I

STATISTICS IN CRIMINOLOGY

Chairman, MARVIN E. WOLFGANG, University of Pennsylvania

PLATO'S r

A NEW FORMULA BASED ON OLD PRINCIPLES

Tadeusz Grygier, University of Toronto

Philosophy and mathematics were always closely related, from Euclid and Pythagoras ("all things are numbers"), through Plato's theory of ideas (all eternal and real things are either ideas or numbers), to Whitehead and Russell's Principia Mathematica (27), and to Leibnitz and the Warsaw School of mathematical logic of Lukasiewicz, Lesniewski and Tarski (29,30).

And yet this relationship is often forgotten in modern science and technology, especially among statisticians, albeit to forget one's relatives is neither gracious nor - if I may be cynical - practical. Maintaining such contact may bring ample rewards: in families through the inheritance of money, in science through the borrowing of old, and the creation of new, ideas. As will be demonstrated, some of the old ideas can lead to most unexpected practical consequences.

The purpose of this paper is to present an extremely simple formula for product-moment correlations which finds its justification in Plato's theory of ideas. As with platonic love, Plato's *r* is more practical than at first appears: both save a great deal of trouble and labour, and help to avoid costly errors, which might otherwise occur despite precautions and technical equipment. Plato's *r* requires no equipment and, even if the "Procrustean Table" (Appendix A) is lost or unsuitable, a new one can be easily constructed once its principle is clearly perceived - and accepted.

According to Plato's theory, most clearly stated in the allegory of the cave at the beginning of Book VII of The Republic (19), we can never perceive reality itself but only its shadow. We are like prisoners in a cave who "lie from their childhood, their legs and necks in chains, so that they stay where they are and look only in front of them". A fire is burning behind them and they can see nothing of themselves or of each other "except the shadows thrown by the fire on the wall of the cave" (19, p.207). Plato anticipated, centuries before Freud, the mechanism of projection not only as a distortion of reality but as the very essence of perception.

As I have discussed elsewhere (11), I believe the point made by Plato to be as valid now, in the light of scientific methodology and of new evidence in social psychology, as it was in Ancient Greece. The scientist never attempts to perceive the total reality around him.* As Wojciechowski says, he "consciously chooses as his formal object a particular aspect of material reality and purposely ignores other aspects which he deems, and rightly so, unfit for his methods and means of investigation. Thus his point of departure, far from being a universal one, is carefully selected and limited" (33, p.30). If he wants to compare his results with those of other scientists he is careful to adopt the same methodological framework; then his conclusions, while never attaining universality, are not entirely individual and subjective. The restriction of his conceptual apparatus is deliberate (*cf*. Ajdukiewicz, 1, p.186), while a layman's is unorganized and unconscious. He is never faced with a heap of factual material and with the task of encompassing it by a theory; on the contrary, he has to hunt for a conceptual apparatus, "for it is this alone which can give rise to empirical sentences" (*ibid*.). And there is no conceptual apparatus without the neurological one.*

The layman's picture of reality is as structured as that of the scientist, but the structure is more complex and is not explicitly stated; it is based on the physical and physiological state of the organism, on the past history of its adaptation and on the expectations built on previous experience. The scientist, as much as the layman, is imprisoned by his organism and experience, but

---

*He often perceives nothing but a single aspect of the complex reality around him and treats it as if it represented the ultimum knowledge. Philosophy often claims to be more universal than science, but what the philosopher chooses as representing all nature also depends on his outlook. A materialist declares that mind is nothing but matter, while a phenomenalist sees matter as mere clusters of sensations, which are psychological phenomena. A Marxist regards morality and character as the outcome of economic processes, while an ethical skeptic sees good and evil as mere projections of likes and dislikes. For a fuller list of "pan-scientific radicals" and their views see Feigl (9).

*As Caws says (7, p.14): "All our knowledge has to be expressed in conceptual terms; we can know nothing intelligently about what is external except as it is mediated to us by the neurological apparatus which originally informs us of its existence. We have no sure way of telling whether the logic, which exhibits itself in every department of enquiry, has its root in that unconscious faculty of man which is interposed between what is delivered to his senses and what is received cognitively, or whether it is, in fact, an ontological characteristic of nature. Whatever may be the truth of this matter, one thing is certain: it is inescapable. It is certainly a condition of our thought, whether as a characteristic of our minds or as a characteristic of a world of which our minds are part, and consequently it is to be found equally in the systems invented by us and in those presented to us."

the layman, unlike the scientist, is not conscious of his lack of freedom from perceptual restrictions. Experimental psychology has shown how much projection there is in any perception, especially if the stimulus itself is relatively unstructured (such as an ink blot). Social psychology has confirmed that prejudice, which feeds on projection, also increases with ambiguity of the stimulus. The lesson that we can draw from both Plato and modern psychology is that, if we are to be prisoners, we should face this fact and have proper bars, so that our data will be neatly arranged for efficient manipulation. Instead of treating our data with reverence as if they were pieces of unattainable reality and then using statistics which could, only approximately, do them justice, we can devise the simplest possible formulae and then impose an order on our observations, so that they would fit our formulae exactly. Not only can this be done; it has been done, and the evidence so far shows that validity (i.e. approximation of assessment to the reality inferred from other observations) is not lost but probably increased. For this reason Q-sorts have, generally, a prearranged distribution which they impose on the sorter.

The formula I suggest is yet another step in the direction of simplicity, speed and accuracy.

The purpose for which the new formula was devised was, originally, very limited. In an investigation in Ontario training schools we wanted to correlate a number of sociometric scores with each other and with behaviour ratings by the staff. The sociometric scores reflected each boy's number of friends (this was the Preference score), his number of enemies (Rejection score), and two derivatives of Preference and Rejection scores, namely Acceptance score (calculated as the difference between the number of preferences and the number of rejections) and Emotional Response score (being the sum of preferences and rejections). Since raw sociometric scores depended on the number of subjects in the group - the larger the group the larger was the possible number of friends and enemies alike - we needed a prearranged distribution which would allow comparisons of relationships from group to group. We also wanted to impose a uniform quantitive frame onto the observers, in this case the staff of the training schools. Finally, we wanted to simplify and speed up the calculations of the correlation coefficient, so that they could be carried out by students with a non-mathematical background (in social work) with a minimum of error and at maximum speed.

First we had to arrange a distribution of data suitable for the calculation of product-moment correlation coefficients. This implied a quasi-normal distribution. The distribution had to be such that it could be used easily by untrained raters; therefore it had to fit their average frame of reference. Since we usually perceive objects as possessing a given quality in a range from very marked through marked and medium to markedly and conspicuously absent or weak, a five-point scale distribution was adopted. This is usual, although seven-point and even nine-point distributions are quite common, especially in Q-sorts. In the distribution we adopted the mean has a value of 3. Values 2, 3 and 4 have class intervals of 1 standard deviation. Values 1 and 5 lie outside these limits. This type of distribution is more satisfactory to the rater than that in which the middle value ranges from one standard deviation below to one standard deviation above the mean observation. Forced normal distributions for groups of varying sizes, from 12 to 100, are arranged in the "Procrustean Table to Stretch Data on". The last column of the Table will be explained later.

For a seven-point scale, values 2, 3, 4, 5 and 6 would have a class interval of two-thirds of the standard deviation; values 1 and 7 would fall outside these limits and the mean would be 4.0. The frequency distribution of values would be such that 5% of the total population would have a rating, score or other measure valued at one, 11% at two, 21% at three, 26% at four, 21% again at five, 11% at six and 5% at seven. The principle is exactly the same, whatever the number of classes, provided it is an odd number.

A standard distribution in which the mean is always the same and always a whole number and all deviations from the mean are also whole numbers allows the calculation of standard deviations and of cross-products of actual deviations from the mean to be simplified. This is usually done, but the new formula goes one step further than is customary: standard deviation is dispensed with altogether. With a mean of 3 and the distributions as shown in the Procrustean Table, the usual formula for product-moment correlations can be reduced to the sum of cross-products of deviations from 3 divided by the sum of all deviations squared. Moreover, the sum of deviations squared is constant for all groups of a given size. It can therefore become a part of the Procrustean Table instead of being calculated from the data; hence the last column of the Procrustean Table.

The derivation of the new formula is very simple. The usual formula for product-moment correlations runs like this:

$$r = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{N\sigma_x\sigma_y}$$

In our case both means are 3 and both sigmas identical; therefore we have:

$$r = \frac{\sum(X-3)(Y-3)}{N\sigma^2}$$

Since sigma square is the sum of deviations squared divided by N, what remains in the denominator is the sum of deviations squared. The formula thus becomes:

$$r = \frac{\sum(X-3)(Y-3)}{\sum(X-3)^2}$$

The numerator can be calculated in one's head and the denominator read off the Procrustean Table.

An example, taken from actual data in one of the training schools in Ontario, is reproduced in Appendix B. At the bottom of the sheet is given the formula for calculating Plato's $\underline{r}$. As can be seen, the number of possibilities for cross-products of deviations is very limited indeed. Whenever the score is 3 (which, of course, often happens), there is no deviation. Thus, whenever one of the pair of scores is 3 (which happens even more often), the cross-product is 0. When both scores are 2 or 4, the cross-product is 1; when one of them is 2 and the other 4, the cross-product is -1. The maximum value for a cross-product is 4 (for both scores of 1 or both of 5) or -4 (for one score of 1 and one of 5). The sum of cross-products can, therefore, be kept in mind as one is perusing the data. Logarithmic tables and desk calculators are unnecessary. In the example given no aids at all are needed, but usually a slide-rule is handy. Since no square roots are involved and only one division (in the final instance), the new formula not only saves time and reduces human errors, but also reduces the errors inherent in the process of division and of taking square roots of any but perfect squares. It also saves time. With the help of the Procrustean Table a person of average ability in arithmetic can calculate the correlation coefficient given in the example within some 20 seconds. A less convenient distribution of data, where it is necessary to use a slide rule, would add another 10 seconds. The last two columns given in the example need not be written down; the original values (in our example, the sociometric scores) suffice, and the rest can be calculated in one's head. The formula suggested can be applied to all five-point scales approximating a normal distribution; it can be easily modified for other scales.

The objection which could be raised against the method adopted here is essentially the same as that frequently used against all statistical procedures: that they distort reality. But what is reality? Russell implies that we cannot know it without inference. As he says, "real" things are not just those that cause sensations; they also "have correlations of the sort that constitute physical objects". Consequently, "A thing is said to be 'real' or to 'occur' when it fits into a context of such correlations" (24, p.185). My contention is similar: the distortion of reality begins at the level of observation, not statistical manipulation. What we handle in statistical procedures are not pieces of reality made into numerical abstractions, but pieces of observation already abstracted. Moreover, as Wojciechowski points out, both objective and subjective factors contribute to the production of number measures. "The cognitive structure, with its definite mode of knowing, sets its own demands, which are conditions <u>sine qua non</u> of knowability and intelligibility of number measures" (34,p.98). Evidence from psychology shows that the cognitive structure sets its demands and causes perceptions to be merely abstractions, irrespective of the use of number measures. Words, which we use as a vehicle of thought as well as a means of communication, are also abstractions; we are more used to words than to numbers, but the latter are more amenable to manipulation.

Patently true of the social sciences, this is also true of physics. Again, Russell supports this claim: "All that physics gives us is certain equations giving abstract properties of their changes. But as to what it is that changes and what it changes from and to - as to this, physics is silent" (26, p.224).

Since Locke and Nicod observation has been the basis for arriving at truth through induction. Nicod regarded simple enumeration as sufficient basis for inductive reasoning (13). Russell (25, p.58) and Whitehead (31, p.5) suggest that we should start from induction, proceed to deduction and then again to induction. Most scientists would regard their mode of operation as both inductive and deductive, and with this view I agree, although I regard the very distinction between inductive and deductive reasoning as somewhat arbitrary and misleading.

Both induction and deduction are merely aspects of ordering observations and concepts. The rules of deduction have been said to provide a regulative mechanism which enforces a consistent language and enables us to express in one form precisely what we have already said in another (Feigl, 9, p.18). The appearance of novelty in deductive inference is only psychological, as it provides sudden insights into the implications of the original set of premises (<u>ibid</u>.).

According to Popper (21), induction is no better; it has no place in scientific reasoning. Indeed, as was well known to Hume, induction by simple enumeration is never a valid form of argument. All statements reached by induction can only be treated as hypotheses which, in the words of Popper, are "tentative for ever". It is only in terms of probability - and therefore of statistics - that hypotheses can be proved or disproved and, as Sellars (28; <u>cf</u>. also Grygier, 10) says, induction is vindicated.

However, <u>pure</u> induction can never be vindicated, not even in probabilistic terms, because it never occurs in practice. The inductive-deductive process does not begin at the point of a scientific experiment or a philosophical discussion. All concept formation - which, as Vigotski (12) and Luria (16) have demonstrated, can be unconscious and entirely non-verbal - shows a mixture of analytic and synthetic processes. Experiments on perception and remembering, for instance by F. C. Bartlett in England (3,4), show the remarkable degree to which our perceptions are affected by previous experience (induction) and the conceptual frame of reference (which surely must imply deduction). More recently Quine has advanced the view that the dichotomy between synthetic and analytic statements is purely conceptual (22), while Wojciechowski (34) implies that the difference between sense data and measurements is also conceptual, and to oppose one way of knowing to the other is to oversimplify the issue. So, even if we agree that all ideas originate in the senses, we must also admit that everything our senses tell us is affected by our ideas. As Polanyi says (20), we learn to perceive only as we develop concepts about the things we experience.

An additional argument comes from the theory of relativity. If we accept its premises, all objects and events in space-time have to be determined in relation to other objects, including the observer. The observer is a part of the total system of relations; he can never be excluded.* Thus he adds to what Polanyi (20) calls "the unaccountable element in science" (the act of personal judgment in the scientific process), which is also subjective and also unavoidable. According to Polanyi every scientific process involves judgment and every act of judgment involves a personal decision; it is not subject to any rules as we cannot have rules to prescribe how judgment is to be applied.

This paper is not attempting to solve a problem which, in principle, is not capable of solution. Subjectivity will always be with us. But we can introduce some rules which can reduce the subjective element in the recording of observations, some order which, even if arbitrarily imposed, can be constant from case to case. Statistical tables, graphs, coefficients of correlation, measures of statistical significance, all help in the task of scientific understanding. As Wojciechowski says, counting is the act of knowing through the medium of appropriate numbers (34, p.90). Anything which facilitates counting makes a contribution to knowledge. Anything which helps to order scientific data makes a similar contribution.

A forced distribution, normal or otherwise, imposes an order on the observations to be made and recorded by the observer, an order which makes him quantitatively comparable to other observers, or to himself at different times. It is akin to the method of hypothesis in scientific enquiry in that it creates a directed process with rational supervision, instead of simple induction, in my view impossible in any case. A hypothesis, according to Barker (2), still makes use of induction, but directs observations towards a workable conclusion. A forced distribution has a similar function.

Any observation, whether expressed in terms of numbers or of statements of quality, can have only an historical significance unless it is reliable. If it refers to a reality which can be assumed to remain basically the same, irrespective of the lapse of time or a shift in the point of observation, it should remain constant, both qualitatively and quantitatively. It is well known that training increases the ability to make observations, especially systematic, reliable observations. Trained anthropologists, psychiatrists and psychologists are known to observe with greater precision. There is sufficient evidence to assume that people trained in the same manner tend to make and record their observations with greater consistency. In that sense their observations tend to be more reliable and, consequently, have more chance of being valid. By training I mean the enforcement - be it through didactic methods, discussions, or conscious and unconscious imitation - of a set of rules.

An important difference between, say, rules of social work, psychiatric or other clinical observation and the instruction to adopt a forced distribution of data is that statistical rules are simpler, more explicit and more precisely defined.* Both sets of rules result in increased reliability of the observations, especially if we define reliability in terms of correlations between the observations made of the same assumed piece of reality by different observers. If different observers adopt entirely different frames of reference it becomes impossible to learn anything about the reality that they are assumed to observe.

If John Smith tells us that he found only 1 in 20 people to be good we remain no better informed about the goodness of people whom John Smith had observed, let alone about the goodness of people in general; but we have learned something about John Smith and the criterion of goodness he employs. If for "goodness" we substitute crime or delinquency the situation will remain the same; if John Smith tells us that 1 in 20 people is criminal or delinquent we can see something of his quantitative perception of the world around him, but we are no better informed about the world.

A most recent example of variation in quantitative frame of reference was given at the 5th International Congress of Criminology in Montreal. Paul F. C. Mueller (18) examined 250 first admissions to a correctional institution and employed two actuarial methods and the clinical judgments of five correctional counsellors to predict parole outcomes. He found that, while actuarial predictions were normally distributed, the distribution of clinical judgments varied from counsellor to counsellor. Many counsellors were unduly pessimistic in their estimates: they ascribed to the offenders a lower probability of success than

*Sir Cyril Burt (6, 235-237) regards tables of measurement, correlations, factor-saturations and the like as comprising a series of mutually equivalent matrices which enumerate "only relations between qualities and not the amounts of those qualities by themselves: just as the coordinates of space and time can only state the position of a star in regard to some other object or observer, and never its absolute position in the universe".

* Thus a sample is generally defined as "a conveniently small portion drawn from a batch to judge the quality of the batch in a certain predesignated respect" (2, p.474). This description is similar to that of a symptom of a disease, but the rules of statistics are more precise and explicit than the rules of medicine.

On the use of oversimplified theories in science, particularly in the "exact" sciences, which make extensive use of mathematical analysis, see Williams (32). He regards simplification as one of the major techniques of science and shows the advantages of a simple "theory of games" in dealing with serious and complex problems.

could be justified on the basis of past experience. Others were so conservative that they predicted a 50% or slightly below chance of success for the majority of the offenders they assessed. Each counsellor saw the offenders, quantitatively, in his own unique way. His observations seemed to have a closer relation to himself as the observer than to the men he was supposed to describe. It is likely that an imposition of the same numerical framework, i.e. a forced distribution, would have increased the interpreter's reliability to the point that at least some validity of clinical judgment would become possible, though by no means certain.*

The only harm done would have been to the observer's self-image. The more we project ourselves onto the reality around us, like Plato's prisoners in the cave, the more we feel that we know intimately what we are trying to observe. What we see then is close to our hearts and therefore regarded as real.** But, as Russell remarks, it is safer to assume that reality exists if it can be consistently perceived by more than one observer (26, p.225). Purely subjective observations are more likely to be mere projections of ourselves and self-deceptions. Complete objectivity implies independence from observer effects or, as Bass (5) puts it, zero variance due to the examiner; this can never be achieved, but we can, at least, eliminate one source of this variance and make the rest more explicit.

An imposition of an objective and explicit frame reduces the observer's chance to project onto all others his own ideas, implicit but nevertheless quantitative, about the world around him.***

* Lack of numerical framework in clinical predictions may be partly responsible for the fact that in a well-known study by Meehl (17) the actuarial method showed superiority over clinical judgments.

** By contrast, if we try to be objective we lose sight of our aims. In psychological measurement, as Loevinger says, "the more one objectifies the nature of the universe from which the sample of items is to be drawn, the less likely is the universe to represent exactly the trait which the investigator wishes to measure" (15, p.655). Any experienced clinician will confirm the wisdom of the above observation; but the consequence of his attitude is that the more he sees what he wishes to see the more inclined he is to accept it as reality.

*** The danger of adopting a frame which is not explicit enough is illustrated by Eddington's (8, p.202) well-known example of the ichthyologist who casts his net into the water, examines his catch in the usual scientific manner, and concludes that all sea creatures are at least two inches long; his conclusion may remain "tentative for ever", but it will never be disproved unless he stops fishing and examines his net - which will never bring up anything that it is not adapted to catch. Unfortunately, a framework may be as visible as a fishing net, but its consequences are often hidden.

A forced distribution is bound, by its very nature, to reduce the amount of projection on the part of the observer. It thus serves a function similar to that claimed for psychoanalysis: it allows new insights through reducing self-deception.

Perhaps the greatest value of all statistical analysis lies in its ability to alert the scientist whenever his notions are clearly incompatible with systematic observations. But the secondary gain is by no means negligible: the science of statistics has contributed to the ability of the scientist, especially of the social scientist, to observe in a more systematic and explicit manner. Our Procrustean Table is harsh, like an iron bed; but it is at least explicit, and its frame is as standard as that of a bed in a modern hospital. Subjective judgments are as cruel to reality as Procrustes was to the travellers he used to rob; and they add another threat, that of the unknown.

The travellers caught by Procrustes at least knew where they were. Subjects assessed by purely subjective judgments are just as distorted, with limbs stretched out or chopped off at the whim of the investigator; and, what is more, they are subject, as in Kafka's The Trial (14), to unknown punishments for unknown misdemeanours. If we cannot avoid the shadow of Procrustes in science, let us at least reduce the sinister, hidden threats of Kafka. Then we can achieve what Anatol Rapoport (23, p.3) calls "the deeper aspect of the freedom of science: the freedom from inner rather than externally imposed constraints".

REFERENCES

1. AJDUKIEWICZ, K. "The scientific world perspective". In Readings in Philosophical Analysis, ed. by Herbert Feigl and Wilfrid Sellars. New York: Appleton-Century-Crofts, Inc., 1949, pp. 182-188.

2. BARKER, S. H. Induction and Hypothesis. Ithaca, N.Y.: Cornell University Press, 1957.

3. BARTLETT, F. C. "An experimental study of some problems of perceiving and imaging". Brit. J. Psychol., 1916, 8, 222-266.

4. --------------- Remembering: A Study in Experimental and Social Psychology. Cambridge: Cambridge University Press, 1932.

5. BASS, Bernard M. and BERG, Irwin A. (eds.). Objective Approaches to Personality Assessment. Princeton, N.J.: D. Van Nostrand Company, Inc., 1959.

