wave represented 345, or 75 percent, of these accounts. After five waves, the remaining respondents represented 269, or 57 percent, of the initially validated accounts.

Framework for the Analysis

Two effects of the panel operation on the reliability of data are considered here--biases in the estimates of the same parameter from one wave to another, and changes in the variance of these estimates over time. Three parameters serve as the basis for this analysis. They are:

1. The extent to which holdings are not reported; in other words, the ratio of nonreporters to total respondents.
2. The average balance per account, first, for the sample respondents, and second, for all sample members regardless whether they are interviewed on a particular wave.
3. The aggregate amount in validated accounts --the product of the number of accounts and the average balance per account. This approximates the statistic that would be sought in estimating aggregate balances in savings accounts.

In addition, information will be provided on the accuracy of the change reports.

Results

Nonreporting of holdings

Table 1 shows that more than one-fourth of the validated accounts were not reported by respondents on the first wave. The table also shows that among those who remained in the panel nonreporting of validated accounts dropped throughout the study to less than 10 percent by the last wave. The drop was most pronounced on the second and third waves and seems to have more or less stabilized by the fourth wave.

Perhaps most significant is the fact that, in an overall sense, the improvement in the reporting of validated accounts served to offset the loss of sample members during the course of the panel. As a result, coverage of validated accounts increased, despite the substantial mortality. This phenomenon is illustrated by Table 2, which shows that on the first wave, 218, or 63 percent, of the validated accounts owned by respondents were reported; on the last wave, 192, or 79 percent, of the validated accounts were reported. In other words, the number of accounts reported per sample member rose substantially, and was much more accurate after five waves than after one wave. A far higher percentage of nonreporters than of reporters on Wave 1 had dropped out by Wave 5 (Table 2).

The manner in which the information is requested--holdings or change--clearly influences the accuracy of the coverage (Table 1). Initially, nonreporting among those asked only for changes was much higher than among those asked for holdings. However, it is interesting to note that the rate of nonreporting fell much more sharply in the former group, so that by the end of the operation nonreporting among those initially asked for change was actually less (though not significantly so) than among those asked for holdings.

Two reasons appear to be mainly responsible for this phenomenon. One reason is the tendency noted in other respects for people not to report an account if little or no change takes place in the balance. In some instances, this is due to misunderstanding and in others to a feeling that such accounts "do not matter" in the context of reporting change. Second, strong evidence exists that the "change approach" tends to retain in the sample people who are not overly cooperative but are willing to cooperate more or less on a marginal basis, and hence are more likely to be nonreporters. The switch to asking these people for holdings appears to be an irritation sufficiently strong to induce many of them to drop out altogether, thus reducing sharply the number of nonreporters who are still active sample members. In the present study, this switch was made on Wave 3, which corresponds with the striking decline in nonreporting on that wave for the "change" group (Table 1), compared with the much smaller decline in nonreporting at the same time among the "holdings" group.

The latter tendency is also brought out when we compared what had happened by Wave 5 to the respondent reporters and nonreporters on Wave 1. Among those who were initially asked for holdings, nonreporters tended to remain nonreporters throughout the study. However, among those asked initially for change, a much larger proportion dropped out during the course of the study, so that by the last wave fewer such people actually remained in the sample. In other words, the Wave 1 respondent nonreporters to the change form were basically less cooperative from the beginning and hence more likely to refuse at a later stage.

Average Balances

The average balance of validated accounts reported by the respondents initially understated the true average balance in all validated accounts owned by the respondents. However, the degree of error declined fairly uniformly throughout the operation, from roughly 10 percent on Wave 1 to virtually zero on Wave 5, as is evident from Column 2 of Table 3. This table includes those panel members interviewed initially for holdings information.

It is interesting to note, from this table, that the average error in balances reported by respondents was mostly negligible throughout the study, registering if anything a slight trend

Table 1

Nonreported Accounts as Percent of Total Validated Accounts of Respondents, by Wave and Type of Initial Form

Wave		Holdings requested	Change requested	Total sample
1		27%	32%	30%
2		16	21	19
3		13	12	13
4		13	9	11
5		11	7	9
Base:	Wave 1	169	176	345
	Wave 5	139	124	263

Table 2

Status of Wave 1 Respondents by Wave 5

	Wave 1 Status	
Wave 5 status	Respondent reporter[a]	Respondent nonreporter
Respondent reporter[a]	74%	24%
Respondent nonreporter	2	37
Nonrespondent or drop-out	24	39
Total	100%	100%
Base	218	127

[a] Includes account reported but balance refused.

Table 3

Reported and Actual Average Balance Per Validated Account of Respondents by Wave

(1)	(2)	(3)	(4)	(5)	(6)
		Average balance			
Wave	Percent understatement of respondents' average balances	Reported	Actual, for balance given	Actual, for balance refused	Actual, for nonreporters
1	9.6%	$2,391	$2,446	$3,102	$2,949
2	15.9	1,975	1,944	3,106	3,078
3	6.3	2,172	1,938	4,162	3,141
4	-4.3	2,172	2,062	1,650	2,439
5	.6	2,213	1,985	2,747	2,920

toward overestimation by Wave 5. The reason for the understatements in the average balances of the respondents is the much higher balances in accounts which were not reported or for which the balance was refused; this phenomenon is illustrated by the last three columns of Table 3.

The disparity between, on the one hand, accounts which were not reported and accounts for which balances were refused and, on the other hand, accounts for which balances were given, continued throughout the study. However, as has already been shown, the number of nonreported accounts declined substantially throughout the operation, much more so than the total number of respondents, with the result that the bias due to the omission of these accounts decreased over time.

In the case of the reports of change, the errors were far more substantial, as one might expect, because of the relatively low bases on which error percentages were calculated. Average errors of 100 percent or more in estimates of average change were common, the direction of the error being invariably toward understatement of change. Moreover, contrary to the case of holdings, no reduction in these errors over time was apparent. On the other hand, in the two instances where accuracy of computed change--the difference between successive holding reports--could be compared with the accuracy of reported change, the former approach was markedly superior, registering average errors of roughly half the size of the average errors in reported change.

Contrary to the situation with respondents' accounts, the average balance of accounts of all the panel members was understated consistently throughout the study. The degree of understatement was actually lowest on the first wave (21 percent), rose to a peak of 34 percent on the second wave, and then declined gradually to 22 percent by Wave 5.

As is evident from Table 4, the reason for this continuing understatement was initially the much higher balances in the accounts held by nonrespondents than in the accounts held by respondents. Since many of the nonrespondents later became drop-outs, the primary cause of the understatement on the later waves of the study shifts to the much higher balances of the drop-outs. In contrast, the balances of the nonrespondents are seen to fall precipitously.

Aggregate Balances

An estimate of the aggregate amount in validated accounts represents the type of statistic that would be sought in estimating aggregate holdings. Such an estimate can be obtained as the product of the average balance in validated accounts per sample member and the number of sample members. (Alternatively, this estimate could be derived as the product of the average balance per account and the number of such accounts.) Two such estimates were made in the present case, one estimate relating to the aggregate balances of the respondents on each wave, and the other estimate relating to the aggregate balances of all the members of the panel.

The error in the estimates of these two aggregate balances is shown in Table 5. Clearly, the error in both of these estimates of the aggregate balances represents a combination of the errors of each of the component parts covered in the preceding tables. For both statistics, the sample estimate is seen to understate substantially the true aggregate. The understatement is largest for the entire panel, as would be expected, the true aggregate being underestimated on the first two waves by nearly 50 percent. On the later waves of the panel, the degree of understatement falls off sharply, though even by Wave 5 the total was understated by more than 20 percent.

It should be noted that the sample estimates were not weighted for the differential sampling fractions used in this study nor was any effort made to allocate balances to nonrespondents and to those who refused balances on a basis other than straight allocation of means. Cursory experimentation with these refinements suggests, however, that the main results in Table 5, and in the preceding tables, would not have been affected appreciably.

Concluding Remarks

The results of this study suggest strongly that, at least in the case of consumer savings, the accuracy of data obtained from consumers improves markedly during the course of a panel operation. The principal reason for this improvement appears to be a substantial decline in nonreporting of holdings, a phenomenon which serves to more than offset any tendency for the bias in data obtained from the sample members to increase as a result of drop-outs.

One additional effect of the panel operation, sometimes overlooked, is on the variance of the estimates of the parameters. As we know from sampling theory, when nonsampling errors are present, the variance of the estimate of a parameter is the sum of the ordinary sampling variance and the square of the bias. It is readily shown that the ratio of this bias to the usual expression of the standard error of the mean measures the extent to which confidence intervals are mis-stated because of this bias. Thus, a .95 confidence interval computed in the usual manner will represent a true confidence interval of the same probability only if this ratio is zero. If the ratio should be 1.0, the probability of the usual symmetrical 95% confidence interval containing the true parameter declines to .83. The true probability declines progressively as the ratio rises--to .45 when the ratio is 2.0, to .15 when the ratio is 3.0, to .02 when the ratio is 4.0, and so on.

For the first wave of the present study, this ratio was 5.5 for the respondents and 9.6 for

Table 4

Estimated and Actual Average Balance Per Validated Account of All Panel Members by Wave

(1)	(2)	(3)	(4)	(5)	(6)
		Actual average balance			
Wave	Percent understatement: all accounts	All accounts	Respondents	Nonrespondents[a]	Drop-outs[a]
1	20.9%	$3,023	$2,446	$3,924	--
2	33.9	2,986	1,944	4,645	$3,438
3	26.5	2,955	1,938	1,419	4,148
4	23.7	2,845	2,062	3,150	3,985
5	22.0	2,839	1,985	2,181	3,650
Sample size[b]					
(W1-W5)	220	220	155-115	65-5	0-100

[a] Nonrespondents include accounts of panel members who could not be contacted on a particular wave but had not previously refused further interviews. Drop-outs for any particular wave are those who refused to grant an interview on a previous wave; the size of this category therefore cumulates over time.

[b] First figure shows sample size as of Wave 1 and second figure as of Wave 5 for each category. For the total sample (Columns 2 and 3), this figure is constant over time.

Table 5

Error in Estimate of Aggregate Balances in Validated Accounts*

Wave	Respondents only	All panel members
1	39.5%	47.1%
2	39.7	48.8
3	27.1	24.8
4	17.7	22.7
5	21.3	21.3

*All percentages are underestimates.

the total sample. In other words, the true probability that the usual .95 confidence interval would contain the actual average balance was virtually zero in each case.

As the panel operation proceeded, this ratio tended to decline. As a result, on the last wave, for the same sample size, the value of this ratio was 1.6 for the respondents and 5.3 for the total sample. Hence, at least in the former case, there is at least a moderate probability that the usual 95 percent confidence interval will include the true value, though the probability is still nowhere near .95; the actual probability is roughly .42.

It is also worth noting that the sampling variance of the mean computed in the usual manner understated substantially the total variance (mean square error) of these data, but the degree of understatement declined consistently through the course of the panel operation.

Essentially similar results were obtained from other panel operations involving time deposits, debt, and life insurance, all of which serve to increase the validity of the present findings. The principal difference between the results for time deposits and for these other holdings was the lower incidence of nonreporting of debt and of life insurance. Nevertheless, the accuracy of the data for these other assets also increased over time, and for the same reasons.

Under the circumstances, these studies would appear to provide a basis for believing that a panel operation serves to increase the reliability of survey-based data, at least in the case of consumer savings.

THE CURRENT POPULATION SURVEY: A CASE HISTORY IN PANEL OPERATIONS

Joseph Waksberg and Robert B. Pearl, Bureau of the Census

I. INTRODUCTION

Most statisticians working in the field of survey methodology sooner or later are faced with the task of constructing a design for a repetitive sample operation. The estimates from the survey on successive occasions are to be used to draw inferences about the underlying population and about changes in the population. If the statistician is lucky, the purpose of the survey will be clear-cut, sharply focused on a single statistic. Frequently, however, a wide variety of statistics are to be collected. These, in turn, will be subject to a variety of uses, ranging from monthly or other short run comparisons, to studies of long-term changes in patterns, or to more or less formal time series analyses. Data from a number of occasions may also be pooled to provide estimates of aggregates or annual averages, or to build up the number of sample cases in order to permit more detailed analyses of the characteristics of the population.

In addition to the usual problems of sample design, questionnaire content, and survey procedures, the statistician is now faced with an additional decision--how much of the sample, if any, should be held constant from one period to the next and how much should be changed or rotated. For most continuing programs, the decision will be to construct some type of panel operation, that is, one in which at least part of the sample will be identical between successive periods. As will be seen later, this decision will usually be dictated by reasons of statistical reliability, budgetary efficiencies, and program advantages. Considerations of respondent cooperation and accuracy in response, however, may often rule against a fixed panel--one in which the entire sample is identical throughout--but rather in favor of one with a systematic rotation of part of the sample.

The Census Bureau's Current Population Survey (CPS), the source of the official Government statistics on total employment and unemployment and of a wide variety of other facts about the U.S. population, is a prime example of a large-scale, rotating panel type of operation involving household interviews. During the 25 years of its existence, a considerable amount of information on the operating characteristics of the CPS has been collected and analyzed. In this paper, we shall focus on those data which be believe are most relevant to reaching decisions on the sample rotation plan for a continuing program.

A brief description of the Current Population Survey may be in order at this point. The CPS is a household survey conducted with a probability sample of about 32,000 occupied units a month in 357 areas of the country. The units are selected from a combination of 1960 Census listings of addresses, area sampling methods where Census listings do not provide clear descriptions of the locations of housing units, and lists of new housing units built since the Census. The detailed sampling methods and general survey procedures used have been well documented, both in Census publications and professional journals, 1/ and will not be repeated here, except for those features relevant to the present discussion. These features are as follows:

Rotation Plan - A new sample, once selected, is subdivided into 8 systematic parts, one of which is introduced into the survey each month over an 8-month period. (Simultaneously, one-eighth of an old sample is dropped out of the sample, leaving the sample size constant.) Each such subsample, called "rotation group" is interviewed 8 different times, once a month for 4 consecutive months and then for the same 4 calendar months a year later. Under this system, 75 percent of the sample units are common from month-to-month and 50 percent from year to year. (Other degrees of overlap exist for other pairs of months, up to those 16 months apart.) The composition of the sample, in terms of age, is identical each month - that is, one-eighth of the units are being enumerated for the first time, the second eighth for the second time, etc. up until the final eighth in their last month in the sample.

Overlap of households - The previous description somewhat overstates the overlap of households or persons in the sample. The stated percentages of the sample units that are common from one period of time to another, represent

---

1/ A concise description of this program can be found in P-23 No. 13, "Concepts and Methods Used in Household Statistics on Employment and Unemployment from the Current Population Survey." A much more detailed discussion is contained in Technical Paper No. 7 "The Current Population Survey - a Report on Methodology" Bureau of the Census, U.S. Dept. of Commerce, 1963.

common addresses, not necessarily families or individuals. If a family or person moves during the period of coverage, the replacement, if any, would be interviewed. Also, nonrespondent households (of the order of 5 percent in a typical month) provide another reason for attrition from the maximum potential overlap. Thirdly, a small amount of new construction is constantly being added to the sample to insure the fact that it represents the total population at each point in time. As a consequence of these factors, the actual sample overlap of individuals is closer to 65 percent from month-to-month and 40 percent from year to year.

Data obtained in each interview - A more elaborate interview is conducted the first month a household is in the sample than in successive months. The first interview starts with a listing of all persons in the household, a series of probing questions to insure that complete coverage of persons is attained, and the collection of a group of demographic items about each person that, for all practical purposes, will not change during the period of time the household is in the sample or for which the change can be predicted. These include age, seg, color, marital status, educational attainment (for adults), and family income during the preceding year, among others. These items are asked only in the first interview, but are used in all eight months as cross-classification variables and as the basis for ratio estimates.

In each month, including the first, an identical set of labor force questions is used. Information is obtained on labor force status during the reference week, occupation and industry, number of hours worked, number of weeks unemployed for those out of work, and other related items. 2/

Kinds of data tabulated - Each month, tabulations are made and results are published for a standard set of statistics on current labor force status, number of hours worked and occupation of the employed, length of period of unemployment for the unemployed, and related items, many of them classified by demographic characteristics such as sex, age and color. One can consider these as one class of statistics produced. They are designed to provide a cross-section of the labor-force situation during each month, and by comparisons with prior periods to measure net changes over time.

A second class of statistics is intended to show gross flows in labor force status. These are produced by comparing the status in each of two months for all persons reporting in both months and measuring the total number of shifts that occur, together with the details of these shifts. To date such tabulations are produced only for pairs of neighboring months although it is possible to prepare similar data for pairs of months a year apart, or for that matter for any pair of months containing identical rotation groups. However, some serious problems and potential biases affect the interpretation of these data. Their publication was discontinued some years ago and they are used internally for only limited analyses. These problems will be described in a later section of this report.

A third class of statistics relate to aggregates for groups of months, such as quarterly or annual averages.

The monthly statistics (the first class described above) are produced by means of a composite estimation procedure. For each item tabulated, two separate statistics are prepared each month. One is based on the information for the current month only. The other is developed from the final estimate for the previous month to which is added a measure of change based on those parts of the sample which are common between the months. The final estimate represents a weighted average of these two estimates. 3/

II. SAMPLING VARIANCES

In any study of the sampling variances for a multi-purpose survey, the analyst is faced with as many different variances as he has separate cells in tables. In the case of CPS, this literally runs into tens of thousands. Some simplifications are obviously required to reduce these to manageable proportions. In our analyses, we have tended to focus attention on a small number of the most crucial statistics developed from the survey, and we base decisions regarding efficiency of sample design, estimation methods, etc. on their effect on these crucial statistics.

2/ Supplemental questions on social or economic characteristics are frequently added on a one-time or annually recurrent basis. When introduced, they are generally asked of all households in the sample in that month. This report does not include any analysis of these statistics.

3/ For information on the theory relating to sample rotation and composite estimates, see Hansen, Hurwitz and Madow "Sample Survey Methods and Theory" Vol. I pages 500-503 and Vol.II pages 272-279; Patterson "Sampling on Successive Occasions with Partial Replacement of Units", J.Roy. Stat. Soc. Series B, 12 (1950); Eckler "Rotation Sampling," Annals of Math. Stat. 26 (1955); "The Redesign of the Census Current Population Survey" by Hansen, Hurwitz, Nisselson, Steinberg, Journal of the Amer. Stat. Assoc. (Sept. 1955); Woodruff "Use of Rotating Samples in the Census Bureau's Monthly Surveys", Journal of Amer. Stat. Assoc. (June 1963).

In this paper, we will concentrate on the items which are generally of greatest concern in the monthly publications - estimates of total civilian labor force, total employment, nonagricultural employment, agricultural employment, and unemployment. For each of these, we will present data for four kinds of statistics. The first is monthly level, that is the estimate of the number of persons falling into the class (e.g. number unemployed) in a typical month. The other three kinds of statistics are month-to-month change, change from the same month a year ago, and annual averages.

Table 1 contains data on the effect of alternative approaches on the variances of the results. The CPS method is compared with a procedure in which independent samples are selected each month, and with one in which a fixed panel is used over and over again. For the CPS method, the effect of using the composite estimate is shown separately from the effect of simply retaining households in the sample. For the independent samples and fixed panels, composite estimates of this type are not applicable. The composite estimate referred to in the table has the form currently used in CPS, that is, with weights of .5 for each of the two separate estimates composing the composite.

Several other comments on table 1 will be useful: (1) First, the data represent approximations over an almost ten year period. During this length of time, the variances fluctuate somewhat, partially with fluctuations in the business cycle (for example, as the unemployment rate varies) and partially a result of seasonal changes, particularly large in the case of agricultural employment. The figures should therefore be considered as indicative of what happens in the long run, rather than as estimates of the situation at any point in time. (2) Secondly, each of the three plans listed assumes that the sampling is done within a fixed set of counties or primary sampling units (PSU's). Consequently, only the within-PSU variance is affected. For monthly level or differences between a pair of months, agricultural employment is the only one of the items for which the between-PSU variance is large enough to have any practical effect on the analysis. However, the between-PSU variance reaches one-half or more of the total variance for all items when annual averages are considered. Thus, even the independent samples do not provide as much of a reduction for an annual average as one would assume from using the equivalent of 12 monthly samples, and the loss of using correlated rather than independent samples is not as severe as would happen if the samples were completely independent. (3) Finally, the table overstates, somewhat, the virtues of a fixed panel. With either a rotating sample or independent samples, it is possible to treat more satisfactorily unexpected large units that occur in the sample and that create huge contributions to the variance. This can be done by identifying all large observations over a period of time in the past, and including them in the current time period. The effect of this is to sample large observations at each occasion at a rate k times that of other observations, where k is the number of independent samples contributing the large observations. This is currently used in the CPS with $k = 7.5$. No computations have been made of the gains resulting from this device, and they are not reflected in table 1.

An examination of table 1 indicates that decisions that appear best for one of the statistics are not necessarily optimum for the others. If one were to concentrate on a single statistic, more often than not, a better rotation or estimation method could be found than the one currently used in CPS. However, when the entire array of statistics is considered, the present method appears to be a reasonable compromise in that it is at or near the optimum for all statistics except year-to-year change.

It may appear to be surprising that fixed or slowly rotating panels do not have greater advantages in measuring change over time than table 1 shows. The gains reflect two things - the rotation pattern and the correlations over time for identical sample segments. For example, the ratio of the variance of month-to-month change for a fixed panel relative to independent samples, is $1-r$ where r is the month-to-month correlation. This ratio is extremely sensitive to high values of r. For values of $r = .98$ (not unusual in establishment surveys) $1-r = .02$. For $r = .80$, $1-r = .20$, or ten times as great as the previous example. Unfortunately the correlations are smaller than might be expected from identical sample units. Typical examples of correlations found in practice are shown below.

Item	Correlation for		
	Civilian labor force	Agricultural employment	Unemployment
Pairs of neighboring months	.80	.90	.50
Pairs 2 months apart	.70	.85	.40
Pairs 3 months apart	.65	.80	.30
Pairs 12 months apart	.70	.70	.30

There are a number of reasons for the relatively low values of these correlations. First, people do change their labor force status from time-to-time. In particular, women and teenagers tend to move into and out of the labor force at a surprisingly high rate. Secondly, identical samples refer to identical addresses rather than persons, and this also affects the correlations. Over the course of a year, about 20 percent of the persons in the U.S. will have moved, thus reducing the correlations appreciably. Thirdly, the five percent of the designated households which are nonrespondents in a typical month have a similar depressing effect on the correlations, in that they reduce the proportion of persons identically in the sample in a pair of months.

A number of modifications in the form of the composite estimate are being studied for possible further reductions in the variance. The current composite estimate can be expressed as:

$$X^*_u = (1-K)\, X'_u + K\, (X^*_{u-1} + X'_{u,u-1} - X'_{u-1,u})$$

where X^*_u = composite estimate for month u

X^*_{u-1} = composite estimate for month u-1

X'_u = regular ratio estimate for month u

$X'_{u,u-1}$ = regular ratio estimate for month u using only those rotation groups that are also in the sample in month u-1

$X'_{u-1,u}$ = regular ratio estimate for month u-1 using only those rotation groups that are also in the sample in month u.

At present K = .5 for all items.

The first and most obvious modification being examined is to use different weights for different statistics. Another possibility is to introduce year-to-year change (as well as month-to-month change) in the estimation procedure by using an average of three different estimates, with weights of K,L, and 1-(K+L). Such an estimate can be expressed as:

$$X^*_u = \left[1 - (K+L\right]\, X'_u + K\, (X^*_{u-1} + X'_{u,u-1} - X'_{u-1,u}) + L\, (X^*_{u-12} + X'_{u,u-12} - X'_{u-12,u})$$

Table 2 shows approximations to the reductions in variance that would result from such modifications. It can be seen that significant improvements are possible, in particular on year-to-year change. A serious operational problem exists, however, in putting either of these plans into effect. With varying values of K (or L), inconsistencies can arise in developing the same statistic from two different tables (e.g. civilian labor force produced by itself in one table is likely to be different from the sum of employed and unemployed in another table). The differences are likely to be not large, but they will trouble the users nevertheless. We are exploring the possibility of developing some compromise solution.

Several other possible developments are also under study. These include drawing in additional months in the composite estimate, making use of the current month's data to revise estimate for the previous month, and finding better methods of adjusting for nonrespondents to dampen their effect on the correlations.

III. COST ADVANTAGES OF PANEL OPERATIONS

One of the prime attractions of panel operations is their clear-cut cost advantage over other survey arrangements. It is rather evident that the use of the same panel on a number of occasions, as opposed to a new selection of respondents in each instance, would markedly reduce the costs of sample preparation. With the present CPS rotation pattern, a full national sample of 35,000 households is needed every eight months. The cost of selection and preparation of sampling materials for this size sample is about $80,000 or 5 percent of the CPS budget. Clearly, it would require an overwhelming technical advantage to justify a new selection of units each month, which would multiply the sampling costs by a factor of 8, resulting in a one-third increase in the total budget.

Savings in sampling costs, however, is only one part of the increase in efficiency made possible by panel operations. There is a substantial cost incurred by interviewers in locating a set of new addresses for the first time. Once located, and assuming continuity in the interviewing staff, these same units are found much more readily in subsequent enumerations.

Aside from the matter of location, the first attempt to survey a household is likely to require more visits than a subsequent one, because the interviewer is unfamiliar with the habits of the occupants and when they are most likely to be home. According to recent CPS figures, the average number of visits to households being interviewed for the first time was 1.6, as compared with 1.3 for those in their second month of enumeration. Even the fifth enumeration, which takes place after an 8-month hiatus and is likely to see much turnover of occupants, required only 1.4 visits, on the average.

Panel operations also afford an opportunity to institute less costly collection procedures once rapport has been established with a group of respondents. In the CPS, it has been found desirable to conduct personal interviews in the first, second, and fifth months in sample to assure continued cooperation. At other times, however, telephone interviews are used wherever significant efficiencies are possible, especially where an entire cluster has telephones or to avoid return visits to households not found at home. In an average month, about 25 percent of the CPS interviews are conducted by telephones; for the rotation groups in which telephoning is permitted, the figure is about 33 percent. Even less expensive methods, such as mail inquiries, could be used effectively in many panel operations, but have not yet been found feasible in the CPS because of the precise time reference of the questions and the extremely tight timetable for conduct of the survey. 4/

4/ Mail inquiries, or "leave-it" self-enumeration forms, have been used extensively, however, in special follow-up studies of subgroups of the CPS sample.

Table 3 contains a distribution of the number of visits required to complete the interviews and the number of cases interviewed by telephone for August 1964. Based on this information, and the fact that interviewer travel represents more than half of total field costs, it is estimated that the enumeration cost of a household in a new sample is probably about one-third greater than for a household in one of the five rotation groups permitting telephone interviewing. For second and fifth-month households, where the telephone is ordinarily not used but prior contacts have been made, the costs are about half way between these two levels.

A point raised earlier--that of recording at the time of first-enumeration many basic demographic facts about the residents of each CPS sample unit--represents still another efficiency of panel operations. Once obtained, these data are automatically available in subsequent periods for cross-classification or other purposes, although some updating is necessary from time-to-time. The preparation of the basic original record takes about 10 minutes per household, on the average, whereas such updating as is needed usually averages only a fraction of a minute. This alone adds 10 to 20 percent to the enumeration cost.

When these individual components of cost are put together, it can be estimated that if a survey of the size and characteristics of the CPS were conducted with an independent sample each month, the cost would be at least 75 percent greater than with the present rotation system, and possibly as much as twice as great. Conversely, of course, a fixed panel would save money, but because of the slow rotation system used in CPS, the difference would be much more modest--possibly of the order of 10 percent.

IV. INCREASED ANALYTICAL OPPORTUNITIES

It would be remiss to omit at least a brief reference to the analytical opportunities offered by panel operations. Already mentioned are the so-called gross flow or gross change data, which provide a cross-classification of the status of an identical group of individuals from one month to the next (or over other periods). In the case of CPS, for example, data are tabulated on the number entering or leaving the labor force from month-to-month, the number shifting in either direction between an employed and an unemployed status, and numerous related flows, thus revealing the dynamics of the labor market which are often concealed in the over-all net changes.

In a broader sense, the use of panels over longer periods provides the framework for the true "longitudinal" study with a much wider range of possibilities. Since a given CPS panel extends over only a 16-month period with interviews conducted in only 8 of the 16, there are obvious limits in this respect. Aside from the gross changes, however, there are opportunities within this cycle for assembling in the same record information collected on different regular or supplementary subjects in different months for the same individuals. This is a means of expanding the detail known about a given group of households without subjecting them to an unduly long interview at any one time.

V. EFFECT ON RESPONSE

With the massive evidence assembled in favor of panel operations, one might conclude that only the uninitiated or the foolhardy could ever consider anything else in establishing a continuing statistical program. In fact, one might ask why CPS does not move even further in this direction and use a fixed panel. Unfortunately, in most human endeavors, benefits are seldom bestowed without exacting a certain price. In this case, the price is the possibility that the information provided by households will be affected or influenced in some way by the fact that they are interviewed on repeated occasions--in more technical terms, that there will be some conditioning of response in the panel.

An obvious example which comes to mind was a one-time proposal to establish a representative panel of individuals who would receive medical examinations at regular intervals, thus providing a measure of the changing health of the population, the effect of aging, and selected facts. A rather awkward problem was envisioned however, arising from the fact that the participants would learn they had previously undetected health problems. (It would presumably be unethical to conceal serious conditions from them.) This knowledge would undoubtedly impel them to seek medical treatment, and while the next survey might show an impressive improvement in the health of the panel, this could scarcely be projected to the general population.

The information collected in the CPS is hardly likely to have as dramatic an impact on the individuals concerned, but strong evidence of conditioning in response nevertheless exists. In particular, a phenomenon has been observed which has loosely been termed the "first-month" bias, since its effect has been most pronounced when comparisons are made between households being interviewed for the first time, and those who had been interviewed in past months. 5/ Illustrations of the effects of the first-month bias are presented in table 4. It can be seen that households in the first month of enumeration show significantly higher levels of unemployment and of part-time and marginal employment than is found for households interviewed for the second or later time. In most cases, the differences are attributable to women, teenagers, and others whose participation in

---

5/ This term has been used for convenience and does not necessarily imply that the first month's data are biased in a statistical sense.

the labor force is often on an intermittent basis. 6/ A somewhat less striking, but rather persistent, downtrend has also been observed in population coverage as the sample ages.

Although there are many theories, the reasons for these sharp differentials have never been satisfactorily established. One hypothesis is that interviewers are more careful in conducting the first enumeration since they have no advance information about the households. Subsequently, they may assume that some persons, such as housewives and teenagers, who were outside the labor force the first time continued in that status without asking the questions completely, and without realizing that these are the persons who account for most of the shifts into and out of the work force. A contrary opinion is that new respondents may over-report employment or unemployment because they are initially uncertain of the purpose of the survey or wish to appear more usefully occupied.

Knowledge of this problem has existed for many years without being a major concern of analysts of labor force data because the extent of the bias did not seem to be a serious limitation on uses of the data. In particular, the effect on over-all changes from month-to-month or over other periods was known to be quite small since the sample at any time consisted of the same mixture of households in different stages of enumeration. Demands in recent years for greater precision in the standard data and for a revival of publication of the gross change data have, however, stimulated a growing concern for finding solutions to the problem. Gross change data are especially subject to downward biases resulting from conditioning since, by definition, they involve a comparison of data for earlier with later periods of enumeration of the same individuals. 7/

As a consequence of these emerging needs, a good deal of research has been initiated on response problems in general and conditioning effects in particular. In an experimental study operated in three areas and involving some 1,500 interviews a month, the Bureau has been testing during the past year and a half several alternative approaches to collection of labor force data. Among the alternatives examined have been a far more detailed questionnaire than the present one, and a procedure in which self-enumeration is attempted after the first month, with telephone and personal follow-up of nonrespondents. A third alternative is a procedure whereby an independent interview is taken but the interviewer also has access to a summary of the information for the previous month and attempts to obtain an explanation and confirmation of any basic changes in employment status and job attachment. The standard CPS questionnaire and procedure is used as a control in the experiment.

Table 5 summarizes data from the experimental study relevant to whether any of the procedures appeared to modify the usual effect of conditioning. Sampling errors are too large to reach any definite conclusions, but our judgment is that only procedure D (which provided for a comparison by the interviewer with the previous month's data) showed any evidence along these lines. Work still must be done to ascertain whether this results in improved data, since availability of prior month's information could create a false stability or a concealed form of conditioning.

During the past six months, another experiment has been going on, investigating the possibility that the first month bias results from interviewers becoming overly familiar with the households in their assignments. This test, which has been taking place in a subset of areas included in the regular CPS sample, provides for changes in interviewer assignments between the first and second, and the seventh and eighth months of enumeration of the same households. These are the stages between which the largest differentials in response generally appear. Table 6 summarizes the findings to date. These data are preliminary since the experiment is not yet over. The results are too fragmentary as yet to render final judgments although, it must be conceded, they do not appear especially encouraging in providing an answer to this problem.