6. BURT, Cyril. The Factors of the Mind. London: University of London Press, 1940.

7. CAWS, Peter. "Definition and measurement in physics". In Measurement: Definition and Theories, ed. by C. West Churchman and Philburn Ratoosh. New York: John Wiley and Sons, Inc., 1959, pp. 3-17.

8. EDDINGTON, A. S. The Philosophy of Physical Science. Cambridge: Cambridge University Press, 1939.

9. FEIGL, Herbert. "Logical empiricism". In Readings in Philosophical Analysis, ed. by Herbert Feigl and Wilfrid Sellars. New York: Appleton-Century-Crofts, Inc., 1949, pp. 3-26.

10. GRYGIER, Tadeusz. "Criminology and the advancement of science". Brit. J. Criminol., 1965, 5, No. 1, 86-92.

11. ---------------- "The dynamic approach to personality assessment" (to appear).

12. HANFMANN, E. and KASANIN, J. "A method for the study of concept formation". J. Psychol., 1937, 3, 521-540.

13. HARROD, Ray. "The general structure of inductive argument". Paper read to the Aristotelian Society, Nov. 14, 1960. Published in The Proceedings of the Aristotelian Society, New Series, 1960-61, 61, 41-56.

14. KAFKA, Franz. The Trial. New York: Knopf, 1957.

15. LOEVINGER, J. "Objective tests as instruments of psychological theory". Psychol. Rep. Monogr., 1957, 3, 635-694.

16. LURIA, A. R. "Experimental analysis of the development of voluntary action in children". In Perspectives in Personality Research, ed. by Henry P. David and J. C. Brengelmann. New York: Springer, 1960, pp. 139-149.

17. MEEHL, P. E. Clinical Versus Statistical Prediction. Minneapolis: University of Minnesota Press, 1954.

18. MUELLER, P. F. C. "The relationship between Be 61A actuarial and clinical estimates of probable favourable parole outcome". Paper read to the 5th International Congress of Criminology, Section Four, Montreal, Canada, Sept. 1, 1965.

19. PLATO. Republic (trans. by Dr. A. D. Lindsay) London: Everyman's Library, J. M. Dent and Sons Ltd., 1945.

20. POLANYI, Michael. "The unaccountable element in science". Philosophy, Jan.1962, 37, 1-14.

21. POPPER, Karl. The Logic of Scientific Discovery. London: Hutchinson, 1959.

22. QUINE, Willard Van Orman. From a Logical Point of View. Cambridge, Mass.: Harvard University Press, 1964.

23. RAPOPORT, Anatol. "The interests of American and Soviet science". Paper presented to the Political Science Section of the Annual Meeting of the American Association for the Advancement of Science. Cleveland, Ohio, Dec. 29, 1963 (mimeographed).

24. RUSSELL, Bertrand. The Analysis of Mind. London: Allen and Unwin, 1921.

25. ---------------- The Scientific Outlook. London: Allen and Unwin, 1931.

26. ---------------- "The world and the observer". The Listener, 1958, 61, 223-226.

27. ---------------- and WHITEHEAD, A. N. Principia Mathematica. Cambridge: Cambridge University Press, 1910-1913 (in 3 vol.).

28. SELLARS, Wilfrid. "Induction as vindication". Paper read to the American Association for the Advancement of Science, Symposium on the Logic of Statistical Explanation and of Induction. Cleveland, Ohio, Dec. 27, 1963.

29. TARSKI, Alfred. Introduction to Logic and to the Methodology of Deductive Sciences. Oxford: Oxford University Press, 1941.

30. TATARKIEWICZ, Wladyslaw. Historia Filozofii. Warszawa: Panstwowe Wydawnictwo Naukowe, 1959 (in 3 vol.).

31. WHITEHEAD, A. N. Process and Reality: An Essay in Cosmology. Cambridge: Cambridge University Press, 1929.

32. WILLIAMS, J. D. The Compleat Strategyst. New York: McGraw-Hill, 1954.

33. WOJCIECHOWSKI, Jerzy A. "Science and the notion of reality". Revue de l'Université d'Ottawa, Section Spéciale, 1961, 31, 25-38.

34. -------------------- "Measurement and understanding". Revue de l'Université d'Ottawa, Section Spéciale, 1962, 32, 88-104.

APPENDIX A.

PROCRUSTEAN TABLE

To Stretch Data On

N	f(X=1)	f(X=2)	f(X=3)	f(X=4)	f(X=5)	$E(X-3)^2$
12	1	3	4	3	1	14
13	1	3	5	3	1	14
14	1	3	6	3	1	14
15	1	4	5	4	1	16
16	1	4	6	4	1	16
17	1	4	7	4	1	16
18	1	4	8	4	1	16
19	1	5	7	5	1	18
20	1	5	8	5	1	18
21	1	5	9	5	1	18
22	1	5	10	5	1	18
23	2	5	9	5	2	26
24	2	6	8	6	2	28
25	2	6	9	6	2	28
26	2	6	10	6	2	28
27	2	6	11	6	2	28
28	2	7	10	7	2	30
29	2	7	11	7	2	30
30	2	7	12	7	2	30
31	2	7	13	7	2	30
32	2	8	12	8	2	32
33	2	8	13	8	2	32
34	2	8	14	8	2	32
35	2	8	15	8	2	32
36	2	9	14	9	2	34
37	2	9	15	9	2	34
38	2	9	16	9	2	34
39	3	9	15	9	3	42
40	3	9	16	9	3	42
41	3	10	15	10	3	44
42	3	10	16	10	3	44
43	3	10	17	10	3	44
44	3	11	16	11	3	46
45	3	11	17	11	3	46
46	3	11	18	11	3	46
47	3	11	19	11	3	46
48	3	12	18	12	3	48
49	3	12	19	12	3	48
50	3	12	20	12	3	48
51	3	12	21	12	3	48
52	3	13	20	13	3	50
53	3	12	21	13	3	50
54	4	13	20	13	4	58
55	4	13	21	13	4	58

N	f(X=1)	f(X=2)	f(X=3)	f(X=4)	f(X=5)	$E(X-3)^2$
56	4	13	22	13	4	58
57	4	14	21	14	4	60
58	4	14	22	14	4	60
59	4	15	21	15	4	62
60	4	15	22	15	4	62
61	4	15	23	15	4	62
62	4	15	24	15	4	62
63	4	15	25	15	4	62
64	4	15	26	15	4	62
65	4	16	25	16	4	64
66	4	16	26	16	4	64
67	4	16	27	16	4	64
68	5	16	26	16	5	72
69	5	16	27	16	5	72
70	5	17	26	17	5	74
71	5	17	27	17	5	74
72	5	17	28	17	5	74
73	5	17	29	17	5	74
74	5	18	28	18	5	76
75	5	18	29	18	5	76
76	5	18	30	18	5	76
77	5	19	29	19	5	78
78	5	19	30	19	5	78
79	5	19	31	19	5	78
80	5	19	32	19	5	78
81	5	20	31	20	5	80
82	5	20	32	20	5	80
83	5	20	33	20	5	80
84	6	20	32	20	6	88
85	6	20	33	20	6	88
86	6	21	32	21	6	90
87	6	21	33	21	6	90
88	6	21	34	21	6	90
89	6	21	35	21	6	90
90	6	22	34	22	6	92
91	6	22	35	22	6	92
92	6	22	36	22	6	92
93	6	22	37	22	6	92
94	6	23	36	23	6	94
95	6	23	37	23	6	94
96	6	23	38	23	6	94
97	7	23	37	23	7	102
98	7	23	38	23	7	102
99	7	24	37	24	7	104
100	7	24	38	24	7	104

APPENDIX B.

EXAMPLE

Showing the calculation of Plato's correlation coefficient between Preference (P) and Rejection (R) sociometric scores in a group of 19 boys in an Ontario Training School

Subject	P score	R score	(P-3)(R-3)	Σ(R-3)(R-3)
A	2	2	1	1
B	3	2	0	1
C	4	4	1	2
D	3	1	0	2
E	4	3	0	2
F	3	3	0	2
G	4	3	0	2
H	3	4	0	2
I	5	5	4	6
J	4	4	1	7
K	3	3	0	7
L	1	2	2	9
M	2	3	0	9
N	2	2	1	10
O	2	3	0	10
P	3	4	0	10
R	3	2	0	10
S	4	3	0	10
T	2	4	-1	9

$$r = \frac{\Sigma(X-3)(Y-3)}{\Sigma(X-3)^2} = \frac{9}{18} = .50$$

A STATE PROGRAM FOR CENTRALIZED CRIMINAL STATISTICS

Ronald H. Beattie, California Department of Justice

During the decade of the 1920's, considerable attention was given to the development of more adequate information on crime and the administration of criminal justice. Several well-known surveys were conducted showing in detail exactly what happened to arrested persons who were charged with a crime. These studies revealed that factual knowledge of the way in which justice was normally administered was often unknown. The proportion of arrested persons who were ultimately convicted of the offenses on which they were charged was very small, much less than had been assumed.

These studies included the Cleveland Survey, directed by Roscoe Pound and Felix Frankfurter, published in 1922; the Missouri Survey, directed by Arthur Lashley, the published report in 1926 being edited by Raymond Moley; the Illinois Survey, directed by Lashley with Moley as consultant, published in 1929; and the New York Crime Commission reports of 1927. In all of these, Professor C. D. Gehlke of Western Reserve University was a primary statistical consultant. Another survey of the same type, made in Oregon by Wayne Morse and Ronald Beattie, was published in 1931.

Each of these studies followed a similar pattern. A schedule was established for each person arrested and each step of the administrative processes that occurred following arrest was recorded up to final termination of the case. This provided a summary record of what happened to persons arrested in a given area in terms of each step of the process and made possible the construction of mortality tables.

The famed National Commission on Law Observance and Enforcement, established by President Hoover, conducted a series of studies which were published during 1931-32. This body, commonly known as the Wickersham Commission, published Report No. 3 on Criminal Statistics on April 1, 1931 and Report No. 4 on Prosecution on April 22, 1931.

The Prosecution Report, authored by Alfred Bettman of Cincinnati, contains a rather complete review and analysis of the crime surveys already mentioned, and in less detail of several others made during the same period. The statistics report considered specifically the problems and needs of criminal statistics. No better description can be offered of current problems than to quote from Page 4 of this report.

"For our purposes in a large view, the statistics which ought to be gathered, compiled, and published authoritatively at regular intervals, fall under three main heads--crime and criminals, prosecution, and penal treatment. We need to know the volume and character of the offenses committed, both in the past and at any specified time in the present, both in the locality in which for the moment we are interested and in other localities, whether like or unlike in their conditions. We need to know what persons or types of persons, if types may be differentiated, commit these offenses. We need to know what happens to them, whether they are arrested, whether they are prosecuted, and, if so, with what result. We need to know how the machinery of investigation and detection operates, how the prosecuting machinery operates, how the machinery of trying and judging operates in each of its parts. We need to know what happens to the convicted offender, what takes place in the course of penal treatment, how the agencies of penal treatment operate, and what happens to the criminal not merely in the course of penal treatment but afterwards."

The report further outlined certain basic principles for development of an adequate criminal statistics system.

1. *Criminal statistics should be gathered, compiled and published in each jurisdiction.*

2. *The publishing and compiling of criminal statistics should not be confined to any bureaus or agencies engaged in administering criminal law.*

3. *Local officials ought not to be expected to do more than to turn in to the appropriate central office exactly what their records disclosed.*

The Commission quoted with approval the following statement of Sam B. Warner who prepared the basic summary of criminal statistics presented in this report.

"The value of criminal statistics in society's struggle with crime may be compared with that of the balance sheet and profit-and-loss statement in a corporation's struggles for profits. Neither the balance sheet nor the profit-and-loss statement show why the business has been successful, yet no corporation would think of operating without them. The balance sheet and the profit-and-loss statement are for the corporation the indispensable tools of knowledge. Similarly, criminal statistics are the indispensable tools of knowledge for any community that is attempting to reduce its crime and improve its administration of criminal justice."

With the emphasis and publicity given to the early surveys and the summarizing and restatement of objectives expressed in the 1931 Wickersham report, it is somewhat disconcerting to realize that today, nearly 35 years later, there has been very little progress made toward establishing centralized criminal statistics of the type outlined.

The fact that every state is sovereign in its criminal administration and its laws and procedures makes it essential that each state assume the primary responsibility for the collection of criminal statistics within its jurisdiction. If we are to have a national picture of crime it can only be accomplished by having individual states do their part by compiling full data in accordance with acceptable uniform definitions and procedures. A central agency would then develop a nationwide picture from the data supplied by each state. No such national information exists today. The Uniform Crime Reports issued by the Federal Bureau of Investigation are almost entirely based on summary information supplied directly by some 8,000 or more local law enforcement agencies and are limited to the number of certain types of offenses reported to the police and an annual summary of persons arrested. These data lack comparability between states and, of course, do not adequately portray the facts as to the administration of criminal justice nor do they touch the correctional area. There is a National Prisoner Statistics' collection published by the Federal Bureau of Prisons, which prior to 1948, had been the responsibility of the Census Bureau. This collection originally accounted for prisoners committed to and released from federal and state penal institutions but in recent years has become much more limited in the information made available than when it was conducted by the Census Bureau.

One of the recommendations of the Wickersham Commission (Page 17) called for a uniform state law to be drafted and enacted centralizing responsibility for collecting desired data on crime and the administration of criminal justice. In 1946, such a law, drafted by Professor Thorsten Sellin, was adopted and promulgated by the National Commissioners on Uniform State Laws. To date, California is the only state that has actually adopted it (in 1955) and created a central bureau for gathering criminal data.

The California development commenced in 1945 when, by executive order, a Bureau of Criminal Statistics was established in the Department of Justice to serve at that time the needs of the Departments of Corrections and Youth Authority as well as provide statewide statistics. In 1955, the bureau had developed to the point of receiving comprehensive reports on crime from all local jurisdictions in the state and at that time the Legislature adopted the Uniform Criminal Statistics Act which in effect gave statutory sanction to the agency already in existence.

The California Bureau of Criminal Statistics today receives criminal data throughout the state as follows:

1. *A monthly summary report from all law enforcement agencies (about 500) of felony offenses known to the police.*

2. *A monthly summary report from all law enforcement agencies accounting for adult felony arrests by offense and police disposition, and a summary count, by offense, of adult misdemeanor arrests.*

3. *A monthly summary report from all law enforcement agencies of arrested persons under the age of 18 years, by offense and disposition.*

4. *An individual report on each person released from a jail sentence from five county jails of the state.*

5. *An individual report on each person prosecuted in California superior courts, accounting by date for type of pleas, trials, and sentences.*

6. *An individual report on each adult referred to probation by the superior courts for 56 counties of the state, and for the other two counties, summary information.*

7. *Reports on the current status of each adult placed on probation until final determination.*

8. *An individual report on each adult referred and placed on probation by the lower courts from 26 counties of the state.*

9. *An individual report on every juvenile referred to juvenile probation departments together with follow-up reports concerning status changes until time of final termination. These data are received on each individual from 56 counties and in summary form from two counties.*

10. *An individual report from 17 counties of the state on each juvenile re-referred to the juvenile court while on probation.*

11. *A report on persons received and released from county juvenile camps, from 11 camps on an individual report form and from 31 on a summary basis.*

12. *Individual reports from 14 counties and summary reports from 44 other counties on persons received into and released from juvenile halls.*

13. *A drug offender file which is kept current on each offender on the basis of information received by the State Identification Bureau including arrest and offense reports involving narcotics, criminal record histories or rap sheets, and disposition reports received relating to these offenders. The file, which was started in 1959, now contains information on some 35,000 individuals arrested on some type of narcotic charge.*

From the above it will be seen that the coverage of information concerning crime and delinquents in California is fairly wide-spread. There are still gaps to be filled in particularly in the area of jail, misdemeanor probation, and re-referral of juveniles. However, the greatest weakness is that all of the reporting at the crime and arrest level is summary and there is no way at the present time to follow individual offenders from the point of arrest through prosecution and treatment.

The objectives of the California bureau for the future are to develop complete information on crime and delinquency within the state and to interrelate all of the known data. This will require first, the extension of the coverage to all areas not now reporting, and second and most important, the establishment of an individual accounting system permitting each person arrested to be followed through the entire criminal-justice process, as was done in the surveys of the 1920's, and through correctional treatment and even beyond to subsequent criminal behavior. Until this is done, the need for information as outlined in the National Commission Report of 1931 will never be met.

The development of an integrated accounting for offenders who come into the processes of criminal justice will necessitate more complete information about the offender and a much more satisfactory classification of offenses. There should be established an individual criminal record history that in itself is complete and comprehensive enough to describe the individual, his characteristics, the status of his criminal career at each and every point in which he appears or reappears, and the effectiveness of the correctional programs to which he has been exposed.

One of the weakest areas in present-day criminal statistics, besides the lack of individualized information, is offense data. Crimes are reported in terms of general groupings such as burglary, robbery, theft, etc. There is no present identification of the degree of seriousness of the offense reported. A most valuable contribution, recently published by Professors Sellin and Wolfgang on "The Measurement of Delinquency", points up this problem and offers the results of a very thorough study of juvenile arrests in the city of Philadelphia and the development of an objective weighting scheme for criminal offenses. This pioneer effort clearly demonstrates the tremendous need for subclassifying offense data in more meaningful terms. Until progress is made in this direction, the general statistical data on crimes reported to the police will continue to be of exceedingly limited value as an acceptable index of crime.

A rapid growth in the number and rate of crimes reported over the past years is revealed in Uniform Crime Reports and in the California data--particularly with regard to property offenses. Many authorities in criminal statistics suspect that more and more of the minor types of criminal property offenses are being recorded than before indicating the tremendous numerical increase may not mean an equivalent rise in serious crime. However, until better classifications can be made of the offenses reported we cannot test even this hypothesis.

In summary, it must be again pointed out that to properly develop criminal statistics in the United States requires each state to assume its primary responsibility for accounting for all of the information on crime, criminal offenders, and the administration of criminal justice under its sovereign jurisdiction. There is a great need for a national picture of crime but the states must produce the basic information on which a national picture can be compiled. It is an unfortunate truth that what we know today about our national crime problem factually is even less than it was 25 years ago and there is little evidence at the moment of any steps being taken to improve this situation.

From the standpoint of the ordinary citizen, crime is a serious problem. It is to be combatted by all means of prevention and control and this includes the concentrated and combined effort of all agencies; law enforcement, prosecution, courts, probation, and correctional institutions. The information as to what happens throughout the states must be made available and if it is not centralized it remains segmented, non-uniform, and subject to ready misinterpretation. The development of an effective coordinated attack by all agencies concerned with this problem rests to a large extent on the creation of adequate and factual information. We still lack the tools of knowledge that Sam Bass Warner, 35 years ago, pointed out must be available to reduce crime and improve criminal justice.

STATISTICAL MEASUREMENTS USED BY THE ADMINISTRATIVE OFFICE OF THE U. S. COURTS

James A. McCafferty, Administrative Office of the United States Courts

The field of judicial and criminal statistics faces demands for meaningful measures with respect to court activity, probation and parole services, and correctional institution programming. In the last few years the States and the Federal government, which are the primary collectors of such statistical data, have changed from merely an accounting system to what might be referred to as a scientific statistical program. Leadership in this trend can be found in several of the States, but primarily the largest effort in terms of personnel, funding, and programming, appears in the State of California and more specifically, in the California Youth and Adult Corrections Agency.

In the Federal Government improved measurements in the field of judicial and criminal statistics are being developed by the United States Bureau of Prisons, the Federal Bureau of Investigation and the Administrative Office of the United States Courts. What follows outlines four major efforts by the Administrative Office in devising improved statistical measurement devices. In the brief time alloted to me it would be impossible to give a complete statement on each of these devices. If you desire more information about them, we will be happy to supply it to you.

Before beginning the discussion I would like to give you a frame of reference with respect to the organization and the responsibilities of our Division. In 1939 the Congress established the Administrative Office of the United States Courts. Although statistics on the work of the courts had been available for some years prior to the Administrative Office it was not until about 1941 when data were collected and compiled in Washington, D. C. The Division of Procedural Studies and Statistics has included among its responsibilities the collection and reporting of the work of the eleven United States Courts of Appeals and the 92 United States District Courts. Statistical matters generally cover civil litigation, criminal cases coming under Federal jurisdiction and bankruptcy.

In the early years Mr. Ronald H. Beattie was associated with the Division, and in 1961 after a successful career in the California Bureau of Criminal Statistics he returned as Chief of the Division. The first three measurements are his contributions and the last one has been preliminarily developed since he returned to California.

Weighted Caseload

In the Federal Court system some 70,000 civil and 30,000 criminal cases are filed each year. Beginning with 1960 an effort was made to develop a weighted measurement of the judicial workload. During the period 1946 - 1958 six special studies were carried out which clearly demonstrated that case accounting, though useful, had little value in attempting to assess the amount of court time and effort required to dispose of different types of litigation. The fifth and sixth studies were regarded to be the best and on the basis of them the so-called weighted caseload concept was developed.

In the study of the courts it was obvious that the amount of trial time and the proportion of cases disposed of varied considerably and in a sense were directly related to the type of case. In other words, some cases might take very little trial time, therefore, very little of the court's time, whereas other cases took considerable trial time and therefore, a considerable proportion of the court's time. In 1962 the weighted case values were published and we have continued to use them with a minor revision in 1964.

The weight system in simplest terms is taking the proportion of court trial time used and dividing this by the proportion of such cases terminated. For example, on the civil side patent cases account for about six percent of all trial effort in the courts, but account for only 1.5% of the total civil cases terminated. By dividing the six percent by the 1.5% the weight for a patent case is 4.0.

It was further determined that trial effort on the part of the judges should be accounted somewhat differently for a jury trial in contrast to a court trial. It was decided that each day of jury trial should be counted as one day in court whereas a court trial (trial without a jury) should be counted as two days. Such trials require considerable more work on the part of the judge in writing opinions.

Turning to the criminal weighting scheme, after many years of experience it was determined that the judges time in the district courts is divided approximately 77% to cover civil litigation and 23% to criminal. We first based our weighting system on defendants, but in 1964 we turned to weighting cases which tended to increase the weight values of the criminal caseload.

In order to obtain the district courts weighted caseload the weighting system for both civil and criminal cases is applied to the filings of new cases for the fiscal year. The weights assigned to the cases are multiplied by the number of cases filed having the particular nature of suit or criminal offense. These totals are separately divided by the number of judgeships available in the respective district court. The word judgeship must be emphasized since this relates to the number of judges allocated by the Congress and does not necessarily mean that the number of judges on the bench during the year would be the same as the number of judgeships.

Each year we publish for all the courts the average weighted caseload for each district. Analysis of these data have provided guidelines in determining the needs of the federal judiciary in preparation for the Omnibus Judgeship Bill now before the Congress.

Caution must be used when making district to district comparisons of the weighted caseload. For example, certain types of criminal offenses or civil natures of suit may, because of a judicial decision or new legislation, bring about significantly more work for a few courts which may not be reflected in any national weighted average. Also statistics on weighted caseload reflect the amount of work which has been filed in the court for each judgeship and therefore, do not indicate the turnover of cases or pending workload. As noted before, not all the judgeships might be filled, and reference must be constantly made to this fact. Finally, any revision of the weighted caseload requires the ability to completely revise trend data for year-to-year comparison purposes.

Considerably more can be done with the weighted caseload such as applying it to dispositions and to the pending caseload itself. Our major effort has been to provide a better measurement instrument for new cases filed. The revised weights used in 1964 appear in the Appendix tables A-1 and A-2. Also table A-3 provides for each district the criminal and civil weights per judgeship for 1964.

Use of Probation

In our Federal offender statistics series in which we publish not only demographic statistics on criminal defendants filed on and disposed of in the United States District courts, we also provide information with respect to the comparative use of probation in the various district courts. In such a diverse country as ours, comparisons on the use of probation in Federal courts are often sought, however, when absolute proportions of probation have been reported the public is not always aware that among the courts there is a marked difference in the type of offenders coming before the courts.

The major difference among the courts is the nature of offense. Thus, greater proportions of liquor law violators may be found in certain Federal districts than in other courts. Certain districts which are on the well-traveled highways between the North and South, or East and West have an inordinate proportion of defendants charged with auto theft.

After close study of the proportionate use of probation among districts with similar types of offense groups it was determined that the overall use of probation within a court can be related in part to the type of offenses coming before the court.

In order to identify the offenses eight separate offense classes were developed. Beginning with Class I which is composed of certain types of fraud, embezzlement and obscene mail the proportionate use of probation was 84.4%, with Class VIII composed of narcotics and robbery offenses showing 11% placed on probation. The proportionate use of probation and imprisonment and other types of sentences by offense class are shown in the Appendix Table A-4.

Having obtained the actual percentage placed on probation the national average proportionate use for the eight offense groups was applied to the separate districts. Taking Class I offenses the proportionate use of probation was 84.4% for the nation as a whole. By applying each proportionate use of probation to the number of defendants disposed of for the eight offense categories we arrive at a figure referred to as the "expected use of probation". By further dividing the actual proportion of persons placed on probation by the percent "expected use of probation" we obtained the percent of those placed on probation above or below percent expected use.

To illustrate, the district with the highest actual percentage of defendants placed on probation had 78.3% placed on probation in 1964. However, when the national average use of probation was applied to this district the percent expected use of probation was 60.9%. Based on the national average this district, in effect, was using probation 28.6% more than what was expected. On the other hand a district with the lowest absolute use of probation, 26.3%, had an expected use computation of 51.6% which meant that this court used probation about one-half of its expected use.

Obviously these comparisons need to be carefully weighed since the number of convicted defendants in the courts ranged from a low of 13 in one District Court to a high of 1,779 in an other District Court. (See Appendix Table A-5.)

Besides trying to compare the use of probation among the courts it is also useful to group courts according to actual and expected use of probation. Further, such comparisons can be related to the proportion of violation of probation. It would appear on the basis of our 1964 data that whether the use of probation was high or low the proportion of major violations was not too different between courts with high use of probation with those with low use of probation. For example, in 1964 in 22 district courts which placed 60% of the defendants on probation the major violation* rate of those removed during the year was about 16%. In 22 districts where probation was given in 40% of the cases only about 10% had a major violation. These figures can be compared to the overall average for the 88 United States District Courts of about 50.2% being placed on probation with 12.6% being removed from probation for a major violation. (See Table 1.)

What we have tried to do here is provide a better measuring tool for quantitatively assessing the use of probation. It appears that the use of probation is closely geared to type of offense. Further, it is to be noted that whether a court has a high use of probation or a low use of probation the major violation rate is only somewhat higher where probation is granted more often.

Sentence Weights

One of the difficult problems facing the statistician when trying to compare the sentences given to groups of offenders is the lack of any means for making such a comparison. The federal courts have available to them several sentencing procedures such as the Federal Juvenile Delinquency Act, the Youth Corrections Act, probation with its variations, regular imprisonment, and a recently enacted statute sometimes referred to as the indeterminate sentence. As we have noted, when there are great variations in the use of imprisonment ranging from a few days in jail to life, as well as the different sentencing procedures it is practically impossible to make comparisons. Therefore, in 1964 a weighting scheme was devised and this appears in Table 2.

Following the publication of the Federal Offender - 1964 report there was some feeling that the category of 1 to 6 months of imprisonment should be given a higher weight value and the category, immediate probation, 13-36 months, should be dropped in value. In our 1965 report we plan to switch the values so that imprisonment will have the weight value of 3 and immediate probation, 13-36 months, a value of 2.

The selection of the weight values might be regarded as arbitrary but the purpose is simply to provide a means of comparison so that groups of offenders having relatively light sentences would have average sentence weights which are less than other groups which would have heavier sentences and consequently, higher average weights. The average weight for a defendant in 1964 was 5.45. In Appendix Table A-6 there is provided a breakdown of the offense classes together with the actual type of sentence categories and the weight used in 1964. It can be seen in the furtherest right-hand column that the average weight for the defendant tends to increase according to the seriousness of the offense.

There are many ways that sentence weights can be utilized. For example, it is possible to study the relative sentence weights obtained on the basis of the type of conviction, that is, a plea of guilty, change of plea of not guilty to guilty, and a conviction by court or jury. Table 3 shows that for persons who pleaded guilty on arraignment, except for offense Classes VII and VIII, the sentence weight is lower than for other types of disposition. For Class VIII, the most serious group of offenses, persons convicted by a jury had a sentence weight of 29.61 in contrast to a 19.15 sentence weight for those who pleaded guilty on arraignment. The high sentence weight for those convicted by jury may reflect the tendency for persons charged with narcotics or robbery to go to trial. Overall, 7% of defendants convicted are convicted by jury trial. For those convicted of robbery or narcotics the proportion going to trial are, 21 and 17 percent respectively. This is one illustration of what can be done with the weighting scale of severity of sentence.

Weighted Caseload and Time Requirements

With the advent of the computers, statistical measurements, such as regression analysis, which have been used in the industrial field, may have application to the social sciences, and more specifically, to the work of the courts. Recently with the aid of the Bureau of the Budget and a computer at the National Bureau of Standards we have made preliminary analyses of the dispositions of civil and criminal cases by grouping such cases with the use of the weighted caseload concept described earlier. We now have measures which show the relative time required to dispose of cases.

One of the by-products of our preliminary studies shows that mass statistics, such as we obtain from the courts, can be computerized. There is some indication that the material has use

---

* "Major violation" is defined when a probationer receives a sentence of 90 days or more or probation exceeding one year. Also included are probationers who abscond with outstanding felony warrants.

Table 1. 88 United States District Courts

Comparison of the Use of Probation in District Courts, by Type of Violation, Fiscal Year 1964
(Excludes violators of immigration laws, wagering tax laws and violators of Federal regulatory acts)

Item	88 District courts	Quartile groups of District Courts: First 22 District courts	Second 22 District courts	Third 22 District courts	Fourth 22 District courts
Average					
Actual percent placed on probation[1]........	50.2	59.9	55.6	49.4	40.0
Percent expected use of probation[2]........	50.2	48.7	50.5	50.2	50.7
Actual proportion placed on probation above or below percent expected use.........	0.0	23.0	10.1	- 1.6	-21.1
TOTAL REMOVED...........	10,983	2,434	2,708	2,642	3,199
No violation............	8,634	1,794	2,098	2,098	2,644
Violated probation......	2,349	640	610	544	555
Minor violation.......	969	255	257	225	232
Major violation.......	1,380	385	353	319	323
Percent					
Violated probation......	21.4	26.3	22.5	20.6	17.3
Minor violation.......	8.8	10.5	9.5	8.5	7.3
Major violation.......	12.6	15.8	13.0	12.1	10.0

1 See Appendix Table A-5. This is the absolute proportion of persons sentenced who were placed on probation.

2 See Appendix Table A-5. This is the expected use of probation when the average use of probation for eight offense classes for the 88 United States District Courts is applied to the actual offense classes in the separate District Courts. See Federal Offenders - 1964.

SOURCE: Persons Under the Supervision of the Federal Probation System, Fiscal Year 1964.

Table 2. 88 U. S. District Courts
Weighting Scale for Severity of Sentence, Type of Sentence and Weight Value, Fiscal Year 1964

Type of sentence	Weight value	Number of defendants sentenced, fiscal year 1964
Average per defendant	5.45	-
TOTAL DEFENDANTS SENTENCED	-	29,170
Suspended sentences[1] and probation without supervision...	0	2,175
Fines only and probation with supervision, one to 12 months...	1	4,399
One to six months sentences to imprisonment, except split sentences[2]	2	1,738
Immediate probation 13-36 months.	3	6,655
Immediate probation over 36 months, split sentences and all delayed probation[3]	4	3,783
Imprisonment (in months)		
7 - 12	5	1,993
13 - 24	8	3,067
25 - 36	10	1,673
37 - 48	12	1,603
49 - 60	14	1,216
61 - 120	25	595
Over 120	50	273

[1] Includes deportation and all sentences where period of imprisonment or probation is four days or less, or fine only, which is remitted or suspended.

[2] Split sentence refers to 18 U.S.C. 3651 which provides that when the maximum sentence for an offense is more than six months, the court may impose a sentence of which up to six months can be served in a jail-type or treatment institution. The balance of the sentence is suspended and the defendant placed on probation.

[3] Delayed probation occurs when the court indicates that probation will begin at the termination of a local or state term of imprisonment or probation, or a period of hospitalization or release from the military service.

SOURCE: Table 13, _Federal Offenders in the United States District Courts, 1964_.

Table 3

88 United States District Courts
Sentencing Weights by Type of Disposition
and Offense Class, Fiscal Year 1964

Offense classes[1]	Total	Plea of guilty on arraign-ment	Plea of not guilty changed to guilty	Convicted by Court	Convicted by Jury
TOTAL SENTENCED DEFENDANTS.......	5.45	5.00	5.22	6.27	10.59
Immigration, wagering tax, and Federal regulatory statutes	1.63	1.59	1.53	2.05	3.19
Classes I and II....	3.20	3.03	3.10	3.98	4.95
Class III	4.04	3.90	3.88	4.60	5.38
Class IV	5.14	4.92	5.25	5.16	7.88
Classes V and VI....	7.35	6.53	7.05	8.22	11.74
Class VII	7.60	7.55	7.25	7.57	9.45
Class VIII	20.25	19.15	17.15	16.92	27.61

[1] See Appendix Table A-4 for offense classes.

SOURCE: Table 15, Federal Offenders in United States District Courts, Fiscal Year 1964.

for projecting the workload of the courts. However, we are still in a preliminary stage and though the results appear of value there is concern about continuing this project because of the time and effort which might be better employed improving our basic statistical indices such as the weighted caseload concept described earlier.

The grave danger in the use of highly sophisticated statistical techniques is the heavy reliance on the technique often losing sight of the reasons for differences. On the other hand, to provide data to the administrator with many qualifications may, for his purposes, make the data difficult to use. It would appear even with expressed limitations the results from this project will give us a better understanding of the use of the large scale computers and their intricate programming requirements. Also the findings, even with all their qualifications, will give us better means for determining what types of studies we should undertake.

Summing Up

In this brief period I have described four measurement devices, the first three of which were developed primarily by Mr. Ronald Beattie. It would appear that the first three would have applicability to state court systems as well as to correctional systems. The fourth device, "regression analysis", by which the work load can be determined, offers an opportunity for continued study. Each represents a major attempt to rise above the "head counting era". They further show that quantitative measurements can be developed for mass statistics collected from many sources by a central Government agency.

APPENDIX

Table A - 1

CIVIL WEIGHTS FOR NATURE OF SUIT ON FILINGS
REVISED JULY, 1964

Nature of Suit		Weight	Number of cases 1964
United States Cases	TOTAL		22,268
Negotiable Instruments and Recovery		0.05	5,823
Other contract		0.50	1,102
Condemnation		1.70	976
Foreclosure and lease		0.10	868
Other real property		1.50	267
Personal Injury:			
Marine		1.00	142
Motor vehicle		1.70	928
Other		3.00	537
Other tort		1.00	504
Antitrust		8.00	59
Civil rights		3.00	64
Prisoner petitions including habeas corpus		0.30	2,182
Penalties and forfeitures		0.30	3,095
Fair Labor Standards Act		0.40	1,440
Other labor		0.70	382
Tax suit		1.20	1,901
All other U.S. Cases		0.50	1,998
Federal Question	TOTAL		18,651
Marine contract		0.40	2,244
Miller Act		1.00	1,053
Other contract		0.50	277
Real property		0.50	187
Employers' Liability		1.50	1,123
Marine personal injury		0.70	3,937
Other tort		1.70	610
Antitrust		4.00	363
Civil rights		1.80	645
Prisoner petitions including habeas corpus		0.30	3,819
Fair Labor Standards Act		0.70	476
Other labor		1.40	889
Copyright		0.30	440
Patent		4.00	890
Trademark		1.50	437
All other Federal Question cases		1.20	1,261
Diversity	TOTAL		20,174
Insurance		1.80	1,697
Negotiable instruments		1.80	333
Other contract		1.80	3,266
Foreclosure and lease		1.80	404
Other real property		1.80	382
Personal injury:			
Marine		0.70	1,438
Motor vehicle		1.20	8,155
Other		1.40	4,094
Other tort		3.00	375
All other Diversity Cases		3.00	30

NOTE: For a complete description of the weighting process, see pages 156-161 in the *Annual Report of the Director of the Administrative Office of the United States Courts, 1964.*

APPENDIX

Table A - 2

CRIMINAL WEIGHTS BY OFFENSE GROUP FOR CASES FILED
REVISED JULY, 1964

Type of Offense	Weight	Number of original cases 1964
TOTAL		29,944
Income Tax Frauds	2.60	605
Postal Frauds	2.20	391
Homicide	2.00	160
Narcotics, except Marihuana Tax Act and Border Registration	1.80	1,221
Sex Offenses	1.80	255
Marihuana Tax Act	1.40	365
Robbery	1.20	750
Counterfeiting	1.20	253
Miscellaneous general offenses	1.20	1,028
Assault	1.00	320
Other Federal statutes	0.80	830
Embezzlement	0.70	738
Other Frauds	0.70	2,116
Obscene Mail	0.70	291
National Defense	0.60	367
Transporting forged securities in interstate commerce	0.60	982
Theft	0.50	2,459
Burglary	0.40	538
Narcotics, border registrations	0.40	178
Liquor, Internal Revenue	0.40	3,529
Auto theft	0.30	4,995
Food and Drug Laws	0.20	344
Forgery	0.20	2,633
Postal Embezzlement	0.20	599
Immigration laws	0.10	2,770
Migratory Bird	0.10	447
Motor Carrier Act	0.10	780

NOTE: For a complete description of the weighting process, see pages 156-161 in the Annual Report of the Director of the Administrative Office of the United States Courts, 1964.

TABLE A-3. UNITED STATES DISTRICT COURTS

WEIGHTED CASELOAD PER JUDGESHIP, FISCAL YEARS 1963 AND 1964
(Based on civil and original criminal cases filed. Weighted caseload for 1963 supersedes previously published data)

Circuit or district	Number of judgeships	1963 Weighted caseload per judgeship			1964 Weighted caseload per judgeship		
		Civil	Criminal	Total	Civil	Criminal	Total
88 Districts.........	289	195	56	251	207	57	264
FIRST CIRCUIT........	11	194	38	232	205	38	243
Maine.....................	1	159	36	195	142	35	177
Massachusetts.............	6	205	36	241	217	32	249
New Hampshire.............	1	111	28	139	109	11	120
Rhode Island..............	1	273	87	360	237	91	328
Puerto Rico...............	2	179	29	208	230	45	275
SECOND CIRCUIT.......	41	212	37	249	207	40	247
Connecticut...............	4	133	55	188	155	36	191
New York:							
Northern................	2	210	39	249	189	47	236
Eastern.................	8	201	43	244	171	38	209
Southern................	24	223	31	254	219	41	260
Western.................	2	231	68	299	267	53	320
Vermont...................	1	326	8	334	331	22	353
THIRD CIRCUIT........	33	177	27	204	189	25	214
Delaware..................	3	122	19	141	75	15	90
New Jersey................	8	140	41	181	153	37	190
Pennsylvania:							
Eastern.................	11	239	19	258	245	15	260
Middle..................	3	144	23	167	164	24	188
Western.................	8	163	29	192	201	31	232
FOURTH CIRCUIT.......	22	204	77	281	213	83	296
Maryland..................	4	264	54	318	235	44	279
North Carolina:							
Eastern.................	2	125	112	237	144	156	300
Middle..................	2	117	86	203	124	91	215
Western.................	2	109	87	196	103	89	192
South Carolina:							
Eastern.................	2	305	130	435	328	130	458
Western.................	2	112	75	187	129	99	228
Virginia:							
Eastern.................	3	379	75	454	406	80	486
Western.................	2	162	65	227	207	58	265
West Virginia:							
Northern................	1-1/2	85	20	105	87	25	112
Southern................	1-1/2	200	70	270	215	80	295
FIFTH CIRCUIT........	44	249	79	328	258	76	334
Alabama:							
Northern................	3	227	59	286	235	64	299
Middle..................	1	166	79	245	194	103	297
Southern.................	1	275	90	365	362	55	417
Florida:							
Northern................	1	175	97	272	231	96	327
Middle *................	3-1/2	218	103	321	233	91	324
Southern *..............	3-1/2	260	100	360	262	116	378
Georgia:							
Northern................	3	207	77	284	250	75	325
Middle..................	2	144	50	194	164	81	245
Southern................	1	191	154	345	267	139	406
Louisiana:							
Eastern.................	4	504	56	560	513	50	563
Western.................	3	234	49	283	237	35	272
Mississippi:							
Northern................	1	308	77	385	325	87	412
Southern................	2	274	47	321	298	39	337
Texas:							
Northern................	5	186	52	238	197	47	244
Eastern.................	2	274	59	333	232	34	266
Southern................	5	224	84	308	206	70	276
Western.................	3	240	164	404	208	176	384

WEIGHTED CASELOAD PER JUDGESHIP, FISCAL YEARS 1963 AND 1964
(Based on civil and original criminal cases filed. Weighted caseload for 1963 supersedes previously published data) - Concluded

Circuit or district	Number of judgeships	1963 Weighted caseload per judgeship: Civil	1963 Criminal	1963 Total	1964 Weighted caseload per judgeship: Civil	1964 Criminal	1964 Total
SIXTH CIRCUIT........	31	187	66	253	208	62	270
Kentucky:							
Eastern.................	1-1/2	282	141	423	309	147	456
Western.................	2-1/2	175	89	264	172	62	234
Michigan:							
Eastern.................	8	180	59	239	211	66	277
Western.................	2	140	38	178	159	33	192
Ohio:							
Northern................	7	166	42	208	157	41	198
Southern................	3	229	105	334	267	84	351
Tennessee:							
Eastern.................	3	242	65	307	277	54	331
Middle..................	2	128	79	207	154	67	221
Western.................	2	192	59	251	257	56	313
SEVENTH CIRCUIT......	23	226	47	273	251	51	302
Illinois:							
Northern................	10	269	54	323	305	55	360
Eastern.................	2	179	32	211	163	35	198
Southern................	2	124	47	171	151	52	203
Indiana:							
Northern................	3	161	35	196	201	30	231
Southern................	3	306	46	352	336	68	404
Wisconsin:							
Eastern.................	2	181	45	226	173	54	227
Western.................	1	150	41	191	142	37	179
EIGHTH CIRCUIT.......	24	164	41	205	179	43	222
Arkansas:							
Eastern.................	2	171	53	224	200	54	254
Western.................	2	122	37	159	120	36	156
Iowa:							
Northern................	1-1/2	113	24	137	102	21	123
Southern................	1-1/2	182	40	222	171	41	212
Minnesota.................	4	212	29	241	213	57	270
Missouri:							
Eastern.................	3	196	63	259	219	59	278
Western.................	4	201	46	247	232	41	273
Nebraska..................	2	196	45	241	223	31	254
North Dakota..............	2	54	30	84	86	28	114
South Dakota..............	2	89	38	127	92	34	126
NINTH CIRCUIT........	43	145	73	218	157	78	235
Alaska....................	2	70	31	101	74	39	113
Arizona...................	3	188	132	320	211	118	329
California:							
Northern................	9	139	57	196	178	60	238
Southern................	13	153	113	266	160	132	292
Hawaii....................	2	68	34	102	79	32	111
Idaho.....................	2	107	37	144	85	43	128
Montana...................	2	144	52	196	145	48	193
Nevada....................	2	63	36	99	102	53	155
Oregon....................	3	213	51	264	248	46	294
Washington:							
Eastern.................	1-1/2	210	39	249	138	29	167
Western.................	3-1/2	163	52	215	148	36	184
TENTH CIRCUIT........	17	187	57	244	203	53	256
Colorado..................	3	171	49	220	217	40	257
Kansas....................	3	285	82	367	282	51	333
New Mexico................	2	180	93	273	206	115	321
Oklahoma:							
Northern................	1-2/3	175	29	204	210	35	245
Eastern.................	1-2/3	165	45	210	164	36	200
Western.................	2-2/3	209	51	260	204	60	264
Utah......................	2	95	37	132	122	33	155
Wyoming...................	1	136	58	194	127	53	180

NOTE: For a complete description of the weighting process, see pages 156-161 in the Annual Report of the Director of the Administrative Office of the United States Courts, 1964.