We hope that the present efforts will provide some insights into this perplexing phenomenon, which casts a shadow on various types of panel operations and frustrates the introduction of more sophisticated techniques in estimation procedures, at least for household surveys such as the CPS where the interviewer or respondent's motivation and attitude towards the interview, may influence the results. Clearly, a wider dissemination of the experience of various public and private survey groups which conduct panel operations should enhance the chances of a breakthrough in this area.

6/ A more detailed presentation of these data appears in a paper prepared by the present authors, entitled "The Effects of Repeated Household Interviews in the Current Population Survey," for the National Conference of the American Marketing Association, Dallas, Texas, June 17, 1964.

7/ Gross change data also suffer from biases resulting from the exclusion of sample cases which are not identical between the periods being compared, such as migrants, nonrespondents, and the like. For a more detailed discussion of gross change data, see paper by Robert B. Pearl, "Gross Changes in the Labor Force: A Problem in Statistical Measurement," Employment and Earnings (U.S. Bureau of Labor Statistics) Vol. 9, No. 10, April 1963.

Table 1.--COMPARISON OF VARIANCES IN CURRENT POPULATION SURVEY FOR ALTERNATE TYPES OF ROTATION PLANS, FOR SELECTED LABOR FORCE ITEMS

(Figures shown represent the ratios of the variances of each rotation plan to a system in which independent samples are used each month. The data assume identical sample sizes and a similar estimation procedure for all plans--the ratio estimation used for the CPS--except the lines for "with composite estimate" which superimpose the composite estimate as used in CPS on the ratio estimates. The data refer to an "average" month, or pair of months.)

Item	Civilian labor force, total and nonagricultural employment	Agricultural employment	Unemployed
Monthly level			
Independent samples	1.00	1.00	1.00
Fixed panel	1.00	1.00	1.00
CPS rotation system			
Without composite estimate	1.00	1.00	1.00
With composite estimate	.83	.90	1.05
Month-to-month change			
Independent samples	1.00	1.00	1.00
Fixed panel	.20	.10	.55
CPS rotation system			
Without composite estimate	.40	.31	.65
With composite estimate	.28	.18	.60
Change from year ago			
Independent samples	1.00	1.00	1.00
Fixed panel	.30	.30	.70
CPS rotation system			
Without composite estimate	.65	.65	.85
With composite estimate	.57	.55	.90
Annual average			
Independent samples	1.00	1.00	1.00
Fixed panel	3.00	1.70	2.00
CPS rotation system			
Without composite estimate	1.50	1.20	1.40
With composite estimate	1.60	1.25	1.50

Table 2.--OPTIMUM VALUES OF WEIGHTS IN ALTERNATE COMPOSITE ESTIMATES, AND RESULTING REDUCTIONS IN VARIANCES FROM CURRENT COMPOSITE ESTIMATE USED FOR CPS, FOR SELECTED LABOR FORCE ITEMS

Item	Optimum weights with current estimation method		Optimum weights using year-to-year change	
	K	Ratio of variance to variance of current estimate	K, L	Ratio of variance to variance of current estimate
Monthly level				
Civilian labor force total and nonagricultural employment	.6	.98	.5, .2	.86
Agricultural employment	.7	.89	.7, .1	.81
Unemployment	.3	.96	.2, .2	.90
Month-to-Month change				
Civilian labor force total and nonagricultural employment	.8	.92	.7, .1	.90
Agricultural employment	.9	.73	.8, .1	.73
Unemployment	.5	1.00	.4, .1	.99
Change from year ago				
Civilian labor force total and nonagricultural employment	.5	1.00	.4, .3	.74
Agricultural employment	.7	.89	.6, .2	.74
Unemployment	.2	.91	.2, .3	.80

Table 3.--PERCENTAGE DISTRIBUTION OF INTERVIEWED HOUSEHOLDS BY METHOD OF INTERVIEW, AND FOR THOSE INTERVIEWED BY PERSONAL VISIT BY NUMBER OF VISITS, AUGUST 1964 CPS

Number of months in sample	Type of interview			For personal visit households, number of visits required to complete the interview				
	Total	Telephone	Personal visit	Total	1	2	3+	Average number of visits
1	100.0	0.1	99.9	100.0	61.8	22.5	15.7	1.57
2	100.0	4.2	95.7	100.0	76.9	16.3	6.8	1.31
3	100.0	30.4	59.6	100.0	77.4	15.7	6.9	1.31
4	100.0	34.7	55.3	100.0	78.1	16.0	5.9	1.29
5	100.0	0.9	99.1	100.0	69.1	20.2	10.7	1.44
6	100.0	29.1	60.9	100.0	78.4	14.9	6.7	1.30
7	100.0	33.3	66.7	100.0	78.8	13.5	7.7	1.30
8	100.0	20.2	79.8	100.0	71.7	19.2	9.1	1.39

NOTE: The second and eighth-month households were involved in an experimental study during August which could have affected the average number of visits and the use of telephone enumeration. These two groups should therefore be omitted from any analysis of the difference between new households and those that have been interviewed in earlier months.

Table 4.--SUMMARY OF CPS EMPLOYMENT STATUS MEASURES, BY ROTATION GROUP: AVERAGE, MARCH 1959 - DECEMBER 1961

(Index numbers: All groups combined = 100) 1/

Measure	Number of months in sample:							
	1st month	2nd month	3rd month	4th month	5th month	6th month	7th month	8th month
Civilian labor force.........	101.1	100.4	100.5	99.9	99.7	99.5	99.6	99.4
Total employed..............	100.7	100.4	100.4	100.0	99.6	99.4	99.7	99.6
Employed in nonagricultural industries............	100.8	100.4	100.3	100.0	99.6	99.5	99.6	99.5
Employed in agriculture......	99.6	99.7	100.9	100.5	99.4	98.5	100.6	101.0
Unemployed..................	107.3	100.3	100.3	98.9	100.7	99.6	96.6	95.0
Private household workers 2/.................	106.4	100.5	97.4	98.8	101.1	98.7	98.0	98.8
Nonagricultural self-employed...............	102.0	100.4	99.9	98.8	101.0	99.9	99.4	98.4
Nonagricultural unpaid family workers.............	102.7	106.6	102.8	99.2	102.7	101.9	99.6	95.3
Regular part-time workers 3/.................	103.7	100.0	98.6	98.1				

1/ Absolute numbers first converted into percentages of population in rotation group and then into index numbers. The ratio estimates to independent estimates of the population by age-sex-color are applied separately by pairs of rotation groups--the 1st and 5th combined, the 2nd and 6th, etc. Consequently, items comprising high proportions of the population generally average out to close to 100 percent by these pairs of rotation groups.

2/ Domestic servants, babysitters, odd-job workers, and the like.

3/ Persons who usually work part time for noneconomic reasons. Information only available by rotation set (1st and 5th months combined, 2nd and 6th months combined, etc.).

Table 5.--COMPARISON OF LABOR FORCE AND UNEMPLOYMENT RATES, BY NUMBER OF MONTHS IN SAMPLE, FOR ALTERNATIVE PROCEDURES USED IN CPS METHODS TEST (3-AREA EXPERIMENTAL PROGRAM): JULY 1963 - JUNE 1964

(Index numbers: average of 4 rotation groups = 100)

Procedure	Index of labor force rates (labor force as percent of population 14 years and over)					Index of unemployment rates (unemployed as percent of civilian labor force)				
	All rotation groups	1st month in sample	2nd month in sample	3rd month in sample	4th month in sample	All rotation groups	1st month in sample	2nd month in sample	3rd month in sample	4th month in sample
A - Standard CPS questions and procedure	100.0	101.6	98.9	100.2	99.1	100.0	109.6	101.5	98.7	90.2
B - Detailed questionnaire	100.0	102.5	99.8	98.9	98.6	100.0	110.4	103.3	101.8	84.1
C - Self-enumeration with follow-up of nonrespondents 1/	100.0	101.6	98.8	100.2	99.5	100.0	116.0	116.1	84.2	83.7
D - Comparison with previous month's responses and confirmation of changes 2/	100.0	100.5	99.6	99.5	100.5	100.0	99.0	93.4	98.6	109.3

1/ The self-enumeration procedure started in the second month; the standard CPS questionnaire was used in personal interview the first month.

2/ Starting in the second month, this procedure involved first an independent interview and then a comparison with the results from the previous month and a check and confirmation of basic changes in status. In the first month, only a direct interview was taken since there was, of course, no comparative data for prior periods. In about half of the cases, the standard CPS questionnaire was used in this procedure and in the remaining half, the detailed questionnaire (used in procedure B) was specified.

Table 6.--EFFECT ON LABOR FORCE STATUS OF CHANGING INTERVIEWER ASSIGNMENTS FOR SECOND AND EIGHTH-MONTH ROTATION GROUPS, FOR SELECTED PSU'S IN CPS

(Data are in the form of index numbers. Each index represents the ratio of the number reported in that rotation group to the average of all rotation groups. For each characteristic, the first line of data is an average for the period July 1963 - June 1964, before the changes in assignment were instituted, the second line covers the period July - November 1964, when changes were made. Both lines refer to an identical set of PSU's, which include about one-third of the total CPS sample.)

Item	Procedure	Number of months in sample							
		1	2	3	4	5	6	7	8
Total males, 14 years and over	Regular CPS	1.009	0.996	1.006	0.997	0.996	0.999	0.997	1.000
	Interviewer change	0.982	0.996	0.991	1.000	1.024	1.016	1.010	0.980
Male, civilian labor force	Regular CPS	1.020	1.000	1.007	0.999	0.993	0.993	0.995	0.994
	Interviewer change	0.991	0.996	0.986	0.998	1.026	1.014	1.000	0.989
Male, employed in nonag. industries	Regular CPS	1.013	0.995	1.005	0.996	0.996	1.000	0.999	0.996
	Interviewer change	0.986	0.994	0.985	0.995	1.017	1.016	1.002	1.005
Male, unemployed	Regular CPS	1.113	1.070	1.042	1.050	0.976	0.900	0.932	0.917
	Interviewer change	1.056	1.015	0.953	1.071	1.010	0.985	0.971	0.940
Total females, 14 years and over	Regular CPS	1.008	1.004	1.015	1.008	0.997	0.999	0.986	0.984
	Interviewer change	0.994	0.997	0.987	0.999	1.012	1.008	1.014	0.989
Female, civilian labor force	Regular CPS	1.050	1.015	1.008	0.998	1.008	0.975	0.977	0.968
	Interviewer change	1.027	1.003	0.974	0.984	1.015	1.006	0.997	0.995
Female, employed in nonag. industries	Regular CPS	1.037	1.011	1.011	0.999	1.008	0.980	0.980	0.973
	Interviewer change	1.010	1.003	0.976	0.984	1.009	1.014	1.002	1.001
Female, unemployed	Regular CPS	1.224	1.062	0.970	0.970	1.003	0.922	0.955	0.895
	Interviewer change	1.222	1.017	0.917	0.978	1.116	0.871	0.966	0.914

DISCUSSION

Philip E. Converse, Survey Research Center, The University of Michigan

I have read and heard our papers today with a good deal of interest, since I have been involved myself in fairly large-scale panel operations. The Survey Research Center at the University of Michigan has done several of these studies, both in the area of consumer economics and in the area of political behavior.

My predominant reaction to most of the material in these papers is one of agreement. That is, over that portion of the findings or observations for which we have any analogous findings or experience, I think there is little that I would be inclined to challenge on any factual grounds.

Where the broad themes of the Waksberg-Pearl paper are concerned, for example, it is certainly true that data on gross change often has a rather different cast than knowledge of net change alone would lead one to deduce. It is also true that there are cost-efficiency arguments for panel studies, although my impression is that much of this gain can evaporate if the logic of the panel design is fully realized. Such a full panel design would certainly require that the investigator follow even those people who change their residences between waves of the panel, unlike the procedure used in the Current Population Survey. This is true simply because gross changes in other attributes among these movers may be quite different than gross changes among those who "never left home." But as soon as one begins to follow movers, of course, the expenses per interview begin to rise very rapidly at the same time so that the cost-efficiency argument for panel studies which is presented in the Waksberg-Pearl paper can be rapidly nullified. Finally, as at least two of our papers argue, people may respond somewhat differently to the same items on the second, third or fourth waves than they did in taking the first interview "cold."

It is on this latter truth--the presence of contamination, practice effect or what you will--that I am choosing to focus in my remarks, in part because it is a matter on which two of these papers make extended comment, taking a rather different view, and in part because we in our panel work have seen similar effects and have still a different impression of them.

Let us start, then, with the common-sense assumption that whatever these effects of repeated interviews may be, they are likely to vary in magnitude and perhaps in nature as a function of various conditions of the investigation. Under "conditions" we might list a number of mechanical design features such as the periodicity of the interviewing, for example. We might list as well features of a less mechanical nature such as the style and the content of the interview. And finally we might even want to include the intentions of the investigator as something which can vary, so that the same reinterviewing effect that delights one investigator may shatter another who has different ends in view.

Now I bring up some of these variable conditions because we know very little about their impact, although these papers are a good start on them. And I bring them up because the panel studies I have been associated with differ rather radically from those discussed here with respect to a number of these conditions. Hence it may be worth trying to put together things that we have noted in our studies with some of the observations in these papers and do a little triangulating to see what reasonable guesses about these effects we may go on to make.

In order for us to understand differences in conditions of investigation I must give at least a brief description of the panel study we have done which departs most notably from those described today. Our panel studies have been election studies, hence the content has been political. The fact that they are election studies also helps to determine the periods of interviewing. Between 1956 and 1960 we conducted a 5-wave national panel study. The waves were phased as follows. The first interview was held just before the 1956 presidential election and was followed by a brief interview after the election. The third wave did not take place until after the off-year congressional election of 1958. There was a similar two-year gap between the third and fourth waves. The fourth wave was taken just before the 1960 presidential election and the final wave involved interviews just after that election.

This was the basic timing of the panel. Now let me suggest some of what I would see to be the primary differences between this design and those which we have heard discussed. First, of course, we are dealing in these three designs with radically different sample sizes from the 300-odd involved in the Ferber study to the enormous case numbers of the Current Population Survey. Our panel involved roughly 1500 cases and hence lies somewhat between the two. Secondly, our design differed from each of the others described in terms of the fact that we attempted to follow people shifting their place of residence between waves.

I think that these two initial differences are largely irrelevant for the comments I wish to make, but there are a number of further differences which are extremely relevant. First, the lapse of time between waves--the periodicity of the interviewing--is much different in our study from any of those described here. Our design involved two lapses of virtually two years rather than the maximum of eight months between the fourth and fifth interviews in the Current Population Survey and the much shorter period described in the Harris and Ferber papers. Secondly, our design was more like the Ferber and Harris studies and quite unlike the current population survey in the lack of governmental suasion, moral if not legal, which is convenient in securing those second, third and fourth interviews from the respondents. Third, our study was like the Harris in subject matter; it was political rather than economic and hence there was a focus on relatively "soft" data--political attitudes--rather than some of what we may think of as relatively "hard" information about employment and personal finances. Finally, our study was not designed primarily as a methodolological investigation. This means that there was no attempt at external validation as in the Ferber study, nor were there the control samples available to the Current Population Survey. At the same time there was a fair amount of internal evidence on contamination and in 1960 for the fourth and fifth waves of the panel there was indeed an independent fresh sample with some overlapping content which permitted comparisons.

Any concern over bias or contamination introduced in the later stages of repeated interviewing naturally divides itself into two parts: (1) the concern over panel mortality or panel attrition, and biases which may thereby result; and (2) contamination of those respondents who stay with the study through all or most of the waves.

With respect to panel mortality, apparently the problem was not a pressing one in the Current Population Survey. However, it was a problem in the Ferber study and it was a problem in ours. In absolute terms, the attrition was very large over this period of time. In our second wave, for example, we were able to reinterview 91 percent of those respondents whom we had interviewed in the first wave. This figure dropped to 70 percent in the third wave, 63 percent in the fourth wave and 61 percent in the fifth wave. Now, of course, death and senility, ignored in these raw figures, take on significant proportions over a four-year period. Hence a better statement of our panel mortality in the sense which it interests us would undoubtedly be closer to two-thirds of the original respondents being successfully reinterviewed in the final two waves.

One way of asking what kind of bias this rather large attrition may have created for us is to compare the 1956 characteristics--soft ones as well as hard--of 1960 survivors with the original pool of 1956 respondents from which they were drawn. You will note here that there is no question of real change in characteristics over the time period since we are referring to the original 1956 characteristics of the 1960 survivors.

With such an examination across a very large number of variables involved in the study, the bias turned out to be remarkably slight in at least one sense. For virtually all variables the survivors were at most one or two percent different with respect to variables (partitioned three, four and five ways) from the original pool from which they had been drawn. There was one major exception. On measures of political involvement, this discrepancy increased to three or four percent; that is, the 1960 survivors included a three to four percent overrepresentation of people who in 1956 had registered themselves as being highly involved and, conversely, was depleted by three to four percent among those least politically involved. Further inspection showed that the other smaller discrepancies which did occur tended to turn up in characteristics correlated with political involvement. That is, there was a slight increase of a percent or two in the proportion of the survivors who had had college education, and involvement and education have some fair positive correlation with one another. This configuration of results suggested a very clear pattern: a systematic loss of those less interested in the subject matter of the study without any noteworthy loss along any other systematic lines save those correlated with the first. Such results, of course, would lead one to suspect that if the subject matter had been different, the nature of selection of respondents who persisted with the study would have looked slightly different as well.

Without further comment on this at the moment, let us turn to the problem of contamination of those who were successfully reinterviewed, for this information sheds interesting light on the dynamics of those whom we did not succeed in reinterviewing. Where such evidence of contamination is concerned there were some interesting differences between the general flavor of our results and those which you have heard reported today. That is, in the Ferber study as well as the Waksberg-Pearl study one of the major changes in response occurs between the first and ensuing interviews. This is true of our panel in terms of sheer proportion of mortality, but where contamination is concerned our findings were just the reverse: the signs of contamination started very slowly if at all through the third interview and only began to be noticeable in 1960 at the time of the fourth and fifth interviews.

I suspect that these are not contradictory results but that they do point up some of the crucial variables involved. First, one might mention the expectation of being reinterviewed as a fairly crucial matter. It is my understanding that in both of the other studies the respondents were apprised that they would be reinterviewed. In our study they were not so forewarned. We did not say that we were not coming back, of course, but we made no point of suggesting that we would. The second variable, which interacts with the first, is the sheer time lapse between interviews (in our case two years), a matter which would tend to reduce expectation of being reinterviewed even further. As I have suggested, we found almost no evidence of contamination in the 1958 interviewing. Perhaps the sole potential instance which came to light was the respondent who remarked to the interviewer in effect, "If I'd a known you were coming I would have studied up." However, by 1960 more than a handful were beginning to say to the interviewer such things as "I bet my husband you'd be back this fall."

Triangulating across these several studies it seems clear that the expectation of being reinterviewed as well as the time lapse between interviews can drastically influence the degree to which the respondent is sensitized to the area of investigation in a manner which prepares him for any reinterviews. How much difference this sensitization makes in behavior patterns or response patterns is almost certain to be a function of the subject matter and, quite reasonably, how sensitive the individual was to the subject matter in the first place. In other words, we would propose that the effects of sensitization are greatest (although they may come on more slowly) in areas where the respondent is normally least sensitized. A standard cross section of the public tends to be less sensitized to politics than to the details of personal saving habits and probably less sensitive to the details of personal saving habits than to whether or not he happens to be employed. A number of studies suggest that inaccurate reports, even on sensitive matters like income, are very often nonmalicious: they represent honest ignorance or at least honest failure to retain details. In studies of personal time use, for example, the person who is to serve as a respondent is often subjected to the forewarning that he will be so asked because it improves the accuracy of his recall. It seems clear that sensitization of this sort is one of the prime effects in the Ferber study.

The question which remains, of course, is whether such sensitization is desirable or undesirable. Ferber is pleased at the phenomenon and we are horrified by exactly the same phenomenon, but we do not have to look far to see why this difference in reactions occurs. For here is where the whole intention of a study becomes crucial. Ferber has a right to be pleased with the effects of sensitization for it pulls his reports closer to the parameters which he hopes to estimate. We, in our work, had cause to be alarmed by exactly the same effects because there was reason to believe that this sensitization was promoting changes in actual behavior in the phenomenon which we wished to investigate. In political studies there are few variables more powerful in mediating crucial differences of behavior than political involvement and political information. In 1958 there was little reason to believe that the political involvement or the political information of our respondents had been affected by the repeated interviewing, but there were quite noteworthy indications that this was true by 1960. In point of fact, after the 1960 election study was completed there was a 5 percent surplus of people reporting that they had voted in the presidential election, a mysterious increment which remained by comparison with the fresh control sample after we had taken account of the differential panel mortality selecting toward those who were more involved. In other words, one effect of our repeated interviewing, and the sensitization which accompanied it, may actually have been to stimulate some people to vote who would not otherwise have voted!

It is clear in this instance, of course, that while a sensitization to politics may have given us better information in those cases where we asked the respondents to recall details of past behavior, analogous to the Ferber results, we were at the same time affecting our sample, and inadvertently drawing its sample characteristics out of line with many of the population parameters which we were trying to estimate. You will note that this contamination of those whom we successfully reinterviewed produces a bias which goes in the same direction as the panel mortality bias, so that the effects of the two put together are joint or cumulative.

I suppose that even in the Ferber case, we must keep in mind the possibility that the study, with its sensitization toward personal bank deposits, might actually have affected respondents' saving behavior, as well as their accuracy of recall of that behavior. Such a change, if it existed, would not of course show up as a discrepancy in the "validation" or "criterion" data used in the study. However, I think we would agree that such effects, if plausible at all, should be slight simply because the behavior in question is less "elastic" than the decision to vote in an election or to pay more rather than less attention to politics in the newspapers. Employment behavior would seem less elastic still: we might reasonably doubt that the Current Population Survey interviewing leads anybody to pick up jobs, drop them, or in other ways change their job-seeking behavior. Nevertheless, Mr. Waksberg assures me that the items employed to determine labor force status inevitably have their "soft" edges involving perceptions as to what a "job" is and the like. Furthermore, we note that response changes as a consequence of repeated interviewing seem most marked in the Current Population Survey case among teenagers for whom indeed labor force behavior is a little more "soft" and definitions more optional.

In closing we might note that it is somewhat simpler to attempt a diagnosis of the effects of repeated interviewing and the contextual conditions which make them greater or lesser than to suggest how we might go about controlling them in those cases, unlike Ferber's, where we find them damaging. If the period between waves is at all short, as it often must be, it is impossible to remove the respondent's expectation of being interrogated again in whatever area of behavior is involved in the investigation. Conceivably, in the long run if we wish to preserve the values of the panel design we may be driven to highly multiple-purpose studies of a sort which will permit sufficiently rapid rotation of content within and across waves that the respondent is kept, so to speak, "off balance," without any clear expectation as to the content areas to be investigated in the next interview.

XI

CONTRIBUTED PAPERS - II

Page

Some Alternative Estimators for a Population Mean - Donald T. Searls, Westat Research Analysts, Inc. 234

The Use of Diverse Sampling Plans for the Collection of Transportation Data - Paul Rackow, Tri-State Transportation Committee.............. 238

Incorrect Inferences Commonly Drawn from Traditionally Designed Surveys - Albert L. Johnson, University of Miami............................. 248

The Problem of Geographic Contiguity - A Monte Carlo Approach - Stephen F. Gibbens and James A. Tonascia, California State Department of Public Health.. 253

The Effect of the Ghetto on the Distribution and Level of Nonwhite Employment in Urban Areas - John F. Kain, U. S. Air Force Academy and The RAND Corporation............................... 260

SOME ALTERNATIVE ESTIMATORS FOR A POPULATION MEAN

Donald T. Searls, Westat Research Analysts, Inc.

The problem considered in this paper is one that most samplers encounter early in their work with sample data. The problem can be illustrated by the question that a sampler frequently asks himself, particularly if he is working with relatively small samples. The question is, "What do I do with large or extreme observations in the sample?" The sampler first attempts to answer this question by a careful review of the data to see if an outlier has somehow appeared or if in fact the offending observation or observations are actually true observations. This paper is concerned only with the latter case where the use of outlier theory would be both unrealistic and unstatistical.

Suppose the data has been collected in order to obtain an estimate of the mean (μ) of the sampled distribution. The types of distributions generally encountered in practical sampling situations can be characterized by the following:

a. Unimodality
b. Positive skewness or symmetry
c. Non-negative values.

Unless the precise form of the distribution is known exactly, $\bar{y}$ will probably be used as an estimate of the distribution mean. An occasional sample will contain one or more observations from the right tail of the distribution due to the sampling process. When this occurs, and the sample size is small, $\bar{y}$ will probably exceed μ by a considerable amount. In this situation the client being a practical man is frequently quick to point out the fact that the one or more large observations are unduly influencing the estimate of the mean. The argument that the procedure is unbiased falls on deaf ears. The client is not interested in what happens in the long run--he wants an estimate as close to μ as possible for this particular case. In fact he is apt to regard any difference between $\bar{y}$ and μ as a bias. By this reasoning any sample estimate is biased unless it coincides with the parameter value.

This is an interesting viewpoint since it points up the fact that merely because these "biases" tend to average out in the long run does not really imply any particular merit for the estimator unless the estimates obtained are accumulated or manipulated in some fashion.

The above leads to the using of mean-squared-error as a criterion for comparing estimators rather than using unbiasedness and variance. There are not always unique minimum mean-squared-error estimators nor is it always possible to determine the MMSE estimator but this does not preclude its use as a criterion for comparing alternative estimators.

Let us examine the estimator generally employed both by statisticians and others in situations of this type. More often than not the large observations are ignored (generally there is only one) and the sample mean is derived from the remaining observations. The question immediately arises: Is this practice tenable with sound statistical theory? Surprisingly, the answer is a qualified yes. And in fact it appears that under proper conditions, if the observation was not ignored, it should have been.

Part of the qualification is that there is some predetermined point t such that if an observation is larger than this point it will be ignored. This point should be explicitly stated in editing instructions before the sample is drawn.

This paper will not provide pat answers for the handling of large observations but it will develop results to indicate that some of the procedures being employed should not necessarily be condemned and that rough guidelines can be derived for future use. Three procedures, in addition to the above, will be considered.

The first estimator formalizes as follows:

$$\bar{y}_1 = \frac{\sum_{j=1}^{r} y_j}{r} , \qquad (y_j < t) \quad (r > 0)$$

$$= t . \qquad (r = 0)$$

The case where r = 0 would be rare but is included for completeness.

For a given $r \geq 1$, $\bar{y}_1$ can be regarded as a simple random sample of size r from the truncated portion of the distribution function and as such it provides an unbiased estimate of the mean (μ_t) of the truncated distribution. Also r is distributed as the binomial with parameters n and F(t) or p. Let q = 1 - p.

$$E(\bar{y}_1) = (1 - q^n)\ \mu_t + q^n t \ ,$$

and

$$MSE(\bar{y}_1) = (1 - q^n)\ [\sigma_t^2\ E(1/r) + (\mu - \mu_t)^2]$$
$$+ q^n\ (t - \mu)^2 \ .$$

When t is equal to the upper limit (b), $MSE(\bar{y}_1) = \sigma^2/n$. If $\frac{d}{dt}[MSE(\bar{y}_1)]$ is positive as t approaches b, then $MSE(\bar{y}_1)$ must be less than σ^2/n for some region of t. This turns out to be true.

$$\frac{d}{dt}\,[\mathrm{MSE}(\bar{y}_1)] = \frac{f(t)}{p}\Big\langle (1 - q^n)\Big[\sigma_t{}^2 + E[\tfrac{1}{r}]$$

$$[(t - \mu_t)^2 - \sigma_t{}^2] - 2(\mu - \mu_t)(t - \mu_t)\Big]$$

$$- np(1 - q^{n-1})\ \sigma_t{}^2\ E[\tfrac{1}{r} \mid n - 1, p]$$

$$- npq^{n-1}\ (t - \mu_t)\ (t + \mu_t - 2\mu)\Big\rangle$$

$$+ 2q^n\ (t - \mu) \quad .$$

If $b = +\infty$ the term $(t - \mu_t)^2\ E(\frac{1}{r})$ will dominate as t approaches b. Thus the first derivative is positive as t approaches b.

A more detailed proof is presented in [1], along with proofs for cases where $b \neq +\infty$.

A second estimator that is sometimes used is one where the large observations are ignored but sample size is kept constant by replacing deleted observations.

$$\bar{y}_2 = \frac{\sum_{j=1}^{r} y_j}{r} \qquad \begin{array}{l} (y_j < t) \\ \\ (r \text{ fixed}) \end{array}$$

For any r, $\bar{y}_2$ behaves as the sample mean of the truncated distribution.

$$E(\bar{y}_2) = \mu_t \quad ,$$

and

$$\mathrm{MSE}(\bar{y}_2) = \sigma_t{}^2/r + (\mu - \mu_t)^2 \quad .$$

The proof for $\bar{y}_2$ is similar to that for $\bar{y}_1$ and can be found in [1].

A third estimator can be formed by replacing all large observations with the value of the cutoff point t.

$$\bar{y}_3 = \frac{\sum_{j=1}^{r} y_j + (n-r)t}{n} \qquad \begin{array}{l} (0 \leq r \leq n) \\ \\ (y_j \leq t) \end{array}$$

where r is distributed as the binomial with parameters n and p.

$$E(\bar{y}_3) = p\,\mu_t + qt \quad ,$$

and

$$\mathrm{MSE}(\bar{y}_3) = \frac{p}{n}\,[\sigma_t{}^2 + q(t - \mu_t)^2] + q^2(\mu_{t'} - t)^2,$$

where $\mu_{t'}$ is the mean of the right truncated portion of the distribution.

$$\frac{d}{dt}\,[\mathrm{MSE}(\bar{y}_3)] = 2q[\tfrac{p}{n}\,(t - \mu_t) - q(\mu_{t'} - t)] \quad .$$

The derivative is positive as t approaches b since the second term in brackets is approaching zero while the first term is approaching a positive constant.

Figure 1 and table 1 demonstrate the gain achieved for the exponential distribution by the use of $\bar{y}_3$ for various values of t/μ . Table 2 presents specified characteristics for $\bar{y}_3$ when optimum values of t are used.

If a near optimum value of t were used for $\bar{y}_3$ over half (56%) of the samples of size five would have one or more observations exceeding t. As the sample size increases this proportion approaches 1. Correspondingly, the expected number of observations exceeding t increases from .75 for samples of size five up to approximately 4 for samples of size 500. The optimum point for t varies from about double the true mean up to five times the true mean over this range.

The preceding results raise the question of whether or not conditions can be found such that one or more of the large observations can be arbitrarily discarded.

Consider the case where only the maximum sample observation (y_m) is discarded.

$$\bar{y}_4 = \frac{\sum_{j=1}^{n} y_j - y_m}{n - 1}$$

$$E(\bar{y}_4) = \frac{1}{n - 1}\,(n\,\mu - \mu_m) \qquad \text{where}$$

$$E(y_m) = \mu_m \quad .$$

$$\mathrm{MSE}(\bar{y}_4) = \frac{1}{(n - 1)^2}\,[n\sigma^{\,2} + \sigma_m{}^2 + (\mu_m - \mu)^2$$

$$- 2n\,\mathrm{Cov}(\bar{y},\ y_m)] \quad .$$

$$\mathrm{MSE}(\bar{y}_4) \leq \ \sigma^{\,2}/n \qquad\qquad \text{if}$$

$$\mathrm{Cov}(\bar{y},\ y_m) \geq \frac{1}{2n}\ [\ (\tfrac{2n-1}{n})\ \sigma^{\,2} + (\mu_m - \mu)^2 + \sigma_m{}^2].$$

For samples of size n = 2 equality holds for the exponential distribution and the strict inequality holds for the pareto. Table 3 presents results for the pareto distribution.

In conclusion it can be stated that the deletion of large observations from a sample may result in the use of an estimator with a smaller mean-squared-error than the sample mean. Perhaps an even more important implication from the above results however is the possibility that even more dramatic gains can be achieved in the estimation of σ^2.

REFERENCES

[1] Searls, D. T., "On the Large Observation Problem," Institute of Statistics Mimeo Series No. 332, North Carolina State College, Raleigh, N. C.

ACKNOWLEDGEMENTS

This paper presents a portion of the research conducted for a Ph.D. thesis written under the direction of Dr. A. L. Finkner. Acknowledgement for helpful suggestions are also due Walter Hendricks, Dr. R. L. Anderson, Dr. W. L. Smith, and Dr. C. Proctor.