* There is a roving judge who serves these districts as well as the Northern District of Florida. His service in the latter district, though, is expected to be virtually nil, and on this basis the Middle and Southern Districts are shown as 3-1/2 instead of 3-1/3.

SOURCE: Table X 1, Annual Report of the Director, Administrative Office, U. S. Courts.

TABLE A - 4

88 United States District Courts
Offense Class and Type of Sentence of Convicted Defendants,
Fiscal year 1964

Offense class[1]	Total convicted defendants	Im-prison-ment	Probation: Total	Probation: Im-medi-ate[2]	Probation: Delayed[3]	Probation: Split sent-ence[4]	Probation: No super-vision[5]	Fine only	Sus-pended sent-ence[6]	Percent: Im-prison-ment	Percent: Proba-tion	Percent: Fine and sus-pended sent-ence
TOTAL	29,170	12,158	12,749	10,429	604	1,115	601	2,689	1,574	41.7	43.7	14.6
Immigration laws	2,588	1,102	170	59	1	11	99	21	1,295	42.6	6.6	50.9
Wagering tax violations	799	100	282	220	2	27	33	410	7	12.5	35.3	52.2
Miscellaneous Federal regulatory statutes	2,702	127	723	469	7	24	223	1,740	112	4.7	26.8	68.5
TOTAL LESS ABOVE	23,081	10,829	11,574	9,681	594	1,053	246	518	160	46.9	50.2	2.9
Class I	2,180	259	1,841	1,669	19	116	37	63	17	11.9	84.4	3.7
Fraud - Group A	666	55	572	522	8	27	15	32	7	8.2	85.9	5.9
Embezzlement	1,231	175	1,037	928	8	84	17	10	9	14.2	84.3	1.5
Obscene mail	283	29	232	219	3	5	5	21	1	10.2	82.0	7.8
Class II	1,178	296	675	516	20	105	34	188	19	25.1	57.3	17.6
Income tax fraud	597	172	334	253	6	65	10	89	2	28.8	56.0	15.2
Other fraud	581	124	341	263	14	40	24	99	17	21.3	58.7	20.0
Class III - Liquor, Internal Revenue	4,445	1,383	2,919	2,508	26	377	8	129	14	31.1	65.7	3.2
Class IV	5,348	2,317	2,924	2,486	180	204	54	62	45	43.3	54.7	2.0
Theft	2,418	993	1,363	1,142	94	96	31	39	23	41.0	56.4	2.6
Postal fraud	413	167	227	195	21	10	1	16	3	40.4	55.0	4.6
Forgery	2,517	1,157	1,334	1,149	65	98	22	7	19	46.0	53.0	1.0
Class V	1,070	524	476	378	20	45	33	59	11	49.0	44.5	6.5
Border registration, addicts	136	67	67	57	6	1	3	1	1	49.3	49.3	1.4
Assault and homicide	233	114	107	90	3	8	6	9	3	48.9	45.9	5.2
Miscellaneous general offenses	701	343	302	231	11	36	24	49	7	48.9	43.1	8.0
Class VI	2,351	1,428	885	745	41	70	29	12	26	60.8	37.6	1.6
Counterfeiting	294	151	140	126	3	10	1	-	3	51.4	47.6	1.0
Burglary	251	150	99	94	2	2	1	-	2	59.8	39.4	.8
Interstate transportation of stolen property	1,043	666	372	298	28	41	5	-	5	63.8	35.7	0.5
Marihuana	353	199	145	125	4	6	10	2	7	56.4	41.1	2.5
National defense laws	275	162	96	77	2	5	12	10	7	58.9	34.9	6.2
Sex offenses	135	100	33	25	2	6	-	-	2	74.1	24.4	1.5
Class VII - Auto theft	5,066	3,349	1,696	1,262	276	125	33	2	19	66.1	33.5	.4
Class VIII	1,443	1,273	158	117	12	11	18	3	9	88.2	11.0	.8
Narcotics	919	796	111	78	9	7	17	3	9	86.6	12.1	1.3
Robbery	524	477	47	39	3	4	1	-	-	91.0	9.0	-

[1] See Appendix for offense classification.
[2] Immediate refers to placing a defendant on supervised probation upon imposition of sentence by the court.
[3] Excludes split sentence. See footnote 4. Delayed probation occurs when the court indicates that probation will begin at the termination of a term of imprisonment or probation, or a period of hospitalization, or release from the military service.
[4] Split sentence refers to USC, Title 18, Section 3651 which provides that when the maximum sentence for an offense is more than six months, the court may impose a sentence of which up to six months can be served in a jail-type or treatment institution. The balance of the sentence is suspended and the defendant placed on probation.
[5] No supervision is where the court determines a period of time during which certain conditions are to be met, such as a payment of restitution. The probation officer is not made responsible for supervision.
[6] Includes sentences of imprisonment or probation of four days or less, deportation, suspended sentence or fine only, which is remitted or suspended.

SOURCE: Table 11, Federal Offenders in the United States District Courts, 1964.

Table A-5

88 United States District Courts
Defendants Placed on Probation by District Courts, Fiscal Year 1964

(Excludes violators of immigration laws, wagering tax laws and violators of Federal regulatory acts)

Circuit and District	Total convicted defendants	Total placed on probation	Type of Probation: Immediate[1]	Delayed[2]	Split[3]	No super-vision[4]	Actual percent placed on probation[5]	Percent expected use of probation[6]	Actual percent placed on probation above or below percent expected use
TOTAL...........	23,081	11,574	9,680	595	1,053	246	50.2	50.2	0.0
First Circuit........	405	261	245	2	13	1	64.4	52.7	22.2
Maine.....................	35	24	21	1	2	-	68.6	51.8	32.4
Massachusetts.............	198	139	129	-	9	1	70.2	56.4	24.5
New Hampshire.............	27	18	15	1	2	-	66.7	53.0	25.8
Rhode Island..............	74	41	41	-	-	-	55.4	48.1	15.2
Puerto Rico...............	71	39	39	-	-	-	54.9	47.5	15.6
Second Circuit.......	1,721	740	614	40	74	12	43.0	50.5	-14.9
Connecticut...............	169	89	61	4	24	-	52.7	50.3	4.8
New York:									
Northern................	127	74	64	9	1	-	58.3	57.1	2.1
Eastern.................	356	133	112	3	18	-	37.4	56.8	-34.2
Southern................	911	348	302	13	29	4	38.2	47.3	-19.2
Western.................	145	88	69	11	-	8	60.7	49.6	22.4
Vermont...................	13	8	6	-	2	-	61.5	49.5	24.2
Third Circuit........	1,011	618	545	32	36	5	61.1	52.6	16.2
Delaware..................	33	14	13	1	-	-	42.4	44.6	- 4.9
New Jersey................	338	195	178	8	6	3	57.7	50.3	14.7
Pennsylvania:									
Eastern.................	266	190	154	14	21	1	71.4	53.5	33.5
Middle..................	102	50	43	5	2	-	49.1	48.8	0.6
Western.................	272	169	157	4	7	1	62.1	56.9	9.1
Fourth Circuit.......	3,449	2,073	1,708	68	282	15	60.1	56.4	6.6
Maryland..................	239	99	84	5	9	1	41.4	50.1	-17.4
North Carolina:									
Eastern.................	657	428	368	3	56	1	65.1	58.6	11.1
Middle..................	387	218	81	1	136	-	56.3	60.1	- 6.3
Western.................	380	216	201	12	-	3	56.8	55.0	3.3
South Carolina:									
Eastern.................	648	408	351	14	43	-	63.0	56.7	11.1
Western.................	274	155	146	8	1	-	56.6	54.3	4.2
Virginia:									
Eastern.................	322	169	131	19	9	10	52.5	53.1	- 1.1
Western.................	226	144	119	2	23	-	63.7	55.0	15.8
West Virginia:									
Northern................	40	20	17	2	1	-	50.0	48.2	3.7
Southern................	276	216	210	2	4	-	78.3	60.9	28.6
Fifth Circuit........	5,095	2,349	1,939	118	183	109	46.1	49.8	- 7.4
Alabama:									
Northern................	385	168	161	1	1	5	43.6	52.4	-16.8
Middle..................	174	78	74	2	2	-	44.8	47.7	- 6.1
Southern................	169	87	78	5	2	2	51.5	53.4	- 3.6
Florida:									
Northern................	171	63	59	1	3	-	36.8	48.7	-24.4
Middle..................	533	190	129	23	25	13	35.6	49.4	-27.9
Southern................	372	186	111	13	60	2	50.0	47.1	6.2
Georgia:									
Northern................	553	254	239	7	8	-	45.9	52.9	-13.2
Middle..................	386	176	168	5	2	1	45.6	54.4	-16.2
Southern................	312	155	152	-	-	3	49.7	52.4	- 5.2
Louisiana:									
Eastern.................	270	115	107	5	2	1	42.6	52.0	-18.1
Western.................	122	59	50	1	8	-	48.4	46.7	3.6
Mississippi:									
Northern................	174	103	89	-	14	-	59.2	55.8	6.1
Southern................	192	100	74	-	25	1	52.1	50.3	3.6
Texas:									
Northern................	348	148	122	10	5	11	42.5	48.9	-13.1
Eastern.................	134	75	58	6	9	2	56.0	49.5	13.1
Southern................	345	186	139	9	9	29	53.9	45.0	19.8
Western.................	455	206	129	30	8	39	45.3	42.4	6.8

Table A-5

88 United States District Courts
Defendants Placed on Probation by District Courts, Fiscal Year 1964 - CONCLUDED

(Excludes violators of immigration laws, wagering tax laws and violators of Federal regulatory acts)

Circuit and District	Total convicted defendants	Total placed on probation	Type of Probation: Immediate[1]	Type of Probation: Delayed[2]	Type of Probation: Split[3]	Type of Probation: No supervision[4]	Actual percent placed on probation[5]	Percent expected use of probation[6]	Actual percent placed on probation above or below percent expected use
Sixth Circuit........	3,275	1,493	1,291	60	129	13	45.6	53.1	-14.1
Kentucky:									
Eastern.................	473	135	124	9	-	2	28.5	55.1	-48.3
Western.................	255	115	81	27	7	-	45.1	46.4	- 2.8
Michigan:									
Eastern.................	734	412	401	3	2	6	56.1	54.3	3.3
Western.................	109	57	55	2	-	-	52.3	58.4	-10.4
Ohio:									
Northern................	401	232	190	8	32	2	57.9	53.1	9.0
Southern................	434	172	154	4	13	1	39.6	48.5	-18.4
Tennessee:									
Eastern.................	418	202	141	7	54	-	48.3	55.0	-12.2
Middle..................	211	105	105	-	-	-	49.8	57.2	-12.9
Western.................	240	63	40	-	21	2	26.3	51.6	-49.0
Seventh Circuit......	1,404	606	522	33	49	2	43.2	45.0	- 4.0
Illinois:									
Northern................	567	238	209	6	21	2	42.0	42.7	- 1.6
Eastern.................	134	64	60	4	-	-	47.8	43.2	10.6
Southern................	141	60	51	9	-	-	42.6	47.7	-10.7
Indiana:									
Northern................	167	73	57	9	7	-	43.7	44.7	- 2.2
Southern................	239	91	72	-	19	-	38.1	45.9	-17.0
Wisconsin:									
Eastern.................	118	61	55	5	1	-	51.7	53.9	- 4.1
Western.................	38	19	18	-	1	-	50.0	43.9	13.9
Eighth Circuit.......	1,371	704	584	42	57	21	51.3	47.4	8.2
Arkansas:									
Eastern.................	207	123	95	7	7	14	59.4	50.3	18.1
Western.................	137	76	66	5	1	4	55.5	51.8	7.1
Iowa:									
Northern................	41	29	28	1	-	-	70.7	55.1	28.3
Southern................	57	33	19	-	14	-	57.9	51.8	11.8
Minnesota.................	194	82	63	8	11	-	42.3	47.0	-10.0
Missouri:									
Eastern.................	244	89	86	3	-	-	36.5	44.6	-18.2
Western.................	244	124	87	13	21	3	50.8	44.6	13.9
Nebraska..................	88	54	51	2	1	-	61.4	46.8	31.2
North Dakota..............	62	36	35	-	1	-	58.1	51.7	12.4
South Dakota..............	97	58	54	3	1	-	59.8	42.6	40.4
Ninth Circuit........	3,806	2,029	1,662	146	165	56	53.3	46.5	14.6
Alaska....................	71	50	43	-	7	-	70.4	55.3	27.3
Arizona...................	441	189	133	42	8	6	42.9	39.1	9.7
California:									
Northern................	649	383	286	19	53	25	59.0	51.7	14.1
Southern................	1,779	965	869	53	29	14	54.2	45.4	19.4
Hawaii....................	60	40	22	7	11	-	66.7	54.8	21.7
Idaho.....................	107	48	44	2	2	-	44.9	43.3	3.7
Montana...................	146	102	78	3	19	2	69.9	45.0	55.3
Nevada....................	136	71	60	6	4	1	52.2	44.6	17.0
Oregon....................	155	60	39	12	8	1	38.7	44.5	-13.0
Washington:									
Eastern.................	74	36	25	1	9	1	48.6	52.2	- 6.9
Western.................	188	85	63	1	15	6	45.2	48.7	- 7.2
Tenth Circuit........	1,544	701	570	54	65	12	45.4	44.6	1.8
Colorado..................	188	84	67	6	11	-	44.7	44.1	1.4
Kansas....................	298	140	120	13	7	-	47.0	43.8	7.3
New Mexico................	345	127	96	8	20	3	36.8	40.7	- 9.6
Oklahoma:									
Northern................	122	75	67	-	8	-	61.5	50.6	21.5
Eastern.................	137	68	55	1	12	-	49.6	53.4	- 7.1
Western.................	247	96	71	15	6	4	38.9	45.5	-14.5
Utah......................	109	63	58	3	1	1	57.8	43.5	32.9
Wyoming...................	98	48	36	8	-	4	49.0	41.7	17.5

[1] Immediate refers to placing a defendant on supervised probation upon imposition of the sentence of the Court.

[2] Excludes split sentence. See footnote 3. Delayed probation occurs when the Court indicates that probation will begin at the termination of a term of imprisonment or probation or a period of hospitalization or release from the military services.

[3] Split sentence refers to U.S.C. Title 18, Section 3651, which provides that when the maximum sentence for an offense is more than six months, the Court may impose a sentence of which up to six months can be served in a jail-type or treatment institution. The balance of the sentence is suspended and the defendant placed on probation.

[4] No supervision is where the court determines a period of time during which certain conditions are to be met, such as payment of restitution. The probation office is not made responsible for supervision.

[5] This is the absolute proportion of persons sentenced who were placed on probation.

[6] This is the expected use of probation when the average use of probation for the eight offense classes for the 88 United States Courts is applied to the actual offense classes in the separate District Courts.

Source:
Federal Offenders in the United States District Courts - 1964
March 1, 1965, Appendix Table 2.

Table A-6. 88 United States District Courts
Offense Class and Type and Length of Sentence of Convicted Defendants, Fiscal Year 1964
(Weight values are in parentheses.)

Table A-6. 88 United States
Offense Class and Type and Length of Sentence of
(Weight values are

Offense class	Total convicted defendants	Suspended sentence and probation without supervision (0)	Fine only and probation 1 - 12 mos. (1)	Imprisonment 1 - 6 mos. (2)	Immediate probation 13 - 36 mos. (3)	Immediate probation over 36 months, delayed probation, split sentence (4)
TOTAL	29,170	2,175	4,393	1,738	6,644	3,800
Immigration laws	2,588	1,394	39	588	22	31
Wagering tax violations	799	40	481	76	138	40
Miscellaneous Federal regulatory statutes	2,702	335	1,910	58	235	95
TOTAL LESS ABOVE	23,081	406	1,963	1,016	6,249	3,634
Class I	2,180	54	471	59	1,002	394
Fraud - Group A	666	22	204	18	277	108
Embezzlement	1,231	26	196	40	600	234
Obscene mail	283	6	71	1	125	52
Class II	1,178	53	295	102	329	205
Income tax fraud	597	12	133	78	160	120
Other fraud	581	41	162	24	169	85
Class III - Liquor, Internal Revenue	4,445	22	327	270	1,774	939
Class IV	5,348	99	566	332	1,588	778
Theft	2,418	54	307	164	712	352
Postal fraud	413	4	45	14	120	77
Forgery	2,517	41	214	154	756	349
Class V	1,070	44	133	112	203	166
Border registration, addicts	136	4	3	4	14	48
Assault and homicide	233	9	37	28	42	31
Miscellaneous general offenses	701	31	93	80	147	87
Class VI	2,351	55	73	73	478	317
Counterfeiting	294	4	13	1	88	38
Burglary	251	3	4	3	65	29
Interstate transportation of stolen property	1,043	10	36	32	186	145
Marihuana	353	17	2	1	58	77
National defense laws ...	275	19	18	34	61	15
Sex offenses	135	2	-	2	20	13
Class VII - Auto theft	5,066	52	86	62	822	757
Class VIII	1,443	27	12	6	53	78
Narcotics	919	26	12	3	36	49
Robbery	524	1	-	3	17	29

Table A-6. 88 United States District Courts
Offense Class and Type and Length of Sentence of Convicted Defendants, Fiscal Year 1964
(Weight values are in parentheses.)--Concluded

District Courts
Convicted Defendants, Fiscal Year 1964
in parentheses.)

Imprisonment							Average weight per defendant	Offense class
7 - 12 mos. (5)	13 - 24 mos. (8)	25 - 36 mos. (10)	37 - 48 mos. (12)	49 - 60 mos. (14)	61 - 120 mos. (25)	over 120 mos. (50)		
1,993	3,067	1,673	1,603	1,216	595	273	5.45	TOTAL
236	261	9	6	2	-	-	1.88	Immigration laws
17	6	-	-	-	1	-	1.71	Wagering tax violations
27	19	11	6	3	2	1	1.38	Miscellaneous Federal regulatory statutes
1,713	2,781	1,653	1,591	1,211	592	272	6.46	TOTAL LESS ABOVE
63	77	32	11	13	2	2	3.16	Class I
13	16	7	-	1	-	-	2.67	Fraud - Group A
44	50	23	8	9	1	-	3.34	Embezzlement
6	11	2	3	3	1	2	3.52	Obscene mail
70	66	36	7	13	2	-	3.28	Class II
45	29	13	1	4	2	-	3.27	Income tax fraud
25	37	23	6	9	-	-	3.28	Other fraud
467	478	112	33	21	2	-	4.04	Class III - Liquor, Internal Revenue
474	670	371	238	164	64	4	5.14	Class IV
217	242	159	110	73	27	1	4.90	Theft
28	54	35	15	11	9	1	5.50	Postal fraud
229	374	177	113	80	28	2	5.31	Forgery
98	132	65	48	33	19	17	5.78	Class V
7	39	15	2	-	-	-	5.63	Border registration, addicts
15	27	13	10	8	5	8	6.53	Assault and homicide
76	66	37	36	25	14	9	5.57	Miscellaneous general offenses
160	357	275	180	238	125	20	8.06	Class VI
25	28	32	19	23	21	2	7.74	Counterfeiting
7	21	34	39	21	21	4	9.37	Burglary
96	182	152	74	78	46	6	7.79	Interstate transportation of stolen property
-	21	13	34	95	32	3	9.84	Marihuana
24	74	21	6	1	-	2	5.23	National defense laws
8	31	23	8	20	5	3	9.52	Sex offenses
370	920	694	980	291	31	1	7.60	Class VII - Auto theft
11	81	68	94	438	347	228	20.25	Class VIII
2	72	47	51	388	197	36	15.39	Narcotics
9	9	21	43	50	150	192	28.75	Robbery

SOURCE: Table 14, Federal Offenders in the United States District Courts, F. Y. 1964.

CRIMINAL STATISTICS: A REFORMULATION OF THE PROBLEM

Stanton Wheeler, Russell Sage Foundation

The history of papers on criminal statistics is rather discouraging. The basis for pessimism lies not in the papers themselves, for there have been many useful, clearly formulated analyses. I need point only to the number and range of contributions by Sellin,[1] the recent work by Sellin and Wolfgang,[2] to Wolfgang's own systematic critique of uniform crime reports,[3] to Ronald Beattie's review of the uses of criminal statistics in the United States,[4] and to discussions by Donald Cressey,[5] Dan Glaser,[6] and many others, not to mention important contributors from other countries. Many relevant problems in the use and interpretation of criminal statistics have been raised, so that we are aware of the shortcomings, the unreliability and lack of uniformity, and hence of the hazards in making valid inferences about crime from the criminal statistics.

The cause for pessimism, therefore, does not lie with the absence of intelligent critical work. Rather, there is an absence of any follow-through that attempts to solve the problems pointed to in the various critiques. Many of the criticisms have been known for a long time, as a recent detailed historical review of the literature of criminal statistics shows. In this paper I want to suggest that our inability to utilize and interpret criminal statistics is a result not of the technical deficiencies that have been pointed to before, but rather is a result of the way in which the original problem has been put. My suggestion is that we need a reformulation of the problem, rather than further refinements in the technology of crime reporting.

The problem has to do with the underlying conception of crime, and therefore with the nature of the materials that are gathered as a result of this underlying conception. Put briefly, the underlying conception is that <u>the data of criminal statistics are mere records of response to the actions of criminals</u>. A person commits an act that is defined as illegal by statute. When the police department is notified of the act we have an offense known to police. If the department also finds someone and arrests him for the act we have a unit that enters arrest statistics. In either case the assumption is that the units reflect the passive responses of officials to the active behavior of criminals. Differential tendencies to report crimes, or failures to catch offenders, are seen as mere unreliability and efforts may be made to stamp out such problems, since unreliability is bad. Efforts are made to achieve uniformity in crime reporting, to assure that all officials are handling the acts in similar ways. And efforts are made to improve the efficiency and reliability of the actual coding and classifying operations themselves, through the work of the research bureau of the police department, of those processing the data at the FBI or elsewhere.

But is it feasible to sustain this conception of the nature of criminal statistics? The assumptions hold true only in very important but extremely rare limiting cases. We can treat the record of criminal acts as the record of criminals only when we have indeed achieved a precise uniformity in the reporting of such acts to the police, and in the processing of such acts by the police. Now of course we approach this ideal more or less closely with differing types of crime, as the early classification into part 1 and part 2 offenses by what the FBI suggests. But the important point is that there is still great room for variability in reporting and processing, and the ideal is only rarely approached. Thus the conception of criminal statistics solely as records of response to the actions of criminals may not be the most useful way to conceive of the underlying problem.

The alternative is to conceive of three elements as <u>inherently</u> a part of the rate producing process, and of the resulting rate as an interaction of all three. The three categories include: 1) the offender who commits an act specified by statute to be illegal, 2) a pool of citizens who may be either victims or reporters of the acts of the offenders, and 3) officers of the law who are formally charged with the obligation to respond to the action. We would then express offenses as a function of the interaction of these three elements, any one of which might be more or less important in a particular instance.

It should be noted immediately that this is no way a radical reformulation of the problem. All who work with criminal statistics are aware of the great sources of variability that lie in differential values of the community and in differential police actions. This proposed change simply introduces these concerns as a legitimate and inherent part of the model of criminal statistics, rather than conceiving of them as external and unwanted sources of error and unreliability. The principal gain from making this transition is that variations in citizen and police actions become important events to be explained, just as we make efforts to explain why some commit crimes and some do not.

Each of these three categories can be looked at both individually and collectively. Thus we have single criminals, or in some cases criminal organizations. We can conceive of the police system as a whole as the responding agency, where variation in police policies, technology, and so forth are the relevant aspects. Or we could concentrate on individual officers, relating their characteristics to their arrest behavior just as we now relate the offender's characteristics to his criminal behavior. Finally, we can think of the community as a pool of separate residents or as an organized whole with shared sentiments in response to crime. But before passing to several specific consequences that would flow from such a reformulation, more should be said by way of an operational and theoretical justification for this shift.

The Operational Justification. Consider how the records of crime are in fact produced. In most cases they do indeed begin with an act of an offender, but they never end there. As is obviously the case, they must be reported by someone or they never end up in our statistics. We have usually assumed, as indicated in a quote by Sellin which is perhaps the most oft-quoted remark about criminal statistics; that "the value of a crime rate for index purposes decreases as the distance from the crime itself, in terms of procedure, increases."[7] But as that very wording suggests, even the immediate reports themselves may be subject to great error, and it is that error which is so troublesome to those who wish to use official data to test theories of crime causation. While there is certainly no reason to quarrel with the general wisdom of Sellin's statement, neither should we let it hide the fact that the greatest gap of all is likely to occur between the crime itself and the initiating procedure.

There are very few detailed accounts of the actual procedures used by police agencies in the processing of cases and the reporting of crime statistics. Where there are really full and detailed statements, (as in the recent Sellin and Wolfgang volume where a full chapter is devoted to the method of reporting delinquency by a division of the Philadelphia police,)[8] two things seem abundantly clear.

First, standardizing decision-making at the initial stage, particularly in areas such as delinquency, is very difficult and requires a great deal of effort and attention to detail. For example, cases may come to the attention of a juvenile bureau either directly in the course of the juvenile officer's duties, or indirectly by referral. In addition to possible effects of these differences, there are different criteria used in the decision to arrest or report. In Philadelphia, these include the juvenile's previous contacts with police, the type of offense, the attitude of the complainant, the offender's family situation, and potential community sources. It seems quite evident that individual officers might resolve these matters in somewhat different ways, despite a good deal of training.

But the second point is more important. These are procedures worked out by the Philadelphia department for the processing of Philadelphia cases. Quite clearly other principles may be utilized in other cities. How, then, are we to compare the figures in any sensible way? Even though each department may end up with reasonably uniform data for its area, comparability across towns, cities, or regions will be missing, as will comparability over time in the same jurisdiction if any further changes occur. The operational justification, then, is that these sources of variability appear built into the problem. It seems a wise course of action to attempt to understand them, since we are unlikely to get rid of them.

The Theoretical Justification. The theo-justification for treating crime statistics as a result of three-way interaction between an offender, victims or citizens, and official agents, is that deviation itself is increasingly recognized as a social process that depends heavily on social definition.[9] Acts become deviant when they are so defined by members of the collectivities in which they occur. Whether a given pattern of behavior will be labelled deviant is itself problematic, and is likely to vary from community to community, or from policeman to policeman, at least within certain fairly broad limits. Why emphasize only the person who might commit an act, and not those who might label it as deviant, or those who might officially respond?

The concept here is close to that suggested decades ago by VanVechten: The tolerance quotient of a community.[10] This has to do with how much "trouble" the community will put up with before it acts, or in other words how much deviant behavior it will permit before either citizens or official agents take offense and respond in some systematic way.

Evidence in support of this orientation is found increasingly in the study of forms of deviation close to but not identical with criminality. Consider for example mental retardation. A recent study shows that the mentally retarded from families with lower educational background spend a shorter time in institutions and are released more readily than are those from higher educational background.[11] This is true even when they are matched carefully by IQ. The suggested explanation is that families of lower educational level are less likely to define their offspring as mentally retarded, and are therefore more ready to accept them back into the home. A related study shows that families of higher socio-economic status are able to get their children accepted into institutions for the mentally retarded more quickly than are those of lower socio-economic status, and this appears in part to be because they are more insistent about the need of the child in question.[12] In other words, they think of this behavior as more deviant than do those of lower socio-economic levels. It appears likely that both entry and release from the hospital are functions of the social characteristics of those who are attempting to get them in or out, and are not mere reflections of intelligence as measured by standardized tests.

Consider further some of the evidence regarding mental illness. Several recent studies suggest that rates of commitment bear a close correspondence to the paths of entry to hospitals. In one instance, that of a child guidance clinic where the concern is for which children are accepted among all those referred, the evidence is that those referred by doctors are more likely to be accepted than those referred by family members.[13] The further evidence is that acceptance is more closely related to the source of referral than to the nature of the symptoms of the individual who is being referred.

Thus in these areas of social deviation, it makes good sense to think of the deviation itself as a social process involving not only the person who commits deviant acts, but also those who choose to label them as deviant and those who are officially charged with acting upon them as such. Indeed, a full understanding of rates of institutionalization or rates of retardation and illness, seems to require that we consider more than simply the mental or intellectual status of the person in question.

The Special Case of Crime. It can be argued that a mistake is made in attempting to treat crime in the same category with the forms of deviance reviewed above. The criminal law is primarily statute law, and the specification of conditions necessary to convict one of the commission of the crime is certainly more detailed and specific than is the case for mental subnormality or mental illness. Criminal statutes typically specify in some detail the nature of the offense, and we have well worked out techniques which, in the case of pleas of not guilty, may be utilized by juries to assess guilt or innocence. Therefore we might expect somewhat more objectivity in the collection and analysis of data on crime than is true for other forms of deviation.

This argument is certainly true to a point, and it would be a mistake to equate crime overly readily with other forms of deviation. There is a sizeable difference between the behavioral specification of acts, for example, of burglary or arson, and the much more general, abstract, and judgmental character of the process of diagnosis of a person as psychotic. But again, two features of crime remain important to note in this context. First, enforcement of all statutes is not attempted. Diligence in some areas is matched by negligence in others. In fact, our policing and detection policies introduce new sources of variation that are not encompassed in the definition of the statutes, as Daniel Bell's article on the myths of crime waves reminds us.[14] Policies to "crack down" on all narcotics users or pushers, while "tolerating" organized prostitution, are likely to be found within the same police jurisdiction. This simply indicates that the clarity of the specification of law violating behavior in the statutes is often not repeated by the policies in fact enforced by the policing agencies.

Even more important than this, however, is the fact that some of the forms of crime that are becoming increasingly important no longer have the clear-cut statutory form of definition. A principal case of course is delinquency. Most legal definitions of delinquency are so broad and vague as to make it roughly synonymous with juvenile trouble making. In addition to including offenses that also hold for adults, there are such things as being truant, willful disobeying of parental commands, and staying out after curfew. The lack of specification in these instances approaches that of the case of mental illness, which of course is not surprising in that many see forms of delinquency and forms of mental illness as synonymous.

For these reasons, I think it can be effectively argued that a model stressing the social definition of crime, and especially the actions of other social agents as well as those of presumed offenders, is pragmatically useful as well as being highly realistic.

Some Practical Consequences

The most immediate effect that would flow from adoption of this rationale is that we might be able to learn something more about systematic variations in the crime rate than we learn by examination of the characteristics of criminals. Consider each of the following four consequences.

1. Improved Understanding of Police and Official Agents. Remembering the distinction between the collective and individual forms, and beginning with the collective, we might ask: What are the characteristics of police systems where high crime or arrest prevail? Here is a problem eminently worthy of study, and we might almost refer to it, especially in the context of recent events, as the Los Angeles police problem. Some years ago Ronald Beattie wanted to argue that the high rate of offenses known to the police in Los Angeles was a result, not of the law-violating behavior of Angelenos, but of the good deeds of Chief Parker and his force. The Los Angeles police department, he argued, was a superior force in terms of efficiency and dedication. The high rate of arrests was a result of efficiency, rather than the result of a high rate of offenses. This example at least suggests that we should be able to find some stable and reliable differences between police departments that report high rates of offenses and those that report relatively low rates. What are those differences? Suppose we introduce controls for the nature of the social composition of the community, would we still find stable differentials based upon differences in the police function?

Clearly, to answer these questions requires that we work hard to establish differential degrees of police efficiency in crime reporting, and differentials in types of police organization. Conceivably the arrest rate is a function of the number of motorcycles versus police cars on the road, a function of the proportion of the total police force that is civilian, a function of the average educational attainment of the individual officers, a function of whether or not there is a police academy that serves to train policemen for this particular department, and so on. The whole point is that introducing the official actions of the police, not as mere passive response to the criminal, but rather as an inherent part of the production of a crime rate, forces us to ask these questions, and hence ultimately to understand better the workings of police organization.

I shall cite two studies, of radically different styles, where this sort of contribution seems to be forthcoming. One is interesting work by the political scientist James Q. Wilson.[15] Wilson has compared a relatively non-professionalized police force in an Eastern city with a highly professionalized force in a West Coast city. His interest was in seeing whether the nature of professional organization of the police is related to modes of handling delinquents and to the rate of arrest of juvenile offenders, and his findings suggest that it is indeed. The old-line force, fraternal in organization, recruits its members largely on grounds of locale, provides little training for them, and little professional esprit. The result is that while they are punitive toward youthful offenders, there is no strong sense of urgency about police work and hence relatively low rates of official actions with regard to youthful offenders. The force in the West Coast city, in contrast, is one that is recruited nationwide, places a high premium on education, pays better, and in other ways appears to fit the model of a professional as distinguished from a fraternal system. In the West Coast city youths are more likely to be picked up for minor offenses, minor offenses are more likely to be treated as major infractions, and the arrest rate tends to be much higher than in the East Coast city. This example merely serves to illustrate that the crime rate may vary in close correspondence with the nature of police organization, and conceivably quite independently from the nature of delinquent activity.

The second example comes from an ecological study by Mr. Greenhalgh of the British Home Office.[16] He had the wisdom to include as a relevant variable in his analysis the number of police per capita in various social units. He finds that the number of officers is related to the number of offenses reported, and while this of course raises a neat problem as to cause and effect, it serves to emphasize the potential role of the structure of the police systems themselves.

We may also find important sources of variation in individual differences within police departments. There is certainly good reason to imagine that there are sizeable differences in policemen in terms of the number of individuals they arrest or take official action upon. A police officer who has had many years of experience once related to me an experience from one of his early days on the force. He was in a squad car when they received a radio call from central headquarters to proceed rapidly to the scene of a particular offense. He was driving the car with his partner in the automobile, a much older and wiser policeman, sitting next to him. As my young, gung ho friend roared to the scene of the crime with the siren wailing, his older colleague turned to him and said "For crying out loud, slow down and turn off the siren. You're makin me noivous." The point is fairly clear: there is little more reason to expect age, training, ethnic background, and other characteristics to be irrelevant in this context than there is to expect them to disappear when we consider offenders.

The necessary first step is to begin collecting data on policemen and police departments similar in form if not in content with what we gather on criminals. This is already done to some extent by the FBI, which annually publishes, for example, the list of the number of uniformed and civilian police employees for every reporting city over 2,500 population. But because this is thought to be relevant for policing but not for crime, there have been no analyses, to my knowledge, of the possible correlation between number of police and either the number of criminals, or the number of offenses cleared by arrest.

The chief practical consequence of adopting a new rationale is that we would begin to understand the dynamics of police systems in relation to offenders. So long as we treat the police as mere reactors to the actions of criminals, this whole area will remain hidden from our view. My argument is simply that if we transform the degree of police efficiency into a variable to be explained, rather than one to be eliminated by the production of uniformity in procedure, we will enhance our understanding of crime.

2. Improved Understanding of Citizens and Social Control. The pool of conventional persons in the community, either victims of crimes or citizens who observe them, often initiate the production of crime rates by being the first reporters of criminal events. There appear to be relatively vast community-wide differences in the rate at which persons call the police for help with problems. In Nathan Goldman's study of differential selection of juvenile offenders for court appearance, he suggests that official agents are highly responsive to the definition of deviation on the part of the citizens of the community, and that some of his communities have high rates of delinquency because the officials feel that the citizens will complain if they don't take official action, whereas other ones have low rates because the citizens simply don't complain.[17]

Another example of the possibilities here is provided in a study by Eleanor Maccoby and others.[18] They interviewed members of two communities, one of which had a high rate of delinquency and the other a low rate, where socioeconomic characteristics were held constant insofar as possible. One of the things they found is that the community that had a low rate of official offenses had a high rate of community cohesion. That is, friends, neighbors and others would intervene when they saw kids getting into trouble. In the community with the high official delinquency, there was very little interaction among members, and little intervention at these early stages. The strong suggestion here is that informal social controls operated effectively in one community to obviate the need for official actions, whereas in the other they did not. The low official rates were due to prompt intervention in cases of incipient deviation; in the other community incipient deviation was not responded to at all, and it grew in seriousness until official actions occurred.

Although the evidence in these two cases is not entirely clear, the general point is certainly not to be debated: different types of neighborhoods and communities may respond to deviant behavior in radically different ways, and their responses become the initiation of the official reporting system. Unless we understand them we will not be led to a full understanding of the rate production process.

As in the case of police systems, we may find individual variation within the neighborhood or community, just as we find systematic variations between them. Either as victim or as observer, we are likely to find many important differences in the role of the citizenry in the production of crime rates. There are of course several studies focusing on the victim, but usually these have been separate investigations that have little relation to routine police reporting. And there is folk knowledge, though little systematic evidence, of individual differences in willingness to report offenses to the police. Older single women living alone are thought by some to be inordinately observant of potential criminal situations. One police captain once told me that the rate of telephone calls reporting crimes in progress dropped substantially with the growth of television. The implication was that people who used to mind other's business, and hence keep their eyes on the street below, were now absorbed watching crime dramas on TV and didn't see the real thing anymore.

Although these examples may be of dubious validity, they serve to illustrate the main point: whether a person gets treated officially as an offender depends on which citizen he happened to meet and which community he happened to be in when the act occurred, and our explanation of variations in crime rates will have to do in part with area variations in the nature of communities and their law-abiding citizens.

3. <u>The Development of Consumer-Oriented Crime Statistics</u>. A third practical consequence is that we could begin to express crime rates in ways that would have more meaning for the public. The police system itself exists for the protection of the community, but so far we have done extremely little to provide data that is directly relevant to community members. This is apparent by examining the denominators that typically are used in construction of crime rates. If one is diligent, one can find arrest rates for Negroes, for Puerto Ricans, for whites. Or one can find age-specific rates of offense. In a handful of cases, one can find cohort analysis tables indicating the probability that a person will ever be arrested between, say, ages 7 to 18.

All of these figures have a curious cast. They tell us much more about who commits the offense than about the person against whom it is committed. Yet if we think now as citizens, and not as persons interested solely in offenders or policing, it seems that we might ask rather different questions. Personally, the risk that my wife or children are assaulted at all matters more to me than whether they are assaulted by a Caucasian, a Puerto Rican, or a Negro. Yet I can find figures on the latter topic but not on the former. Similarly, one may wonder what New York City residents would make of the fact that the reportedly rising crime rate in the City could be explained as a function of the increased number of persons of juvenile age, which is of course the age at which most crimes are committed (so far as we can tell from official statistics). Certainly it is important theoretically to understand that the rising rate does not appear to be a response to new forces and fears in mass society, but rather can be explained fairly directly as a function of the age structure of the population. But for the typical resident, the important question would seem to be whether or not the rate has gone up for victims in his category.

This is simply to suggest that a useful way of reporting crime data would be to use as a denominator not some characteristic that might describe offenders, but one that will describe their victims. Apartment dwellers might well want to know what the probability is that their apartment will be burgled within the next five years. Others might want to know what the probability is that they will be robbed. In principle, it should not be difficult to prepare such statistics. We take the number of offenses appearing in a particular area against a particular type of victim, and express it as a proportion of all persons who have the social characteristics that the victim happens to hold. In this way we have victim-specific rather than offender-specific crime rates--in effect, a box score which the citizen can use to keep tabs on differing areas in his community, and hopefully on differing communities. It would become abundantly clear, for example, which areas of the city are most dangerous at night, and for what categories of persons they are most dangerous. Such consumer-oriented statistics would seem to be more important as a public service than are offender-oriented statistics such as those we now produce.

The issues are clearly more complicated than suggested here. One problem is the necessity of correcting for the daytime and nighttime populations of the areas. And in order to get detailed victim-specific rates, we would have to learn more now than we normally do about the nature of the victim. In the latest Uniform Crime Report available to me (for the year 1963) only one out of some 49 tables tells us anything about the victim. This one has to do with the victims of homicides, and classifies the victims according to their age, sex, and race.[19] At least, I would argue, it is an effort in a much needed direction.

4. Improved Understanding of Criminals and Criminal Acts. The fourth and final consequence is that adopting the frame of reference outlined here might enable us to approach what we have always traditionally desired, namely, better descriptive and explanatory accounts of the actions of criminals. Paradoxically, it is only by first directing our attention to the citizens and the police that we can begin making headway on the initial problem of sources of variation in crime rates.

At the moment, any community-wide comparisons of crime data are subject to possible unreliability, and certainly debate as to the interpretation of meaning, because of possible differentials in the functioning of the citizens and the police. A higher rate of crime for community a than for community b cannot be guaranteed to tell us something about the actual level of law violations in the two communities, for all the reasons we have already reviewed. Any efforts aimed at assessing the actual rate of legal violations, or differentials in the rate that are related to differential characteristics of the offenders, must of necessity take into account the variation due to citizens and policing. We can do so, of course, only if we have studied such variations and have evidence with regard to them.

The Necessary First Step. The most essential first step is that there must be new sources of input to the official collections of data. If the position argued here is correct, it will no longer be enough for the established reporting agencies such as the FBI to collect data simply on the number of crimes reported in the various jurisdictions. It will be essential that they also collect systematic data on a) the complaining witnesses, b) the social characteristics of the community, c) the reporting or arresting officer, and d) the nature of the police system as a whole. Just as there is a reporting form for crime, there must be a reporting form for complainants, for the community, for officers, and for police departments. This would enable us to gather systematic data on the other possible sources of variation in crime rates. The details for such reporting forms would of course have to be worked out, and problems of uniformity would be sure to arise. But there is no reason why they should be any more severe than those now plaguing the reporting of crimes. Also, it would be necessary for us to think about more creative denominators for crime rates, along the lines suggested above. But here too, the technical task is not overwhelming, and much of the work has already been done by the Bureau of the Census.

If prestigious organizations such as the FBI were to begin collecting such data routinely, we could begin to close in on the haunting problems of biases in criminal statistics. We could at least compare jurisdictions whose police procedures were roughly similar, and where the types of complaining witnesses were not simply a function of the demographic structure of the community. More importantly, we could begin to examine the interactions between the three major sources of variation: the offender, the citizenry, and the police system.

Conclusion. Most of the questions raised here concern the uses of crime rate data. Implicit throughout is the question: what is a useful rate? Assuming that crimes or arrests enter the numerators, the question concerns the sorts of denominators that are most important and relevant. The suggestion is that the received wisdom, so far, leads us to construct denominators reflecting the nature of the crime-committing person. The principal suggestion of this paper is that we ought to broaden the conception of the relevant denominators to include characteristics of the police system and the nature of victims or the citizen population. To adopt such a view systematically would, I feel, greatly broaden the richness and relevance of our understanding of crime and its control, and would have the further advantage of making more meaningful the very data we now complain about in our critiques of criminal statistics.

Footnotes

1. Sellin, Thorsten, "The Basis of a Crime Index," Journal of Criminal Law and Criminology, Vol. 22, pp. 335-356, September 1931.

2. Sellin, Thorsten, and Marvin E. Wolfgang, The Measurement of Delinquency, New York: John Wiley & Sons, Inc., 1964.

3. Wolfgang, Marvin E., "Uniform Crime Reports: A Critical Appraisal," University of Pennsylvania Law Review, Philadelphia, Pa. Vol. 111, No. 6, April, 1963, pp. 708-738.