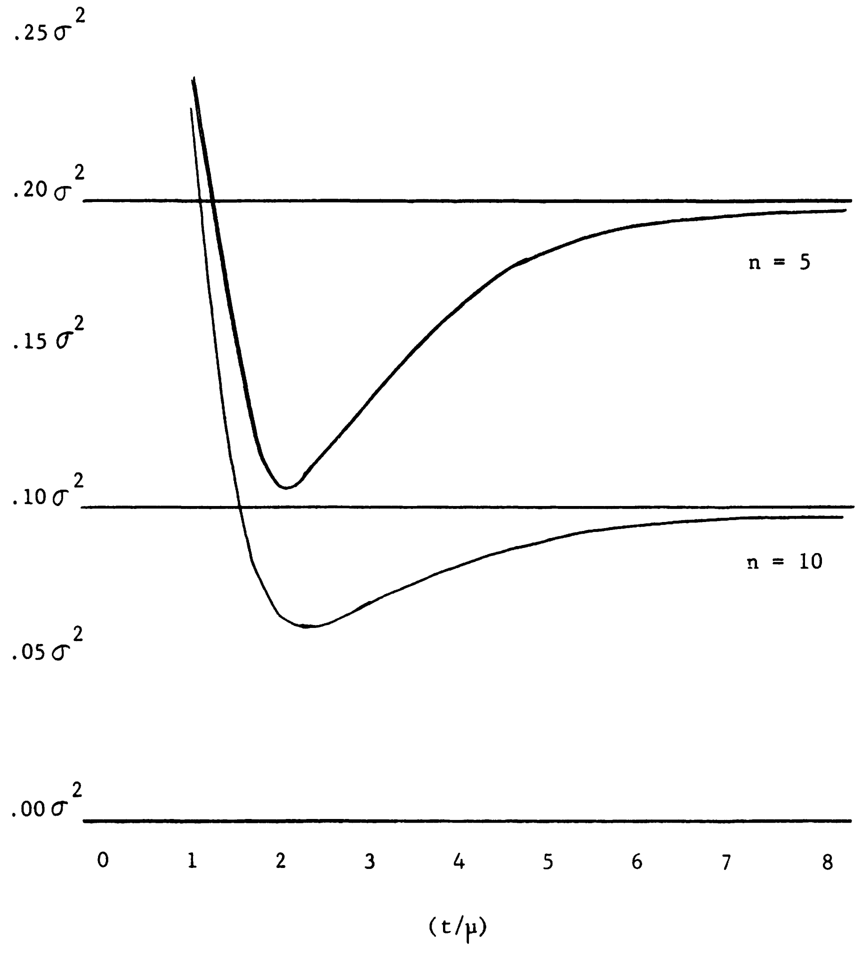


Figure 1. Comparison of MSE($\bar{y}_3$) and σ^2/n for the exponential distribution

Table 1. Relative efficiencies (%) of $\bar{y}_3$ for samples from the exponential distribution.

Values of t/μ	Sample Size 5	10	50	100	500
1	83.4	44.1	9.2	4.6	.9
2	183.1	153.3	66.7	39.1	9.1
3	140.5	138.0	120.6	104.2	49.9
4	117.0	116.7	114.9	112.7	97.7
5	107.2	107.2	106.9	106.7	104.7
6	103.1	103.1	103.0	103.0	102.7
7	101.3	101.3	101.3	101.3	101.3
8	100.5	100.5	100.5	100.5	100.5
9	100.2	100.2	100.2	100.2	100.2
10	100.1	100.1	100.1	100.1	100.1

Table 2. Exponential distribution; characteristics for $\bar{y}_3$ when optimum values for <u>t</u> are used.

Sample size	Values of t/μ	Exp. no. $>t_o$	Exp. % of samples with one or more $>t_o$	Relative eff. (%)
5	1.9	.75	55.5	184.9
10	2.2	1.11	69.1	156.7
50	3.2	2.04	87.5	121.3
100	3.6	2.73	93.7	113.7
500	4.9	3.72	96.3	104.7

$$f(y) = \frac{1}{\mu} e^{-y/\mu} \qquad (0 < y < + \infty)$$

Table 3. Pareto distribution; $\bar{y}_4$ with n = 2 .

Value of α	Value of $r^2\ (\bar{y}, y_m)$	Relative Efficiency (%)
3	.964	250.0
4	.947	197.5
5	.938	150.0
6	.931	137.5
10	.918	118.8

$$f(y) = \alpha\, y^{-(\alpha + 1)} \qquad (\alpha > 2)$$

$$(1 < y < + \infty)$$

THE USE OF DIVERSE SAMPLING PLANS FOR THE COLLECTION OF TRANSPORTATION DATA

Paul Rackow, Tri-State Transportation Committee

I INTRODUCTION

The problem of improving transportation systems for the growing number of urbanized regions throughout the United States has been receiving increased attention in recent years. Great impetus to Urban Area Transportation Studies was given through the passage of the Federal Highway Act of 1962. This legislation provided for the use of federal-aid highway money in the planning of integrated region-wide transportation systems. A region is analogous to the definition of a Standard Metropolitan Statistical Area (SMSA) as used by the Bureau of the Census in the 1960 Census of Population.

In August 1961 an informal arrangement between the Governors of New York, New Jersey, and Connecticut led to the formation of the Tri-State Transportation Committee. The three State Governments in conjunction with various Federal Agencies mutually agreed upon appropriate financing arrangements. Membership on the Committee included not only representatives from each of the three State Governments but also included participants from the U. S. Bureau of Public Roads, the U. S. Housing and Home Finance Agency, the Federal Aviation Agency and the New York City Planning Commission. Local cooperating committees representing cities and counties within the Tri-State Region are important adjuncts to the Committee. Following the structuring of the Organization, the next step was the completion of a Prospectus for the Committee which would outline the projects to be undertaken in order to fulfill its mission. In its own words the purpose of a transportation plan is to provide for "the expeditious movement of persons and goods throughout the region which is essential for the continued economic growth of the 17 million people who presently reside within it."

As might be expected, the first step in the process of transportation plan development is the large scale collection of relevant data. Previous urban area studies have shown the existence of an intimate functional relationship between trip generation and population, socio-economic, vehicle ownership and land use characteristics of a region. The projects assigned to the Data Collection Section were the design and field work associated with a number of travel inventories. These inventories are to determine the current functional relationships among the variables mentioned above for use in the projection of trips by mode and purpose in future periods. Hence while the survey results themselves are answering enumerative questions, the use to which the data will be used in transportation planning involves an analytic study. Thus, the survey is both enumerative and analytic.

II SURVEY SAMPLING PLANS

The Travel Inventories comprise three areas of Data Collection:

1. Home Interview Survey
2. Truck-Taxi Survey
3. External Cordon Line Survey

The aim of the Travel Inventories is to describe the totality of trips made by persons and vehicles via all possible modes of travel on an average weekday within a defined Cordon Area. The Cordon Area is defined as that portion of the total Tri-State Region which contains now or is expected to contain by 1980 most of the urban development. The line encircling this area is the frame within which we wish to estimate the totality of trips. Each of the three inventories provide a segment of the total trips made within the Cordon. The Home Interview Survey provides those trips originating at the living places within the Cordon by all modes of travel except trucks and taxis. These trips are estimated through the Truck-Taxi Survey by sampling from Motor Vehicle Bureau files. The remaining segment of trips made within the Cordon Area originate outside the Cordon. The External Cordon Line Survey samples trips made by automobiles, trucks, and taxis crossing the Cordon that are destined inside or outside the line. Those trips found to be garaged within the Cordon Area are duplicated by results from trips made either by the Home Interview or Truck-Taxi Survey. These must be eliminated before the results from the three surveys can be added together.

This paper does not propose to discuss the survey methods associated with the External Cordon Line Survey. That survey was the responsibility of each of the three concerned State Highway Departments. Basically, the sampling plan involved the collection of at least 10% of the trips made across the Cordon Line via all roadways with an average of 2,000 vehicles or more per day. For safety and other practical reasons, it was felt that a probability mechanism was not applicable to the type of roads concerned. It is the opinion of the author that a modification of a probability sampling technique described in a paper he and others presented to the American Statistical Association in 1960 may be applicable even here.[1]

[1] Leslie Kish, Warren Lovejoy and Paul Rackow "A Multi-State Probability Sample for Traffic Surveys" Presented at the 1960 Annual Meeting of the American Statistical Association.

A. Sources Available for Sample Selection

There are a number of sources available for selection of the Home Interview Survey in the New York-New Jersey-Connecticut Tri-State Region. These range from various field listing techniques to the use of Sanborne or other maps, Reverse Telephone Directories, or the records of the various Utility Companies that cover the region. Typically, Urban Area Studies have used the results of a Land Use Inventory which collects data on the development of the land including a detailed listing of all living places within the area as a means for a selection of a sample of households. However, the situation at Tri-State required that both the Land Use Inventory and the two Travel Inventories be conducted almost simultaneously.

In the case of the Truck-Taxi Survey, the source for selection was the records of Truck and Taxi Registrations within the Cordon Area of the three concerned States. Each State presented unique problems of sampling due to both the form in which their records were kept and their willingness to cooperate in processing the data.

B. Requirements of the Sample Design

The sample design for the two travel inventories had to be economical and practical in order to meet the necessary requirements for speedy field work and reasonable workloads, conditions of work and overall cost.

There were a number of specific requirements which had to be met by the Survey Design:

(1) Sample estimates had to be produced for a great variety of characteristics. These included such things as trips between geographic areas, trips by mode of travel, purpose of trip or land use category. In additions, such items as car ownership rates and incomes distributions for various household types by varying population density groups in the region was necessary as well. These estimates were for trips made on an average weekday over the period of the survey.

(2) The sampling frame included all housing units and other special dwelling places within the defined Cordon Area for the Home Interview Survey and all trucks and taxis registered with the three State Motor Vehicle Departments within the Cordon Area for the Truck-Taxi Survey.

(3) The sample was a probability sample requiring that every housing unit (or special dwelling place) or truck-taxi registration have a known probability of selection achieved by randomized selection procedures.

(4) Built into the sample design were procedures whose purpose was to measure the sampling errors associated with the various sample estimates to be made.

(5) The statistical reliability required in the survey specified that with the expected overall sample of 450,000 trips, a sample estimate of 1% or 4,500 trips, would have a coefficient of variation of the order of 2.5% or less.

The design had to be tailored to the practical problems inherent in the actual conduct of a Home Interview or Truck-Taxi Survey, and these considerations necessitated the use of the following procedures:

(6) The survey budgets permitted the use of some 250 interviewers, control clerks, and other administrative personnel to collect the necessary data over a relatively short time period. The Home Interview Sample was randomly assigned to 80 travel dates while the Truck-Taxi sample was spread over 91 dates.

(7) The survey designs included a 1% probability sample of housing units and other special dwelling places for the Home Interview Survey and a 3% probability sample of appropriate truck and taxi registrations.

(8) To improve coverage of different origin and destination patterns and mode of travel usage, both samples are randomly spread over time in such a way as to insure geographic dispersion for each weekday assignment.

(9) Average workload per interviewer ranged from about eight home interview assignments in the City of New York per weekday to a lesser amount in the more suburban areas. Assignments usually averaged 7 sample units for the Truck-Taxi Survey.

C. Field Survey Procedures

The Home Interview Survey was designed to obtain weekday trip information and other socioeconomic data from a sample of persons residing in housing units and other special dwelling places. As a means to this end, the frame for sampling purposes was defined as living places located within the Cordon Area. Sample units selected were randomly allocated to unique weekday travel dates over an 80 day period. The field interviewer was assigned to obtain the necessary data from each person 5 years of age and older residing at the living place for the randomly chosen travel date only. This was accomplished by visiting the residence of the day following the travel date after a letter announcing the intended visit had been sent a week earlier. Up to three days after the travel date assignment was allowed for completion of the interview.

Similarly, the Truck-Taxi Survey was designed to obtain weekday trip and other data from a sample of trucks and taxis registered within the Cordon Area.

The sampling frame comprised trucks and taxis registered with each of the three State Motor Vehicle Departments. After selection of the sample and random allocation to 91 travel dates, field interviewers were required to complete their questionnaires by oral interview of truck or taxi owners, dispatchers, drivers, etc. within three days after the assigned travel date. As in the Home Interview Survey, letters to prospective respondents were sent well in advance to insure a maximum of cooperation.

D. Types of Sampling Plans

1. Home Interview Survey

A number of sampling frames were used as sources for selection of a one-percent probability sample of living places in New York-New Jersey-Connecticut Cordon Area.

New York City

In cooperation with another public agency which had already completed the necessary preparatory work a one-percent clustered area probability sample was selected. The sampling frames were defined by two strata:

a. The civilian, non-institutional population living in housing units and other special dwelling places in existence according to the 1960 Census of Population and Housing.

b. Housing units built since the 1960 Census up to and including February 28, 1963 as represented by occupancy certificates obtained from the New York City Department of Buildings

Blocks, as defined by the Bureau of the Census, are assigned measures of size approximately equal to one-eighth the number of housing units found in the 1960 Census. These measures of size are always whole numbers. A block with a given measure of size is said to contain that many clusters of housing units which, hopefully, average approximately eight housing units.

Some of the defined Census blocks received no independent measures of size. A block with less than six housing units recorded in the 1960 Census was arbitrarily joined to an adjacent block that contained six or more housing units. Such linked blocks were treated as single pseudo-blocks.

Certain blocks were excluded in whole or in part. Census blocks wholly excluded are those within military areas and blocks along the waterfront which yield a largely transient population depending on the number of boats docked. Blocks excluded in part are those portions containing institutional population as determined by the Census. However, housing units for institutional staff members were not excluded.

Non-institutionalized persons living in group quarters were converted into measures of size on the basis of twenty-five people per measure. These measures were added to that for housing units to obtain a total measure of size for the block.

After all conversions to measures of size were completed, geographic area sequence was randomized as follows: First, a random arrangement of boroughs and within each borough a random ordering of health districts was defined. Finally, within health districts, a random sequence of health areas was determined. As a rule, a health area contains more than five census tracts. Upon completion of this randomization procedure, every 100th measure of size was selected following a random start. In the borough of Richmond, every 50th measure was selected and later, after definition of the interviewing cluster, a random half per cluster was taken so that each cluster averaged four housing units.

It is felt that the randomization procedure used reduced to insignificance the likihood that final ordering of the areas contained any periodicity in the occurrence of characteristics of the population that would be a multiple of the sampling interval. Hence, this systematic sample of measures of size attached to randomly ordered areas may be considered as a simple random sample of measures of size.

In New York City, whenever a new residential structure is constructed or an old structure is converted or modified, occupancy permits must be secured from the Department of Building before such residential quarters may be occupied. These permits are a source for determining new construction since some base period such as the April 1960 Census.

Data on occupancy permits issued since the 1960 Census up to an including February 1963 were secured and included physical location(address and apartment number) and the number of residential units involved.

The structures involved were arranged in tax block sequence within each borough. Measures of size, approximately equal to one-eighth the number of units covered by occupancy permits, were assigned to each tax block, and every 100th measure was selected after a random start.

If one or more of the structures on the census block were included in the occupancy permit frame, whether or not they were selected from that frame, the census block was considered to contain all housing units outside the occupancy permit structures. Thus, if a sample hit were made in the selection of measures from stratum A (1960 Census) turns out to be in a structure which is a member of the occupancy permit frame, that measure is given a "no units" designation and is not covered in the field interviewing. Also, "no units" designations can result for measures selected from stratum A but which were demolished since the 1960 Census with no subsequent construction of residential quarters until after the interviewing period.

Those Census blocks containing selected measures of size had to be listed in detail by field personnel. After listing all housing units within the block, they were grouped into segments equal to the number of measures previously assigned to the block. A great effort was made to define the measures with a maximum of physical contiguity to insure unique identification of the segment assignment. Then preceding in a logical, pre-established sequence, the appropriate selected measure was determined.

Definition of the sample from the frame of occupancy permits was relatively simple. No comparison had to be made to exclude census block population. The unique physical limits of the selected measure had to be determined.

A modification of the sample design described above was introduced to reduce sampling error in that portion of the sample containing hotel and motel living places. This was done because of the Census Bureau definition of housing unit and our desire to sample persons living in transient hotel and motel rooms.

Another sampling frame was defined consisting of hotels and motels in New York City having 100 or more rooms assigned to transient guests. Measures of size equal to approximately one-tenth the number of transient guest rooms were assigned to each hotel and motel in the frame. After geographic stratification of the frame, every 200th measure was systematically selected in two independent replications. A precise definition of transient status was given to each interviewer and all assignments were allocated to travel dates at random over a time period. Interviewers were instructed to interview only in rooms within the assigned measure containing transient guests. The introduction of this new sampling frame had an effect on the major sample described previously. The effect was as follows:

(1) All previous sample measures that fell in hotels or motels included in the transient sampling frame are to be treated specially. Only rooms within the defined measures which contain permanent guests were to be included in the sample and interviewed. Others were to be considered out-of-scope.

(2) All other previous sample measures that fell in hotels or motels which are not members of the transient sampling frame were to be interviewed in entirety as a catch-all measure. That is, all persons living in rooms defined by the measure are to be interviewed whether they are transient or permanent guests.

In summary, there are then three types of sample measures defined for hotels and motels in New York City:

(1) Transient Measures
(2) Permanent Measures
(3) Catch-all Measures

Outside New York City

The necessity for a speedy, inexpensive sample selection scheme required the use of the records of the various Electric Utility Companies as a sampling frame for most of the remainder of the Cordon Area outside New York City. Although initial costs were insignificant (all utilities cooperated as a public service with no charge to Tri-State) and the sample was selected quickly so that processing for field assignment could proceed rapidly, other methodological problems arose which led to considerable expenditure of time and money.

Utility company records are broadly split into two groups which are of interest to us in sampling living places:

(1) Residential rate customers usually representing individual housing units in the Census definitional sense.

(2) Commercial or General Service rate customers representing multi-unit housing units in the Census definitional sense.

A residential account or meter may represent two or more housing units (living places) in some cases but generally are in bi-unique correspondence to one housing unit.

However, commercial rate accounts range from representing an office in a commercial building with no housing units to a motel representing 50 separate family units. Commercial or General Service rate accounts do not include a source for sampling large housing complexes such as Public Housing or Private cooperative Housing which is supplied electricity through demand meters (master-metering). However, many of the utility companies did supply us with information about non-public large housing complexes supplied electricity in this manner. Further assistance in this problem was obtained through the cooperation of the various County Planning Boards that comprise the Tri-State Region. Information about large public housing complexes with demand meters was secured from the various Public Housing Authorities.

Each utility company was asked to select one in every one hundred residential rate accounts following a random start. Records of each company are in meter-reading cycle order which is, in effect, a geographic stratification of the universe. A meter cycle lasts either 21 or 42 days. In addition, a 100% listing of all commercial or general service accounts were obtained as well. The sample frame included both active and inactive meters; hence, vacant units as well as occupied ones are reflected in the sample. However, since each sample was selected from utility company records at an instant in time, there is no reflection of housing unit growth from the time of selection to the end of the field interviewing period.

The commercial rate accounts although numbering about 225,000 represent much less than 5% of the final residential units sample. Nevertheless, to insure complete coverage of the living place frame, a sample from this portion is necessary though costly. It was decided to stratify this frame into three groups based on the number of living places likely to be associated with an account.

(1) Accounts likely to have no housing units or other special dwelling places.

(2) Accounts likely to have from one to twenty-five housing units or other special dwelling places.

(3) Accounts likely to have more than twenty-five housing units or other special dwelling places.

Group (1) was sampled at a rate of one in every one hundred; group (2) at one in every ten; and group (3) was covered one hundred percent. All accounts selected from groups (1) and (2) were assigned to field personnel whose job was to determine the number of housing units or special dwelling places associated with each account.

These results were then sampled at a second stage to provide an overall probability of 1 in 100. That is,

Group (1) 1/100 = 1/100 X 1/1

Group (2) 1/100 = 1/10 X 1/10

Accounts in group (3), such as hotels, motels, rooming houses, hospital staff quarters and school dormitories, were sampled at a rate of 1 in 100 at the 1st stage to obtain a 1% sample. Information on the number of housing units and/or other special dwelling places was obtained for each in-scope account by mail or telephone contact. After this data is obtained the 1% sample is selected systematically following a random start.

Data on the number of housing units associated with structures supplied electricity through demand meters (master-metering) was obtained through the cooperation of the various utility companies, county planning boards or public housing authorities. Again, a 1% systematic random sample was selected following a random start.

Some towns within the Cordon Area outside New York City were not covered for one reason or another by sampling the records of the various utility companies. These enclaves were sampled by means of a block field listing procedure analogous to the area probability design in New York City. Estimates of the number of housing units in existence by block were available to us through previous work of the Tri-State Land Use Inventory. With this data as a frame, a 1% random sample was selected systematically in five independent replications of 0.2% each. All unduplicated blocks in which the sample hits fell were then again field listed in detail with great care taken to physically identify each individual housing unit. Each sample hit was then uniquely associated with its appropriate physically defined housing unit. Additional sample units or a reduced number was a function of the accuracy of the original block estimates of the number of housing units. The replication feature of the survey design reduced the field listing work as well as providing a simple procedure for calculating sampling errors of the estimated to be produced.

Military Sample

All branches of the Armed Services (Army, Navy, Marines, Coast Guard and Air Force) were contacted and asked to provide data on the number of family housing units and barracks living quarters (rooms, cots, beds, etc.) available for officers and enlisted men within each base in the Tri-State Cordon Area. A systematic random sample of 1% of these living places were selected and assigned for interviewing. Additional sample units were interviewed when original estimates were found to be low in relation to actual numbers of living places. Conversely, a reduction in the original assignment occurred when estimates from the military overstated the true number.

2. Truck-Taxi Survey

The sample frames used for selection of the three percent probability sample of trucks and taxis were the records of the Motor Vehicle Departments in New York, New Jersey and Connecticut. The records are kept in a different manner by each of the three states ranging from micro-film on cards to IBM punch cards to 1401 computer tape files. The sampling plans were tailored to fit each of these three record forms without sacrificing the requirements for the proper selection of a probability sample. The basic design incorporated the selection of three independent subsamples of one percent each with varying degrees of natural geographic stratification. Replication of the sample in the form of independent subsamples has many statistical advantages the primary one being the ease of calculating the standard errors of the survey. Each subsample was selected from a random start followed by systematic selection of every one-hundreth license plate thereafter. This procedure was accomplished in New York and Connecticut with selection taking place in a serpentine fashion through each District office or tax town thereby providing initial natural geographic stratification. In New Jersey, an initial ten percent sample was selected by associating a random number with the last digit of the in-scope truck or taxi registration. After geographic stratification of this initial sample by county was effected, three independent subsamples of one percent each was selected by a systematic random sample of 30% of the initial sample.

A separate selection of three one percent samples of in-scope Post Office Department and Military Base vehicles was effected manually after lists of appropriate license plate numbers were received. Stratification by municipality and base respectively was accomplished before selection.

E. Audits Applied to the Data

Immediately after selection of the home interview sample, various checks were applied to the sample to insure its reasonableness when compared to published sources. Outside New York City, for each municipality, town or other political subdivision within the defined Cordon Area, building permit data was obtained from appropriate state agencies covering the time period from the 1960 Census up to as close to the survey periods as was available. Building permit data includes information on the number of housing units that are expected to be built in the political subdivision. While, in some cases, the housing unit may not in fact be built by the contractor, this is rare. Usually a unit is built within three to six months after a building permit is taken out by the contractor. All building permit data was available up to at least the end of December 1962. This is six months before the beginning of the field survey period for the Home Interview Survey. Building permit data added to the 1960 Census of housing figures for each political subdivision gave us reasonable estimates of the number of housing units in existence as of the beginning of the survey period. The one percent sample after expansion by the inverse of the sampling fraction was compared to these independent estimates. Any large discrepancies were checked in detail to account for the difference.

In New York City, a different type of audit was conducted after the completion of the sample survey. The primary sample unit in New York City was a uniquely defined cluster of about eight housing units. To check the accuracy of interviewers in carrying out instructions to interview all units within their assigned cluster, an independent assessment of the number of units within a sample of clusters was made. A simple random sample of clusters was selected after stratification by health districts and week of work assignment. The reason for health district stratification was the apriori knowledge of varying difficulty in carrying out assignments in the different parts of the city. Week of assignment was considered an important mode of stratification since, apriori, it was felt that quality interviewing takes some time to achieve. Hence, after these two modes of stratification, a simple random sample of clusters was taken; one of 20% for the first four weeks of assignment and 10% for the remaining weeks. The results of the audit indicated an overall understatement of about 1% of the housing units that should have been interviewed but were missed. This varied considerably by geographic area in the city ranging from a 0% to 4% understatement.

III SURVEY RESULTS

The objectives of the survey consists of making many estimates of trip production for various geographic couplets by mode of travel and purpose of trip. This data is related to the use of the land at the couplet as well as car ownership rates and income category at households generating trips. Estimates are projected for an average weekday throughout the survey period although analysis of peak and non-peak hour periods will also be ascertained.

A. Sample Estimates

Any simple estimate from the survey such as some aggregate of a characteristic may be estimated simply in the form X" = Kx where K is the inverse of the sampling rate and x is the weighted count (including adjustment for non-response) in the sample with the specified characteristic. An analogous estimate for a ratio would be of the form X^1 = = p where p is some proportion made up of two characteristics x and y. In the usual case, x may represent the number of trips made by y household or p is a statistic of trips per household.

B. Sampling Errors

Two types of error arise in sample surveys: sampling errors and non-sampling errors. Non-sampling errors come primarily from errors of response in collecting data, errors contributed during the processing of the results and any bias in the sample due to non-response. Errors of response were kept to a minimum by careful training of interviewers who were supervised by people experienced in home interview techniques or familiar with the trucking industry. Processing errors were kept within bounds through the use of quality control methods. The possibility of bias due to non-response of a small proportion of the sample assignments remains. The method for dealing with non-response is discussed in the next section under non-sampling errors.

Sampling errors of the survey estimates arise from the fact that the characteristics as mirrored by the sample do not exactly coincide with the characteristics that would emerge from an equal complete coverage of the entire frame.

For computing the sampling errors in New York City we use a model in which the entire selection for the Health District within a borough consists of two random and independent halves. Actually, by this method of collapsed strata[2] we create the two computing units. Since this disregards further stratification actually accomplished, this results in slight over-estimation of the variance.

The two halves for the jth Health District may be represented as follows in estimating the proportion for some characteristic:

$$p_j = \frac{x_j}{y_j} = \frac{x_{j1} + x_{j2}}{y_{j1} + y_{j2}}$$

For the entire city of 30 Health Districts the simmilar estimator is the ratio of the sums of the Health Districts.

$$p = \frac{\bar{x}}{\bar{y}} = \frac{x}{y} = \frac{\sum_{j=1}^{30}(x_{j1} + x_{j2})}{\sum_{j=1}^{30}(y_{j1} + y_{j2})}$$

The "relvariance" (the square of the coefficient of variation) of p can be estimated by:

$$C_p^2 \doteq C_{\bar{x}}^2 + C_{\bar{y}}^2 - 2\, C_{\bar{x}\bar{y}}$$

$$C_{\bar{x}}^2 = \frac{1}{x^2} \sum_{j=1}^{30} (x_{j1} - x_{j2})^2$$

$$C_{\bar{y}}^2 = \frac{1}{y^2} \sum_{j=1}^{30} (y_{j1} - y_{j2})^2$$

$$C_{\bar{x}\bar{y}} = \frac{1}{xy} \sum_{j=1}^{30} (x_{j1} - x_{j2})(y_{j1} - y_{j2})$$

Thus from the sum of 30 Health District contrasts for the City's estimates the relvariance can be computed with 30 degrees of freedom equal to the number of Health Districts within N. Y. C. These computations are made for a large number of characteristics which are then plotted and average values, subject to smaller variations, are used for estimating standard errors.

The relvariance for a simple estimate (X" = Kx) is just $C_{\bar{x}}^2$ defined above.

---

[2] Nathan Keyfitz, "Estimates of Sampling Variance Where Two Units are Selected from Each Stratum", Journal of the AMerican Statistical Association, 52, (1957), Pp. 503 - 510.

Outside New York City the method of random group estimation of variance will be used. The entire sample is divided into t random groups of K units each. An estimation of the variance[3] may be expressed as

$$S_K^2 = \sum_{g=1}^{T} \frac{(x_g - \bar{x}')^2}{K(t-1)}$$

where x_g = the value of the characteristic in the gth group

$\bar{x}'$ = the mean per group of the t group totals

For both estimates inside or outside New York, sampling errors of various size cells for a representative group of important characteristics can be plotted on a graph. Sampling errors for other estimates not plotted can be read off the graph.

The design of the Truck-Taxi Survey as a replicative sample in three subsamples provides a simple procedure for producing estimates and calculating sampling errors.

Simple inflation estimates are of the form $X' = Kx$ in which K is the inverse of the sampling fraction and is the value of the characteristic for the three subsamples added together. Ratio estimates of the form $f = \frac{x}{y}$ where x is the value of some characteristic in the sample and y is the value of some other characteristic in the sample, both for the three subsamples combined.

The variance of the two types of estimates may be expressed as follows:

$$\text{Var. } X' = \left(1 - \frac{R}{K}\right) \frac{K^2}{R(R-1)} \sum_{i=1}^{M} \sum_{j=1}^{R} \left(x_{ij} - \bar{x}_{i.}\right)^2$$

$$\text{Var. } f = \left(1 - \frac{R}{K}\right) \frac{B}{y^2} \frac{1}{R-1} \sum_{i=1}^{M} \sum_{j=1}^{R} \left[\left(x_{ij} - \bar{x}_{i.}\right) - f\left(y_{ij} - \bar{y}_{i.}\right)\right]^2$$

where $R = 3$ = the number of independent subsamples

K = sample inflation factor

x_{ij} = value of characteristic x in the jth subsample and ith stratum

y_{ij} = value of characteristic in the jth subsample and ith stratum

$\bar{x}_{i.}$ = mean value for characteristic in the ith stratum

$\bar{y}_{i.}$ = mean value for characteristic in the ith stratum

C. Non-Sampling Errors

There is no consistent or unbiased way of adjusting for non-response in surveys of a human population. The magnitude of the biases resulting from any subjective procedure is not known. Hence, great efforts were taken to achieve small non-response rates within the limits of the budget. The effort resulted in a non-response rate of about 10% for the Home Interview Survey and less than that proportion for the Truck-Taxi Survey. These rates are computed on the base of in-scope housing units or licensed trucks and taxis for which trips are possible.

Criteria for representing Home Interview non-responding sample units by factoring completed samples include:

(1) Geographic area on a Census Tract grouping level in New York City and a Municipality grouping outside of New York City.

(2) Composite of structure types and housing density grouping.

The strata so defined are felt to be appropriate with respect to the characacteristics being measured by the survey.

The similar criteria for the Truck-Taxi Survey sample include:

(1) Geographic area on a county group level for the truck or taxi base of operations.

(2) Body type (truck) or vehicle type (taxi).

Each of the three independent subsamples are to be treated separately throughout.

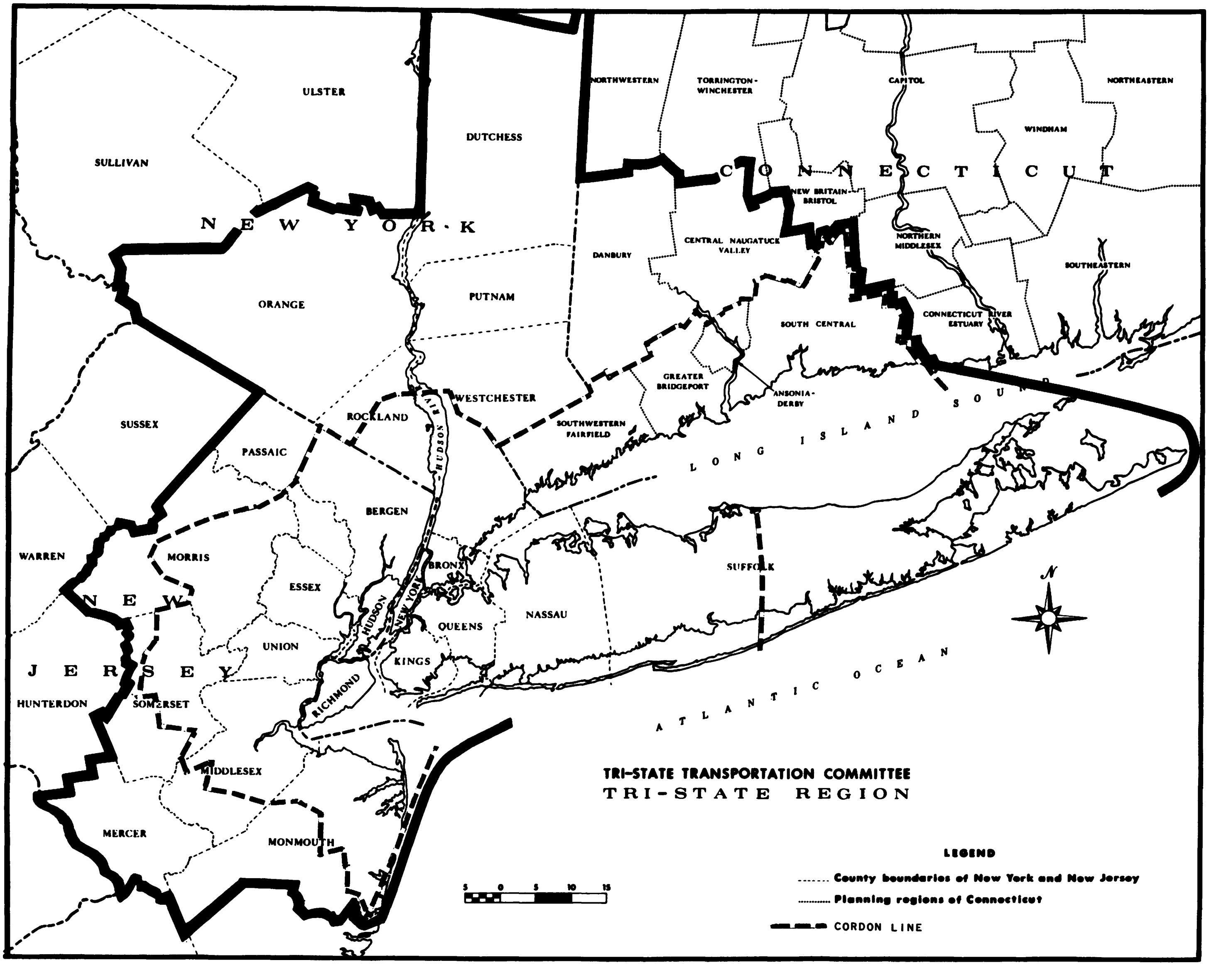
---

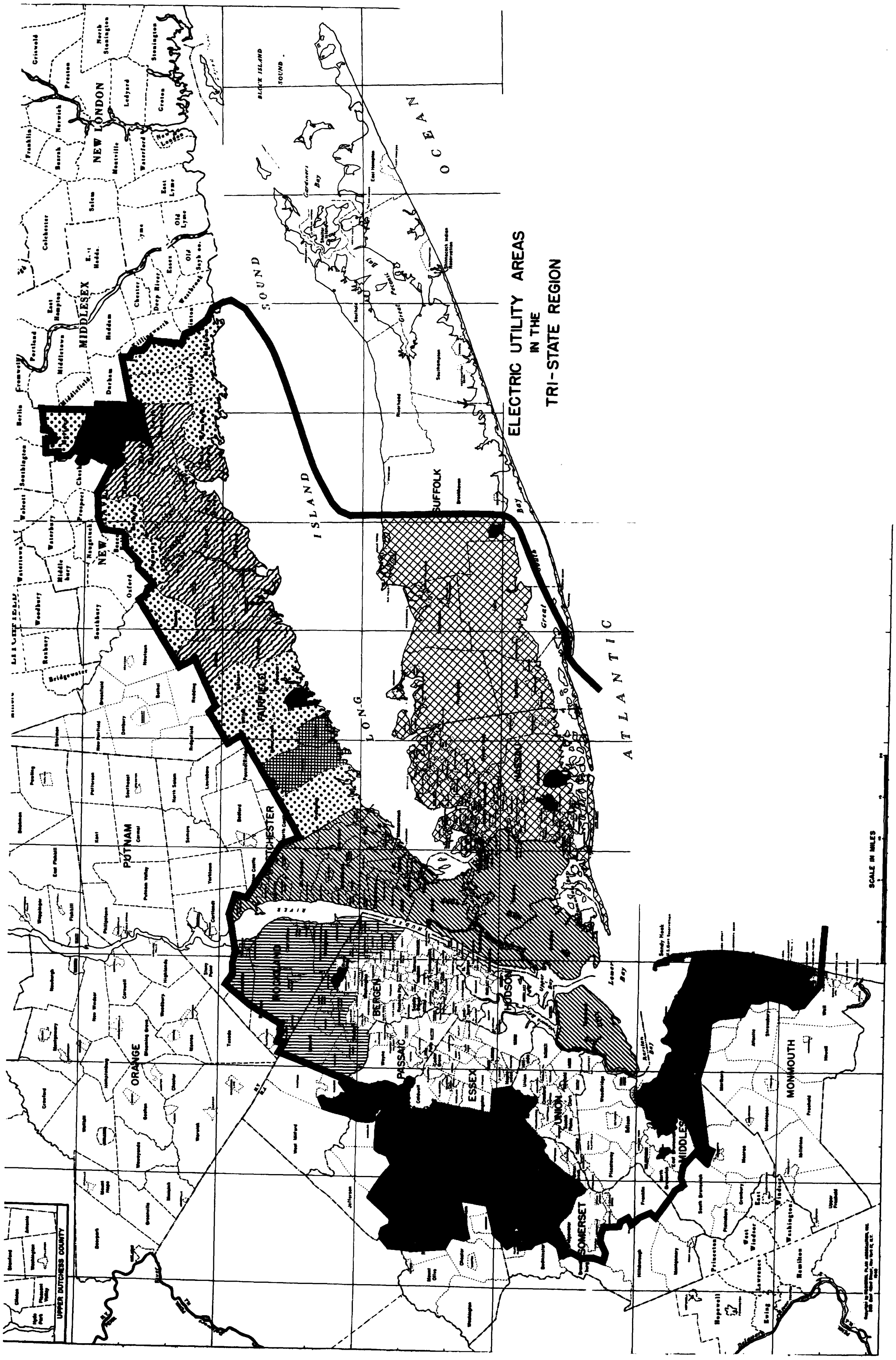
[3] M. H. Hansen, W. N. Hurwitz, and W. G. Madow, "Sample Survey Methods and Theory," Volume I, Pp. 440 - 444.

For each of the strata defined above, the ratio Interviewed Households (Trucks, Taxis) Plus Non-Interviewed Households (Trucks, Taxis) Interviewed Households (Trucks, Taxis) is computed.

These ratios are applied to all data for interviewed households in the corresponding strata, except in groups where the ratio exceeds 2. In such cases, the process of combining strata is continued until the ratio becomes 2 or less.

APPENDIX





ELECTRIC UTILITY AREAS
IN THE
TRI-STATE REGION

INCORRECT INFERENCES COMMONLY DRAWN FROM TRADITIONALLY DESIGNED SURVEYS

Albert L. Johnson, Department of Medicine University of Miami

Introduction

Those of us sitting in on the sessions sponsored by the Social Statistics Section for the past three days have heard many excellent suggestions regarding sampling schemes, cost reduction methods, analytic techniques, measurement problems in panel operations, and some able presentations contributing to our better understanding of certain current social problems. Through all of this, we - or at least I - did not hear specific reference to the problems which serve as content for this paper. Perhaps such omission is justified in meetings such as these because those in attendance at such gatherings are far too sophisticated to commit such errors in their own work. Or perhaps our recent concerns with the more technical aspects of design, collection, and analysis have overshadowed certain of the basics in survey research which we learned so long ago in our apprenticeship training. Whatever the reason for omission of these points in prior discussions, it is the intent of this paper to belabor the obvious in the hope that the future will find no evidence in the survey research literature that such practices exist.

The Nature of the Critique

It seems indicated at the outset to set the stage for the development of this critique by clarifying some terms used in the title of this paper. For instance, the title refers to "traditionally designed surveys" and the intent here is to include all types of data gathering procedures which seek information from people - no matter how sophisticated the techniques used to gather this information - after those people have been exposed to some related decision making experience. Examples of interest in my work would be surveys conducted to gather information from people who have and have not performed some health related piece of behavior like the taking of a preventive medication, or surveys conducted to gather information about patients suffering from a specific type of disease such as coronary heart disease. In both of these examples, traditionally designed refers to the fact that the behavioral act or the disease state existed in fact before the survey probe was made.

"Incorrect inferences commonly drawn," refers to the tendency of authors of reports based on these types of surveys to infer a causal relationship between two or more of the variables included in their studies. In the latter example just cited - surveys of characteristics of patients with coronary heart disease - several authors have noted the existence of very high levels of anxiety, repressed hostility and other personality dimensions. They then infer that these traits are causally associated with this disease state and fail to point out that it is equally possible for these traits to be the result of the disease. For an excellent review of this type of literature and a cleverly designed study to untangle the time dimension between these variables one is referred to Dr. Michel Ibrahim's (1) "The Role of Certain Psychological Factors in Coronary Disease, Blood Pressure, and Serum Cholesterol" in which he casts considerable doubt on the hypothesis that these personality traits preceed - therefore could be causally associated with - this state of ill health.

The former example cited - surveys conducted after people have been exposed to an experience requiring them to act with regard to a preventive medication - will be used to supply the data upon which this paper is founded. While these two examples seem rather different in terms of problem formulation, research design, data gathering, and analytic techniques, they were chosen to illustrate the generic aspects of the critique being developed in this paper. A great many other examples could be cited from marketing research, opinion polls, etcetera but to cite these would only detract from the main point of interest which is to substantiate the critique that has been made.

The Data

In 1960, Dade County, Florida (Miami) was selected as the site for the first large-scale field test of a new polio vaccine in the United States, and we were given the opportunity to study the public's reaction to this new preventive measure. The sponsors of that program, however, did not provide this opportunity until two weeks before the public distribution of the vaccine was to begin. This obviously precluded any before-study of the target population and all we could hope to achieve was a chronology of events leading up to the decision to carry out the program in this county of 1,000,000 persons, and the efforts made to secure approval and support of all the official and non-official agencies and organizations necessary to reach nearly 500,000 persons with the vaccine in a period of about twelve weeks. That aspect of our research endeavor was referred to as the community organization phase and contains some extremely interesting data in its own right, but is not particularly germane to the topic under discussion today. While the community organization phase of the research endeavor was being conducted, attention was focused on design of an instrument to gather data regarding the characteristics of the target population, recruiting and training of field staff and all those other details that go into a large-scale survey. We were ready at the end of the vaccine program to initiate our house-by-house interview procedures as soon as the vaccine program was ended. In drafting our questionnaire for use in the field we drew very heavily upon the prior work done with regard to the introduction of Salk vaccine into this country about 1954 and the rash of studies conducted in relation thereto. This earlier work suggested, among other things, that people's social status as well as their perceptions of risk to and severity of poliomyelitis were major determinants of their acceptance or rejection of such medical innovations. Consequently we built in measuring devices to tap these dimensions and were able to support the association between social status level and likelihood that the vaccine was taken, but unable to find evidence supporting the perceived risk-severity hypothesis. The data from that study have been widely circulated (2) therefore, will not be reintroduced here.

In 1962 a similar program was initiated in another county in Florida (Hillsborough-Tampa) and we were again invited to study the populace's participation patterns (3). This time, however, the request, was received sufficiently in advance of the initiation of the vaccine program as to permit a before-after design which would permit ascertainment of selected characteristics of a sample of the target population both before they were informed that they were going to be asked to take a new vaccine and after they had acted by either taking or not taking one or both doses offered them. In the 1960 study we were not certain whether the inability to obtain a significant association between respondents' perceptions of risk and severity regarding poliomyelitis and behavior with regard to taking or not taking the vaccine resulted from an inadequate tap of these perceptions or the fact that they were being measured after the act had been committed. In the 1962 study, clarification of this point was possible by improving the measuring instrument and by obtaining readings both before and after the behavioral act of taking or not taking the vaccine. These data are used to support the critique conveyed in this paper.

In the Hillsborough study measurements of perceived susceptibility to poliomyelitis and perceived severity of the disease if contracted were made on a modified semantic differential scale developed by Jenkins (4) as part of a larger piece of research he was then developing. With regard to perceived susceptibility the respondent was asked to mark a point on a forty-one point scale anchored at one end with: "The chance you have of getting it (poliomyelitis) is": "Big Chance." and at the other end with "No Chance." There were no intervening labels on this scale but it was marked at quarter inch intervals to guide the respondents' estimate of relative distance between the two extremes. This same type of scale was used in both the before-program and after-program interviews. Table 1 shows the responses obtained in the pre-program interview according to the respondent's then current status relative to having or not having previously taken any Salk Vaccine injections against this disease. Table 2 shows the responses obtained from all respondents at two different points in time. These and subsequent tables are taken from an earlier publication (5).

The data in Table 1 reveal a significant difference in level of perceived susceptibility to poliomyelitis, at the .05 level, whether tested by the non parametric Komozorov test or the traditional parametric two sample t-test. This difference was encouraging in the sense that this scale had not previously been validated, and there was a priori reason to believe that those who had taken the injected polio vaccine should feel somewhat less at risk to this disease. Along the same line of reasoning, the data in Table 2 were reassuring in that a significant decrease in susceptibility following the oral vaccine program was detected. On the basis of these data, plus the fact that respondents' scores on this scale with respect to other diseases properly arrayed these diseases in terms of current incidence rates among young adult populations, it is believed that the scale was adequately measuring perceptions of susceptibility to the disease of interest.

With this assurance that the scale seemed to be measuring levels of perceived susceptibility to poliomyelitis we can look at the data obtained both before and after the vaccine program for groups of respondents who had acted in various ways during the interval. Those data are given in Table 3. If only the after-data, as reflected in the column headed "as of May" in Table 3, were available the situation would be like that pre-

Table 1. Cumulative Percentage Distributions of Level of Perceived Susceptibility to Poliomyelitis in January by Respondents' Previous Vaccine Experience.

Injected Vaccine Status	Total Persons	Your Chance of Getting Polio									
		No Chance								Big Chance	
Non-Takers	175	3	3	3	11	15	27	77	82	85	100
Takers	198	3	5	7	25	29	38	80	85	91	100
Total	373[a]	3	4	5	19	23	33	79	84	88	100

[a] 7 Respondents refused to answer this question.

Table 2. Cumulative Percentage Distributions of Level of Perceived Susceptibility to Poliomyelitis for All Respondents in January and in May.

Time Period	Total Persons	Your Chance of Getting Polio									
		No Chance									Big Chance
Prior to Oral Program	373[a]	3	4	5	19	23	33	79	84	88	100
After Oral Program	373[a]	8	14	21	26	49	73	79	84	90	100

[a] 7 Respondents refused to answer this question.

Table 3. Mean Levels of Perceived Susceptibility to Poliomyelitis in January and May By Respondent's Vaccine Acceptance Category.

(The lower the score the greater the perceived risk)

Vaccine Acceptance Pattern	Total Persons	Mean Levels of Perceived Susceptibility		
		As of January	As of May	Amount of Change
Non-Takers, Any Vaccine	57	15.26	16.32	1.06
Oral Vaccine Only	118	13.13	18.21	5.08
Injected Vaccine Only	41	15.70	18.44	2.74
Both Types of Vaccine	157	16.13	21.64	5.51
All Oral Vaccine Takers				
February Only Dose	33	14.39	16.23	1.84
April Only Dose	35	14.45	21.52	7.07
Both Doses	207	14.93	20.56	5.58

vailing in the traditionally designed survey which permits the type of inference being criticized in this paper. Such an inference might state that there is an inverse relationship between level of perceived susceptibility and likelihood of taking preventive action since the data reveal that the behavioral group perceiving themselves as most susceptible to this disease had taken neither the earlier injected vaccine or the new oral vaccine in the recent program, whereas those who had taken both types of vaccine had the lowest level of perceived susceptibility. These differences are statistically significant when tested at the .05 level by means of the two-sample t-test. All too often this is the type of data which is available and the type of inference which is drawn.

If we shift our focus, however, to the data obtained from these same people as of January, before they were aware that a new vaccine was to be offered them, we find that mean levels of perceived susceptibility do not vary significantly, at the .05 level. These data suggest that prior levels of perceived susceptibility were not very useful predictors of the type of vaccine behavior that was to follow. On this basis we would infer that earlier writers were probably in error in emphasizing perceived susceptibility as a major determinant of this type or preventive health behavior.

The bottom half of Table 3 adds some insight into the dynamics underlying people's perceptions and relevant behavioral acts. Notice, for instance, that the mean scores for the three types of behaviors relevant to the taking of the oral vaccine were essentially the same before the program. But notice the difference in shifts in these perceptions dependent upon the type of behavior which had been emitted as indicated by the data in the column headed "Amount of Change." Those who took the first of the two doses offered showed only a slight and not statistically significant decrease in their level of perceived susceptibility to poliomyelitis following that

Table 4. Mean Levels of Perceived Severity of Poliomyelitis in January and May by Respondent's Vaccine Acceptance Category.

(The higher the score the greater the perceived severity)

Vaccine Acceptance Pattern	Total Persons	Mean Levels of Perceived Severity		
		As of January	As of May	Amount of Change
Non-Takers Any Vaccine	58	25.05	19.35	5.70
Oral Vaccine Only	117	21.43	20.09	1.34
Injected Vaccine Only	39	23.58	15.79	7.79
Both Types of Vaccine	151	22.84	18.01	4.83
All Oral Vaccine Takers				
February Only Dose	32	22.09	18.95	3.14
April Only Dose	34	24.26	19.35	4.91
Both Doses	202	21.91	19.84	3.06

act. Compare this change with that observed among people who refused the first dose but took the second dose when offered. That group gives evidence of a significant decrease in their level of perceived susceptibility even though in fact they were no better protected against the disease than those who had taken only the first dose in February. We submit that this type or perceptual shift following a behavioral commitment is best interpreted by dissonance theory as developed by Brehm and Cohen (6, p. 96) when they describe the tendency toward overvaluation of a chosen alternative when the alternatives are of approximate attractiveness in the initial decision making process. From the data in the bottom half of Table 3 it would seem that the group of people who had deterred taking the first dose of the vaccine in February but switched their decision and took the April dose give evidence of this overvaluation of the efficacy of the vaccine. Notice that as a group they expressed feelings of less risk to poliomyelitis following their one dose of the vaccine than did those who had apparently not gone through this conflict in decision and took both the first and second dose.

Neither time nor usefulness to the discussion today permits development of essentially the same points when the focus is shifted to levels of perceived severity of poliomyelitis should it be contracted, but the data are given in Table 4 for completeness of presentation as we turn to the second type of fallacious inference drawn from survey data. It will be recalled that it was stated earlier that one of the types of incorrect inferences resulted from failures to recognize the time dimension underlying the variable chosen for study and the resultant confusion regarding cause and effect relationships. The second deficiency results from failures to recognize interaction effects between and among variables selected for study through the survey process.

We saw in Table 3 that prior levels of perceived susceptibility were not particularly good predictors of the behavior being studied and the same comment holds for the data in Table 4 where levels of perceived severity of the disease are used as a predictive variable. Table 5 presents data regarding the effects of various levels of perceived susceptibility *and* severity before the vaccine programs on acceptance rates of the oral vaccine when controls are maintained for prior behaviors with regard to another vaccine for the same disease. Using a Chi-square test on the data it is found that these prior perceptions were not significantly, at the .05 level, associated with subsequent vaccine behavior among those who had already taken one type of vaccine against this disease, but were significant, beyond the .05 level, among those who had no prior experience with polio vaccine. The data further suggest that the relationship between levels of perceived susceptibility-severity and subsequent behavior is not linear as suggested by previous writers on the subject. Notice, for instance, that the highest levels of vaccine acceptance do not occur when both perceived susceptibility to the disease *and* severity of the disease if contracted are high, and this is true regardless of prior experience with a preventive measure for the same disease.

Table 5. Oral Vaccine Acceptance Rates At Various Levels of Perceived Susceptibility-Severity as of January by Previous Vaccine Status.

(Numbers in parentheses indicate the denominator on which the rate is based)

Injected Vaccine Status	Total		January Perceptions			
			Low Susceptibility and		High Susceptibility	
	Persons	Oral Rate	Low Severity	High Severity	Low Severity	High Severity
Takers of Injections	188	.79	(30) .77	(38) .84	(53) .85	(67) .73
Non-Takers of Injections	173	.67	(22) .50	(26) .56	(56) .86	(69) .61
	361	.73	.65	.73	.85	.67

Discussion

These data, while restricted to a specific behavioral act primarily of interest to those doing health related research, are illustrative of more general problems commonly trapping survey researchers. The danger of confusing cause and effect through failure to consider the time relationship between variables chosen for study was illustrated by use of data which showed that the so-called dependent variable (the behavioral act) was influencing the so-called independent variable (the perceptual set). Further demonstrated was the fact that the nature of this reversal would not have been apparent without before and after data. Such data not only permit appropriate tests of the predictive power of variables but also inferences about underlying dynamic processes which may prove to be at least as useful in furthering our understanding of human behavior as are predictive variables which cannot be understood or interpreted.

The data also provide evidence supporting the contention that interaction effects between and among the several variables selected for study should be considered. Far too often the author leaves the reader with nothing more than a series of p-values, all "highly significant," and no clues as to what happens when two or more of these are pitted against each other to see how they jointly influence the dependent variable.

While these data do not speak directly to this point, all of the background work which led up to these two surveys and much that has happened since suggests a need for curtailment of the prevailing "coveat emptor" philosophy with regard to social-behavioral science literature in particular and survey research literature in general. Long over due are some self imposed disciplinary standards based on an awareness that poor theory begets weak or inappropriate variables built into survey designs, and these in turn when improperly designed and analyzed complement and speed up the cycle by begetting still more tidbits of inadequate theories of behavior which become grist for still more surveys.

References

1. Ibrahim, Michel A.: The Role of Certain Psychological Factors in Coronary Disease, Blood Pressure, and Serum Cholesterol. A doctoral dissertation submitted through the Department of Epidemiology, School of Public Health, University of North Carolina at Chapel Hill. 1964.

2. Johnson, A. L. et al.: Epidemiology of polio vaccine acceptance. Florida State Board of Health Monograph No. 3. Jacksonville, Florida 1962.

3. Northcutt, Travis J., et al.: Factors Influencing Vaccine Acceptance, Chapter IV in Hillsborough County Oral Polio Vaccine Program. John S. Neill and James O. Bond eds. Florida State Board of Health monograph No. 6. Jacksonville, Florida. 1964.

4. Jenkins, C. David: Identification of public beliefs about health problems as a basis for predicting use of health service. Department of Epidemiology, School of Public Health, University of North Carolina at Chapel Hill. 1964.

5. Johnson, A. L.: An emerging theory to explain health behavior using a reward-cost analysis. Ph.D. dissertation. University of North Carolina at Chapel Hill. 1963.

6. Brehm, Jack W., and Cohen, Arthur R.: Explorations in cognitive dissonance. John Wiley and Sons, Inc. New York. 1962.

THE PROBLEM OF GEOGRAPHIC CONTIGUITY - A MONTE CARLO APPROACH

Stephen F. Gibbens, James A. Tonascia
California State Department of Public Health

This paper describes a method for determining whether patterns of counties, census tracts or other types of geographic units depart from a configuration that might have occurred by chance. The method presented is conceptually simple and can be applied with comparative ease through the use of a computer program available through SHARE[1]. This paper describes the problem, the method of solution and gives some useful probability distributions.

Suppose, for example, a rate of some type was calculated for each of the N counties in a particular state and a map was shaded, showing the N/4 counties having the highest rates. This top quartile of counties might distribute themselves in a number of ways - varying from a single cluster forming a large clump to the opposite situation, where most of the selected counties were isolated, surrounded by counties with low rates. Whatever pattern was formed, most observers would make a judgment on the meaningfulness of the pattern manifested. If the high rate counties tended to be those containing large metropolitan centers, this fact would probably be noted rather quickly. If, however, a judgment was made whether the high rate counties showed a pattern of contiguity - or geographic contagion - the problem might become much more difficult. The point at which a pattern departs from simple randomness is obscure indeed, when it must be based upon visual inspection.

A number of criteria could be formulated defining nonrandomness. A tendency for the chosen counties to form a few clusters, a preponderance of chosen counties falling in a single quadrant of the state, a "least squares" solution based on distance from some focal point - all might be developed into probability sampling distributions for departures from randomness. In this paper, two criteria are developed: (1) the largest cluster formed by the selected counties; and (2) the number of pairs of contiguous counties found among those counties selected. These will be referred to as the "clusters" and "pairs" methods.

Clusters Method

From a shaded map we have a visual picture of geographic contiguity. A nonrandom pattern is characterized by clusterings of shaded counties. However, our data could be more easily quantified if we described the clustering in terms of the number of counties found in the largest cluster formed by the N/4 sample. With this criterion, it is possible to conveniently describe distributions of many samples, and to obtain probabilities of the occurrence of clusters of various sizes. In order to effect this quantification, a listing is required for each state showing each individual county with its contiguous counties.[2]

With the availability of a "county contiguity list" (See Table A), it is possible to select a sample of N/4 counties, from a particular state and note the largest cluster formed by these selected counties.

Briefly described, say we have a state with 100 counties and by use of random numbers select a sample of N/4 counties, notated as A, B, C, . . . Y. We search the county contiguity list for the counties contiguous to A. If counties B and J are contiguous to A, we know we have a cluster of at least three counties. Next, we scan the list of counties contiguous to B and find counties A and F. County A was already counted but F is a new county and so our cluster is known to include at least four counties; A, B, F and J. We proceed similarly to search the lists of counties contiguous to F and J, incrementing the cluster size by one for each sample county found. We continue to county C (the first of the remaining counties sampled that is not already known to be in our first cluster of contiguous counties), repeating the above procedure, and hence on to county Y, the last county included in our sample. When we have finished, we record the number of counties in the largest cluster formed. If we were to repeat this procedure for many samples, accumulating the results, we would have an approximate sampling distribution for the largest cluster size found for that state when N/4 counties are selected by chance. For example, the State of Iowa has 99 counties from which 3,000 independent samples of 25 counties each were selected. In about 5 percent of these 3,000 samples a cluster as large as

---

[1] The computer program is available through the SHARE users group (identified as GO BC GEOG), with specifications.

[2] It should be noted here that the determination of contiguity or noncontiguity is often quite arbitrary. In this paper, two counties were required to share a common border in order to be considered contiguous. If they were separated by water narrow enough to be bridged, they were considered contiguous. Hence, the Mississippi was not a barrier to contiguity but Lake Michigan was. However, different investigators may want to apply different rules for county connections. For example, when four counties meet at a point, so that their borders form a "cross" (+), a decision must be made as to whether or not the diagonal counties are to be considered contiguous. If "point contiguity" is accepted, i.e., if the diagonal counties are considered contiguous, does this carry over to the situation where the borders describe pairs of obtuse angles, rather than right angles as was the case for the cross? These decisions should be made only after careful consideration of these problems, since alternatives produce significantly different results.

10 counties was found; in 1 percent, a cluster as large as 13 counties was found, and in .1 percent, a cluster as large as 16 counties was found.

Pairs Method[3]

The use of the largest cluster size found is not always an ideal measure of geographic contiguity. If there were several smaller clusters, there might be a departure from randomness even though the largest cluster had only a moderate probability level. Another approach would be to obtain the sampling distribution of the number of pairs of contiguous counties found when selecting many random samples of N/4 counties. A large number of pairs of counties could be found either in the situation where there was a single very large cluster, or, several smaller clusters.

The number of pairs of contiguous counties found in a sample would be ascertained, again by recourse to the county contiguity list, accumulating a tally for every border common to two counties included in the sample. In the above example for the State of Iowa, in 3,000 samples, 19 pairs of counties were found in 3½ percent of the samples, 21 pairs in 1 percent of the samples, and 24 pairs in .1 percent of the samples.

Discussion

The "clusters" method and the "pairs" tend to give similar results when applied to live data. The pairs method is more sensitive to the situation where several clusters of moderate size are formed by the sampled counties, sometimes showing a significant probability level where the clusters method is not significant in terms of the number of contiguous counties in the largest cluster. However, this gain in sensitivity in the pairs method, is in equal measure lost, when contiguity is concentrated in a single cluster. In this situation, the clusters method may show a significant probability level while the pairs method shows a very moderate probability.

Figures 1 and 2 illustrate these different situations. Figure 1 shows the lowest quartile of counties in terms of live birth rate for Iowa in 1962. The largest cluster contains 13 counties. This occurred less than once in every 100 trials when using chance methods. However, among these same counties there were 17 pairs (instances of 2 low quartile counties sharing a common border). This occurred about 12 times in every 100 trials when using chance methods. Figure 2 shows the contrasting situation where the cluster of 8 counties was shown to have occurred over 11 times in every 100 trials whereas the pairs method occurred only twice in every 100 trials.

Since the various states tend to contain different numbers of counties and since the configuration of counties tends to vary as well, clearly, each state has its own unique sampling distribution. Even where two states have the same number of counties and the appearance of the counties is similar, their sampling distributions are often distinctly dissimilar when subjected to a Kolmogorov-Smirnov test. The general appearance of the sizes and shapes of the counties in a state gives little indication of the probability pattern that will emerge.

Table 1 shows the Estimated Cumulative Probability Distribution for Cluster Sizes of Contiguous Counties for Selected States based on 3,000 samples drawn from each state. Each Sample included about 25 percent of the counties in the state. Referring again to the State of Iowa, the table shows that in .001 of the samples, the largest cluster formed by the 25 sample counties was 2 counties; that in about half the samples (.492) the largest cluster found was 5 counties, or less. Similarly, in 95.2 percent, the largest cluster formed was 10 counties or less. The largest cluster of contiguous counties found in the 3,000 samples drawn was 16 counties. From a sample of 25 counties in the State of Iowa, it would be rare indeed to find the largest cluster to be as few as 2 counties, or as many as 16 counties.

Since these probability distributions were developed by the Monte Carlo method, they are qualified as estimated distributions. Barring a bias in the pseudo-random number generator used by the computor (BC RNDY), the true probability limits could be estimated by the usual method.

For a 10-county cluster in the State of Iowa which has an estimated cumulative probability of .952, a .05 level confidence interval would be calculated as:

$$.952 \pm 1.96 \sqrt{\frac{(.048)(.952)}{3{,}000}} = .952 \pm .008$$

There appear to be two main problems in the practical application of this methodology. Let us say an investigator is attempting to ascertain whether a particular type of congenital anomaly found in newborn infants shows a nonrandom geographic distribution. Usually the investigator would not know in what percent of the counties the anomaly should show high rates - perhaps in N/4 of the counties or perhaps N/10. This would depend on the distribution of the factors which he hypothesizes gives rise to high incidence of the anomaly. Since each sample size (N/4 or N/10) would have a different sampling distribution, the task of selecting an appropriate sample size becomes critical. The only solution to this would seem to be to rank the counties according to incidence rates and if some of the high incidence counties were tested to show significant departures from a norm - to cut off below these counties and use that number as the sample size. The use of "fortuitous" distributions would not appear to invalidate the ultimate probabilities since these are based on the geographic pattern manifested, rather than on the statistical significance of the rates in the selected counties.

---

[3] The authors are indebted to Donald Loveland for suggesting this approach.

The second problem is more subtle. The probability distribution obtained is valid and useful in an a priori sense. However, usually an investigator has the data on a shaded map and is interested in knowing whether the distribution is nonrandom. When viewed in an a posteriori sense, it may appear invalid to use the tabled probabilities because of gross irregularities in the particular counties included in the shaded portion of the map. For example, if a shaded map of California showed a number of the coastal counties in the high rate quartile, the tabled probabilities would be inappropriate for comparison because the probabilities were obtained from all counties. Since the coastal counties have no contiguous counties on their ocean side, they would have less opportunity to be included in large clusters and no opportunity to contribute pairs for a portion of their borders. Therefore, clusters of counties which included coastal counties would occur less often than suggested by the tabled probabilities. At the other extreme, an unusually large size county would tend to border on more counties than the average size county and would therefore be found in clusters more often than suggested by the tabled probabilities. In atypical situations such as these it might be reasonable to compare the obtained pattern with a large number of samples which included one or more specified counties in each sample selected. The resulting conditional probabilities would reflect this bias introduced by the inclusion of a constant county.

The computer program mentioned in the introduction allows for the inclusion of constant counties for obtaining these biased probability distributions.

Analysis of county contiguity within a state often shows clusters of counties along state boundaries. The question arises as to the results that might have been obtained had the bordering state been included in the analysis. The computer program allows for grouping a number of states into a single analysis and gives probability distributions of clusters and pairs of counties drawn from this larger parameter.

While geographic contiguity is not a common statistical problem, it is frequently encountered in epidemiology and geography. The "shaded map" seems to present a strong stimulus for "closure" or resolution, compelling the observer toward interpretation. In the experience of the authors, several observers may interpret a shaded map as showing various degrees of contagion. The most articulate, or, the senior observer present, wins the argument. The methodologies presented in this paper, when properly applied, provide the investigator with a simple, objective determination of contiguity.

REFERENCES

Geary, R. C. "The Contiguity Ratio and Statistical Mapping", Incorporated Statistician, Volume 5, 1954.