4. Beattie, Ronald H., "Criminal Statistics in the United States--1960," Journal of Criminal Law, Criminology and Police Science, Vol. 51, No. 1, May-June, 1960.

5. Cressey, Donald R., "The State of Criminal Statistics," National Probation and Parole Association Journal, Vol. 3, No. 3, July, 1957.

6. Glaser, Daniel, The Effectiveness of a Prison & Parole System, Indianapolis: The Bobbs-Merrill Company, Inc., 1964. See also Glaser, Daniel, "Criminal Career Statistics," in Proceedings, American Correctional Association (1957), pp. 103-106; Glaser, "Released Offender Statistics: A

Proposal for a New National Program," American Journal of Correction (March-April 1957), pp. 15-17, 25.

7. Sellin, Thorsten, "The Basis of a Crime Index," Journal of Criminal Law and Criminology, Vol. 22, September, 1931, p. 346.

8. Sellin and Wolfgang, op. cit., pp. 82-114.

9. This orientation is most fully stated in Becker, Howard S., Outsiders, Glencoe, Ill.: Free Press, 1964. Earlier contributors include Edwin Lemert and John Kitsuse, but this is not the place for a detailed treatment.

10. VanVechten, C., "The Tolerance Quotient as a Device for Defining Certain Social Concepts," American Journal of Sociology, Vol. 46, July, 1940, pp. 35-42.

11. Mercer, Jane R., "Social System Perspective and Clinical Perspective: Frames of Reference for Understanding Career Patterns of Persons Labelled as Mentally Retarded," Social Problems, Vol. 13, No. 1, Summer, 1965.

12. Sabagh, Georges, Richard K. Eyman and Donald N. Cogburn, "The Speed of Institutionalization: A Study of a Preadmission Waiting List Cohort in an Institution for the Retarded," unpublished manuscript.

13. Teele, James E., and Sol Levine, "The Acceptance of Emotionally Disturbed Children by Psychiatric Agencies," Ch. V, forthcoming book.

14. Bell, Daniel, "The Myth of Crime Waves," in The End of Ideology, Glencoe, Ill.: Free Press, 1960, pp. 137-158.

15. Wilson, James Q., "The Police and the Delinquent in Two Cities," Ch. II, forthcoming book.

16. Greenhalgh, W. F., "A Town's Rate of Serious Crime Against Property and Its Association with Some Broad Social Factors," Home Office, Scientific Advisers Branch, February, 1964.

17. Goldman, Nathan, The Differential Selection of Juvenile Officers for Court Appearance, National Council of Crime and Delinquency, 1963.

18. Maccoby, Eleanor, et. al., "Community Integration and the Social Control of Juvenile Delinquency," Journal of Social Issues, No. 3 (1958), pp. 38-51.

19. Uniform Crime Reports, 1963, Table 18, p. 102.

II

CONTRIBUTED PAPERS I

Chairman, DAVID L. KAPLAN, U. S. Bureau of the Census

ON THE RELATIONSHIP BETWEEN CHILDREN'S AND PARENT'S POLITICAL PREFERENCES (Abstract)

Stanley Naparst, University of California, Berkeley

Prior work has discovered that a relationship exists between children's and parent's political preferences. This paper takes the work one step further by postulating the exact relationship. In analogy to Mendelian heredity it is assumed that each parent "gives" the child one half of his political preference. Preferring the same preference as the parent is dominant over not preferring the same preference.

Children can either have the same or different preferences than the parent, but those that have the same preference as the parent may be hybrids or pures. The hybrids will have some offspring that have different preferences than themselves. The dominant pures' offspring will not differ from the parents. The recessive pures' offspring will not differ from the parents. A simple heuristic model (and then a Markov model of this idea) are developed.

Some testable consequences of this model are:

1. One half the children of parents with mixed preferences will be for one party preference, one half for the other.

2. Children of parents with the same preference will not be divided up in this way.

3. In the absence of any tendency to change (or of immigration and/or emigration) strong inbreeding will produce successively stronger preference for one party.

Similarities are developed between mathematical learning theory and the model. It is suggested that it is fruitful to consider further the parallel between learning and evolution to see if the same model will fit both.

DISTRIBUTION OF PATENTS ACCORDING TO NUMBER OF INVENTORS

Irving H. Siegel
The W. E. Upjohn Institute for Employment Research*

I

Sole Inventors Dominant

A striking feature of contemporary U. S. patent annals is the continuing importance of individual contributors -- of "sole inventors", as they are designated in legal parlance. Indeed, in firm after firm and in many technologically progressive or economically strategic industries, individuals still account not only for many more new patents than pairs, trios, or any other size-group of "joint inventors" but also for more new patents than all size-groups combined. Exceptions are, of course, also evident -- for example, in the chemical and pharmaceutical field, where collaborative patent activity has long been common. But it is the hardiness of the sole inventor that invites attention because the phenomenon may seem at variance with a complex of well-publicized trends.

Let us briefly note some of these trends. In recent decades, the proverbial garret inventor has largely been displaced by another stereotype -- a species of organization man. The typical modern inventor is a school-trained engineer or scientist. He works, not for himself, but for an employer -- say, a corporation, a government agency, a foundation, a university. He also relies on his employer for needed apparatus and instruments. He usually has an assigned task in a larger project, which in turn fits into a larger research and development program. Indeed, he frequently operates nowadays as a member of a team, performing a defined role in a joint mission with colleagues of the same or other disciplines.

Institutional and other Factors

Any serious effort to explain the continuing prominence of single patentees would have to range widely, to encompass psychological, legal, economic, and sociological, as well as technological, factors.[1] The topic of this paper does not require a methodical treatment of such factors, but several comments are offered for the benefit of readers who may wish to look behind the statistics. These comments suggest that the cards are not "institutionally" stacked in favor of sole inventors.

Surely, competitiveness is no less characteristic of creative people than of other kinds; and an individual is more likely than a group to be interested in patent recognition, to respond to incentive awards, to persevere in a quest for honors. Even in a team environment, one person often stands out in performance of his specialty; and, since the unit of invention is not legally rigid, a motivated individual whose colleagues are not patent-oriented may isolate his own creative contribution to a joint project, cast it into legally appropriate form, and seek public credit for it.

Companies and other organizations that have positive patent policies and adequate patent counsel probably encourage and facilitate joint application to a greater degree than they promote filing by sole inventors. Such organizations especially assist the prosecution of joint applications from teams completing their work or already broken up and redistributed among other projects; and they are also well equipped to act in instances in which some co-inventors either refuse to participate in a joint filing or cannot be located readily after a change in employment. Even companies that offer awards for disclosures and for issued patents

---

* The author's views do not necessarily represent positions of The W. E. Upjohn Institute for Employment Research. This paper, furthermore, is based primarily on studies conducted by the author for the Patent, Trademark, and Copyright Research Institute of George Washington University.

1/ For earlier discussion and references, see the following papers by I. H. Siegel: "Persistence of the Sole Inventor", Patent, Trademark, and Copyright Journal of Research and Education (later renamed IDEA), Summer 1961, pp. 144-149; "Individual and Joint Patent Production," ibid., Summer 1962, pp. 241-260; and "Dominance of Sole Patentees in Computer-Related Technology", IDEA, Spring 1964, pp. 45-50. In the preparation of the present paper, account was also taken of more recent information, such as that provided in Patent Counsel in Industry (Studies in Business Policy, No. 112), National Industrial Conference Board, New York, 1964.

are eager to minimize divisive staff competitiveness and accordingly prefer as generous a diffusion of credit as is feasible. Note should also be taken of the fact that the criteria of patentability, such as "usefulness", may in some fields (e.g., the chemical and pharmaceutical industries) oblige the division of labor, the distribution of tasks to persons best able to pursue them in the interest of the sponsoring company.

It is sometimes alleged that organizations attempt to "save" a joint contributor for possible use as an informed but "disinterested" witness in interference proceedings. If this strategy is indeed employed, it would seem more practicable when there are more than two actual inventors, in which case the proportion of recorded sole inventors is not affected. On the other hand, a patent may be voided if information supplied in an application is incorrect or incomplete; the U. S. law and its administrators frown on the malpractices of "nonjoinder" (improper omission of a inventor) and "misjoinder" (improper designation as a coinventor). 2/

II

We now turn to statistical evidence on the distribution of patents according to the number of recorded inventors. Advantage is taken here especially of material already presented in reports prepared under the auspices of the Patent, Trademark, and Copyright Research Institute of the George Washington University.3/ For convenience in presentation, the data are arranged in two ways, according to technological fields and by companies.

2/ See, for example, G. M. Naimark, A Patent Manual for Scientists and Engineers, Charles C. Thomas, Springfield (Ill.), 1961; and Rules of Practice of the United States Patent Office in Patent Cases, Washington, June 1960, pp. 20-22.

3/ See papers of I. H. Siegel mentioned in footnote 1; and Edgar Weinberg and I. H. Siegel, "Analysis of 203 Transistor Patents," Patent, Trademark, and Copyright Journal of Research and Education, Fall 1960, pp. 201-207. Additional material was obtained for the present paper, as the text indicates, from recent issues of the IBM Journal of Research and Development and from Index of Patents Issued from the U. S. Patent Office for 1963 and 1964.

Some Evidence for Fields

Let us first consider patents that largely fall into the electric-electronic category -- or simply "electrical", as it is designated by the Patent Office. About four-fifths of the items classified as "electrical" in the December 17, 1963 issue of the weekly Official Gazette were credited to sole inventors. Approximately the same ratio (79 percent) was derived in an analysis made during 1960 of the contents of the Patent Office file for transistors and related devices. An examination of the new patents selected for listing in 1962-63 issues of Computer and Automation, a monthly trade and technical magazine, revealed that single inventors accounted for about two thirds of the total, pairs of inventors for slightly more than a quarter, and larger groups for roughly one sixteenth. The share of single inventors in the patents listed in the same publication in an earlier year, 1956, was still higher -- about three quarters.

On another technological frontier, atomic energy, individuals also dominate, though not so overwhelmingly as in the electric-electronic area. An examination of 1266 patents released by the Atomic Energy Commission for royalty-free, nonexclusive licensing showed 56 percent credited to sole inventors, 31 percent to pairs, and 13 percent to larger groups of joint inventors (up to 6). A distribution of 250 additional patents released subsequently showed approximate corresponding percentages of 51, 30, and 19.

A frequency analysis of 888 chemical patents reported in the four weekly issues of the Patent Office's Official Gazette for June 1962 showed individuals still dominant but accounting for slightly less than half the total. Thus, sole inventors were credited with 48 percent of the patents, pairs for 36 percent, trios for 12 percent, and larger groups of joint inventors (up to 10) for only 4 percent.

Some Evidence for Companies

Proceeding to company data, we note first some electric-electronic examples. An updated calculation for International Business Machines Corporation, referring to the 2740 patents listed in the company's Journal of Research and Development for the period January 1957-July 1965, shows 69 percent attributed to sole inventors, 23 percent to pairs, and 8 percent to larger groups of joint inventors (up to 7). Fairly similar distributions are indicated by IBM's figures for shorter periods, and by figures for fewer patents relating to the Western Electric Company and the International Telephone and Telegraph Corporation. A higher ratio for sole inventors -- 73 percent -- is derived from information for patents assigned to General Electric Company, as reported in the 1963 Index of Patents.

The dominance of sole patentees in companies classifiable in the Patent Office's "mechanical" category also seems decisive. For example, individuals are identified with 77 percent of the patents listed for Ford Motor Company in the 1963 Index of Patents and with 64 percent of the patents listed for General Motors Corporation. In two tire companies, Goodyear and B. F. Goodrich, the individual shares in 1964, according to the Index, were 84 and 73 percent, respectively.

Results obtained for chemical and pharmaceutical firms are more equivocal. Individuals there still contribute more patents, as a rule, than any size-group of joint inventors; but conspicious exceptions are evident, and individuals frequently account for less than half the total patent output. For example, in 1964, individuals contributed about the same number of assigned patents as did paired inventors in W. R. Grace, Merck, Hooker, and Allied Chemical; and they contributed fewer than pairs did in Esso Research and Engineering, American Cyanimid, and Rohm and Haas. On the other hand, they showed clear dominance, accounting for over half the total, in Pfizer, Pennsalt, Air Products, Upjohn, and Norwich; and they also led all other size-groups in some companies, such as Monsanto and Air Reduction, where, however, they failed to produce,or barely exceeded, a majority.

The statistical variation from firm to firm (like the fluctuations also observed in data for the same company over time)4/ probably reflects important differences in (1) the mix of product and process research and (2) the way in which product research is organized, either for daily operations or for *ad hoc* exploitation of a chance discovery. To establish the usefulness of a "composition of matter" for patent purposes may require skills and backgrounds different from those that are effective in discovery. Legal ability, which may score tellingly in assuring the "novelty" and "unobviousness" required of a product or a process in patent law, also varies from firm to firm.

III

Variety of Distributions

The different company distribution patterns reported in this paper do not correspond to a single probability model, although it is possible, of course, to give a specious unity to the various imputable stochastic processes by writing a very general formula and then liberally manipulating the parameters.

All the observed distributions are skewed. They commonly resemble a reversed J, with the frequency for sole inventors equaling or exceeding half the total. Sometimes, as the data for the chemical and pharmaceutical companies show, the single-inventor bar of the histogram is shorter than, or about equal in height to the frequency column for paired inventors; and, in such instances, the frequency mass is still concentrated to the left, with the average number of inventors per patent remaining below 2. The tail seldom extends beyond 7 on the right, although larger groups of joint inventors are occasionally encountered. (Perhaps the largest number in patent history, 21, was recorded in 1963 -- for a compact computer system of the National Cash Register Company.)

Most of our frequency distributions follow what M. G. Kendall has called "the-higher-the fewer rule". In this connection, he cites the Zipf "least-effort"formula, the special case of this formula that is celebrated in Pareto's income law (which H. T. Davis has shown to be applicable also to many non-income phenomena), and the more comprehensive frequency function that H. A. Simon has christened the "Yule distribution". Among the other eligible models are the

---

4/ For example, sole inventors in Upjohn accounted for 50 out of 70 patents in 1964, while paired inventors contributed 20; they accounted for 48 out of 111 in 1949, while pairs contributed 42; and they accounted for 8 out of 13 in 1949, while pairs contributed 3.

truncated Poisson distribution (omitting the zero class)[5] and the geometric distribution.[6]

Geometric Law: IBM Data

As an empirical probability law, the geometric distribution appears fairly applicable to the electric-electronic category. In particular, a one-parameter version gives a close fit to the frequency data for International Business Machines Corporation, as may be seen from the accompanying table. The single parameter is P_1, the observed proportion of patents for sole inventors; and the percentage shares for this class and for larger size-groups are given by the formula

$$P_i' = P_1 \ (1-P_1)^{i-1},$$

where P_1 occurs a second time and $i=1,2,3, \ldots$.

In the accompanying table, the distribution observed for 2740 patents reported in the complete file of the IBM Journal through July 1965 is compared with the percentages computed according to the formula. Since $P_1 = 69.09$ percent (the observed proportion for sole inventors), the computed percentage for pairs of inventors is 21.36 (=69.09 x 30.91); for trios, 6.60 (=21.36 x 30.91); for quartets, 2.04 (=6.60 x 30.91); etc.

The closeness of the observed and computed ratios is impressive. An agreeable property of the computed figures is that they sum to 1 in the limit; the sum for the only occupied IBM classes, 99.97 percent, is virtually exhaustive. The derived theoretical mean number of inventors per patent, 1.45 ($=1/P_1$), is very similar to the observed (weighted) mean, 1.42. Of course, still better results are obtainable (for the variance as well as the mean) if refinements in the basic formula are introduced; but refinement means the addition of parameters, the reduction of degrees of freedom. Even though one would hardly claim that the geometric probability law represents the purified model of IBM's actual experience, there is at least an aesthetic charm in a one-parameter formula that permits good mental estimates to be made for all size-groups of joint inventors once the contribution of sole inventors is known.

Distribution of IBM Inventors According to Number of Inventors*

Inventors per patent	Actual Patents		Computed Patents, Percent (P_i')
	Number	Percent	
1	1893	69.09	69.09
2	624	22.77	21.36
3	164	5.99	6.60
4	43	1.57	2.04
5	13	.47	.63
6	2	.07	.19
7	1	.04	.06

* Data for actual patents were obtained from the entire file of IBM Journal of Research and Development, January 1957-July 1965. Computed percentages in last column were estimated from the formula presented in the text.

5/ An intriguing alternative to truncation is to regard the whole corpus of inventive activity as a Poisson system including a very large zero class. This class would then represent non-inventions -- e.g., ineligible rediscoveries and discoveries that are screened out by corporation committees and patent counsel or that fail for other reasons to progress to patent status. In this Poisson universe, the probability of of occurrence of invention is small, as it is in the real world. Against such a complete Poisson model, a truncated Poisson distribution refers only to the "tail" of successes (i.e., to patents awarded).

6/ On this paragraph, see M. G. Kendall, "Natural Law in the Social Sciences," Journal of the Royal Statistical Society, Part 1, 1961, pp. 1-16; H. A. Simon, "On a Class of Skew Distribution Functions," Biometrika, December 1955, pp. 145-164; G. M. Kaufman, Statistical Decisions and Related Techniques in Oil and Gas Exploration, Prentice-Hall, Englewood Cliffs, 1963, pp. 107, 113-114; H. T. Davis, Theory of Econometrics, Principia Press, Bloomington, 1941, pp. 23-51; and E. A. G. Knowles and D. S. Stewart, "Characterisation of the Flow of Events -- A Problem of Simulation," Applied Statistics, June 1963, pp. 113-128.

THE STATISTICAL DESIGN AND ANALYSIS OF AN EXPERIMENT TO MEASURE THE EFFECTIVENESS AND COSTS OF A HEALTH AND WELFARE PROGRAM

Julius A. Jahn, University of Pennsylvania

As a part of a research and demonstration project concerned with Special Services for the Aging, in the Community Service Society of New York, an experiment was planned to answer questions concerning the effectiveness and costs of some alternative programs. These questions were concerned with the relative effectiveness of several alternative service programs that could be experimentally tested, and also, with ways in which the effectiveness of these programs could be increased.[1]

The major steps or phases involved in this experiment were as follows:

I. Definition of the Problems for Experimentation.
II. Formulation of Hypotheses to be tested by Experimentation.
III. The Statistical, Experimental Tests of the Hypothesis.
IV. Derivation of Hypotheses from the Experiments.
V. Application of Hypotheses that have been Experimentally Tested or Derived.

I. DEFINITION OF THE PROBLEMS FOR EXPERIMENTATION

A. THE GIVEN CONDITIONS

Definition of the Population. The effectiveness and costs of programs of "Services for the Aging" were measured for a population consisting of persons 60 years of age or over in two major subpopulations or strata. The first, referred to as the "Applicant Stratum," included persons in the population for whom applications were made for service to either the Special Services for Aging Office or in the Central Office of the Community Service Society. The other stratum, referred to as the "Non-Applicant Stratum," includes all other persons in the population for whom such applications were not received. The term "Participant" was used to refer to any member of this population. In the "Applicant Stratum" there were four criteria defined for identifying "Participants:"

(1) Persons who are 60 years of age or over at the time of application.

(2) Persons who are making an inquiry or request (or for whom an inquiry or request is made) for service to the Special Services for Aging Office, either for themselves or for another person, and whether in-person, by telephone or by letter.

(3) Persons whose permanent place of residence at the time of application is not outside Manhattan, Bronx or Queens Borough of New York City, or is not in a "congregate" residence for aged persons; and, who do have a residence within these three boroughs (except for applicants for "Ward Manor").

(4) Persons who are not included in cases currently opened for service in a Community Service Society district office or cases which have been closed within the current month in the Special Services for Aging office.

In the "Non-Applicant Stratum" only two of these criteria were applied: (1) The ages of persons as of the date they were first listed for sampling during any given month of the period for study, and (2) place of residence.

Definition of Alternative Programs. In this paper only certain distinctive conditions involved in the alternative programs will be indicated, as more complete definitions are provided elsewhere.[2]

Program 0, the "Basic" program, included those health and welfare services that would be available to persons in the population, excluding those in the two alternative programs.

Program 1, "Short Term Service" included services by either a caseworker or a public health nurse within the Special Services for the Aging unit of the Community Service Society. The service was to be completed within two months from the date of case opening and was not to require more than four in-person interviews

Program 2, "Collaborative Service," included the possibility of services by

[1] Footnotes are appended to end of this paper.

either or both caseworker and public health nurse within the Special Services for the Aging unit. No restriction on the time or number of in-person interviews was given.

Measures of Effectiveness of Programs. Of the many possible criterion variables that might have been selected, the final choices involved the use of both personal judgments and statistical methods of analysis. Two major steps were involved in making the selection. The first stage involved selecting questions to be included in the research schedule. For purposes of both interviewing and analysis, the various questions were classified into seven "content areas," identified as: (1) Housing, (2) Occupation, (3) Financial, (4) Physical and Mental Health, (5) Personal Adjustment, (6) Interpersonal Relations, and (7) Recreational or Social Activities and Interests.

The selection of questions for inclusion in the research schedule was primarily based on the judgments of the research staff. These judgments were made with reference to three sources: First, the particular goals and policies of the Community Service Society with respect to providing services for aging persons. Second, the kinds of problems indicated by the requests for service by persons applying to the Community Service Society. And, third, research and professional literature on problems of aging and programs for aging, and reports of public meetings of professional and lay persons especially concerned with problems of aging persons. The specific questions were selected to provide objective definitions of problems of aging persons which seemed to be explicitly referred to in the various statement found in these sources.

The second stage was carried out in order to reduce the number of criterion variables from the approximately 600 included in the final research schedule to a relatively small number to use for measuring effectiveness of alternative service programs. The second stage was carried out using procedures to avoid the obvious error of selecting those criterion variables for which the differences were in the "right direction."