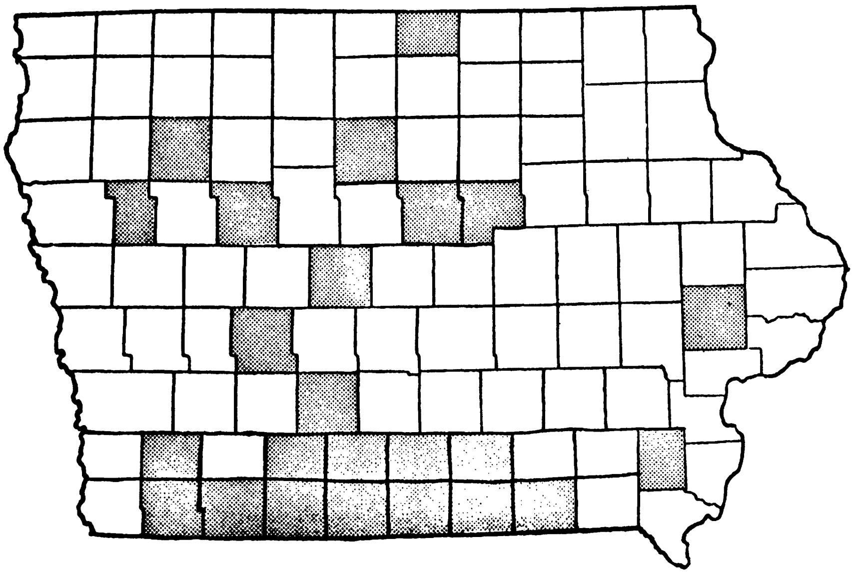
Ederer, Fred; Myers, Max H. and Mantel, Nathan. A Statistical Problem in Space and Time: Do Leukemia Cases Come in Clusters? 1963 (Unpublished).

Table A

SAMPLE COUNTY CONTIGUITY LISTING

(A portion of the State of Iowa showing counties according to numeric code.)

INDEX COUNTY	COUNTIES CONTIGUOUS TO INDEX COUNTY					
01	15	39	61	88	02	
02	69	15	01	88	87	
03	96	22				
04	93	68	26			
05	83	14	39	15		
06	86	07	10	57	48	
07	38	12	09	10	06	86
08	37	94	40	85	77	25
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
99	46	41	40	35	94	



LIVE BIRTH RATE IN IOWA 1962
(lowest quartile)

Largest cluster - 13 counties

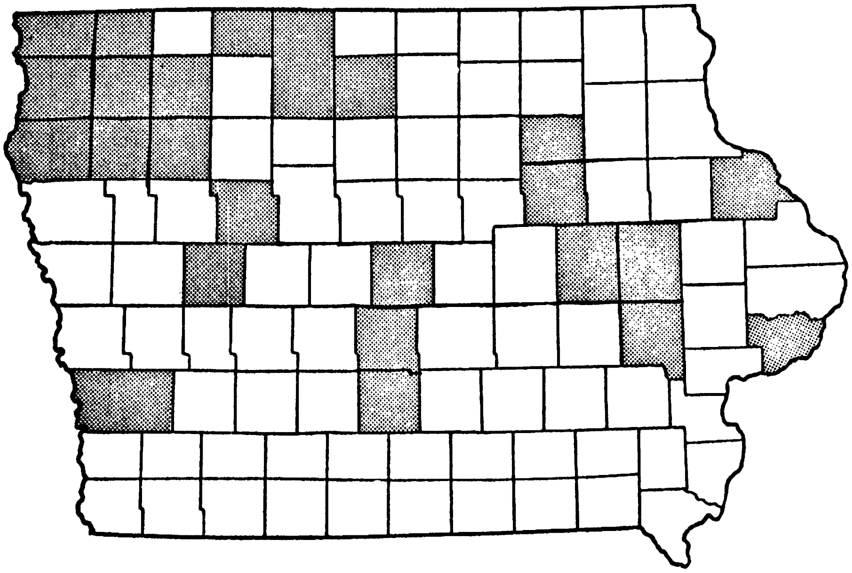
Probability that a cluster as large as 13 counties
could have occurred by chance =.008

Number of pairs of counties - 17

Probability that as many as 17 pairs of counties
could have occurred by chance =.117

Figure 1. A Configuration Favoring The Cluster Method

CRUDE DEATH RATE IN IOWA 1962
(lowest quartile)



Largest cluster - 8 counties

Probability that a cluster as large as 8 counties
could have occurred by chance =.116

Number of pairs of counties - 20

Probability that as many as 20 pairs of counties
could have occurred by chance =.018

Figure 2. A Configuration Favoring The Pairs Method

Table 1

ESTIMATED CUMULATIVE PROBABILITY DISTRIBUTION FOR CLUSTER SIZES OF CONTIGUOUS COUNTIES
SELECTED STATES - 25 PERCENT SAMPLE

(Based on 3,000 trials)

NUMBER OF COUNTIES IN LARGEST CLUSTER	ALABAMA[1] (17)	ARKANSAS (19)	CALI-FORNIA (14)	COLO-RADO (16)	IOWA (25)	KEN-TUCKY (30)	LOUI-SIANA (16)	MINNE-SOTA (22)	MIS-SOURI (29)	MONTANA (14)	NEBRASKA (23)	PENN-SYLVANIA (17)	SOUTH DAKOTA (17)	TEN-NESSEE (24)	VIR-GINIA (24)	WIS-CONSIN (18)
1	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
2	.005	.003	.037	.014	.001	.000	.146	.003	.000	.028	.002	.012	.014	.001	.003	.011
3	.103	.076	.255	.137	.062	.016	.164	.059	.018	.227	.080	.139	.142	.040	.073	.122
4	.319	.265	.543	.369	.254	.122	.426	.265	.134	.493	.306	.391	.397	.183	.270	.354
5	.538	.499	.767	.583	.492	.312	.646	.476	.327	.700	.536	.507	.607	.419	.497	.568
6	.701	.672	.879	.736	.688	.515	.793	.651	.511	.834	.725	.762	.769	.610	.687	.726
7	.820	.804	.947	.838	.806	.674	.913	.773	.672	.915	.826	.863	.866	.748	.814	.831
8	.896	.882	.976	.908	.884	.778	.937	.860	.785	.963	.898	.921	.923	.851	.882	.905
9	.946	.933	.991	.952	.931	.865	.966	.914	.860	.984	.938	.953	.960	.916	.928	.953
10	.974	.959	.997	.975	.952	.910	.988	.947	.912	.993	.966	.979	.979	.954	.957	.971
11	.988	.980	.999	.989	.977	.948	.994	.972	.947	.999	.982	.993	.990	.973	.978	.987
12	.995	.989		.995	.986	.968	.998	.985	.968		.986	.998	.995	.985	.986	.994
13	.998	.994		.998	.992	.982	.999	.993	.980		.995	.999	.999	.992	.990	.998
14	.999	.998		.999	.996	.991		.997	.989		.998			.995	.995	.999
15		.999			.998	.995		.998	.993		.999			.998	.998	
16					.999	.998		.999	.997					.999	.999	
17						.999			.999							

[1] Number in parentheses is the number of counties used in the 3,000 trials for that particular state and is approximately 25 percent of the total number of counties in the state.

Table 2

ESTIMATED CUMULATIVE PROBABILITY DISTRIBUTION FOR THE NUMBER OF PAIRS OF CONTIGUOUS COUNTIES
SELECTED STATES - 25 PERCENT SAMPLE

(Based on 3,000 trials)

NUMBER OF PAIRS	ALABAMA[1] (17)	ARKANSAS (19)	CALI-FORNIA (14)	COLORADO (16)	IOWA (25)	KENTUCKY (30)	LOUI-SIANA (16)	MINNE-SOTA (22)	MISSOURI (29)	MONTANA (14)	NEBRASKA (23)	PENN-SYLVANIA (17)	SOUTH DAKOTA (17)	TENNESSEE (24)	VIRGINIA (24)	WISCONSIN (18)
1	.000	.000	.001	.000	.000	.000	.000	.000	.000	.001	.000	.000	.000	.000	.000	.000
2	.000	.000	.004	.001	.000	.000	.001	.000	.000	.003	.000	.001	.000	.000	.000	.000
3	.001	.000	.028	.003	.000	.000	.004	.001	.000	.018	.000	.003	.001	.000	.000	.002
4	.004	.001	.083	.015	.000	.000	.013	.001	.000	.064	.000	.012	.009	.000	.000	.009
5	.016	.004	.196	.048	.001	.000	.053	.005	.000	.155	.002	.032	.033	.000	.001	.023
6	,052	.016	.353	.104	.004	.000	.137	.011	.000	.298	.010	.080	.089	.001	.003	.061
7	.122	.050	.523	.216	.012	.000	.260	.031	.000	.470	.027	.170	.191	.004	.007	.128
8	.234	.109	.685	.361	.031	.000	.415	.064	.001	.629	.069	.310	.320	.012	.017	.245
9	.368	.214	.815	.517	.074	.002	.566	.139	.005	.758	.147	.464	.471	.031	.045	.371
10	.523	.344	.894	.657	.144	.004	.721	.231	.009	.860	.245	.615	.625	.073	.094	.528
11	.664	.495	.948	.778	.238	.012	.822	.338	.024	.925	.359	.745	.754	.137	.164	.674
12	.779	.637	.976	.863	.354	.031	.896	.478	.052	.964	.501	.842	.852	.227	.272	.786
13	.868	.746	.991	.926	.493	.060	.941	.597	.096	.982	.631	.912	.923	.345	.404	.871
14	.930	.835	.996	.962	.632	.111	.967	.715	.161	.993	.756	.952	.958	.467	.521	.923
15	.967	.901	.998	.982	.752	.183	.983	.805	.256	.997	.851	.977	.978	.592	.641	.956
16	.982	.943	.999	.990	.842	.277	.991	.884	.356	.999	.906	.990	.988	.703	.748	.978
17	.992	.970		.997	.883	.382	.996	.922	.480		.943	.994	.996	.804	.832	.987
18	.996	.986		.999	.938	.508	.998	.954	.587		.969	.996	.998	.868	.897	.995
19	.998	.994			.965	.617	.999	.974	.690		.984	.999	.999	.921	.935	.999
20	.999	.998			.982	.719		.985	.770		.993			.949	.964	
21		.999			.993	.798		.993	.846		.997			.972	.980	
22					.996	.861		.995	.899		.998			.985	.991	
23					.998	.909		.999	.935		.999			.994	.995	
24					.999	.942			.961					.997	.997	
25						.943			.976					.997	.999	
26						.983			.987					.999		
27						.989			.991							
28						.995			.996							
29						.998			.998							
30						.999			.999							

[1] Number in parentheses is the number of counties used in the 3,000 trials for that particular state and is approximately 25 percent of the total number of counties in the state.

THE EFFECT OF THE GHETTO ON THE DISTRIBUTION AND LEVEL OF NONWHITE EMPLOYMENT IN URBAN AREAS

John F. Kain, U.S. Air Force Academy and The RAND Corporation*

Numerous researchers have evaluated the effects of racial discrimination and racial segregation on the operation of the housing market, on the cost and the quality of nonwhite housing, and on several other aspects of nonwhite welfare.[1] The author has commented extensively on the effects of residential segregation on the travel behavior of both whites and nonwhites and on the resulting urban transportation demands.[2] He has also pointed out that many current urban problems, the demands for urban transportation subsidies, the near failure of urban renewal programs, and the demands for many other urban programs have their roots deeply implanted in the desire to avoid facing up to the problem of housing segregation.[3] Still other researchers have investigated discrimination in employment and have attempted to determine the extent to which fewer job opportunities and the higher unemployment rates among nonwhites are attributable to factors other than racial discrimination, such as the lower levels of educational attainment.[4]

To the author's knowledge, this paper is the first to link discrimination in the housing market to the distribution and level of nonwhite employment in urban areas. The hypotheses evaluated here are that racial segregation in the housing market: (1) affects the distribution of nonwhite employment, and (2) reduces nonwhite job opportunities. These hypotheses are tested empirically, using origin and destination data obtained from the 1952 Detroit Area Traffic Survey and the 1956 Chicago Area Transportation Study (CATS).[5] To understand how housing-market segregation affects the distribution and level of employment in United States metropolitan areas, it is first necessary to understand the distribution of nonwhite housing — the spatial characteristics of the ghetto.

THE GHETTO

The means for perpetuating racial segregation in the housing market are well documented in both popular and scholarly literature.[6] Both legal and extra-legal means have been used, including racial covenants; racial zoning; violence; threats of violence; preemptive purchase; various petty harassments; implicit or explicit collusion by realtors, banks, mortgage lenders, insurance companies, and other lending agencies; and, in the not so distant past, various Federal agencies. Since the fact of racial segregation is well established and the ways it is maintained are well documented, no extensive documentation of these practices is attempted here. This paper is limited to a discussion of historical patterns of racial occupancy in Detroit and Chicago, and a brief description of the general patterns in other United States cities.

The nonwhite ghettos in Detroit and Chicago, as in most other U.S. metropolitan areas, lie mostly within the central city near the Central Business District (CBD). In some instances secondary nonwhite housing areas are found in the central parts of older suburbs and sometimes in previously rural areas. In the postwar period, the latter variety have often been surrounded by new housing developments during suburbanization. Both kinds of outlying nonwhite residence areas may importantly affect the distribution of nonwhite employment and the actual or potential job-loss that may result from racial segregation.

Figures 1 and 2 indicate the broad outlines of the Detroit and Chicago ghettos. The geographic areas shown are the zones used for the empirical analysis that follows. The two figures clearly show the central location of the principal ghetto and the boundaries of the smaller outlying nonwhite residence areas, or "ghettoettes." The Detroit area has "ghettoettes" in three places: one about 16 miles from the CBD, far to the north; one about 8 miles from the CBD along Woodward Avenue; and one to the southwest about 11 miles from the CBD.

Detroit's principal ghetto lies within the central city. The ghetto has a slight sectoral pattern, with the heaviest concentrations of Negro workers residing along Woodward Avenue. The extent of segregation is indicated by the proportions of the white and nonwhite labor forces residing within the principal ghetto. If the principal ghetto is defined to include only those contiguous central zones with a population more than 10 per cent Negro, 89 per cent of Detroit's Negroes live there, but only 28 per cent of its whites. If the zones adjacent to this concentration containing between 2 and 10 per cent Negroes are included, the area houses approximately 93 per cent of Detroit's nonwhite work force. Nearly all of the remaining 7 per cent live in one of the small outlying nonwhite residence areas. The geographic size and quantitative importance of these outlying residence areas are badly exaggerated by the size of the zones used in tabulating the data collected in the 1952 study. Zone 76 houses only 280 Negro workers as compared to over 5,000 white workers. Zones 59 and 60 house about 3,600 Negro workers and 17,000 workers. Zones 94 and 95 have approximately 2,800 Negroes and 11,300 whites.

The preponderance of Chicago's Negroes live in the notorious South Side. In addition, fingers of the ghetto extend due west and due north from the

---

*This paper was prepared for presentation at the Annual Meeting of the American Statistical Association in Chicago, December 27-30, 1964. The author wishes to acknowledge the many useful suggestions and comments of Professor John C. Ries. Any errors or oversights that remain are, of course, the author's sole responsibility.

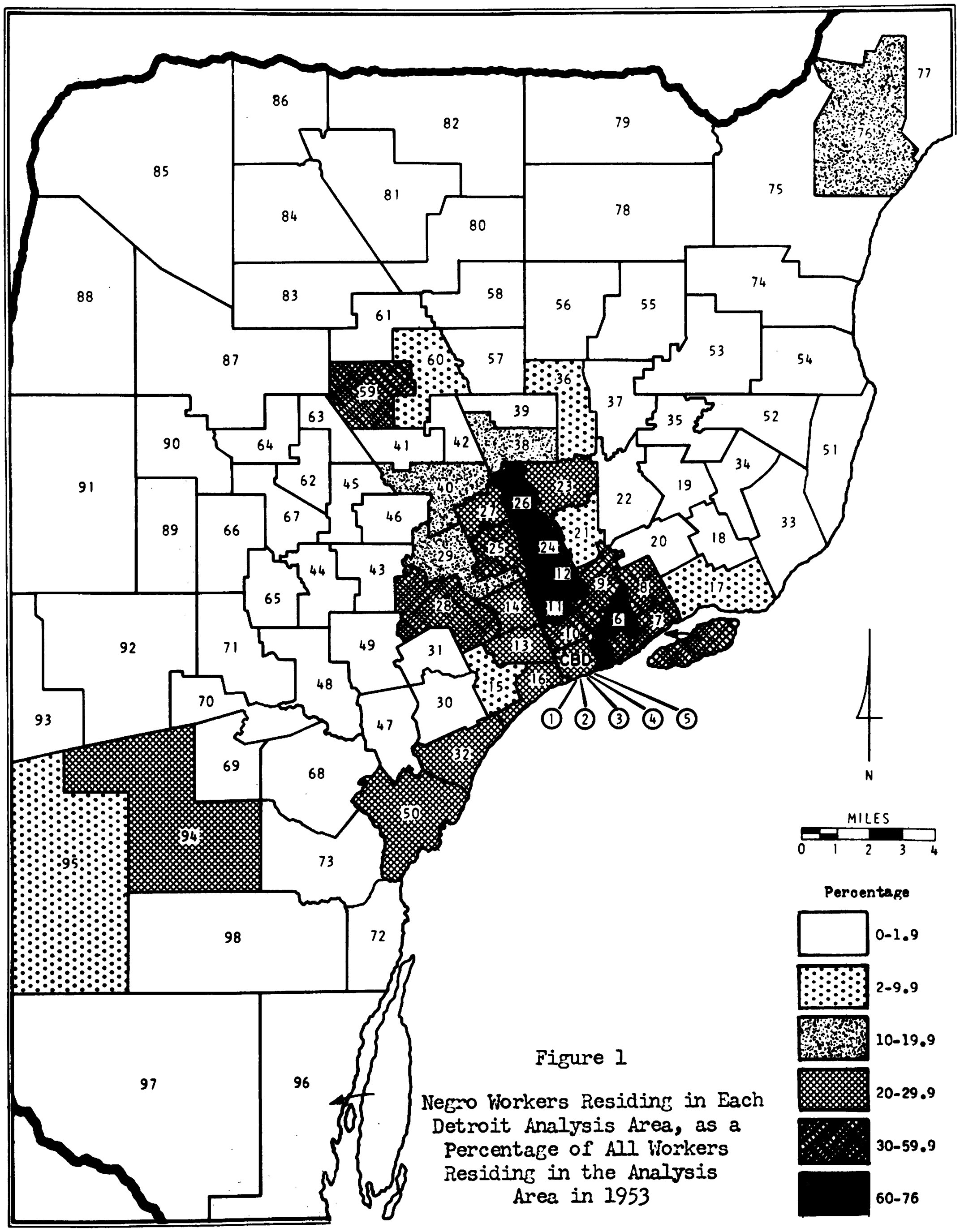


Figure 1

Negro Workers Residing in Each Detroit Analysis Area, as a Percentage of All Workers Residing in the Analysis Area in 1953

CBD. The only other significant nonwhite settlement in the Chicago area is far to the north in the suburb of Evanston; it houses about 1,900 nonwhite workers, or just under 1 per cent of the nonwhite labor force. The size of the aggregation areas used in Fig. 2 also badly exaggerates the size of the Negro ghetto. If smaller areas were used, the southern tail of the ghetto would be separated from the major Negro concentration by white residence areas and would appear as several islands.

Chicago's principal ghetto houses an even larger percentage of its nonwhite workers than does Detroit's: 94 per cent of Chicago's nonwhite workers reside there, but only 20 per cent of its white workers. If contiguous central areas having between 3 and 10 per cent of nonwhite residents were added, nearly 96 per cent of Chicago's Negro workers live there.

The patterns of Negro residence in Chicago and Detroit are by no means of recent origin. In their study of the growth and geographic expansion of the Negro population in Chicago, Otis and Beverly Duncan conclude that the spatial outline of the Negro community in Chicago had been established by 1920, if not by 1910, and that further expansion of the Negro community occurred within areas which already had been accommodating a

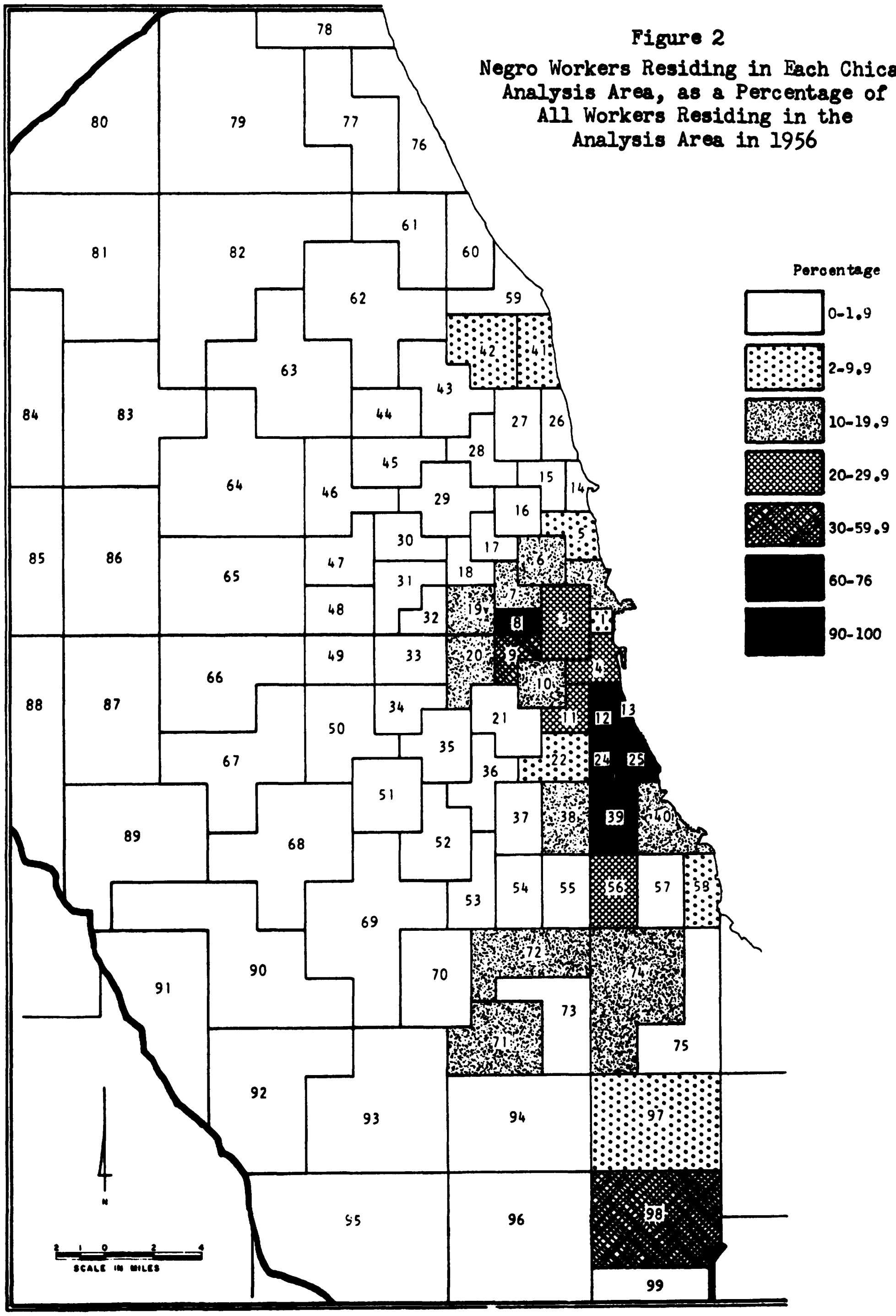


Figure 2
Negro Workers Residing in Each Chicago Analysis Area, as a Percentage of All Workers Residing in the Analysis Area in 1956

nucleus of Negro residents in 1920.[7] Insofar as expansion of the Negro residential areas has occurred, it has been in terms of adding areas contiguous to existing Negro concentrations, with most of the apparent exceptions reflecting the locational pattern of publicly financed housing developments.[8] The authors conclude, "Apparently, this pattern of expansion has resulted in an increase of residential segregation for Negroes in Chicago and a consolidation of the Negro community."[9]

Observers have also commented extensively on Detroit's Negro ghetto; Robert Weaver traces it back to 1942. The growth of Detroit's Negro population occurred later than Chicago's; thus the development and expansion of its ghetto occurred later. Even so, Weaver reports that in 1940 approximately 85 per cent of the nonwhite residents of Detroit were concentrated in census tracts with 20 per cent or more Negro residents and that over two-thirds of them were in tracts 50 per cent or more Negro. He further notes that of the city's total 19,500 blocks, less than 1,900 had Negro occupants despite a rather widespread dispersion of the Negro settlers prior to 1910. Also in 1940, 416 blocks were from 10 to 49 per cent Negro, 335 blocks were from 50 to 99 per cent Negro, and 139 blocks were 100 per cent Negro. Most of these blocks were concentrated in the principal Negro ghetto, where at least 80 blocks were solidly Negro and 125 blocks were from 90 to 99 per cent Negro. In all, the main area of Negro concentration included almost half the total number of blocks in the city occupied by any Negroes.[10]

Weaver's findings for Detroit were corroborated by the Committee on Race and Housing studies. In these studies, McEntire reports that in 1950, 55 per cent of the Negroes living in Detroit and adjacent areas resided in tracts more than 75 per cent nonwhite, 19 per cent in tracts between 50 and 74 per cent nonwhite, 13 per cent in tracts between 25 and 49 per cent nonwhite, and only 1 per cent in tracts less than 1 per cent nonwhite.[11] In 1940 and in 1950 Detroit appears less segregated by census tracts than does Chicago, although the concentration of both seems greater in 1950 than in 1940.

These patterns of racial segregation are by no means limited to Detroit and Chicago. They are typical of nearly all U.S. metropolitan areas, as Tables 1 and 2 illustrate by the statistics they contain regarding Negro and white population composition since 1900. Table 2 is especially graphic in showing how more and more Negroes are taking up residence in the central city, while the white population is shifting to the ring.

Table 1

NEGROES AS A PERCENTAGE OF TOTAL STANDARD METROPOLITAN AREA, CENTRAL CITY, AND RING POPULATION

Year	Total	Central City	Outside Central City
1960	10.8	16.8	4.6
1950	9.4	12.4	5.2
1940	8.0	9.6	5.5
1930	7.5	8.4	5.7
1920	6.8	6.9	6.5
1910	6.7	6.3	7.5
1900	7.4	6.5	8.9

SOURCE: U.S. Bureau of the Census, U.S. Census of Population: 1960, Selected Area Reports, Standard Metropolitan Statistical Areas, Final Report PC(3)-1D, Government Printing Office, Washington, D. C., 1963, pp. 1-5.

Davis McEntire summed up the findings of the study he directed for the Committee on Race and Housing by saying, "Characteristic of all cities studied is a principal area of nonwhite concentration near the business center of the city. This area consists of a 'segregated' core surrounded by successive zones of 'concentrated,' 'mixed,' and 'dispersion' tracts."[12]

Donald Cowgill has made a comprehensive attempt to measure racial segregation in American cities, calculating segregation scores from block statistics for 197 cities in 1940 and 209 cities in 1950.[13] His index gives Chicago a segregation score of .893 in 1940 and .880 in 1950. These statistics suggest that Chicago's nonwhite population was less concentrated in 1950 than in 1940. Its rank among 187 cities in 1940 was 13th; in 1950 its rank among 209 cities was 42d. This also suggests that Chicago's nonwhite population was less concentrated in 1950 than in 1940.

Detroit's segregation index was smaller than Chicago's in each year and also declined between 1940 and 1950. In 1940, Detroit's segregation score was .841 and its rank was 28th; in 1950, the score was .831 and its rank was 89th. The unweighted mean segregation score for the 185 cities having block statistics in both 1940 and 1950 increased by .033, from .734 to .767. Since an unweighted mean gives equal weight to cities of all sizes, Cowgill computed a combined weighted index based upon the sums of all of the original figures. The obsolute change in this weighted index was exactly the same as for the unweighted one, increasing from .830 to .863.

The differences in the weighted and unweighted means indicate that, using this measure of segregation at least, larger cities were more segregated than smaller ones in both 1940 and 1950. Table 3 gives the segregation scores and ranks for Chicago and Detroit as well as 12 other large U.S. cities. When interpreting these scores, one should remember that they refer only to central cities where the overwhelming preponderance of the Negro population lives and where the greatest increases in Negro population have occurred. Had the indexes for each year been calculated for the entire metropolitan area, it seems highly unlikely that those for Chicago and Detroit would have exhibited declines in 1950. If the idea had been calculated for the entire metropolitan area, it certainly would have been both higher in each year and would have increased between 1940 and 1950. Thus,

Table 2

PERCENTAGE OF NEGROES AND WHITES LIVING IN THE CENTRAL CITY AND RINGS

	White			Negro		
Year	Total	Central City	Ring	Total	Central City	Ring
1960	100.0	47.8	52.2	100.0	79.6	20.4
1950	100.0	56.6	43.4	100.0	77.2	22.8
1940	100.0	61.6	38.4	100.0	74.6	25.4
1930	100.0	63.9	36.1	100.0	72.8	27.2
1920	100.0	65.9	34.1	100.0	67.2	32.8
1910	100.0	64.9	35.1	100.0	60.4	39.8
1900	100.0	62.8	37.2	100.0	54.5	45.5

SOURCE: U.S. Bureau of the Census, U.S. Census of Population: 1960, Selected Area Reports, Standard Metropolitan Statistical Areas, Final Report PC(3)-1D, Government Printing Office, Washington, D. C., 1963, pp. 1-5.

Cowgill's method provides an extremely conservative measure of both the extent of and increase in racial segregation in U.S. urban areas.

Since these data on segregation in other U.S. cities indicate that racial segregation is the norm in American cities, the findings of this paper have relevance beyond the Detroit and Chicago metropolitan areas. The data are not ideal, however, because they do not describe precisely the compactness of the segregated area — and it makes a great deal of difference whether the pattern of racial segregation is one of a single massive ghetto or whether it consists of several geographically dispersed ones.

Table 3

SEGREGATION SCORES OF 14 AMERICAN CITIES, 1940 AND 1950

City	Segregation Score		Rank	
	1940	1950	1940	1950
Baltimore	.847	.910	45	18
Boston	.853	.836	36	93
Chicago	.893	.880	13	42
Cleveland	.874	.855	20	72
Detroit	.861	.838	28	89
Los Angeles	.821	.798	61	121
Miami	.974	.969	1	2
Minneapolis	.781	.789	92	130
New York	.798	.794	81	122
Philadelphia	.813	.821	67	103
Pittsburgh	.789	.809	88	115
St. Louis	.813	.857	68	68
San Francisco	.516	.693	64	174
Washington, D. C.	.624	.540	150	192
Composite Index (185 cities)	.830	.863		

SOURCE: Donald O. Cowgill, "Trends in Residential Segregation of Nonwhites in American Cities, 1940-1950," Amer. Soc. Rev., Feb. 1956, Table 1, p. 45.

THE DISTRIBUTION OF NEGRO EMPLOYMENT

This section investigates the effect of the ghetto on the distribution of employment in urban areas. To evaluate this subject empirically, special tabulations were made from data contained in the Detroit and Chicago origin and destination studies. The data were obtained from home interviews — approximately 50,000 conducted in Detroit (1952) and 60,000 conducted in Chicago (1956). Home-interview origin and destination data are the only source giving detailed place of residence and place of work by race.

The tabulations shown in Fig. 3, by concentric rings in Chicago and Detroit, suggest an interrelationship between workplace and dwelling. In both cities, Negroes tend to work and reside in central sections of the city, while the white population tends to work and reside in outlying or suburban rings.

It might by tempting to try to explain these differences in white and Negro residence patterns by the differences in their employment patterns. Such an explanation might contend that the jobs held by nonwhites are more centrally located for historical and technological reasons than are those of whites and that, in turn, Negro residences are more centrally located than those of whites because, ceteris paribus, workers prefer to live near their places of work. Such an analysis would argue that causation runs from place of work to place of residence and that decisions about where to work are major determinants of residential location. Indeed, this is the causation assumed in most of the author's empirical work on residential and travel behavior.

The work-to-residence causation is an obviously unsatisfactory framework, however, for evaluating nonwhite travel behavior and choices of residential and employment locations. For example, given the significant constraints on the Negro's choice of residential location described earlier, it seems probable that the choice of a workplace location might be significantly be affected by the limitations on his residential choice. Some job locations are so far from Negro areas as to impose prohibitive costs on those who might seek employment there.

There are several reasons why housing-market segregation may limit Negro employment opportunities. The most obvious are: (1) the distance and difficulty of reaching certain jobs from acceptable Negro residence areas may impose costs on Negroes high enough to discourage them from seeking employment there. (2) Negroes may have less information about and less opportunity to learn about jobs distant from their housing areas.[14] (3) Employers may discriminate against Negroes out of real or imagined fears of retaliation from white customers for "bringing Negroes into all-white residential areas," and there may be little pressure for him not to discriminate. (4) Similarly, employers in or near the ghetto may discriminate in favor of Negroes.

Given the knowledge that there are great disincentives for Negroes to locate outside of the ghetto, the data presented in Fig. 3 seem to indicate that the centrally located Chicago and Detroit ghettos cause nonwhite jobs to be more centrally located. While this evidence is highly suggestive, a more rigorous test is desirable. Such a test can be obtained by fitting a multiple series of regression models for each city. The percentage of total Negro employment in each of 98 workplace areas is the dependent variable; the explanatory variables are a series of proxy variables that measure the factors causing Negroes to be underrepresented in distant workplaces.

Transportation costs from the workplace area to the ghetto, and the effect of distance on knowledge of job opportunities, are proxied by two variables: the airline distance from the workplace to the nearest Negro residence area (the nearest residence zone having more than two per cent Negro residents), and the airline distance from the workplace to the nearest point in the major ghetto. The residence zones shown in Figs. 1 and 2 have the same boundaries as the workplace zones. The percentage of Negro residents for each zone is a proxy for the employers' propensity to discriminate in favor of or against nonwhites

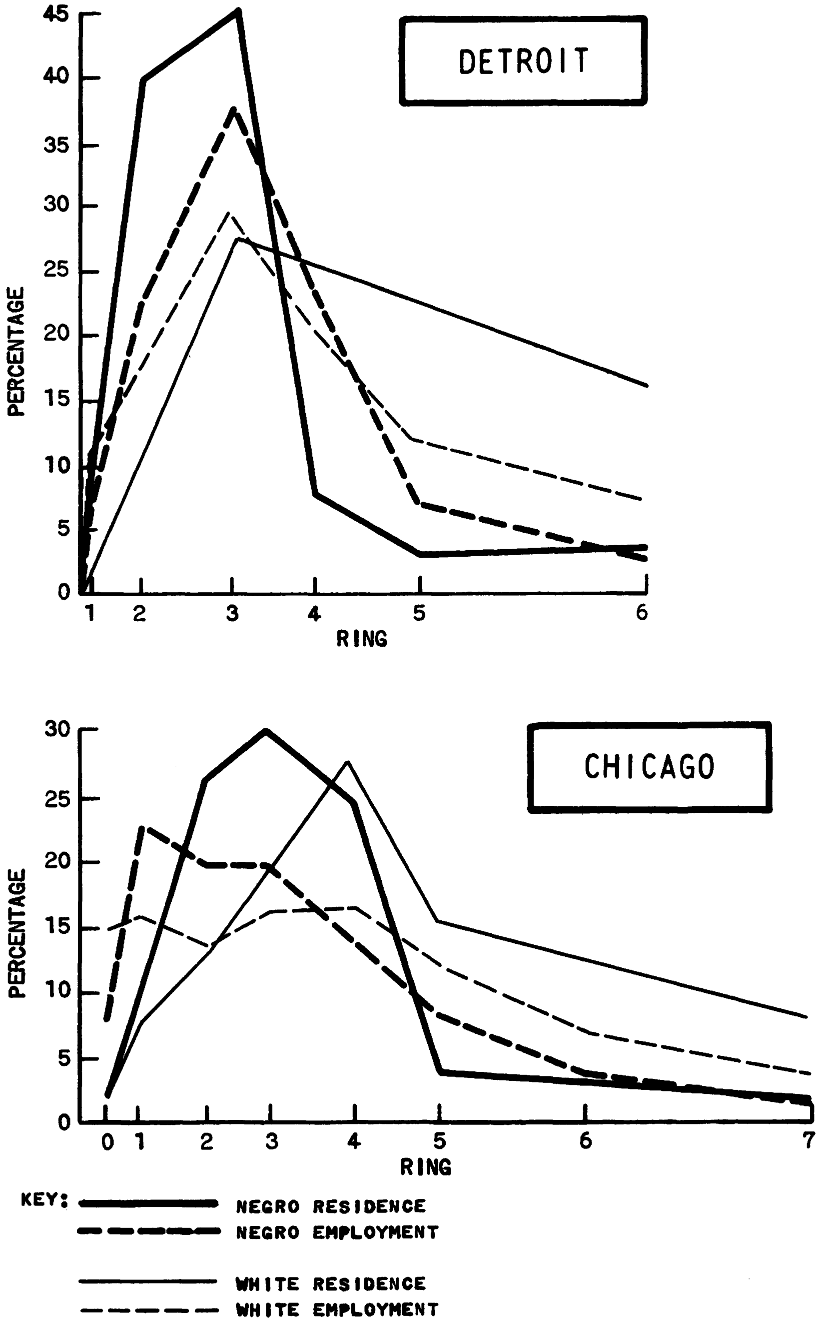


Fig. 3 — Percentages of Detroit and Chicago Workers Residing in Each Distance Ring

because of attitudes the resident population has toward the employment of Negroes. It seems likely that this measure also picks up some of the transportation-cost and information-cost effects. (Its correlation with the distance from the major ghetto variable is r =-.34 in Chicago and r =-.45 in Detroit. Its correlation with the distance from the nearest ghetto variable in Detroit is even higher, r = -.52.) The regression equations, the correlation matrices, the means, and the standard deviations of the variables are presented below for each city.

Since distance from the major ghetto and distance from the nearest ghetto are highly intercorrelated, using them both in the equation does not add much to the explained variance. Thus three equations are fitted for each city. The nonwhite residence variable is included in all three. The first equation for each city, R1 and R4, includes the distance variable from the nearest ghetto; Eqs. (R2) and (R5) include the distance variable from the major ghetto; and Eqs. (R3) and (R6) include both distance variables. Including both distance variables in the same regression equation greatly reduces the statistical significance of their coefficients. This is especially true for the Chicago models. When only one distance proxy is used, the coefficients of all variables are highly significant.

Equations

Chicago $\qquad R^2$

$$\text{(R1)}\quad W = 9.62 + \underset{(0.036)}{0.465R} - \underset{(0.014)}{0.565D^n} \qquad .720$$

$$\text{(R2)}\quad W = 9.75 + \underset{(0.036)}{0.464R} - \underset{(0.011)}{0.447D^m} \qquad .719$$

$$\text{(R3)}\quad W = 9.83 + \underset{(0.036)}{0.462R} - \underset{(0.313)}{0.331D^n} - \underset{(0.251)}{0.209D^m} \qquad .722$$

Detroit

$$\text{(R4)}\quad W = 12.78 + \underset{(0.032)}{0.091R} - \underset{(0.262)}{1.141D^n} \qquad .359$$

$$\text{(R5)}\quad W = 12.64 + \underset{(0.034)}{0.100R} - \underset{(0.158)}{0.758D^m} \qquad .382$$

$$\text{(R6)}\quad W = 13.45 + \underset{(0.035)}{0.082R} - \underset{(0.345)}{0.563D^n} - \underset{(0.212)}{0.524D^m} \qquad .400$$

W = % of zone i's workers who are Negroes.

$$\left(\frac{\text{Negro employment}\ \ i}{\text{Total employment}\ \ i}\right) 100$$

R = % of zone i's residents who are Negroes.

$$\left(\frac{\text{Negro residents}\ \ i}{\text{Total residents}\ \ i}\right) 100$$

D^n = Airline distance in miles to nearest Negro residence area.

D^m = Airline distance in miles to nearest point in the major ghetto.

Means and Standard Deviations

Chicago			Detroit		
W	11.15	12.18	W	10.44	6.84
R	8.22	19.40	R	10.25	18.66
D^n	4.06	5.07	D^n	2.87	2.58
D^m	5.39	6.36	D^m	4.26	4.00

Simple Correlation Matrices

	Chicago				Detroit			
	W	R	D^n	D^m	W	R	D^n	D^m
W	1.00	.82	-.48	-.49	1.00	.47	-.56	-.57
R		1.00	-.33	-.34		1.00	-.52	-.45
D^n			1.00	.91			1.00	.75
D^m				1.00				1.00

The most obvious difference between the Detroit and Chicago models occurs in the proportion of total variance explained by the equations. All three Chicago regressions explain more than seven-tenths of the variance in the dependent variable, while the Detroit regressions explain only about four-tenths. This difference occurs because racial segregation is greater in Chicago than in Detroit. As mentioned earlier, Detroit's major ghetto is larger and more dispersed than Chicago's, and Detroit also has more and better-placed outlying Negro residential areas. Thus it is natural that the model explain less about Detroit, where Negro residences are not so concentrated geographically. These differences are indicated by the means and standard deviations of the distance variables. The mean distance from the 98 Chicago workplace areas to the major ghetto is 5.39 miles and the standard deviation is 6.36 miles. The same mean distance for Detroit is only 4.26 miles and the standard deviation is only 4.00 miles. The mean distance from the 98 Chicago workplace areas to the nearest ghetto is 4.1 miles and the standard deviation is 5.1 miles. The same mean distance for Detroit is only 2.9 miles and its standard deviation is only 2.6 miles.

The coefficients for the two cities also differ considerably. The percentage residence coefficients for Chicago are much larger than for Detroit. A 1-per-cent increase in the number of Negro workers living in a Chicago residence area is associated with nearly a one-half-of-1-per-cent increase in Negro employment. In Detroit, by contrast, a 1-per-cent increase is associated with an increase in employment of only about one-tenth of 1 per cent.

The distance variables, however, are far more important in the Detroit models. With each 1-mile increase in distance from the major ghetto, the percentage of Negroes employed in a workplace area declines by eight-tenths of 1 per cent in Detroit, but only four-tenths of 1 per cent in Chicago. There is a similar correspondence for the distance-from-the-major-ghetto coefficients in Eqs. (R2) and (R5). In Chicago, the Negro percentage of the workforce declines by six-tenths of 1 per cent with each 1-mile increase in distance from the major ghetto; in Detroit the decline is 1.1 per cent. A similar relationship exists between the distance coefficients in the two cities when both are included in the model. Finally, the intercept in the Detroit equation is approximately one-third larger than in each of the comparable Chicago equations, despite the fact that the mean percentage of Negroes in the total of zone employees is higher in Chicago than in Detroit. I am not prepared, at this time, to speculate on the reasons for the differences in these coefficients and would welcome any views on the question.

The hypothesis that the ghetto, or the spatial distribution of Negro residence areas, importantly affects the distribution of Negro employment is strongly supported by the findings presented here. This is especially true in Chicago, where the segregation of Negroes is much greater.

THE LEVEL OF NONWHITE EMPLOYMENT

This section investigates the second of the paper's two hypotheses — that racial discrimination in the housing market reduces nonwhite employment opportunities in metropolitan areas. As mentioned earlier, loss of opportunity may be attributed to several things: prohibitive transportation costs from the ghetto to the place of employment, lack of knowledge about distant job opportunities, or greater discrimination by employers who are distant from the ghetto.

The regression equations for Chicago and Detroit can be used to estimate Negro job losses caused by housing segregation. To do so, it is assumed that the same proportion of Negro workers live in each residence area; this figure is identical to the percentage of Negro workers living in the metropolitan areas as a whole. These percentages are 13.6 for Detroit and 14.6 for Chicago. They are substituted in Eqs. (E1) to (E6) for R, which in Eqs. (R1) to (R6) denoted the percentage of zone i's workers who are Negroes. Since there would be no ghetto if there were no segregation, the distance from the major and nearest ghettos would be zero. Thus, the distance variables in Eqs. (R1) to (R6) drop out of Eqs. (E1) to (E6). Equations (E1) to (E6) thereby provide estimates of what percentage of Negro workers there would be in each workplace zone if there were no racial segregation, and if the proportion of the Negro population was the same for every residence zone. Since all zones included in Eqs. (E1) to (E6) have identical characteristics, these equations give the percentage of total metropolitan employment that would be Negro if there were no racial segregation in the housing market.

Once these figures are obtained, Eqs. (I1) to (I6) are used to convert the estimates of the percentage of total Negro employment to estimates of the loss of Negro employment. The estimated Negro proportion of total metropolitan employment is multiplied by the total metropolitan labor force. The loss of Negro jobs is obtained by subtracting the actual number of Negro jobs from this estimate of the number of Negro jobs. For Chicago, the loss estimates range from 31,662 to 34,654. The estimated losses in Detroit are much smaller, ranging from a low of 3,556 to a high of 9,113. The much smaller estimates for Detroit, like the smaller explanatory power of the Detroit models, are consistent with the lesser degree of racial segregation there. The ghetto in Detroit is larger and more extensive and there are more and better-situated secondary ghettos, housing is less of a constraint on nonwhite job choices in Detroit than in Chicago. Furthermore, Chicago's labor force is nearly twice as large as Detroit's; thus, the larger estimates of nonwhite job losses there are entirely plausible.

<u>Estimates of nonwhite employment assuming a uniform distribution of Negro residents</u>

<u>Chicago</u>

(E1) $W = 9.62 + 0.465\ (14.61) = 16.41$
(E2) $W = 9.75 + 0.464\ (14.61) = 16.52$
(E3) $W = 9.83 + 0.462\ (14.61) = 16.58$

(I1) $LNJ = 16.41\ (1{,}760{,}148) - 257{,}178 = 31{,}662$
(I2) $LNJ = 16.52\ (1{,}760{,}148) - 257{,}178 = 33{,}598$
(I3) $LNJ = 16.58\ (1{,}760{,}148) - 257{,}178 = 34{,}654$

<u>Detroit</u>

(E4) $W = 12.78 + 0.091\ (13.59) = 14.01$
(E5) $W = 12.64 + 0.100\ (13.59) = 14.00$
(E6) $W = 13.45 + 0.082\ (13.59) = 14.56$

(I4) $LNJ = 14.01\ (937{,}555) - 127{,}395 = 3{,}556$
(I5) $LNJ = 14.00\ (937{,}555) - 127{,}395 = 3{,}863$
(I6) $LNJ = 14.56\ (937{,}555) - 127{,}395 = 9{,}113$

LNJ = Loss of nonwhite jobs.

While these estimates must be considered highly tentative, they do suggest that housing-market segregation and discrimination may significantly affect the level of Negro employment in metropolitan areas. If this is true, it has grave welfare implications, since the costs that housing segregation imposes on Negroes may be even larger than is generally believed. The constraint placed upon job opportunities by housing-market discrimination may also partly explain the much higher unemployment rates of Negroes. Part of what is usually charged to employment discrimination may be an indirect effect of housing discrimination. This illustrates how pervasive various types of discrimination may be and how the indirect costs of discrimination may greatly exceed the direct costs.

<u>CONCLUSIONS</u>

This paper has empirically tested the effect of racial segregation in the housing market on the distribution and level of nonwhite employment. There is very strong evidence that racial segregation is an important determinant of the distribution of nonwhite employment. Negro workers, for example, are significantly underrepresented in employment zones distant from the ghetto, and the underrepresentation increases as distance from the ghetto increases. There is less overwhelming but still highly suggestive evidence that segregation patterns in U.S. metropolitan areas affect nonwhite employment levels. Racial segregation may cost Negroes as many as 35,000 jobs in Chicago and 9,000 jobs in Detroit. Chicago's larger labor force and greater racial segregation account for the difference in the figures for the two cities.

In addition to the obvious effects that loss of job opportunities has on nonwhite welfare, several other considerations relate to these findings. If the dispersal and suburbanization of employment characterizing the past few decades continues, the loss of nonwhite employment opportunities resulting from patterns of housing segregation will probably increase.

The findings also bear on estimates of the amount of nonwhite residence relocation that would result from a lessening of racial prejudice. Many current estimates are based upon existing distributions of Negro employment. Findings presented here indicate that a reduction in housing segregation would lead to a dispersal of Negro residences and to a more even distribution of Negro employment.

The study further indicates that some findings, such as those of Anthony Pascal, are extremely conservative regarding the amount of racial segregation that can be explained by socio-economic factors.[15] Using multiple regression models, Pascal finds that socio-economic variables explain only 46 per cent of the variation among Chicago residence areas in the proportion of all households headed by Negroes. The proportion explained by socio-economic variables in Detroit was even smaller — 33 per cent.[16] Access to nonwhite jobs is included among the socio-economic variables in Pascal's equations; this variable is highly significant for Chicago, but not for Detroit. This is consistent with one of the findings presented above: that Detroit's lesser residential segregation reduces the effect of segregation on the distribution of nonwhite employment. Omitting the access variable from the Chicago model would considerably reduce the amount of residential segregation that can be explained by socio-economic differences.

In his estimating equations, Pascal recognizes and comments on the possibility of reverse causality in the use of the access variable. He argues, however, that "attributing the maximum degree of nonwhite residence patterns to job location . . . makes for the more conservative estimate of non-socio-economic segregation."[17] The findings of this paper indicate that the causation is the reverse of that assumed in Pascal's models, but confirm, as he anticipated, that his estimates of non-socio-economic segregation are conservative.

The importance of the findings of this paper depends, in part, on the degree to which the historical patterns of racial segregation persist in the future. There has been continual pressure toward relaxing the housing-market barriers. It appears, however, that progress has been slow. In an analysis of census data for 1930, 1940, 1950, and 1960, Mayer and Hoult conclude, "Negro Detroiters are more segregated in their housing today than they were three decades ago."[18] Discussing maps of the distribution of Negro residences, they state, "The most notable feature of these maps is their clear demonstration that Negroes in the 1960 decade live in essentially the same places that their predecessors lived during the 1930's — the only difference is that, due to increasing numbers, they occupy more space centered about their traditional quarters."[19] Thus, as late as 1960 there was no appreciable change in the patterns of nonwhite residences in Detroit. Maps of 1960 census-tract data for Chicago indicate a similar state of affairs in Chicago. Consequently, it seems unlikely that the findings of Mayer and Hoult will soon be rendered obsolete by changes in the pattern of racial occupancy in U.S. metropolitan areas.

FOOTNOTES

1. Davis McEntire, Residence and Race: Final and Comprehensive Report to the Commission on Race and Housing, Univ. of Calif. Press, Berkeley and Los Angeles, 1960; Eunice and George Grier, Privately Developed Interracial Housing: An Analysis of Experience, Univ. of Calif. Press, Berkeley and Los Angeles, 1960; Nathan Glazer and Davis McEntire, Studies in Housing and Minority Groups, Univ. of Calif. Press, Berkeley and Los Angeles, 1960; Luigi Laurenti, Property Values and Race: Studies in Seven Cities, Univ. of Calif. Press, Berkeley and Los Angeles, 1960; Chester Rapkin and W. C. Grigsby, The Demand for Housing in Racially Mixed Areas: A Study of the Nature of Neighborhood Change, Univ. of Calif. Press, Berkeley and Los Angeles, 1960; O. D. Duncan and Beverly Duncan, The Negro Population of Chicago, Univ. of Chicago Press, 1957; Morton Grodzins, The Metropolitan Area as a Racial Problem, The University of Pittsburgh Press, Pittsburgh, 1958; Beverly Duncan and P. M. Hauser, Housing a Metropolis — Chicago, The Free Press, Glencoe, Illinois, 1960.

2. J. F. Kain, "Commuting and the Residential Decisions of Central Business District Workers," National Bureau of Economic Research, Special Conference 17, Transportation Economics, Conference Monograph (forthcoming); J. R. Meyer, J. F. Kain, and Martin Wohl, The Transportation Problem, Harvard University Press, Cambridge, Mass. (forthcoming, Spring 1965).

3. Kain, op. cit.

4. H. J. Gilman, "Economic Discrimination and Unemployment," paper presented at the RAND Conference on Urban Economics, August 24-25, 1964, Santa Monica, Calif.; idem, "The White/Nonwhite Unemployment Differential," in Human Resources in the Urban Economy, Resources for the Future, Inc., The Johns Hopkins Press, Baltimore, 1963 , pp. 75-113; Jacob Mincer, "On the Job Training: Costs, Returns, and Some Implications," J. Polit. Econ., Vol. 70, Supp., October, 1962, pp. 50-79; R. H. Turner, "Foci of Discrimination in the Employment of Nonwhites," Am. J. Soc., Vol. 58, November, 1952, pp. 247-256.

5. Detroit Area Traffic Study, Report on the Detroit Metropolitan Area Traffic Study: Part I — Data Summary and Interpretation, Speaker-Hines and Thomas, Inc., State Printers, Lansing, Michigan, July 1955; Chicago Area Traffic Study, Vol. I, Western Engraving and Embossing Co., Chicago, Ill.

6. G. S. Becker, The Economics of Discrimination, University of Chicago Press, Chicago, 1957; Robert Weaver, The Negro Ghetto, Harcourt, Brace and Co., New York, 1948; Glazer and McEntire, op. cit.; Grodzins, op. cit.; McEntire, op. cit.; T. J. Woofter, Jr., Negro Problems in Cities, Doubleday, Doran and Company, Garden City, New York, 1928; Rapkin and Grigsby, op. cit.; Laurenti, op. cit.

7. Duncan and Duncan, op. cit., pp. 87-107.

8. Ibid., p. 95.

9. Ibid, p. 106. Similar observations have been made by other researchers viewing the phenomenon of Chicago's ghetto. In 1947, for example, Robert Weaver noted that, "Spatial separation of the races increased too. By 1934, slightly over nine out of every ten Negroes in Chicago were in blocks predominantly occupied by Negroes." At still another point he commented that ,"The black belt in Chicago was also expanding. Patterns of racial occupancy had long been set, and the delayed and inadequate accretions to the areas of established Negro occupancy occurred within the framework of the ghetto pattern. An island of whites surrounded by Negroes was wiped out. Smaller centers of coloured residences expanded slowly, but the principal effect of the steady flow of Negroes to the city was to increase the congestion that had long plagued the Negro South Side." Weaver, op. cit. p. 54, 61.

Similarly, Davis McEntire provides the following description of the 1950 pattern of residential segregation in Chicago: "In Chicago in 1950, there were six segregated areas detached from the major Black Belt, each with adjacent tracts of concentrated, mixed, or dispersion types." Segregated, concentrated, mixed, dispersion, and exclusive types refer to five classes of census tracts defined by McEntire according to the proportion of Negroes to the total resident population. Segregated tracts were 75 per cent or more nonwhite; concentrated tracts were 50 to 75 per cent nonwhite; mixed tracts were 10 to 49 per cent nonwhite; dispersion tracts were from 1 to 9 per cent nonwhite; and exclusive tracts had fewer than 1 per cent Negroes in their population. McEntire, op. cit., p. 34.

10. Weaver, op. cit., p. 115.

11. McEntire, op. cit., p. 362.

12. Ibid., p. 34.

13. D. O. Cowgill, "Trends in Residential Segregation of Nonwhites in American Cities, 1940-1950," Amer. Soc. Rev., Vol. 21, No. 1, February 1956, pp. 43-47.

14. Labor mobility studies show that few jobs are located from newspaper advertisements, employment offices, and the like. Workers most frequently learn of jobs from friends, by passing the place of work and seeing help wanted signs, and by other casual associations. Since nonwhites have few associations with white areas distant from the ghetto and since few of their friends and neighbors are employed there or make frequent trips there, the chances of their learning of distant job opportunities may be significantly lessened.

15. A. H. Pascal, Summary: The Economics of Housing Segregation, paper presented at the RAND Conference on Urban Economics, The RAND Corporation, Santa Monica, California, August 24-25, 1964.

16. Ibid., pp. 6-7.

17. Ibid., p. 6

18. A. J. Mayer and T. F. Hoult, Race and Residence in Detroit, Urban Research Laboratory, Institute for Urban Studies, Wayne State University, Detroit, August, 1962, p. 1.

19. Ibid., p. 2.

XII

STATISTICS AND CIVIL DEFENSE

Chairman, Louis Schwalb, U. S. Office of Civil Defense

Page

Civil Defense Statistical Programs - Joseph Romm, U. S. Office of Civil Defense.................. 272

The Application of Statistics to the Resource Management Program - Joseph D. Coker, National Resource Evaluation Center...................... 282

USDA Surveys for Emergency Preparedness - Trienah Meyers, U. S. Department of Agriculture. 289

Discussion - Harry B. Sheftel, U. S. Office of Statistical Standards................ 293

CIVIL DEFENSE STATISTICAL PROGRAMS

Joseph Romm, Office of Civil Defense

The civil defense function has been in the Federal establishment in one form or another since shortly after the start of World War II. Rather than attempt to provide a broad review of many studies over the past two decades, I will cover two active programs: first, the National Fallout Shelter Survey begun shortly after the civil defense functions were transferred to the Department of Defense in August 1961, and second, the damage assessment and vulnerability analysis program started by our predecessor agencies and now carried on by the Office of Emergency Planning and the Office of Civil Defense.

NATIONAL FALLOUT SHELTER SURVEY

The National Fallout Shelter Survey was established as a result of a determination that within feasible expenditure levels, a system of fallout shelters would provide a greater saving of lives than any other system. The first step, designed to make use of existing assets, was a survey to: (1) Locate suitable fallout shelter facilities; (2) mark them with distinctive signs; and (3) stock them with food and water, medical and sanitation kits, and radiation measuring instruments. The location of suitable fallout shelter facilities involved a large scale survey with all its inherent problems, plus additional administration problems arising from the use of trained architects and engineers to perform the survey, the use of the military engineer organizations as contract administrators for the survey, and the fact that civil defense responsibilities are divided between the Federal government and State and local governments.

The survey consisted of two distinct phases:

Phase I recorded structural and geographic data on all buildings other than single family housing units in the continental United States, Alaska, Hawaii and the possessions, that were estimated to have a protection factor of at least 20 and potential shelter space for 50 or more persons. This means that the shelter had to be capable of reducing the radiation intensity of the fallout outside the shelter by a factor of 20, and have 10 square feet per person in adequately ventilated spaces or 500 cubic feet per person in unventilated space. Further, the requirement includes 1 cubic foot of secure storage space per person.

Phase I operated as follows:

1. Data were gathered and recorded by architectural and engineering firms under contracts negotiated, administered and supervised by the Army Corps of Engineers and Navy Bureau of Yards and Docks field offices.
2. Field information was transmitted to the Bureau of Census Office in Jeffersonville, Indiana where it was microfilmed and sent to Suitland, Maryland for conversion into computer tape.
3. The computer tapes were transferred to the National Bureau of Standards for processing on its 7090 computer. Printouts were returned to the field offices and the contractors and summary reports were furnished to the various levels of local, county and state governments.

Information was recorded in Phase I on forms using the FOSDIC process "Film Optical Sensing Device For Input To Computer." This process, developed for the Bureau of Census, was used in recording data in the 1960 census. The procedure eliminates the card punching operation. The FOSDIC form is microfilmed, read on a FOSDIC head and converted into a computer tape.

COVERAGE

To make sure that every area of the United States was covered, we used a system of geographic identification that has been in existence for some time. This system divides the country into about 43,000 "Standard Locations" that have an average population of around 4,000. The standard location number designates the OCD region, state, county and sub-division of that county where appropriate. The areas are comparable to the Census Tracts and Minor Civil Divisions. This information, coupled with the name and address, adequately locates a structure for local needs and computer requirements.

The survey covered about 5 million buildings. About 3.7 million of these were rejected as patently unsuitable. Another million buildings were eliminated by closer examination and the data for 381,902 buildings were completely processed.

TRAINING

In order to develop competent enumerators, it was necessary to train some 2,700 architects and engineers in shielding analysis courses given at nine universities and two military schools.

PROCESSING

More than 500,000 FOSDICS forms were prepared, as some buildings required more than one form. A form was required for each wing or major section of a building.

The 7090 computer calculated the protection factor and capacity on a floor by floor basis, building by building. It required about 1/20 of a second to compute and check the results as compared with about two hours needed by a professional engineer to do these same calculations. The results were as follows:

Protection Factor	Buildings	Spaces (Millions)
PF 100 or better	112,899	55.8
40-99	103,467	68.1

Now for Phase II:

Based on the results of Phase I, those buildings which provided better than 40 protection factor were revisited by the architect and engineer firm to make detailed studies. Phase II also included the survey of selected special facilities, such as caves, mines, and tunnels, for shelter suitability. Phase II included the identification of:

1. The specific areas which provided the required shelter, and
2. The improvements required to increase the amount of protection and ventilation, to improve habitability and to increase shelter capacity.

Pertinent information for each building or facility surveyed was sent to appropriate State and local civil defense officials for use in shelter planning. Similar information was sent to each of the military services relative to facilities under their jurisdiction. This information included the shelter capacity of existing buildings and special facilities suitable for shelter use, the computed fallout protection factors, estimated cost of upgrading substandard shelter space to a protection factor of 100, and estimated cost of increasing shelter space by improvements such as ventilation.

For program control and for further use in shelter planning, data compilations included the identification of shelter space available according to various types of structural categories; e.g., 33 classes of physical vulnerability, 5 types of ownership, 41 categories of current usage, and 9 kinds of special facilities. In addition, shelter space data were summarized by standard locations to show the findings for the entire Nation, OCD regions, states, counties, cities of 25,000 population or more, and standard metropolitan statistical areas.

The second step in the survey is the marking of the shelter facilities with distinctive signs. This is accomplished by the Corps of Engineers in cooperation with local civil defense directors. Obtaining of permission of the building owner to allow the facility to be used as a shelter is the function of the local civil defense director. This is done through a licensing agreement signed by the building owner, which includes permission to place shelter supplies in his building. This last step in the program, the stocking of the shelter with food and water, medical and sanitation kits, and radiation measuring instruments, is a joint task. The Federal government is responsible for procurement, warehousing and distribution of these supplies. The local civil defense director is responsible for getting these supplies into the shelter facility and for inspection to make certain that they are properly stored and maintained.

The initial program and its continued updating has now produced the following statistics as of October 25, 1964.

	Facilities	Spaces (000)
Located	146,332	124,671
Licensed	74,333	68,135
Marked	82,128	69,127
Stocked	50,552	26,852

Now for a related, but quite different program, - damage assessment and vulnerability analysis. Damage assessment is a postattack computation of the effects of an attack on population and other resources. It may be based upon actual damage reports, on aerial reconnaissance of attacked areas or on on-site survey reports. It can make use of manual or computer calculations of the effects of the attack using reports of the attack, our knowledge of weapons effects and physical vulnerability factors and precise locations of resources of interest. Vulnerability analyses, on the other hand, involve preattack estimates of the nature and effects of a wide range of possible attacks. Damage assessment models run the gamut from those which are designed to provide rapid national postattack situation estimates, to estimates designed to provide extensive detail at other levels of government. The model which we usually use for this latter type of assessment is known as JUMBO. Chart 1 provides the symbology for this model. I believe that the first line of the chart is self-explanatory so I shall limit my discussion to the squares identified as DUSTY and FLAME. DUSTY simulates the lifting of the radioactive particles by the mushroom cloud from each surface burst weapon, the transportation of the particles by the upper winds, and the depositing of these particles on the surface.

THE ATTACK

First, we must feed into the system the attack data; i.e., the ground zeros or coordinates of points of burst, the yields or sizes of the weapons, the heights of burst (or at least an indication of whether the burst is low enough for the fireball to touch the surface of the ground, because only such bursts cause serious fallout problems) and the time of burst.

Next, we must feed in the wind data; i.e., the directions and velocities of the upper winds, for all areas of the country. The computer then plots a path downwind from each surface burst weapon, the path curving in accordance with changes in wind directions, area by area.

The pattern by which the computer simulates the deposition of fallout along and on both sides of this path depends on the yield of the weapon and on basic assumptions concerning such things as the configuration of the fallout cloud and the rates of fall (which in turn depend on size and weight of the particles containing radioactive materials) -- so these inputs must be provided. Furthermore, computations of the densities of the radioactive fallout require information or assumptions about fission/fusion ratios -- the greater the percentage of fission, the greater the radioactive intensity. Radiation intensities decline rapidly with the decay of radioactivity, and in order to be additive, intensities must be computed as of some base time after detonation. The standard base used is H + 1, i.e., one hour after detonation.

The model estimates the intensity of radiation in a series of 2-minute trapezoids from each surface-burst weapon, and then sums the intensities of fallout in each trapezoid from all surface burst weapons.

The model also computes the radiation dose one might receive if standing, unprotected in the open, in each 2-minute trapezoid. It can do this on the basis of some stipulated period of accumulation, taking radiation decay rates into account. Usually, it follows a more complex procedure of taking into account both radiation decay rates and also rates of biological recovery from radiation damage. The dose thus computed is called Equivalent Residual Dose, usually referred to by the initials ERD.

DUSTY

The DUSTY model provides two forms of output. One consists of fallout maps, including both fallout intensity maps showing intensities at one hour after detonation, and "outside" dose maps showing, for example, 36-hour dose or equivalent residual dose.

The other output is a magnetic tape record of radiation intensities and "outside" doses for each 2-minute trapezoid. This record is used as an input to the attack environment model.

FLAME

Now for the FLAME model. As the name implies, this model simulates the spread of fires from points of ignition by direct thermal radiation. The basis inputs are the attack data and weapons data on thermal radiation. The computer then identifies the 2-minute trapezoids in which 2-minute trapezoids in which ignitions might start. Then it considers in turn: firespread barriers (both absolute and probable barriers), fuel densities (i.e., densities of combustible materials), and weather and moisture conditions, area by area. It estimates the potential spread of fire through contiguous 2-minute trapezoids beyond the area in which fires might be ignited by direct thermal radiation from the fireball.

The burned-out areas are delineated on an electronically-printed map, and also a list of burned 2-minute trapezoids is recorded on magnetic tape as another input to the attack environment computation.

VULNERABILITY ANALYSIS

Calculations based on a single set of attack inputs are fairly straight forward. However in vulnerability analysis we apply a range of similar inputs; and I should like to cover briefly the analytical model which we use most commonly in our vulnerability studies.

Analytical models are used to simulate almost any conceivable activity or operation, e.g., an air battle, the effects of a nuclear attack, or even the operation of a national economy. Repeated trials or operations of the analytical model, with stipulated changes in assumptions or parameters, enable us to determine the sensitivity of the results to changes in any of the variables in a problem or operation. Using the results of many trials in such controlled "experiments", we can perform our analyses, understanding the consequences of alternate courses of action.