In this study it has not been possible to arrive at the selection of one particular criterion variable or group of criterion variables which were acceptable to all persons involved. Instead, four different groups of criterion variables were selected by somewhat different conditions and methods:[3]

C.V. Group 1. Adaptive and Adjustive Status. This is a modification of scales developed and used in previous studies by the Institute of Welfare Research of the Community Service Society.[4] In the present study, the ratings were based on the direct observation, interviews and judgment of research interviewers, and not on previously recorded service records. The ratings were made and recorded at the time of an initial research interview and again at the time of the follow-up interview six months later.[5]

In order to simplify analysis, "The Overall Level of Adjustment and Adaptation," was selected as a criterion variable in this group. The ratings were recorded using number 1 through 11 to indicate ratings from "Very Poor" to "Excellent." For uniformity in analysis the order of the numerical ratings has been reversed so that 11 represents "Very Poor," and 1 represents "Excellent." This rating was selected as a criterion variable in this experiment because the "Movement Scale," from which it was derived, has been previously used in research studies to measure effectiveness of casework services in the Community Service Society and other social agencies.

C.V. Group 2. The Basic Problems of Aging. The criterion variables in this group were selected by personal judgments of the research staff that these variables would be accepted as problems by most of the persons expected to support programs dealing with "problems of aging." The variables are coded so that "1" represents the occurrence of a "problem" for the aged person, and "0" absence of a problem.[6]

C.V. Group 3. Problems of Aging Applicants for Services. The third group of criterion variables was selected statistically, by comparing Applicants assigned to Program 1 or 2, with Non-Applicants at the time of the initial research interview. The problems which lead persons to apply were expected to occur more frequently among the Applicants than among the Non-Applicants. The variables were selected from all those which had been coded for IBM tabulation from the research interviews or from combinations of such coded variables. Variables were considered as "potential criterion variables" if the following considerations were satisfied: First, the variable provided a basis for ranking persons from more or less "favorable" conditions; second, the variable was included in both the initial and the follow-up research schedules; and third, the variable was so defined that changes could occur in the rank order of persons between the initial and the follow-up research interview.[7]

Minimum Statistical Conditions for the Selection of Criterion Variables. In order for any of the potential criterion variables to be selected for the specified group, five statistical conditions were defined that were to be satisfied according to the results of the statistical analysis. (1) The percentages of persons within some of the subcategories of the potential variable must be higher for the applicants than for the non-applicants. Those subcategories for which this condition is satisfied will be referred to as "problems" of the aging applicants. The presence of such problems was coded by the number 1 and their absence was coded by the number 0. (2) The frequency of persons in the "problem" sub-categories among the non-applicant sample must be lower than 50%. (3) The frequency of the "problem" sub-categories must be higher for persons 65 years of age or over than for persons 60 through 64 years of age in the population sampled. (4) The frequency of the "problem" subcategories must be higher for the applicants than for the non-applicants within either one of the two groups, those 60 through 64 or those 65 or over. (5) The frequencies of the "problem" sub-categories based on the initial interview must be equal to or less than the corresponding frequencies of these problem subcategories based on the follow-up interview for the non-applicant sample.

These five conditions were applied to select one criterion variable in each of the seven content areas. If only one of the potential criterion variables in an area satisfied all the conditions, it was selected as the criterion variable for measuring effectiveness in that area.

The Optimal Statistical Condition for Selecting Criterion Variables. If more than one of the potential criterion variables in an area satisfied these five conditions, one was chosen for which the percentage of applicants included in the "problem" sub-category was highest. In case of ties using this condition, the one chosen was that for which the percentage of non-applicants who were in the "problem" sub-category in the initial interview but were not in this sub-category in the follow-up interview, was the lowest. In order to allow for errors due to sampling, differences of five percent or less were considered to be ties.

C.V. Group 4. Problems of Aging for Non-Residential Services. The fourth group of criterion variables was statistically selected using the same methods as for the third group just described. However, the methods were applied to information obtained for applicants selected for Program 0, who were referred to the district offices of the Community Service Society, and Non Applicants. The persons referred to the district office were selected by random sampling from all applicants after excluding certain of the applicants from the referral because of agency policies. Any applicant for whom the request for service was concerned with the residences for aging operated by the Society or the summer camping program for aging persons, was not referred to the district office. Since persons selected for Program 0 were interviewed only at the time for follow-up, it was necessary to use data based on the follow-up rather than on the initial interview.[8]

With respect to each of these criterion variables, persons were classified into two sub-categories depending on the presence or absence of "problems." The symbol P_{jst} will be used to represent the proportion of persons in the population who would be included in the "problem sub-category" with respect to a specified criterion variable "j" assuming that all of the persons had been selected for a particular program "s" during a particular period for experimentation "t;" each of the measures of effectiveness can be defined in terms of the following statistical index:

$$E_{jst} = \frac{P_{jot} - P_{jst}}{P_{jot}}$$

Measures of Costs of Programs. In this experiment two measures of cost have been used: (1) The mean length of time between the beginning and termination of services. (2) The mean number of in-person interviews by a member of the professional staff. The symbol C_{kst} will be used to represent any defined measure of costs of programs. The value of this measure will depend on the way in which cost is defined, represented by the subscript "k", as well as upon the particular program represented by the subscript "s", and upon the period of time represented by the subscript "t."

Uses of Measures of Result of Experimentation. These measures of effectiveness and costs had three major purposes: To select one of these alternatives as the "best" for most persons in the population until further experimentation may change this conclusion; to select any one or more of the programs for repeated experimentation; to suggest improved programs which may be expected to satisfy certain minimum levels of effectiveness.

In this experiment the minimum level of effectiveness was defined to be equal to the value computed for a specified criterion variable, assuming the frequency of problems among Non-Applicants 60 and over is reduced to the level occurring for those 60 to 65, and that the frequency of problems among those in the Applicants' sub-population is reduced to the level occurring in the Non-Applicant sub-population.

B. THE QUESTIONS TO BE ANSWERED BY EXPERIMENTATION

In order that the defined measures of effectiveness and costs could be used for the purposes of selecting and improving programs of services, it was necessary for the values of these measures to be determined by hypotheses which have been or can be experimentally tested. The unknown values of the measures of cost and effectiveness that are required can be expressed in the form of questions:

First, with respect to the purpose of selecting alternative possible programs for maximum utilization, two questions had to be answered. (a) Which of given possible alternative programs is most effective with respect to all selected measures of effectiveness (E_{jst})? (b) If more than one of the given programs is most effective, which one is least costly with respect to the measures of cost (C_{kst})?

Second, with respect to selection of programs for further experimentation, which of the experimental programs is more effective than the "Basic" program with respect to reducing the mean number of "problems" for a specified group of criterion variables?

Third, with respect to the development of improved programs the following questions had to be answered. (a) Do any of the existing programs satisfy minimum conditions specified with reference to each of the measures of effectiveness? (b) If not, how can these programs be modified or developed to increase the degree of effectiveness with respect to defined measures?

II. HYPOTHESES TO BE EXPERIMENTALLY TESTED

The hypotheses that were formulated prior to experimentation were so formulated that acceptance of the stated hypothesis would lead to the selection of the "Basic" program pending further experimentation.

1. Hypothesis for the Selection of the "Best" Program

Hypothesis 1(a) - The effectiveness of Programs 1 or 2 will not be greater than that for Program 0 with respect to all of the defined criterion variables.

2. Hypothesis for Selection of Programs for Further Experimentation.

Hypothesis 2(a) - In reducing the mean number of "problems" in criterion variable Group 1, Group 2 and Group 4, the effectiveness of Program 1 will be equal to or greater than Program 0.

Hypothesis 2(b) - The effectiveness of Program 2 will be equal to or greater than Program 0.

3. Hypotheses Concerning the Development of Programs

Before some experimentation was carried out, it was not possible to formulate hypotheses with reference to specific ways in which programs can be improved in effectiveness and costs.

III. STATISTICAL-EXPERIMENTAL TESTS OF HYPOTHESES

Major Stratification: For the purposes of sampling, the defined population was sub-divided into major strata; (1) The "Applicant" stratum consisted of those persons in the defined population for whom an application was made for services in the Community Service Society. This strata also included any other person in the population living in the same household with such "Applicants." (2) "Non-Applicants" consisted of those persons in the population for whom no such application was made at the time selected.

The Master Samples. Within each of these strata, "master samples" were selected by random-probability sampling methods.

The first step in selecting an "Applicant" Sample was listing all persons for whom an application is received in the Special Services for the Aging office during a given month. All persons were selected for whom a case was opened for service between the date listed and the final date for the follow-up interview about six months later. A sample for all other Applicants was selected by simple random sampling within each ten consecutive listed names.

Three Non-Applicant samples were selected using lists of blocks, block segments, or dwelling units available from the United States Bureau of Census reports and maps provided by the New York City Housing Authority.

Multiple Samples for Months of Study. The specified sampling methods were applied to select a series of samples for each of the eight months during which the study was carried out, from April 1 to December 1, 1957. The number of persons sampled in any given month depended on the number of persons who could be served by the available staff. For the Applicants, samples were selected each month from all those for whom applications were received during the month. For the Non-Applicants, some samples of blccks, block segments, or dwelling units were selected each month and the "Participants" residing in the sampled area were listed and interviewed by the research staff.

Random Selection of Persons for Alternative Service Programs. Within each of the "master samples" of Applicants and Non-Applicants, three sub-samples were selected by random-probability sampling. Each of three alternative service programs were preassigned to one of these three sub-samples. Consequently, each of the programs were made available to persons in one of the "sub-samples" selected from the "master samples" previously selected from the defined population.

Random Selection for Interviewing. Not only were "Participants"sampled for alternative programs, but in addition they were randomly sampled for assignment to various persons in the interviewing staff. Each month a number of persons were selected and preassigned to each one of the interviewers. Some changes in assignment were made after the random assignment under special conditions. No substitutions were made.

Size of Samples and Probabilities of Selection. The size of samples and probabilities of selection were predetermined for each of the monthly periods of interviewing on the basis of two general conditions: (1) The total number of persons sampled during the twelve month period was to be sufficiently large to apply statistical tests that assume large samples: (2) The maximum number to be sampled and interviewed in any given month was not to exceed a number which could be interviewed and served by the research and service staff in conformity to the standards and procedures defined for the experiment.

The number of persons sampled for the initial interviews in the Applicant Sample was 119 and in the Non-Applicant Sample was 133. The number sampled for the follow-up interview in the Applicant Sample was 217, and in the Non-Applicant Sample was 84.

The information used to classify persons according to the criterion variable for measuring effectiveness was obtained by the use of standardized research schedules. The two basic schedules were the Program Evaluation Schedule, used for interviewing the Participant to obtain selfratings and objective information concerning the Participant and others in the household, and the Research Interviewer Schedule used by the interviewers to record their observations and judgments.

Definitions and instructions for the administration of the research schedules were used in the training and supervising the interviewing process. In general, they required that the interviewing be attempted and completed within the first month after selection and again within six months after selection. For Applicants sampled for Program 0 (District Office Service) no interviewing was attempted until the time for the six month follow-up. The place for interviewing was to be within the home of the Participant or in the office of the Community Service Society.

An unlimited number of attempts could be made within a specified time period for interviewing. If the interview could not be completed within one month, two weeks additional time was allowed in order to complete the contact if possible. It was not expected that all persons be interviewed. However, it was expected that at least three attempts be made to interview. If the interviewer was unable to complete the interview within the specified time period, the reasons for noncompletion were specified.

The research interviewers were selected from persons with previous social work training and experience, but in this study they were selected, trained and used only for interviewing and not for any service functions.

The general policies for the administration of the service program were those already in practice within the Community Service Society, but some modification of these policies was required by the experimental design. Written instructions were developed for these policies. The service workers for Programs 1 and 2 were selected, trained and supervised in carrying out the services according to these policies for approximately one year before the experimental study was begun. The same persons were involved in carrying out both Programs 1 and 2 in the Special Services for the Aging unit. A different service staff carried out Program 0 in the district offices.

In order for any person to actually receive any service, it was first necessary for an application to be made and for the case to be opened for service according to the usual policies and practices of the Community Service Society. Persons initially interviewed by research staff who were sampled for Program 1 or 2 from the "Non-Applicant" stratum of the population, were "informed" about the availability of these services at the Community Service Society by the research interviewer. However, the research interviewer was not to enter into a service relationship with the persons interviewed nor to encourage them to apply, except in case of special emergencies.

The hypotheses about the effectiveness of alternative programs in this experiment were applied in making two kinds of decisions. (a) The first decision was to select one of the alternative programs or the "best" for maximum utilization by the agency in providing services for persons in the defined population. (b) The second decision is to select one or more of the alternative programs for further experimental testing. Accordingly, two somewhat different rules for testing the hypotheses will be defined and applied.

<u>Statistical Tests of Hypotheses in the Selection of Programs for Maximum Utilization</u>. As the basis for testing this hypothesis, results of research interviews carried out in the six-month follow-up period for Participants sampled within the "Applicant" stratum during April 1 to December 1, 1957 were used. Those Participants identified as "exceptional" cases because they could not be referred to the Program 0 - District Office Service Program were excluded from comparison of Program 1 and Program 2 with Program 0. Results of interviews for the "Non-Applicant" samples were not included because only one sampled Non-Applicant requested service within the six-month period for the follow-up research interview. The numbers of sampled Participants included were: For Program 0, $n_0 = 68$; Program 1, $n_1 = 44$; and Program 2, $n_2 = 22$.

In order to classify persons on the basis of information recorded in the research schedule, it was necessary (1) for the persons to be located, and alive and (2) for a research interview to be at least partially completed. It was expected that the frequencies of persons for whom information was not obtained for these reasons would be sufficiently large to make possible separate analysis, and that, these frequencies could be effected by the service program as well as by the procedures used in the follow-up interviews. In order to simplify the analysis, information about these reasons were coded in terms of the following: (1) The Participant was located and alive up to a date three months from the date sampled. (2) Research data was obtained that was sufficient to permit classifications of the sampled Participant on some of the possible criterion variables.

Program 0 is to be recommended for maximum utilization unless a Null-hypothesis 1 (a) 1; that the effectiveness of Program 0 is equal to or greater than that of Program 1, and of Program 2 is rejected. This hypothesis was to be rejected if the outcome of the experiment is as follows:

(1) For comparisons between Program 1 and Program 0: (a) The estimated percentage of persons located and alive three months after being sampled for Program 1 is greater than the corresponding percentage of Program 0. That is, $P'(v)_1 > P'(v)_0$; and (b) Among those located and alive, the percentage for whom research interview data was obtained for Program 1 is greater than the corresponding percentage for Program 0. That is, $P'(u/v)_1 > P'(u/v)_0$; and (c) Among those persons for whom research data was obtained, the estimated percentage of persons with a specified problem for Program 1 is less than the corresponding percentage for Program 0, for each of the selected criterion variables in Groups 1, 2 and 4. That is, $P'(j/u)_1 < P'(j/u)_0$; and (d) The computed value for the "Normalized t-test" for all criterion variables in a selected group is equal to or greater than a specified value "C"= 1.28[10]; or

(2) For the corresponding comparisons between Program 2 and Program 0:

(a) $P'(v)_2 > P'(v)_0$,

(b) $P'(u/v)_2 > P'(u/v)_0$, and

(c) $P'(j/u)_2 < P'(j/u)_0$

for all criterion variables in a selected group; and

(d) The computed value of the "Normalized t-test" is equal to or greater than specified value "C" = 1.28.[10]

If this hypothesis was rejected, Program 1 was to be recommended for maximum utilization if a Null-hypothesis 1(a)2, (that is, if the effectiveness of Program 1 is equal to or greater than that of Program 2), was rejected.

Program 2 was to be recommended for maximum utilization if both Hypotheses 1(a)1 and 1(a)2 were rejected by the results of the experiment.

Comparisons between Program 1 and Program 2 were to be made with reference to criterion variables in Groups 1, 2 and 3; whereas, comparisons between Program 1 or Program 2 with Program 0 were to be made with reference to criterion variables in Groups 1, 2 and 4.

Statistical Tests of Hypotheses in the Selection of Programs for Further Experimentation. The choice of programs for further experimentation was not restricted to the choice of only one of the alternatives. Any one or more of the alternative programs could be selected for further testing by the following rule:

(1) Whichever program is selected as a result of the application of the rule defined in the preceding part will be initially selected for further experimentation.

(2) Each of the remaining programs will be excluded from further experimental tests if the Null-hypothesis 1(a)3, (that is, if the effectiveness of this program is equal to or greater than that of the initially selected program), is rejected for each program.

This Null-hypothesis will be rejected if the results of the experiment are as follows:

(a) The estimated mean number of problems among persons sampled for the alternative program is greater than corresponding estimated mean for persons sampled for the initially selected program: That is, $M'(x)_s > M'(x)_{s*}$; and (b) The estimated percentage of persons for whom research data is not obtained for the alternative program is greater than the corresponding estimated percentage for the initially selected program: That is, $P'(u)_s > P'(u)_{s*}$; and (c) The computed value for the Normalized t-test applied to this difference in percentages in condition (b) is greater than "c" = 1.28.[10]

(3) Each of the remaining programs will be included for further experimental testing, if the Null-hypothesis 1(a)3 is

rejected for that program, according to results of the experiment that have been carried out using the rule for rejection defined in the preceding section.

Results of Statistical Tests of Hypotheses for Selection of the "Best" Program for Maximum Utilization. The results of the application of the defined rules for testing hypotheses to the data obtained in this experiment were as follows:

(1) When effectiveness was measured by C.V. Group 1, (Overall Level of Adjustment and Adaptation), the Null-hypothesis that the effectiveness of Program 0 is equal to or greater than that for both Program 1 and Program 2 was not rejected.

(2) When effectiveness is measured by the criterion variables in Group 2, (Basic Problems of Aging), this Null-hypothesis was also not rejected.

(3) When effectiveness is measured by the criterion variables in Group 4, ("Problems of Applicants for Non-residential Services") this Null-hypothesis was also not rejected.

Accordingly, Program 0 would be selected as the "Best" Program for maximum utilization.

Tests of Hypotheses for Selection of Programs for Further Experimentation. The results of the application of the defined rules for testing these hypotheses are shown in Tables B1 and B2.

(a) The Null-hypothesis, (that the effectiveness of Program 1 is equal to or greater than that of Program 0), was not rejected when the criterion variables in Group 1, Group 2 or Group 4 were selected to measure effectiveness. Accordingly, Program 1 would be selected for further experimentation, based on the criterion variables in Groups 1, 2 or 4.

(b) The Null-hypothesis, (that the effectiveness of Program 2 is equal to or greater than that of Program 0) was rejected when criterion variables in Groups 1, 2 or 4 were selected to measure effectiveness. Accordingly, Program 2 would not be selected for further experimentation based on the criterion variables in these Groups.

FOOTNOTE REFERENCES

1. The work in preparing this report was supported by funds from U.S. Public Health Service Grant 2M-6358. For more complete reports of the experiment see: Blenkner, Jahn and Wasser (1), Blenkner and Sibulkin (2) and Jahn and Blenkner (4). The research was supported by a grant from the Rockefeller Brothers Fund and by funds from the Community Service Society of New York.

2. See Blenkner, Jah and Wasser (1), and Blenkner and Sibulkin (2).

3. See Jahn (4) pp. 37-39 for a discussion of sources, procedures and results of the statistical selection of criterion variables. The sources from which items were selected or derived as potential criterion variables are given in Appendix A-5. For a more generalized discussion of the problems, principles and methods involved see Jahn (5).

4. See Kogan (6) for a discussion of the development of the "Adjustive Status Scales."

5. See Jahn and Blenkner (4) for the form used in making these ratings.

6. See Jahn and Blenkner (4) for definitions of the criterion variables included in Group 2.

7. See Jahn and Blenkner (4) for definitions of the criterion variables included in Group 3.

8. See Jahn and Blenkner (4) for definition of the criterion variables in Group 4.

9. For more complete presentation of basic principles and methods for sampling, estimation and testing hypothesis see Hansen (3) Neyman (7) and Jahn and Blankner (4)

10. See Jahn and Blenkner (4) for a definition of the "Normalized t Test" applied in this experiment. The value .05 has been selected as the probability for rejecting any "Null-hypothesis" assuming it is true, based on the specified research design and procedures.

BIBLIOGRAPHICAL REFERENCES

1. Blenkner, Margaret, Jahn, Julius and Wasser, Edna. Serving the Aging: An Experiment in Social Work and Public Health Nursing. New York: Community Service Society of New York, March 1964.

2. Blenkner, Margaret and Sibulkin, Lillain. Experimental Research in Social Work: A Report of a Study in Services to the Aging. Community Service Society of New York, June 1959.

3. Hansen, M.H., et al. Sample Survey Methods: Vol. I. New York, John Wiley & Sons, Inc., 1953.

4. Jahn, Julius and Blankner, Margaret. Serving the Aging: Methodological Supplement - Part I. Experimental Design and Statistical Tests of Hypotheses.New York: Community Service Society of New York, March 1964.

5. Jahn, J.A. "The Statistical Definition and Selection of Criteria for Evaluating Effectiveness in Social Work." University of Pennsylvania School of Social Work, Research Center, 1961.

6. Kogan, Leonard and Shyne, Ann. A Study of Components of Movement. New York: The Community Service Society, 1957.

7. Neyman, Jerzy. First Course in Probability and Statistics. New York: Holt, Rinehart & Winston, Inc., 1950.

VARIATIONS IN RESPONSE ERRORS INDUCED BY CHANGING INSTRUCTIONS TO ENUMERATORS

Charles H. Proctor, North Carolina State University

0. Abstract

This article reports on some research into response errors that was done on the Agricultural Enumerative Surveys in North Carolina. The data on response errors were collected by comparing farm information obtained by follow-up enumerators with that turned in by initial enumerators. These initial enumerators varied their interviewing procedures in ways that were suspected to produce varying amounts of discrepancy. Three kinds of discrepancies were distinguished and a probability model of the data collection operation is introduced. The model has three parameters: π_s-stability, π_v-vigilance, and α-scatter, to represent the three kinds of discrepancy.