What we aim for in vulnerability studies is to obtain estimates of the probability that certain effects may take place in areas of interest to us given a specified set of attack parameters.

The model which we use for these studies is known as RISK. This model computes the distribution of effects that a specific point will be exposed to, in terms of ranges of blast and radioactive fallout intensities as well as ranges of losses and availability of specified resources.

In developing input parameters for the RISK system, major uncertainties such as the following must be considered:

a. A hypothetical attack may include military, population and industrial targets, or any combination thereof.
b. The specific targets within each of these categories may vary, as may the number of weapons assigned to each.
c. Delivery vehicles may be missiles or aircraft; they may also vary as to type and number and may have varying reliability. Additionally, our defense capability against enemy vehicles will vary.
d. Variations will occur in weapons yield, fission yield, height of burst and aiming error.
e. Our assumptions on the damage resulting to various resources from blast, radiation and thermal effects are approximations.
f. Meteorological considerations have infinite variations. To meet these uncertainties, we require the following:

Input Data for Each Target

a. The coordinates of the designated ground zeros (DGZ). (The DGZ represents the aiming point. The actual ground zero (AGZ), or actual point of detonation, may differ because of weapon inaccuracy.)
b. Numbers of weapons, yield, height of burst, and time of detonation.
c. Abort rate - (weapon does not arrive over U. S.)
d. Attrition rate - (Weapon arrives over U. S., but is destroyed prior to reaching target.)
e. Circular error probable (CEP). The CEP represents the weapon aiming error. It is construed as the radius of a circle around a target in which it would be expected that 50% of the weapons would land.

f. Mean wind speed and direction, and the corresponding standard deviation. These data are furnished for each season of the year. Now let us examine the RISK model (CHART 3).

In the attack selection process, we start with statements of the various alternative combinations of the objectives and capabilities of the potential enemy along with pertinent data on the U. S. resources against which these capabilities might be applied. The output of this process is a number of enemy attack options, each related to one of the alternative sets of objectives and capabilities.

The attack options lead to the next process, attack gaming, which takes into account military operational factors such as abort rates, attrition rates, and aiming errors.

A Monte Carlo Program generates random numbers and applies them to wind and weapons data for each target. The result is an attack 'trial' which contains AGZ's, weapons delivered, wind speed and direction. All the variables in the problem, including the operational factors and the weather, are taken into account in many separate 'trials' for each of the attack options. The output consists of one hundred or more separate attack 'trials' which then become inputs to the RISK analysis. The results of these attack trials applies to resource points are shown as a distribution table of attack effects.

OUTPUTS

There are two general types of output derived from the RISK analysis:

1. Estimates of hazards at particular places in terms of probabilities for -
    a. Various ranges of blast overpressure intensities, and
    b. Various ranges of radioactive fallout, including variations in arrival time as well as intensity.

2. Estimates of hazard to population and resources expressed in terms of probabilities for -
    a. Various ranges of casualties
    b. Various ranges of resource losses, and
    c. Various ranges of resource availabilities during specified postattack time periods.

Tables 1, 2, and 3 illustrate some RISK outputs.

USES

There are many applications of these 'probability' data with varied levels of sophistication. One relatively simple use is to provide vulnerability advice to all levels of government, the military and industry, for dispersion planning to provide protection in existing and new structures. At the other end of the scale, these data are used in government-wide studies of the magnitude of postattack problems and these studies provide a tool for developing preattack plans to meet such postattack problems.

As I mentioned previously, models can be developed to simulate almost any conceivable activity or operations; e.g., an air battle, the effects of a nuclear attack, or even the operation of a national economy. Development work on many models has been carried out by the National Resources Evaluation Center. I shall now turn the discussion over to my friend, Dr. Joseph D. Coker, the Director of the NREC, who will discuss models for resources management.

THE JUMBO DAMAGE ASSESSMENT MODEL (Part I)

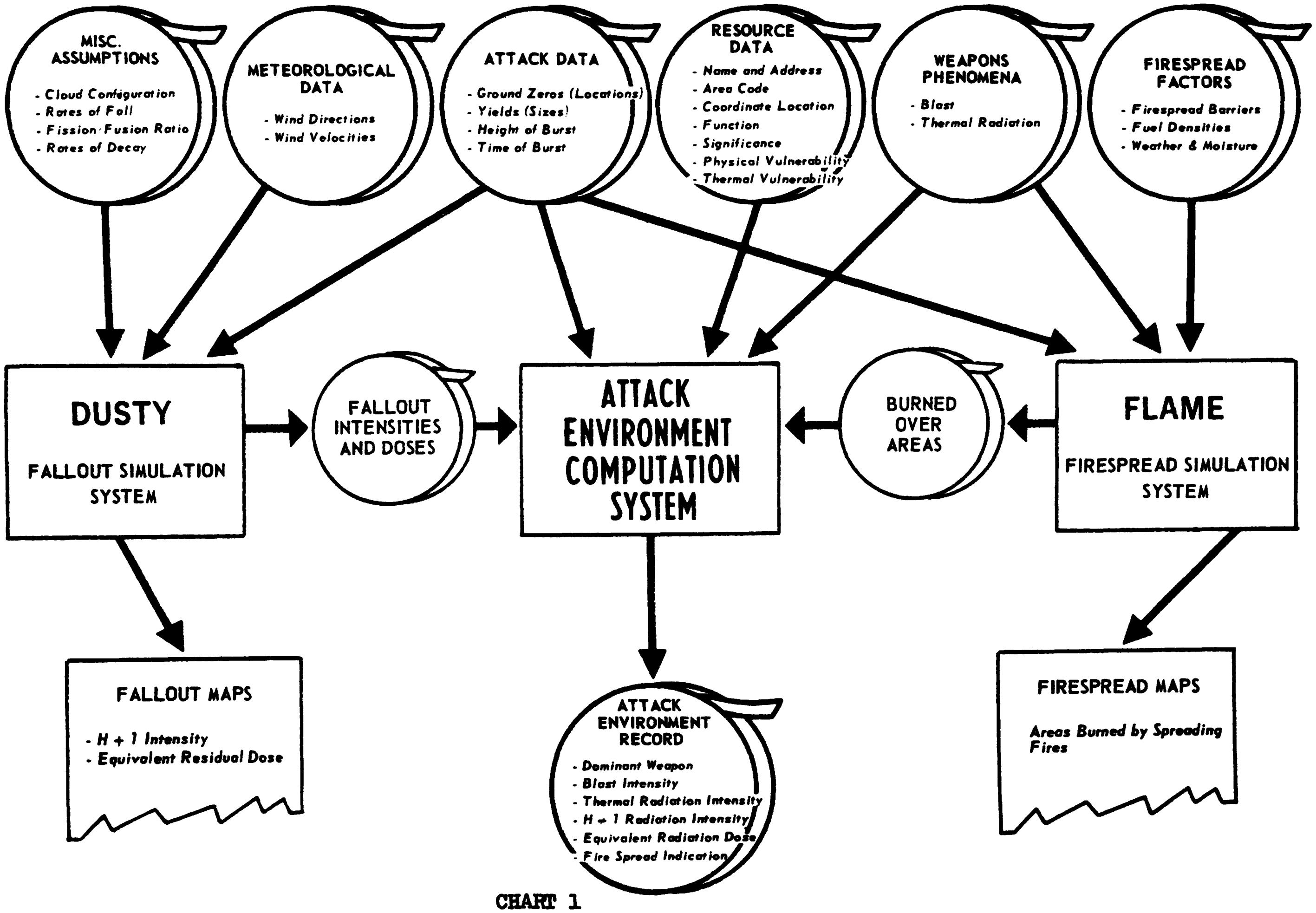


CHART 1

THE JUMBO DAMAGE ASSESSMENT MODEL (Part II)

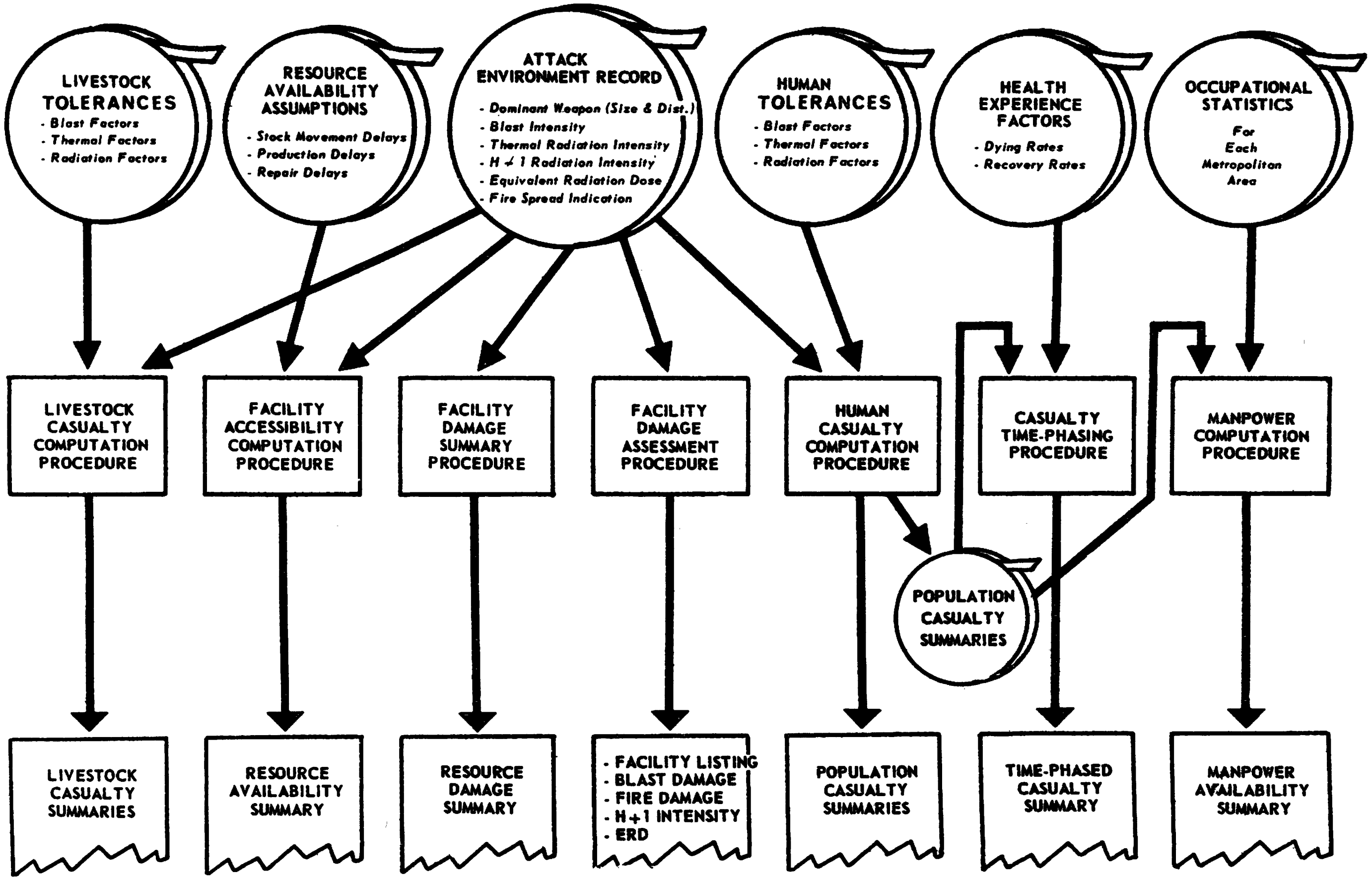


CHART 2

THE RISK MODEL

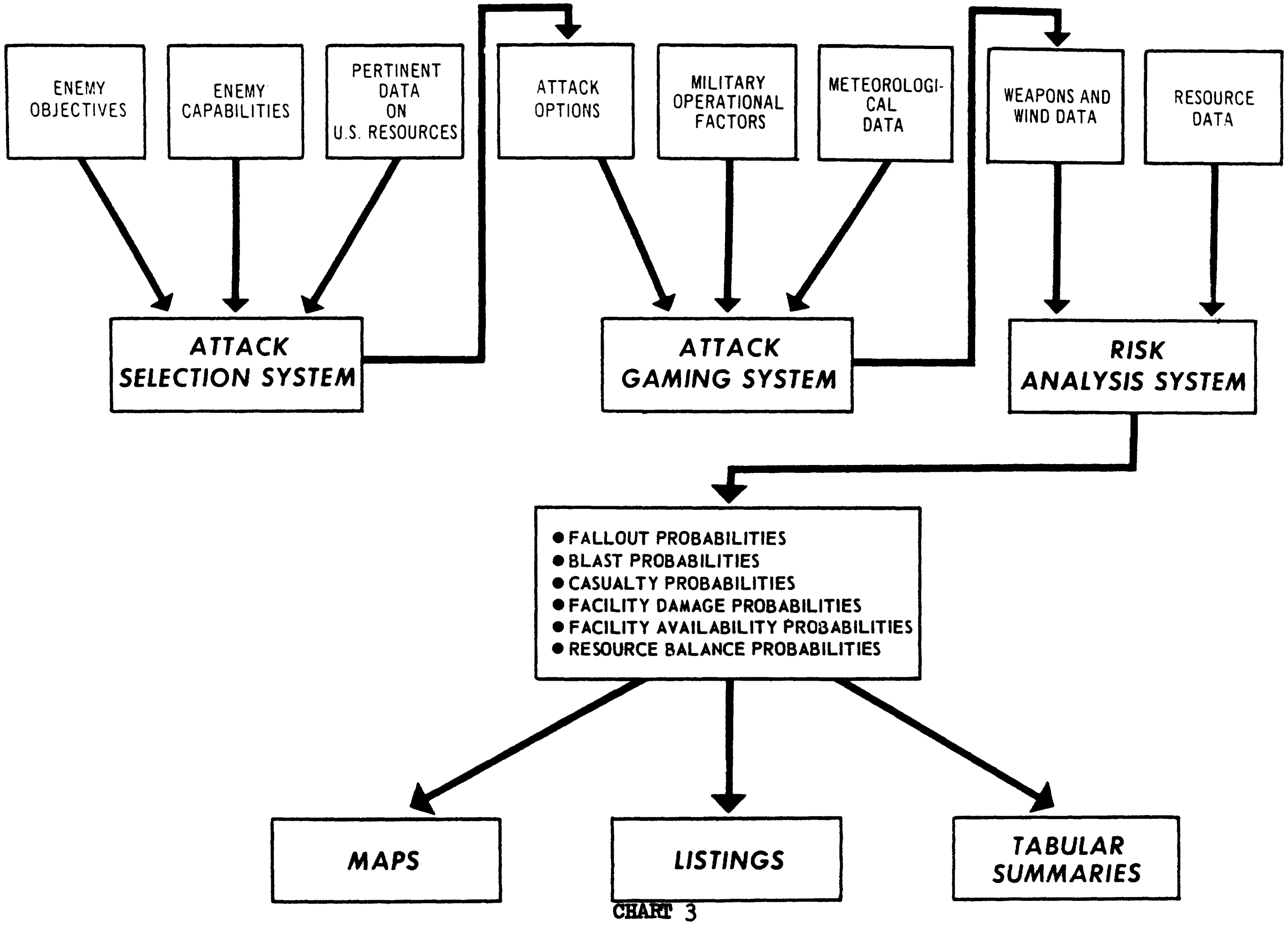


CHART 3

UNCLASSIFIED

Probabilities of Weapon Effects Based on Attack AR2 Run 1 100 Trials Season Winter Computed 12 Apr 62

Probabilities of Experience in Percent

Northwest Washington Washington DC Class TR 2211000000 VN 2201 2 UTM 18 43127 3198

. . . .

Blast and Radiation Dose Probabilities

Peak Overpressure (P.S.I.)

Radiation Dose Roentgens	0-1	1-2	2-3	3-5	5-7	7-10	10-15	15-20	20-25	25-50	50-100	100-200	200-300	300-500	Over 500	Maximum ERD Prob.	Cumulative Maximum ERD Prob.
Over 100,000	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100
25,000- 100,000	..	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100
10,000- 25,000	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100
7000- 10,000	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100
4000- 7000	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100
2000- 4000	2	1	.	1	.	.	.	.	.	.	.	.	.	.	.	4	100
1000- 2000	2	1	.	1	.	.	.	.	.	.	.	.	.	.	.	4	96
300- 1000	3	1	.	.	.	.	1	1	.	.	.	.	.	.	.	6	92
100- 300	3	3	2	.	.	.	.	.	.	.	.	.	.	.	.	8	86
Under 100	70	6	.	2	.	.	.	.	.	.	.	.	.	.	.	78	78
Overpressure Prob.	80	12	2	4	.	.	1	1	.	.	.	.	.	.	.	100	
Cumulative Overpressure Prob.	80	92	94	98	98	98	99	100	100	100	100	100	100	100	100		100

UNCLASSIFIED

FORMAT NO. 6: Point Experience. Overpressure and Dose

Table 1

UNCLASSIFIED

```
Probabilities of Weapon Effects     Based on Attack    AR2  Run 1    100 Trials Season Winter   Computed 12 Apr 62

Independent Probabilities of Experience in Percent

Blast Overpressure (PSI)                                                Radiation Dose Rate (R/Hr Normalized to H+1 Hour)
Does Not Exceed                                                         Does Not Exceed
   1   2   3   5   7   10   15   20   25   50  100  200  300  500     100     300    1000    3000   10000   30000

Northwest Washington                               Washington  DC   Class TR 2211000000  VN 2201 2  UTM 18 43127 3198
. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 

  80  92  94  98  98   98   99  100  100  100  100  100  100  100      86      91     100     100     100     100

Washington Monument                                Washington  DC   Class TR 2211000000  VN 6500 2  UTM 18 43063 3234
. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 

  85  94  96  97  98   99   99   99  100  100  100  100  100  100      83      91     100     100     100     100

Southeast Washington                               Washington  DC   Class TR 2211000000  VN 2201 2  UTM 18 43028 3291
. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 

  91  97  98  99 100  100  100  100  100  100  100  100  100  100      80      86     100     100     100     100
```

UNCLASSIFIED

FORMAT NO. 2: Condensed Point Experience

Example 2. Overpressure and Dose Rate

Table 2

Summary Analysis of Vulnerability Based on Attack AR2 Run C2 Season Spring 100 Trials Computed 20 Mar 62

Percent of Pre-Attack Population	Probability Percentile											Population
In Each Casualty Class	1%	5%	10%	15%	25%	50%	75%	85%	90%	95%	99%	
Category UP1 Test Population Category												
National Total												389825
Blast Killed	0	0	0	0	0	0	0	0	1	8	41	
Blast Casualties	0	0	0	0	0	0	0	1	8	23	41	
Total Killed	0	0	0	0	0	0	0	0	1	8	41	
Total Casualties	0	0	0	0	0	0	11	22	23	41	44	
Total Non-Casualties	56	58	77	78	86	100	100	100	100	100	100	
400 Maryland												88280
Blast Killed	0	0	0	0	0	0	0	0	0	0	48	
Blast Casualties	0	0	0	0	0	0	0	0	0	34	82	
Total Killed	0	0	0	0	0	0	0	0	0	0	48	
Total Casualties	0	0	0	0	0	0	0	0	62	96	96	
Total Non-Casualties	4	4	18	100	100	100	100	100	100	100	100	
410 Montgomery County												72100
Blast Killed	0	0	0	0	0	0	0	0	0	0	59	
Blast Casualties	0	0	0	0	0	0	0	0	0	41	100	
Total Killed	0	0	0	0	0	0	0	0	0	0	59	
Total Casualties	0	0	0	0	0	0	0	0	59	100	100	
Total Non-Casualties	0	0	0	100	100	100	100	100	100	100	100	

FORMAT NO. 19: Casualty Summary Formats

Percent of Population - Probability Percentiles

Table 3

THE APPLICATION OF STATISTICS TO THE RESOURCE MANAGEMENT PROGRAM

Joseph D. Coker, National Resource Evaluation Center

ABSTRACT

Resource Management implies appropriate action to promote the availability of resources for all important requirements. Prerequisites of resource management include estimates and comparisons of resource availabilities and requirements under conditions of interest. This paper examines briefly some of the analytical models developed to support resource management in a post-nuclear attack emergency. These models include one which focuses attention on questions of adequacy of certain types of consumption goods for survival of the population and another which focuses attention on the longer-term problems of economic and industrial recovery. Also noted are some of the important statistical elements of both models and the sources from which they are derived.

Introduction

Before discussing the application of statistics to resource management, let me first explain what we mean by "resource management." This term implies appropriate action by government to promote the availability of resources under all contingencies to meet all important requirements for resources.

Before any action may be determined to be appropriate for such a purpose certain types of estimates must be available. These include estimates of resource requirements and estimates of resource availabilities. Resource requirements in a national emergency depend on policy determinations at the highest levels of government.

From such levels, however, it is not reasonable to expect explicit statements of detailed requirements. Such detailed statements must be deduced from generalized high policy determinations of national objectives and priorities. For translation of such policy determinations into explicit requirement schedules, detailed statistical data and computing procedures and equipment are necessary. For determination of the extent to which such requirements can be satisfied, additional detailed statistics and procedures are needed in order to estimate resource availabilities and to compare these with the stated requirements.

The assembly of the necessary statistical data and the development and application of the necessary analytical models constitute the principal mission of the National Resource Evaluation Center, usually called NREC. Located in the Executive Office of the President, NREC is an interagency activity staffed and otherwise supported by upwards of 25 federal departments and agencies. Operating as a part of the Office of Emergency Planning, its prime objective is to meet as many as is feasible of the analytical requirements of that Office and the civil agencies of the federal government in connection with emergency planning and preparedness in the general areas of continuity of government and resource management.

Analytical Models

These analytical requirements are met, for the most part, by operation of several types of analytical models which apply advanced data processing techniques to large arrays of statistics. NREC has developed and used models for hazard analysis, situation analysis, damage assessment and other resource evaluation purposes. The damage assessment models make use of more than 750,000 detailed resource records. NREC's most important supplier of basic information is the Bureau of the Census, which provides detailed statistical data on population, housing, business, and manufacturing; but most of the other statistical collection programs of the federal government funnel information into the NREC system.

In order to provide support for broad resource program determinations we need analytical models for providing detailed estimates of resource requirements and for comparing resource supplies and requirements. No single model appears to be sufficient for all such resource evaluation purposes. Let me suggest two quite different types of resource supply-requirement analyses for which different models appear to be required.

On the one hand, we need comparatively short-term analyses of supplies of, and requirements for, survival items on a fairly detailed product and geographic basis, to support the immediate responses to nuclear attack on the United States. During the initial post-attack period, each area may be dependent upon locally available supplies, either because transportation may not be available from areas in which surpluses exist or because planned mechanisms for transferring surpluses in certain areas to cover deficits in others have not yet become effective. As time passes and as regional and national control systems become more effective, supply-requirement analyses for increasingly larger areas become appropriate and meaningful.

For long-term resource management problems, and especially for problems which involve the economic recovery of the entire country, the supply-requirements analysis is not so much concerned with local balances as with national balances. Moreover, it is concerned, not simply with production of survival items, but with all goods required for operation of the economy. It must take into account the interdependencies among industries and total supply-requirement balances.

Survival Items Analysis

To meet the first of these two types of requirements for supply-requirements analyses, NREC developed an analytical model called SURVIVAL.* A schematic representation of this model is presented in the first of the two charts. There are many parts to the analytical process and the way in which they fit together is moderately complex. But if we follow through the analytical process step-by-step the whole system turns out to be fairly simple.

But, first, a few words about the symbology of the diagrams. The analytical processes are identified in rectangular boxes. The inputs are on magnetic tape, as their circular labels suggest. Some of the outputs which become inputs for succeeding processes also are on magnetic tape. The final output is printed on paper which is suggested by rectangles with a ragged lower edge.

In the first analytical step in the operation of the NREC SURVIVAL model, estimates are computed for various classes of casualties, by cause and severity of injury (as calculated in the NREC casualty model)** and also health experience factors (that is, dying rates for fatal casualties and recovery rates for the various classes of non-fatal casualties) are applied in the casualty time-phasing computation. Thus, the surviving population is distributed among eight health categories, during each of five post-attack time periods in each state and region and in the nation as a whole. To the resulting time-phased distribution of survivors are applied per capita requirements of persons in each health category for each type of survival item. In this computation, the computer builds up a statement of total requirements. The requirements side of the picture is completed when these item requirements are summarized by area and time period.

The survival analysis now moves into the evaluation of the supply side of the problem. First, the analysis considers inventories. The NREC damage assessment model** has provided an evaluation on a plant-by-plant basis of the post-attack status of inventories in each class of survival items. These estimates, together with resource availability assumptions -- that is, assumptions as to how soon inventories which have been subjected to various ranges of blast and radioactive contamination will be available for use -- are used in making a time-phased estimate of available inventories.

Next, the SURVIVAL model considers the question of what quantities of each type of survival item can be expected from production during each time period in each area. Here, the problem is quite a bit more complicated. The SURVIVAL model does not consider all the complexities involved. It does, however, make a simple check to determine whether production will be limited by the availability of plant capacity or by the availability of manpower.

The controlling limitation is used in estimating time-phased supplies available from production. These are added to supplies freed from blast and radiation "frozen" inventories, in the development of total time-phased supply estimates.

Now the SURVIVAL model is ready for its final operation -- the comparison of requirements and supplies and the printing, for use by analysts and decision makers, of estimates of surpluses or deficits, item by item, period by period, and area by area. In this fashion, it would provide -- very early in a post-attack situation -- the first rough basis for coordinating and guiding national and regional responses to survival resource imbalances. But this type of survival item analysis during pre-attack periods can also provide a rational basis for decisions on stockpiling survival items, and for pre-attack preparations for controlling, conserving, producing, allocating, transporting, and distributing critical survival items under various emergency contingencies.

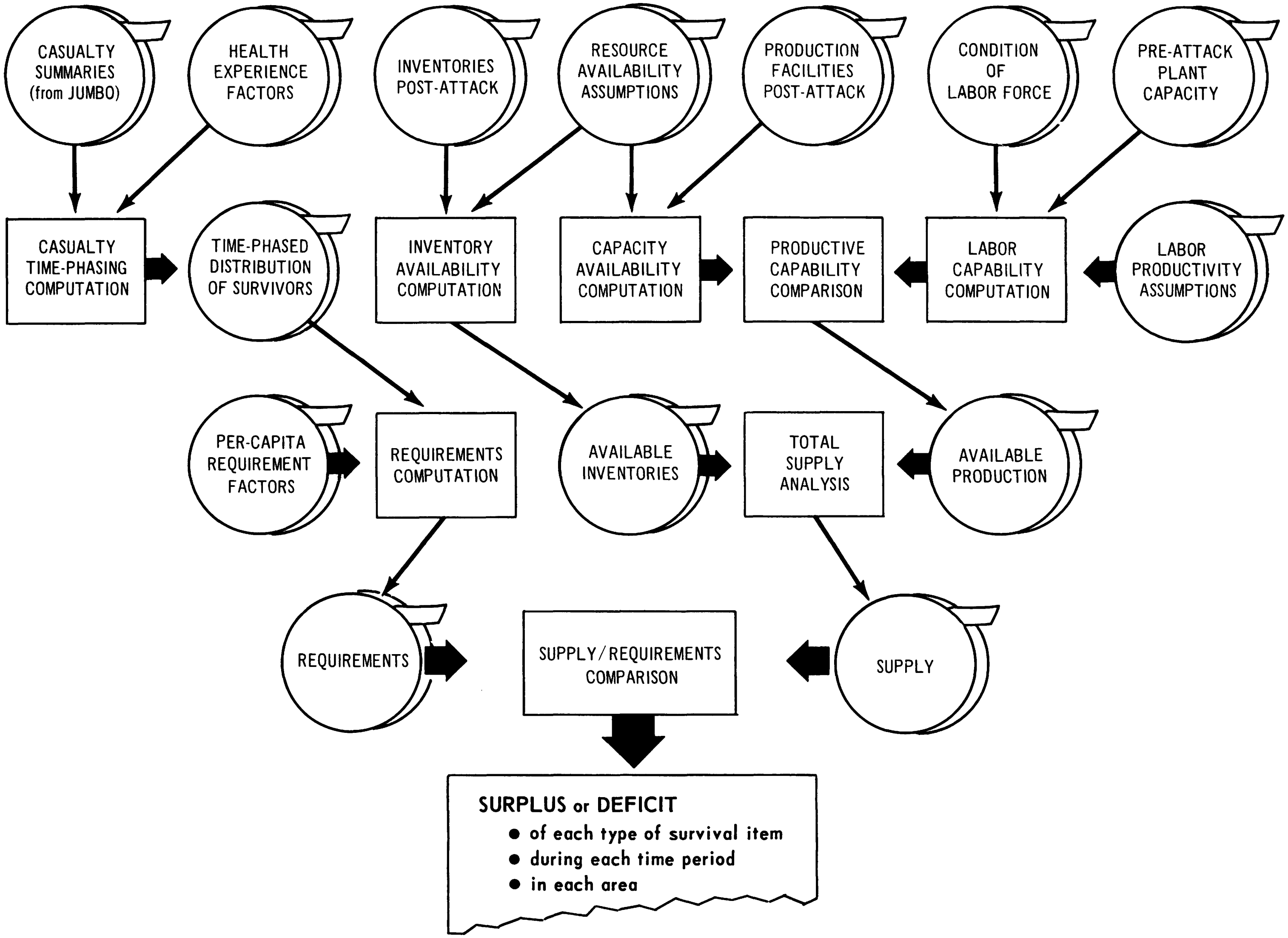
What can be said in the way of summary appraisal of the SURVIVAL model? It is fairly versatile and sophisticated in some respects but it affords only superficial consideration of vertical supply problems. By vertical supply

---

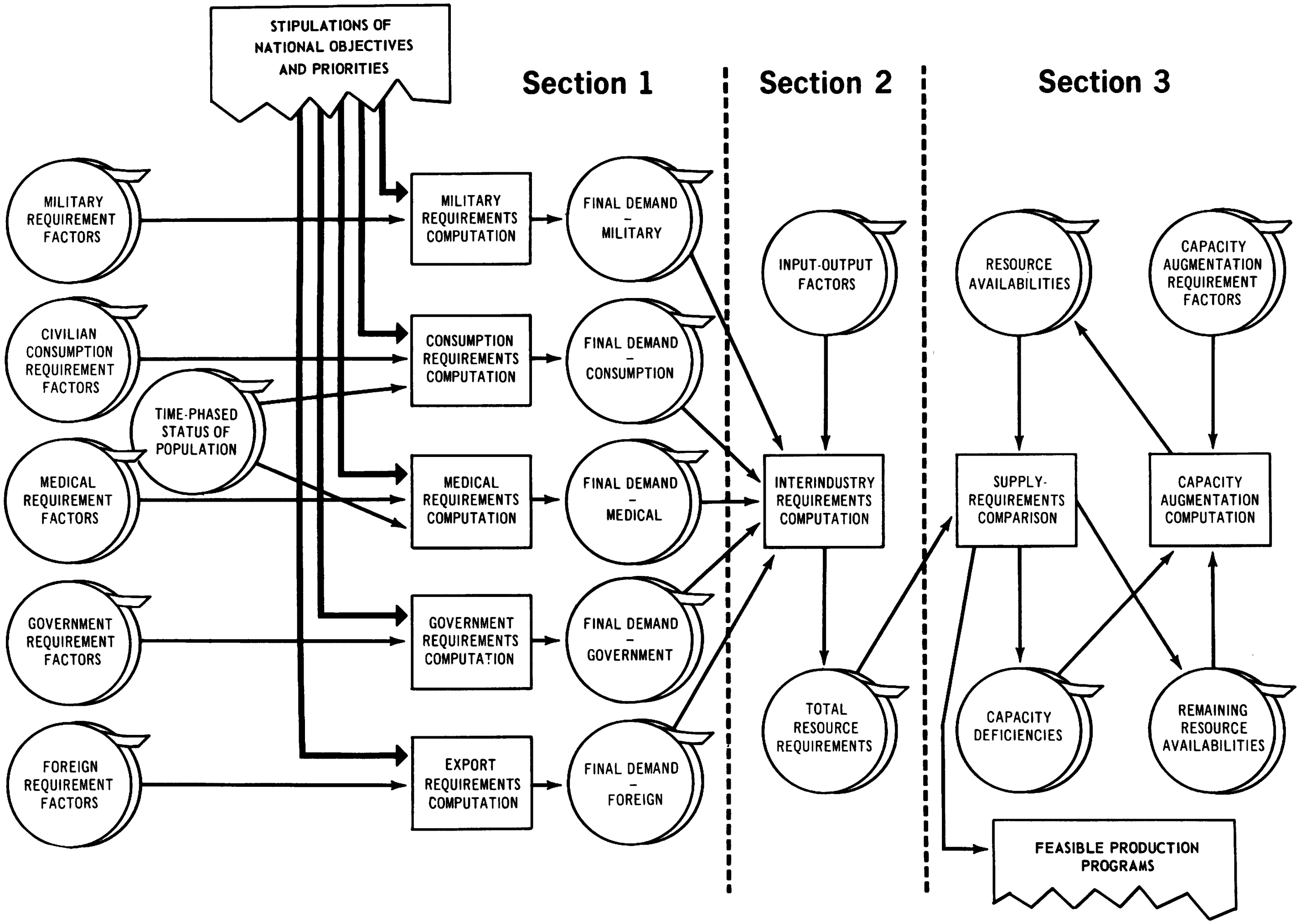
* Survival Item Analysis Program, NREC Technical Report No. 4 (Rev.), Nov. 1960

** Ready I (Third Draft), NREC Technical Report No. 24, July 1963

THE SURVIVAL MODEL



The PARM Model



problems, we mean difficulties which may arise in securing the various inputs required in production of survival items.

The superficial consideration afforded by the SURVIVAL model to these problems is simply that the model offers a choice of estimating production on the basis of either the assumption that there will be a continuing supply of all necessary inputs or the assumption that total post-attack production will be limited by the recorded pre-attack inventory of inputs. The SURVIVAL model makes no analysis of the availability of necessary inputs.

The principal strength of the SURVIVAL model is its power to examine product detail and geographic detail. Its principal weakness is the superficiality of its vertical analysis, that is, its lack of depth in the analysis of the availability of inputs required for production.

It is obvious from even this cursory description of the model that its operation depends upon a very extensive program of data development and derivation. Roughly 100,000 resource records are used in the SURVIVAL model. These include data on inventories, capacity, production and production "runout" (which means production achievable without replenishing plant supplies of input materials). Data for these survival item records were collected by the Census Bureau in consultation with the Business and Defense Services Administration of the Commerce Department, and the Departments of Interior and Health, Education and Welfare. The survival requirement factors were deduced by specialists in those departments from pertinent statistical records developed in their own statistical programs. The important manpower productivity factors used in SURVIVAL were developed by the Bureau of Labor Statistics.

Long-Range Problems of Recovery

Now we shall turn abruptly from survival resource problems to the consideration of recovery management tasks, which demand a very different time dimension and perspective in supporting analyses. In the short-term and primarily local analysis of supply-requirement balances for individual survival items, it is not necessary to give extensive consideration to the vertical and time-phased constraints that may be imposed upon resources used in production. But for long-range problems of recovery management program feasibility must be tested on a total, integrated economy basis, with full consideration of vertical interindustry relationships.

Realistic analysis of these long-term recovery problems calls for a dynamic economic model which will simulate the operation of an entire national economy -- with primary attention focused on the interdependencies and associated lead times between each industrial and service component of the economy and all others at each stage of production. Such a model must consider not only all the direct requirements for finished products; it must be able to compute the indirect requirements for all the resources required for production of such finished products; and it must determine when each increment of all these resources are needed in the production process. Finally, it must compare these total, time-phased resource requirements with time-phased resource availabilities.

The formulation of such an analytical requirement by NREC led to the development (under a research contract with the National Planning Association) of an economic model called PARM. The letters PARM stand for Program Analysis for Resource Management. PARM may be described as a dynamic analytical model, or system of analytical models capable of tracing the time-phased effects on the economy of various emergency situations over a two-year period. PARM can be used to simulate the effects on the economy of either real or hypothetical courses of action by changing data inputs and various conditions of the model. A highly simplified schematic representation of the model is presented in the second chart. PARM projects the impact of proposed programs, endeavors to "break" bottlenecks which would occur if preventive action is not taken, identifies those requiring intervention of program managers, and then spells out the impacts of alternative courses which they choose.

From the analyst's standpoint, there are three major sections of the PARM system. These three sections develop explicitly the economic programs to be tested, generate detailed requirements, and then compare requirements with availabilities under a variety of postattack conditions and adjustments.

The first section of PARM is concerned with developing detailed statements of final demand which are consistent with national objectives and priorities. It also converts them into the form necessary for evaluation by the model. These programs are developed separately for civilian consumers, military requirements, other governmental operations, and foreign trade on a month-by-month basis for a two-year period. A military requirements submodel generates detailed statement of requirements for military materiel on the basis of generalized statements of requirements.* A civilian consumption submodel automatically generates detailed and internally consistent estimates of requirements for surviving consumers based upon numbers of survivors and general stipulations. A medical submodel generates estimates of demand for emergency hospital and medical care by comparing time-phased casualties with surviving doctors and hospital beds and general stipulations. A government submodel estimates the detailed requirements of government operations based upon general decisions as to which agencies and functions would be emphasized. Routines also exist for introducing stipulated foreign trade programs and for estimating their resource requirements.

The second section of the model is an interindustry model -- that is, a model that examines simultaneously the input relationships or requirements among all industries and related services. These requirements are expressed as simple equations. Each industry is represented by an equation which defines the quantities of inputs required from other industries to produce one unit of its output. Through simultaneous solution of these equations an interindustry model determines the production necessary in all the supporting industries to achieve the finished products called for by the programs being tested. Throughout this process and in the section to follow, PARM takes account of all lead-time considerations.

The third, and most important, section of PARM automatically compares the specific program requirements with resource availabilities. This section of the PARM system also includes a variety of submodels and routines designed to simulate postattack conditions and decisions. The initial requirements are balanced against available stocks and are reduced to the extent that these can be met by stocks. Labor force casualties and constraints are computed for each plant or resource point. If available capacity exceeds the extent to which it can be operated by available labor, it is reduced to an effective capacity level. If capacity rather than labor is a bottleneck, then the model will automatically specify the additional capacity required. Also, the model will compute the requirements on other industries for this capacity augmentation and, if the necessary resources are available, the model will proceed under the assumption that the augmentation will be made. In computing the resource requirements of capacity augmentation the model will consider, in order, requirements for conversion of similar plants, decontamination of plants that are unavailable because of fallout, repair of damaged facilities, and finally through construction of new plant. The model keeps a running account of requirements and supplies by source. Finally, it compares these totals to determine the ultimate feasibility of achieving -- time-period by time-period -- production sufficient to meet stipulated requirements.

Against this very general account of the PARM structure, let me say a few words about the data now at hand for operating the system.

The PARM model in its prototype form makes use of approximately 100,000 resource records. Individually, these are less detailed than those developed for survival item analysis. But as a group they are much more comprehensive. The PARM simulation of our economy requires representation of all its significant components. The PARM resource file is therefore drawn from the basic NREC Resource Library, which is an integration of data carefully siphoned from numerous statistical collection streams. These resource data are reasonably current and are subject to ambitious improvement programs in which many governmental and private institutions are participating.

In addition to resource records the PARM system needs a very extensive collection of input-output coefficients and a large series of requirement factors. The PARM interindustry factor file is composed of some 94,000 input-output factors which describe requirements in both production and capital-formation processes. These coefficients have been deduced from many collections of data. These include industrial and capital coefficients derived from the Census of Manufactures and the Census of Business; manpower relationships developed from records of the Bureau of Labor Statistics; construction inputs from the Corps of Engineers and data from the Defense Department, AEC and NASA concerning military end-items, nuclear and space age activities. In far too many cases these input-output factors as now constituted are outdated. These will soon be replaced, however, by factors deduced through multiple regression analyses of data collected in recent economic census operations. PARM requirement factors representing the various elements of final demand also have been drawn and deduced from numerous source materials. These factors have been the subject of extensive cooperative research but are still subject to much further refinement. Elaborate documentation is now available on the data development aspects of the PARM system. Work continues in these areas with the active cooperation of several Federal agencies.

---

* This submodel is incomplete. It lacks military requirement factor details.

Let me conclude these remarks on the data problems and prospects associated with our PARM project by reference to the use of such resource management models in a real emergency. In a post-attack situation, we would face a massive task of data acquisition and reduction as soon as field survey activities commence. This information flow problem would continue throughout the recovery period. In order to minimize communication loads as well as data processing tasks, it would be necessary to use new and improved statistical procedures of many kinds. For example, various sampling techniques may provide feasible means for rapid updating of resource inventory records and for monitoring program performance. These statistical aspects of readiness for such a contingency pose many challenges and opportunities for the inquiring and innovating statistician.

Summary and Conclusion

We have reviewed briefly some of the analytical processes that are prerequisite to national resource management in an emergency. We have reviewed still more superficially some of the voluminous arrays of statistical data upon which these analytical processes depend. The job of selecting, adapting, and organizing the necessary statistics is often tedious. The realistic clarification of concepts and logical relationships which must be achieved in model formulation is often a painful process. But once the rigorous demands of mathematical model-building have been met and a complete system of computer instructions and input data are ready, the analyst and the decision-maker have at their service truly unprecedented capabilities for the orderly anticipation of experience and the testing of plans and programs. Through the sophisticated and discriminating application of statistics, these analytical systems can provide powerful tools for the management of resources in any emergency.

USDA SURVEYS FOR EMERGENCY PREPAREDNESS

Trienah Meyers, U.S. Department of Agriculture

I'm pleased to have this opportunity to tell you about some of our USDA studies in the area of civil defense.

One of the most important concerns of civil defense--after precautions to save as many people as possible from the immediate effects of nuclear attack--is to provide for the continued existence and economic recovery of our people in the ensuing days.

Foremost among these concerns is food. By Executive Order, the President has delegated to the USDA broad responsibility for planning for food and agriculture during an emergency. In order to carry out these broad responsibilities it is particularly important to make periodic assessment of food resources available from all sources in a National emergency. Realistic assessments of this type are vital to emergency planning for the management of food supplies and for estimating the need for food stockpiling both from the standpoint of the Nation as a whole and the various localities within it.

We all know that the U.S. has an abundance of food. But is our food supply so distributed as to be readily available to our people in an emergency? What would happen to people in Chicago, Detroit, New York, Washington, Los Angeles, or Miami? What would happen in Dubuque County, Iowa, or on the farms of downstate Illinois?

In our attempts to answer these questions we have published a number of civil defense studies done at the request of the Office of Civil Defense. A list is available for those of you who would like to obtain them. These studies include estimates of food supplies in households, in retail stores, and in wholesale warehouses, with special estimates of total food supplies by counties, and a report of fallout facilities and fuels on farms. We are presently working on estimates of food supplies in institutions and away-from-home eating establishments.

I know we are all interested in the findings of these studies, because they affect our possible survival both as a Nation and as individuals. But as you are also interested in techniques I'll discuss both.

First, I'd like to talk about our study of "Fallout Facilities and Fuels on Farms in 24 Central and Southern States." Information was collected in 1962 as a supplemental part of the Department of Agriculture's annual December survey of farmers. Personal interviews were obtained from a probability sample of about 3,000 farm operators in the 24 States of the North Central, South Atlantic and South Central regions. These States have over three-fourths of all U.S. farm-operator households and slightly more than 60 percent of the U.S. milk cow herd.

In these 24 States, 86 percent of the farm-operator families used gasoline on their farms. Forty-eight percent used LP gas for either household or production use; and 40 percent used diesel fuel, fuel oil or kerosene.

On the farms that use and store the various fuels storage capacity averaged just over one-sixth of annual use for gasoline and about one-third of annual use for LP gas and for diesel fuel, fuel oil and kerosene. Storage capacity as a percent of annual fuel use was quite uniform by regions for each fuel.

Nearly 60 percent of the farm-operator families in these 24 States had facilities of a type that provided some fallout protection, the most common being a basement or cellar under the house, with storm cellars ranking second.

Facilities providing protection of any type for farm families varied sharply by regions. About 85 percent of farm-operator families in the North Central States reported protection of some type, while fewer than one-third of those living in the South Atlantic and South Central States had any underground or specially designed fallout facilities.

Enclosed structures offering some protection were available for three-fourths of the milk cows, but again differed sharply by regions. Shelter was available for only about one-third of the milk cows in the four South Atlantic States, and about half in the South Central States, compared with nine-tenths of the milk cows in the North Central States. This interest in shelter for animals is both from our concern for food supply and for breeding stocks.

Our other civil defense studies are concerned with food in various distribution channels. We obtained data for the study among homemakers through a questionnaire administered by the Bureau of the Census in conjunction with the Current Population Survey of June 1962. Using a sample composed of 3 out of 8 rotation groups included in the CPS for that month, we secured over 11,000 schedules, mostly by personal interview.

Homemakers were asked to estimate the number of days the food currently on hand would last if all household members were at home all the time and were eating the kind of meals they usually eat. Then they were asked how many more days, if any, this food could be made to last if household members ate only enough to get by on. Estimates used in the report were obtained by combining the answers to these two questions. When making their "usual meal" estimates, homemakers tended to give stereotyped responses, such as 7 and 14 days. The combined estimates of the total number of days food could be stretched appeared to be less stereotyped, although there were peaks.

We decided to obtain the data in this manner because the results of an unpublished pilot study indicated it was a useful method. In the pilot study, time estimates were also computed for the same households by dividing the total calorie value of the food on hand, as inventoried by the interviewer, by the family's daily nutritional needs. While total days' supply as computed from inventory data was likely to be higher than that estimated by the homemaker, the figures derived from the two methods correlated fairly well. The inventory method, besides being costly and cumbersome, has other disadvantages. For example, the homemaker's ingenuity or lack of it in making the best use of food supplies on hand is not taken into account.

In our report, variations in food depletion patterns are presented by degree of urbanization, family income, civil defense region, such family characteristics as size, age of homemaker, and presence or absence of children, and by the day of the week the interview took place. Because of the preponderance of once-a-week shoppers and because the latter part of the week is the busiest time in most food stores, we thought the day of the week might have a pronounced effect on homemakers' estimates of how long their food on hand could last. However, it made little difference. Possibly, the stereotyped answer of number of days may have contributed to this. Or perhaps many households have enough food on hand at all times to minimize the influence of the day of the week on such estimates. They might run out of certain foods earlier or later in the week--but they do have a supply.

Our findings indicate that if an emergency should cut off outside food supplies, large numbers of American households would be in trouble after a few days. Homemakers grouped themselves roughly into thirds--those who would run out of food in about a week or less, those who could go more than a week but not more than two, and those who estimated their supply would last over 2 weeks.

Seven out of 10 downtown big city households would run through their food supplies in 2 weeks. This would be true for 6 out of 10 homes in suburbs, but only 4 out of 10 on farms.

Homes in the Northeast and Southeast would run out of food fastest. Those in the western North Central States and the Pacific Northwest could hold out the longest.

The lower the income, the faster food would run out. Where income was under $4,000, 38 percent of the homemakers said food on hand would last a week or less. In the $4,000-$6,000 range this figure drops to 29 percent. In the $10,000-and-up category, it falls to 22 percent.

Middle sized families--those with three, four and five members could make their foodstocks last longer than the very small or very large households.

Now let's turn to the retail food store survey. The basic data were obtained by mail questionnaires. The study was conducted during the spring of 1957 among a National sample of about 7,000 stores engaged primarily in selling food for off-premises consumption. Our sample was based on the Census Bureau's retail trade sample. It was composed of 2 of the 12 area sample panels plus a sample of large independent food stores and stores of large food chain organizations. Each establishment was asked to supply inventories of commodities, selected from a master list of 239 items. The number of retailers chosen to report for any one commodity was in direct proportion to the estimated importance of that commodity in total retail food sales. Data were obtained for both food and nonconcentrated beverages.

Three publications have resulted from this survey. One presents estimates of total and per capita man-days' supply of foods and beverages in retail food stores by the four census regions, States, and counties for 1957. To produce estimates of the number of man-days' supply on hand, the physical volume of food and beverages was first converted to caloric values. Next, we found the total calories available per person. Then we divided this by the usual daily per capita calorie requirement. A similar analysis in terms of fluid ounces was made for all nonconcentrated beverages. Individual county estimates of population from the publication, Sales Management, for May 10, 1957, were used since the Bureau of the Census publishes no county population figures except for census years. As a measure of verification, we compared aggregates of county estimates shown in Sales Management with midpoints between 1956 and 1957 census estimates of population by regions. The percentage differences were minor.

A second publication presents revisions of inventories by civil defense regions, States, and counties as of 1962, using population estimates in accordance with the 1960 Census of Population. This indicates that there is a 15.5 days' supply of food at the level of two thousand calories per day in inventories of retail food stores for each person in the continental United States and that slightly more than four-fifths of the total supply can be stored for relatively long periods without special handling. There is a 3.4 days' supply of nonconcentrated beverages.

A third publication presents the retail stores data in pounds for the continental United States. This was prepared because various groups dealing with specific commodities want to know the physical quantities of supplies available rather than caloric values.

So much for the retail phase.

We obtained data for the survey of wholesale establishments in 1962 through questionnaires mailed out by the Bureau of the Census through its regional offices. The National sample of approximately 5,000 wholesale establishments was selected from a universe of about 45,000 classified in the 1958 Census of Business in 6 categories. Establishments from each kind of business were selected on a probability basis proportionate to size, with 1958 inventories as the measure of size. A sample of firms entering business since 1958 was drawn from the records of firms obtaining new Federal Social Security employee identification numbers from the Bureau of Old Age and Survivors Insurance.

We minimized the burden of response by using the random-part sampling technique to establish inventory estimates. This technique was discussed in a paper given by Ralph Woodruff at your convention in September 1957. Each establishment was asked to furnish physical inventory information for sample commodities taken from a master list of 187 products and groups of products rather than to furnish data for all 187 products and groups of products. The probability of selection of each commodity varied by kind of business. The number of commodities selected varied from 6 to 150 for each establishment, depending largely on its size. As in the retail store survey, inventories were converted to caloric values to permit summarization on a uniform basis.

Data are shown for kind of business and for civil defense regions by storability type and by major food group. In addition, total and per capita supply (on a calorie basis) and man-days' supply of food and liquids are shown for standard metropolitan statistical areas. For each person in the United States there is a 16.1 days' supply of food and a 4 days' supply of beverages in warehouses at the wholesale level of distribution, where most food can be kept for a long period of time without refrigeration. A supplementary report presents the data in pounds for the benefit of members of the food trade. Inventory data are listed for major food groups by wholesaler's kind of business and type of operation and for individual food products.

The last publication I'd like to call to your attention is the most comprehensive of all. It utilizes data from the household, retail store and wholesale studies already mentioned. It also draws on information from other sources, such as independent surveys of public cold storage warehouses and estimates by commodity experts within the Department of Agriculture of stocks at food processing plants based on surveys and mandatory reports. The report was prepared in response to a request from the Office of Civil Defense to assess total available supplies of food and their geographic distribution, with particular emphasis on household, retail, and local wholesale stocks. Food was defined as that ready for household or restaurant use. For example, meat was defined as food whereas live meat animals were not.

The report presents figures for man-days of food stocks available by source, the number of days home-retail-wholesale food stocks will last, the number of days all food stocks will last, man-days of food that can be made available by 30 days' production and by stocks plus production and finally, the number of days stocks plus production will last.

This report estimates maximum and minimum food supplies as of 1963 by county, State and civil defense region. Alaska, Hawaii, and offshore possessions were excluded from the analysis. Differences between maximum and minimum supplies are accounted for by the fact that inventories in food processing and cold storage plants are relatively high in January and low in July as well as by January-July differences in food production. Food inventories in homes, retail food stores, and wholesale warehouses were estimated to remain about the same from month to month in terms of total food value.

To make the report more meaningful we adopted certain assumptions. We assumed that enemy attack might seriously disrupt transportation of food and other agricultural commodities between counties, States, and regions for up to 90 days. We assumed further that our people could survive without serious health hazard at the level of 2,000 food calories per person per day--slightly under two-thirds the usual average diet. We made no provision for assuring a balanced diet.

In communities not severely damaged but cut off from outside supplies, such as fuel or feed for livestock, we could count on some local food production. We assumed that for the first 90 days local food production might be continued in substantially undamaged communities at about one-third the usual rate, without inshipments of additional fuel, feed, or raw materials. This explains our use of the 30 days production category which I mentioned a moment ago.

We made these assumptions for the purpose of the report only. They should not be considered as our appraisal of what would likely occur in a National emergency.

The number of man-days' supply of food other than in homes was calculated by dividing the food availability data--converted to calories as a common unit for all foods--by the specified 2,000 calorie level. For home food supplies, we used the median estimates of homemakers as to how long available food supplies could be stretched in an emergency.

The procedures for converting data into county units varied somewhat by source of food stocks. For example, food stocks in homes and in retail food stores were computed for individual counties by apportioning United States and regional inventories on the basis of population estimates. Food stocks at the wholesale level were computed for individual counties by distributing United States inventories in proportion to the number and estimated size of wholesale warehouses physically located in the counties. This method was used because food stocks at wholesale are not necessarily related to size of population.

In addition, our USDA commodity experts pooled their knowledge and provided estimates of the stocks of various kinds of food on hand in food processing plants and other sources before the wholesale level throughout the country. We then allocated these estimates to individual counties, except where county data were already available, on the basis of the number of employees by type of establishment.

Let me stress that the data in the report are not precise. They are intended to give a general idea of where available food supplies might be large, and where they might be small, in a National emergency.

Now for the findings. On the basis of January food inventories and food production levels (when food supplies are at a maximum in most counties) 7 States, the District of Columbia, and over 40 percent of the counties in the U.S. would have less than 90 days' food supplies at the 2,000 calorie level.

The major deficits would occur in parts of New England; the Washington, D.C. trading area; the Pittsburgh, Cleveland, Dayton and Detroit trading areas in the Midwest; the central Appalachian region; the Miami and Mobile trading areas in the South; and the Los Angeles and San Diego areas in the Far West. Some of the less densely populated parts of the country also would face deficits--because less than 90 days' supply of food would be available from local stocks and local food production in an emergency.

As you would expect, the food stocks in processing plants and cold storage warehouses reveal some imbalances. In certain trading areas in the Southeast and Southwest, for example, peanuts in shelling and manufacturing plants constitute a large element in the total stocks structure. In parts of Maine, potatoes bulk heavily in the food supply. Dry edible beans are a considerable part of total stocks in certain areas of Michigan and in other dry bean producing States. Dairy products account for much of the total supply in some areas, and fats and oils tend to dominate in others. However, food stocks in most large urban centers appeared to be in fairly good balance, except that seaport cities tended to have large supplies of sugar.

But what's the overall picture for the U.S.--still in January, when food stocks are at a maximum? Our report shows that, under the conditions assumed, all food stocks could be made to last for roughly four months, if distributed equally. The amount of food available in homes would last a little less than two weeks, on the average. Retail food stores could provide about a two-weeks' supply, and wholesale establishments, it is estimated, contain about another two weeks' supply. The remainder of the food supply surveyed (about two-thirds of the total, or 82.6 man-days) is located in distribution channels before the wholesale level, and would, therefore, be more difficult to distribute equitably.

If we consider the 90-day period to be realistic --that this much time might be needed for the food industry to recover from an attack and to resume the supply of food in quantities approaching normal--then we have a problem. Obviously, the amount of emergency food supplies readily available to the average household--that is, available in the home and in local retail and wholesale warehouses--is inadequate.

Since retail and wholesale inventories are generally fixed and subject to more or less rigid controls, it appears that the more logical place to increase the emergency food larder would be in the home. While civil defense authorities thus far have not had much success generally in getting consumers to stockpile additional food in the home, such an increase could conceivably be attained through additional educational and promotional activities directed to the housewife. Such an undertaking if successful would yield two important benefits. The American household would be better able to meet the immediate emergency situation through more bountiful rations and have a more balanced diet if the inventory were purposeful--rather than just what happened to be on hand.

Bibliography:

Estimated Number of Days' Supply of Food and Beverages in Retail Stores, 1962. Marketing Research Report No. 577, Dec. 1962.
Estimated Number of Days' Supply of Food and Beverages in Warehouses at Wholesale, 1963. Marketing Research Report No. 632, Oct. 1963.
Fallout Facilities and Fuels on Farms in 24 Central and Southern States. Statistical Reporting Service-3, 1963.
Food Supplies Available by Counties in Case of a National Emergency. Agricultural Economic Report No. 57, July 1964.
Homemakers' Estimates of How Long Food on Hand Could be Made to Last. Marketing Research Report No. 669, July 1964.
Inventory of Food Products and Beverages in Retail Food Stores. Supplement to Marketing Research Report No. 286, April 1960.

Discussion

Harry B. Sheftel, U. S. Office of Statistical Standards

It is my understanding that the function of the OCD is planning for coordinated action to protect the population in periods of emergency using governmental structures adapted to the emergency. Civil defense aims at the preservation of life and the restoration of the postattack economy.

It is also my understanding that the statistical programs of OCD aim at developing coordinated plans to achieve postattack solutions, to develop plans that will increase chances of population and industrial survival.

It is encouraging to note the variety and depth of the statistical studies presented. The mass handling of data is heavily involved. There are also administrative, legal, production, military and scientific data to be collected, understood and used. There are human factors involved in cooperation with State and local governments and with industry. There is need for managerial efficiency and for coordination.

I am not certain whether some of the following remarks are within the scope of the papers or the session. But I am taking advantage of a discussant's prerogatives and shall probably offer more questions than answers. Since we do not have all the answers, I trust that questions may be useful in provoking light by discussions, rather than heat by provocation.

There is insufficient time to discuss the technicalities of the papers presented in any detail. Moreover, I am more interested in matters of program and statistical coordination. I do have a few comments, or questions.

Romm Paper

It has been said that the fallout shelter program is the best single way to save the maximum number of lives and the program described has located a great many public fallout shelter spaces. But discussion of some added factors would be helpful to our understanding. How many more spaces are needed. Are there any scientific possibilities for shelter against blast and fire. Do public shelters complete the program. What about home shelters and the use of space in schools and hospitals. Statistics available in these areas would round out the discussion.

The review of damage assessment and vulnerability analysis again illustrates the difficulties and the wide range of military, geographic, weather, and scientific information needed for the various models... JUMBO, ERD, DUSTY, FLAME, RISK, etc. I would appreciate some evaluation of these models, to understand how they are put together and what the gaps and problems may be. What part does the maintenance and improvement of these data play in the overall OCD statistical sphere. How are decisions made based on the output of these models.

Coker Paper

The second paper illustrates the use of computer systems to develop concepts of inter-relationships between various elements in the economy, and provides some idea as to the method for solving postattack resources management problems.

The Survival model is intended to provide supply-requirements data by geographic areas on the kinds of items that would be desperately needed. For my better understanding, I would ask how up-to-date do the inputs have to be, what are the statistical processes that would account for constant shifts in population, stocks, facilities and the shelf life of goods. How are these complex factors kept in view so that the model remains alive and alert.

The outline of PARM has been well presented. It is used to simulate the effects on the economy of either real or hypothetical actions. This is a massive task. Recognizing the difficulties that faced PARM, and the high order of capability that went into its development, it is still pertinent to ask: What kinds of statistical problems have been encountered in the development of PARM. What kinds of techniques have been used. Are there any problems remaining. What are the nature and extent of any gaps. Have there been any test runs of PARM. What were the results.

Myers Paper

The third study shows how one agency outside of OCD is carrying out its civil defense responsibilities. It has been well stated that the problem centers on the geographic distribution of the available food in relation to people. The Department of Agriculture has been most active in assessing available food resources, and in providing indications of the amounts and locations of food that might be expected postattack.

I note that the data produced by these surveys have resulted in some overall conclusions, and that we have some basis for food stockpiling as well as a basis for special food plans in areas where food reserves may be low.

Many facts have been presented, but what is the status of the program. Is it complete. Some added words would be helpful regarding any further specific programs to further assure adequate supplies and the kinds of statistics needed to implement these solutions. How-up-to date are the data. A survival model has been discussed in a previous paper. To what extent has the food situation been brought together with the other items in a survival plan. Is food a resource under the PARM system.

Other Programs

I would like to mention a phase of civil defense statistics and research not discussed here, public opinions surveys aimed at investigating public attitudes towards civil defense activities. OCD has been active in this area. I can understand that OCD finds it necessary to engage in educational activities, and to assess public opinion on many aspects of its work. It is possible to scatter one's resources in the name of research. I wonder if it would be advantageous for OCD to consult more widely, within its own organization as well as both in and out of Government to establish a technically sound public opinion research program geared to the specific objectives of civil defense statistics.

There are other important statistical series not discussed at this session that merit more than passing mention. For instance, the construction equipment area, the warning system, the executive reserve, utilities, water systems, power systems, telephones, communications, the whole area of industrial defense and the like. Some or all of these factors may be included in the models discussed, I'm not certain. However, these areas also appear to play a part in the civil defense statistics program.

Aside

Now as a sort of aside, let me put what I'm trying to say regarding OCD programs in another way. Civil defense exists to preserve life and industry. Then the OCD programs and the statistical subject matter that ties to civil defense objectives should be understood easily by the citizen who is to be saved. At the risk of opprobrium by philosophers, I don't think civil defense and its statistical formulations should be as obscure as philosophical explanations.

I had read about Sartre so I wanted to know what existentialism was. I selected a book supposed to define existentialism. I read about phenomenology, ontology, solipsistic, etc., each an important component, but I never did find out what existentialism was about. If the complex details of the material presented here, the forest and the trees, were readily understood in relation to everyday life, I wouldn't be raising so many questions about phenomenology, ontology and existentialism. End of aside.

Summary

It is evident that no single model or set of statistics is sufficient unto itself to provide for all of OCD statistical needs. A key word is coordination. The papers presented, excellent as they are, give us a picture of many sets of important large scale discrete areas. How are these data put together to fashion an overall program. What are the statistical goals and how far along is OCD in meeting these goals. What added data are needed to complete evaluation, program and progress. How are these statistics used in furtherance of the OCD mission.

The surveys and models did not spring full bloom into existence; nor will the data derived from many separate series automatically bloom together into a coordinated whole. I have raised many questions, and would like to raise a final question. What is the mechanism by which all these data are coordinated, analyzed, interpreted and funneled up to a decision-making process. To what extent are the data, the statistics, the models, ready to operate as a system if there should be an emergency tomorrow.

MINUTES OF ANNUAL MEETING OF THE SOCIAL STATISTICS SECTION

Chicago, Illinois, December 29, 1964

Over 50 members attended the meeting and heard representatives make their reports.

First, the following resolution on race-color designation submitted by the Social Statistics Section was presented.

"Race-Color" classifications are important in the statistical classification of the population and of events, such as births and deaths, that affect the composition of the population. They are essential in the development of statistical information as a basis for public health and social programs, for scientific research, and for public policy determination and administration.

We agree completely with the purpose of protecting individuals from invidious labels in documents used for personal identification and public record, and emphasize that this objective can be met without the impairment of the use of information, such as "race" or "color," for statistical purposes.

Without in any way approving of social distinctions based on race or accepting any particular notion or definition of "race," we must, as social scientists, take cognizance of the fact that groups are set apart in this way and that they have differing economic, social, and demographic conditions of great significance for them and for their future development. Without basic information concerning these groups, social scientists are gravely handicapped in gaining an understanding of the changing group structure of our society and responsible leaders are left to base intergroup policy on rumor and ignorance rather than knowledge.

We, therefore, join with other social science and statistical groups in urging that "race" and "color" data continue to be included in vital statistics, census and other public records which are used to provide data for statistical and scientific purposes. It is not the controversial or noncontroversial character of a social phenomenon, nor its approval or disapproval, but rather its social importance that determines the necessity of obtaining information about it.

Second, Edwin Goldfield reported that the distribution and sale of the Proceedings of the Social Statistics Section have approached the break even point.

Third, Eli Marks presented for approval, amendments to the charter of this Section to read as follows:

AMENDMENTS

Under "Organization" - Revise third paragraph to read:

The officers of the Section shall consist of a Chairman, Chairman-Elect, Vice-Chairman and Secretary. The term of office for the Chairman and Chairman-Elect shall be for one year, and for Vice-Chairman and Secretary two years, and for all officers until newly elected officers have been qualified. No member shall be eligible for re-election to the same office for more than two consecutive years, except that this provision does not apply to the position of Secretary. Election shall be by majority vote of members of the Section by mail ballot, subject to approval of procedures by the Board of Directors and Council of the Association. The management of the affairs of the Section between annual meetings shall be entrusted to a Section Committee composed of the officers, the Section's representatives on the Board of Directors and Council of the Association, the Program Chairman and the Editor of the Proceedings of the Social Statistics Section.

Between elections, the Chairman shall designate persons to fill any vacancies in the officers or the Section representatives on the Council and the Board of Directors of the Association, and shall designate alternates to serve as Section representatives to the Board of Directors and Council in the absence of the regularly elected representatives. The Section, or the Section Committee, may establish such subcommittees as they wish to carry out the functions of the Section.

Amendments

This Charter may be amended by the Section with the approval of the Council.

The meeting adjourned at 9:00 A.M.

Respectfully submitted,

Philip C. Sagi, Secretary
Social Statistics Section
American Statistical Association

1964 Officers of the Social Statistics Section

Chairman:	Leslie Kish
Chairman-Elect:	Eli S. Marks
Vice-Chairmen:	Daniel O. Price (1963-64) Jacob J. Feldman (1964-65)
Secretary:	Philip C. Sagi (1964-65)
Proceedings Editor:	Edwin D. Goldfield