We calculate relative mean square error as a function of π_s, π_v and α and then suggest cost functions to use in optimizing the levels of the three. This reasoning goes beyond our data in that the cost parameters have not yet been estimated. We then return to the data of the experiment-in-a-survey and indicate those variations in interviewing techniques which seemed to result in more favorable levels of the response error parameters. These, in short, are the official rather than the friendly approach and the field-also rather than the house-only location instruction.

1. Introduction

As large-scale surveys become more widely used, the response error portion of the variance is getting more attention in survey design. This is quite understandable since large sample sizes have reduced the uncertainty arising from sample selection, while the use of supplementary information, as in ratio estimates, has further cut down the sampling error. Thus, although response error or measurement error may have been only of minor importance in early stages of survey work or with small samples, now it has grown in relative prominence. This is true even in factual surveys such as agricultural enumerative surveys of crop acreages and production and of livestock numbers.

The present study, an experimental survey, grew out of the interest of the Standards and Research Division, Statistical Reporting Service, USDA and the North Carolina State Statistician's office who wanted to learn about how the data were gathered during the June and December Enumerative Surveys and how this affected its quality. Two of us at the Institute of Statistics, starting in 1961, talked with supervisors, accompanied enumerators and tried interviewing using the forms and instructions of these surveys. We gradually came to recognize an "optimum interviewing style". This may be defined as a course of action by the enumerator which would somehow squeeze the most information from the situation presented at each sample segment.1/

This style we found embodied in supervisory enumerators and other "successful" enumerators who were fully acquainted with the schedules and the importance of their work, and who were lucky enough to adopt just the right way of handling problems in conversation with farmers.2/ Such a subjective conception is, of course, too hazy for precise experimental control but we decided to use the results of supervisory enumerators as a standard to compare to those of other enumerators working under a variety of instructions. Thus in the discrepancies between the results of two enumerations of the same tract we planned to measure "response error", and here we met quite a problem.

At each tract two numbers were recorded.3/ One, to be written x, was obtained by the non-supervisory, initial enumerator under particular instructions and the other, to be written x', was found by the supervisory, follow-up enumerator under optimum interviewing style. The problem was to characterize these differences in some convenient, and hopefully, suggestive way so as to reflect the main components of response errors. The following "solution" may look technical, and it is, but it is also very close to the data - a grossly empirical approach.

1/The sampling unit is a land area for these surveys. This area is used either in the "closed segment" approach or the "open segment" way. Under the open segment definition all land operated by only those persons living in the sampled area is covered in the questionnaire, while under the closed segment approach all acreages and livestock on only the sampled area are covered. A tract is defined as that portion of a closed segment which is operated by one person.

2/Our findings at this stage were reported in Progress Report 31, of the Institute of Statistics and USDA. This report covered the work done from August 1961 to February 1962. These semi-annual mimeographed reports, which we will refer to only by number from now on, are available upon request to the Institute of Statistics, North Carolina State University, Raleigh, N. C.

3/In point of fact, two collections of "pencil patterns" were recorded and we in the office translated these to the numbers x and x'. We did not use in full the editing instructions for these surveys and omitted "don't know" responses from many of the comparisons. Progress Report No. 33 described our early attempts to tabulate data from initial and follow-up interviews.

2. Response Error Model

To be specific let's consider livestock numbers, e.g. "cattle and calves of all ages". In some cases the x and x' values were identical- a stable response. Where they were not equal, two types seemed worth distinguishing: those having $x = 0$ with $x' \neq 0$ (the reverse case $x \neq 0$ and $x' = 0$ was rare but present, although we will ignore it for now) and those having $x' > x$ and both non-zero (again the reverse case of $x > x'$ will be ignored for now since it was infrequently observed). When both x and x' were non-zero the variation in the difference was greater the greater the level of the x and x' values.

These three types of discrepancies can be represented by a random process with three parameters, π_s, π_v and α.4/ The following fanciful story shows the process. Suppose the farmer has some cows. When the enumerator starts to talk with the operator about a tract he takes out a short roofing nail, flips it and if it lands point up, he writes "no cows" without bothering to ask the farmer. If the nail lands on its side (which it does with "high" probability, π_v) he then gives the farm operator a thumbtack. The farm operator flips the thumbtack and if it lands point up (which it does with probability π_s) he tells the enumerator the "correct" number but if it lands on its side he selects a number r say in the range 0 to α (where α is something like .8) and multiplies the "correct" number by 1 - r and tells the enumerator this "shrunken" result.

An initial enumerator who operates with $\pi_v = 1.00$ and either $\pi_s = 1.00$ or $\alpha = 0$ (or both) will produce results as free from response error as a follow-up enumerator. Deviations of these parameters from their extreme values shows deterioration of interviewing. Using rather naive techniques of estimation and a sample of 108 tracts we estimated $\pi_v = .95$, $\pi_s = .60$ and $\alpha = .8$ for "cattle and calves of all ages on (this tract)".

3. Relative Mean Square Error

Having such a model of response errors one can calculate the influence of changes in the parameter values on, say, the mean square error of estimates and also design and conduct experiments to see how easily or painfully the parameter values can be improved. We began work in both of these directions and will describe the findings shortly. Another more pressing problem would seem to be to find alternative, possibly simpler and more realistic models. Although we feel that this is an important job we left it undone.

To calculate the mean square error of the mean of n x-values from farms with non-zero x' value we partition the square as follows:

$$(1)\quad E(\bar{x}-E(\bar{x}'))^2 = E[\bar{x}-\bar{x}') + (\bar{x}'-\mu')]^2$$
$$= E(\bar{x}-x')^2 + 2E(\bar{x}-\bar{x}')(\bar{x}'-\mu') + E(\bar{x}'-\mu')^2 .$$

Here we are using $E(\bar{x}') = \mu'$. The last term is the sampling variance of the optimum interviewing value mean, which we may call σ^2/n. The quantity $\bar{x}-\bar{x}'$ can be written as $\frac{1}{n}\sum_{j=1}^{n}(x_j-x_j') = \frac{1}{n}\sum_{j=1}^{n} Z_j x_j'$ where the Z's are random quantities, sort of percent bias quantities, whose distributions are given by the response error model and its parameters. In evaluating $(\frac{1}{n}\sum_{j=1}^{n} Z_j x_j')^2$ we will require to calculate $E(Z_j)$ and $E(Z_j^2)$ when x_j' is non-zero.

Notice that $Z_j = -1$ if there is lack of vigilance (because $x_j = 0$ while $x_j' \neq 0$), $Z_j = 0$ if there is stability and vigilance and $Z_j = -r$ if there is vigilance but not stability. We took the distribution of r to be "parabolic" rather than uniform and in this case $E(r) = .40\alpha$ with $E(r^2) = .32\alpha^2$. That is, the frequency function for r was chosen as $f(r) = 1.5(r-2)^{1/2}\alpha^{-3/2}$ for r in the range 0 to α. We used the maximum observed value of r to estimate α.

$$(2)\quad E(Z_j) = 1(1-\pi_v) + 0(\pi_v\pi_s) + E(-r)\pi_v(1-\pi_s)$$
$$= [1 - .40\alpha(1-\pi_s)]\pi_v - 1 = B \text{ say.}$$

$$(3)\quad E(Z_j^2) = (1-\pi_v) + .32\alpha^2\pi_v(1-\pi_s) = C \text{ say.}$$

$$(4)\quad E(\bar{x}-\bar{x}')^2 = (\tfrac{1}{n})^2 E\left[\sum_{j=1}^{n} Z_j^2\, x_j'^2 + \sum_{i\neq j}\sum Z_i Z_j x_i' x_j'\right]$$
$$= (\tfrac{1}{n})^2[Cn(\sigma^2+\mu'^2) + B^2 n(n-1)\mu'^2]$$
$$= \tfrac{1}{n}[C(\sigma^2+\mu'^2) + (n-1)B^2\mu'^2] .$$

The middle equals sign in (4) depends on the Z_j's and x_j''s being independent and doing the sampling with replacement - neither of which are true strictly, but actual cases will not be too far from this. By a similar computation the middle term of equation 1) becomes $2B\sigma^2/n$ so that the mean square error becomes $(C+2B)\sigma^2/n + C\mu'^2/n + B^2(n-1)\mu'^2/n + \sigma^2/n$ .5a/

4/Progress Report 34 describes a more elaborate model of which the present one is a special case.

5a/For the general case when only a proportion, P say, of the farms have non-zero x' values this mean square error is $P\left\{(C+2B+1)\frac{\sigma_+^2}{n} + (C-PB^2+2BQ+Q)\frac{\mu_+^2}{n} + PB^2\mu_+^2\right\}$ where σ_+^2 and μ_+ are the variance and mean of the non-zero x' values.

The relative mean square error appears when we divide this by μ'^2. Thus rel-MSE = $(C+2B)V^2/n + C/n + (n-1)B^2/n + V^2/n$, where V^2 is σ^2/μ'^2, the population rel-variance of the optimum interview results.

4. Cost and Worth Functions

The quantity B^2 so dominates the result, when n is even moderate in size, that in practice we would be concerned to reduce it first.5b/ The further B^2 is reduced the more valuable are the results, while the closer π_v and π_s are to 1.00 and α to zero the more costly, presumably, is the interviewing. If we could provide these cost functions then maximizing the value of the survey's worth would give us some notion of what level of vigilance, stability and scatter we should aspire to.

In our experience vigilance, π_v, was .95 and to reduce the .05 of non-vigilance remaining to .025, say, would seem to me to require doubling the cost. Getting $\pi_v = 0$ is almost free and to obtain $\pi_v = 1.00$ would be priceless. The function $\$.05/(1-\pi_v)$ which becomes very large as π_v approaches 1 and doubles from \$1 to \$2 as π_v goes from .950 to .975 may serve to represent this. A value more appropriate than \$.05, call it ν in general, may be found after experimentation, such as we will describe shortly, has been carried out. Similar cost functions can be designed to include π_s and α.

The worth of the estimate would increase upon decrease in B^2 but not without bound for values of B near zero. At the other extreme it could become negative, I suppose, if the results were dangerously misleading. A worth-per-interview function with some claim to applicability is $w_o - w|B|$ where w_o and w are appropriate values. If an interview in a survey which provides an unbiased estimate is valued at \$10 and one in a survey with a 5% bias is deemed worthless then w_o may be near \$10 and w near \$200.

5. Optimizing π_v

To illustrate the application of this representation consider determining the economic optimum level of vigilance when α and π_s are fixed, at α^* and π_s^* say. We wish to maximize, by judicious choice of π_v, worth per interview minus cost per interview for the survey or $(w_o-w|B|) - \nu/(1-\pi_v)$ or $(w_o-w+\gamma\pi_v-\nu(1-\pi_v)^{-1}$, where $\gamma = w[1-.4\alpha^*(1-\pi_s^*)]$. The answer by differentiation is to set $(1-\pi_v) = \sqrt{\nu/\gamma}$. This argues for increasing vigilance if it costs less (i.e., ν decreases) or if reduction in bias becomes deemed more worthwhile (i.e., w increases) or if bias from other sources is introduced (i.e., $\alpha^*(1-\pi_s^*)$ increases).

Plugging in the values suggested, namely $\nu = \$.05$, $w = \$200$, $\alpha^* = .8$ and $\pi_s^* = .60$, we find that π_v should be .95. This may only verify that we have chosen the parameter values to be reasonable in the light of actual practice. The net worth per interview in this case is -\$5.3, however and this does not sound too reasonable. It appears that bias is very large in this example. If vigilance drops to .90 then the net worth per interview becomes -\$33.5 while if it is increased to $\pi_v = .99$ then the net worth per interview is -\$21.3. Thus we see that the optimum is "sharp" to this extent.

6. Styles of Interviewing

Our speculations about cost functions are, we believe, of some interest and may argue for further work on the numerical values of the constants. The investigation which we did in fact carry out, was an experiment-within-a-survey designed to find out what effect changing the style of interviewing would have on the response error parameters.6/ We drew a sample of 144 tracts in a way so as to parallel the sample drawn for the December 1962 Enumerative Survey in North Carolina. Our sample was not used to make estimates of livestock and crops, although the data were collected in exactly the same way.

There were six different initial enumerators, two in each of three sub-regions of the state. An observer accompanied each initial enumerator; he timed sub-tasks of the interview and "coded utterances."7/ The observers' data were collected to allow us to see whether the various styles of interviewing differed greatly in time requirements. Except as we will mention below, they did not and thus we can concentrate directly on the variation in response errors by differences in style of interviewing.

Each initial enumerator was trained to interview in 8 different ways. These formed a 2^3 factorial arrangement. The three factors were called location instruction, respondent instruction and approach instruction. The two levels of location were fields-also and house-only. Under the fields-also instruction the enumerator asked the respondent if he might go to the fields as the data were collected on crops and livestock, while when he interviewed under the house-only instruction he carried on the conversation in the yard or house.

5b/If bias were removed by some kind of adjustment, say a subsample was reinterviewed and a correction made then one would focus on the quantity C.

6/Progress Report 34 carries the more detailed description of this work.

7/During the June 1962 Enumerative Survey we carried out an "Observational Study of Interviewing Technique..." which is discussed in Progress Report 32.

Variations in the respondent instruction were called best and first. Under first, the enumerator began asking the schedule questions as soon as he found himself talking to some one connected to the tract, while under best he was to determine, by each section of the questionnaire if it looked possible, who would be best informed and best able to answer and talk only with them.

The approach instruction had two alternatives, friendly and official. Under the friendly approach the enumerator looked at the respondent's eyes while asking the questions and down at his papers while the respondent answered. This was reversed under the official approach. Also under the friendly approach the enumerator moved closer and side-by-side versus more distance and head-on for the official. Extra topics of conversation were discouraged under the official but pursued more naturally under the friendly approach.

7. Effects on Behavior during the Interview

The effectiveness of these instructions in producing changes in the enumerators behavior is open to some question. For example, the interviewer may have attempted to ask questions of the housewife or young son when he was under the first instruction but in a majority of cases he would have been referred to the operator himself and ended up talking with the same person that he would have under the best instruction. However, we can tell by the increased number of don't know responses under the first respondent instruction, that there was some effect on the interview.

Although for six of the eight treatments, 25 to 29 minutes was the average time per interview there was a decided difference between the official and friendly approaches within the fields-also and best instructions. There the official approach "cost" only 22 minutes while the friendly approach cost 37 minutes per interview.8/ This shows some influence of the instruction and also it is important effect in that the fields-also with best is an attractive combination.

8. Effects on Response Errors

We examined the response errors on four kinds of items: field crops and livestock by the closed and open segment basis. We also distinguished, where applicable,9/ the three kinds of response errors. Thus we looked at many cross-classifications and, when dealing with scatter, at many analyses of variance. We found relatively few showing significance, and our chances of making a type I error by reporting them all is high. But their significance level is not the only reason that we think these relationships should be considered as suggestive. Thus we'll report on all of them and try to guess which ones may hold up in future survey experience.

The respondent instruction was practically never shown to be a source of change in level of response error. There was a suggestion that first responses were more stable than best responses on hog items for the entire farm (open segment). The data showed that of 17 unstable responses 13 were made under the best instruction, while the 34 stable responses were split evenly. This result is so contrary to what would be expected that it cannot be taken too seriously. There, however, is a suggestion that the respondent who is selected under the first instruction may turn out to be the "best" informant on hog items; that is, the farm operator may not know the hog enterprise as well as he does the rest of the farm.

The location instruction, fields-also, reduced response error in a number of places relative to the house-only instruction. Fields-also seems to cause higher stability, larger π_s, for acreages of "other land" in the tract. This category includes house lot, "woodland not pastured, swamp, pond, idle land...". Of 47 fields-also interviews, 37 were stable while of 45 house-only interviews only 24 were stable on the other land acreages. It is reasonable to expect that looking at the land itself would enable the interviewer to better classify the areas in accord with the schedule instructions and particularly when the category is as broad as this one. In practice the reported acreage under "other land" suffers heavy office editing so perhaps improvement here is not worth attempting to get.

The number of tracts with non-zero hog numbers, 28, was too small to allow us to estimate all three response-error parameters so we used a combined index of discrepancy, namely the difference in the two reports divided by the largest number reported. The fields-also

8/Table 4 of Progress Report 34 contains these data and also shows the breakdown of time by subtasks within the interview. This shows that the extra time was largely taken up in "introductory conversation" rather than in "other topics". Other topics did show a slight increase under the friendly approach, however.

9/The final formulation of the response-error model came while we were doing this analysis so that much of our work would have to be redone if we were to be faithful to the current model. We were using what were called D-scores as measures of response errors. We distinguished "stable", "mid-scores" and "high" D-scores and this corresponds to the stability, scatter and vigilance breakdown of the current representation. See Progress Reports 33 and 34.

instruction had a smaller mean value for this index than did house-only. The fields-also instruction seems to bring out more vigilance and higher stability on the principal crop item. This question asks for the principal crop and its acreage for the entire farm, an open segment definition. In both of these cases the fields-also effect may be present and desirable but we do not find any compelling reason for it. There may be a generalized effect from asking to see the fields that puts the respondent in a more serious frame of mind.

In general on the crop items the official approach showed up as more beneficial than the friendly. Data were collected on 584 fields and 251 of these were done under the official approach. Of these 251, there were 13 of them with a different crop reported initially than at the follow-up enumeration. Of the 333 fields done under the friendly approach 33 were reported as a different crop. This kind of error is of the vigilance type. Under the official approach 27% of the reported yields per acre (on 112 fields) were stable while this was only 16% under the friendly approach (used on 134 fields). There was also some evidence, based on the index of discrepancy mentioned above, that the official approach is superior to the friendly on hog items but inferior on cattle items. Thus while the picture is mixed for livestock items the official approach seems to work on the crop items.

In the course of searching for effects on response errors we uncovered some apparent interactions among the instructions. There was more stability in acreages by tenure status when best was paired with fields-also and first with house-only than when best went with house-only and first with fields-also. The latter two combinations would be expected to be more awkward to handle and to cause a loss of rapport, so that when the questions on land rented out and so on arose they could have more easily been misunderstood.

III

THE EDUCATION EXPLOSION: ITS MEASUREMENT AND MANAGEMENT

Chairman, ARNOLD A. HEYL, U. S. Office of Education

EDUCATIONAL OUTPUT BY LEVELS, KINDS, AND NUMBERS

Marjorie O. Chandler, U.S. Office of Education

In this paper, I should like to discuss certain output measures obtained by the U.S. Office of Education.

At the outset, a brief statistical overview of the educational system may be in order. At the lower end of the scale, we find that about two-thirds of 5-year olds are now enrolled in school. During the usual age span for elementary grades (6 to 13) about 99 per cent of our youngsters are attending school. During the usual high school age span (14 to 17) about 93 per cent are enrolled in educational institutions.

At present, about two-thirds of our youth graduate from high school. A little over half of the high school graduates--or about one-third of the age group--may be expected to enter college. The number of college graduates will probably be a little over half the size of the group that entered four years before, or about a fifth of the age group. The number of Master's degrees will be about a quarter as large as the number of Bachelor's two years before, or about one-twentieth of the age group. The number of Doctor's degrees will be about one-sixth as large as the Master's three years earlier, or under one per cent of the age group.

Now I should like to give you some more detailed facts and figures on outputs which will also serve to indicate the sorts of data that USOE collects. I shall also mention some areas where we do not collect data.

It seems reasonable to begin our consideration of outputs at the point of high school graduation. So far as lower levels are concerned I might note that USOE does collect data on the number of students enrolled in each grade in elementary and secondary school.

Since we do not yet have a firm count for the academic year just ended (1964-65) I am going to use as a base the academic year 1963-64. Table 1 gives a quick summary starting with that year's group of high school graduates.

In 1963-64, the total number of high school graduates was about 2 1/3 million. This figure represents an increase of about 18 per cent from the immediately preceding year, reflecting the great rise in birth rate following World War II. It is estimated that graduates in 1964-65 (the year just past) will be up 18 per cent from 1963-64. To indicate how atypical these bulges are, I might point out that from 1961-62 to 1962-63 the rise in graduates was about 1 per cent, and for the previous year, there was a decline of about 2 per cent. Further, the current estimates suggest changes of less than one per cent a year for each of the next three years.

The 1963-64 high school graduates represent about 69 per cent of their age group. It would be beyond the scope of my paper to discuss in detail the problems in defining and identifying an appropriate denominator for a figure such as this. Let me just note that USOE has been using as a denominator (the base population) one-half of the 17- and 18-year-old population, since these are typical ages at the time of high school graduation.

One thing we don't know enough about is the curriculum pursued by the high school graduates, a point of special importance for those not going on to college. How many were trained in business courses, in auto mechanics, and so forth? Some data have been collected on curriculum for a recent class, and the analysis is in progress. Meanwhile, we do know specific course enrollments for certain years--how many students were enrolled in typing classes, in mathematics, and so forth--but this is probably not as helpful as information classified by over-all curriculum.

As noted earlier, about 69 per cent of young people finish high school. When did the other 31 per cent leave, and where did they go? We can tell something about when on the basis of our grade-by-grade enrollment statistics. While some few get lost every year the number is slight until about 8th grade when several per cent start to drop off each year. The biggest single loss is between 10th and 11th grade; about 10 per cent of 10th graders do not show up for 11th grade. As it happens, a student who progresses at the usual rate should reach 11th grade at age 16; school attendance laws often relate to age 16, so perhaps drop out at this grade is not very surprising. What happens to the drop-outs is a question which USOE does not survey routinely, although some information is available elsewhere on this very interesting point.

Turning now to the college level, we find that there were about one and a quarter million (1,225,000) first-time students in degree credit programs in fall 1964. This figure is about 53 per cent as large as the number of high school graduates for the immediately preceding year. It should be noted that not all of the 1964 college entrants came from the 1963-64 graduating class, although the majority undoubtedly did. Another point to keep in mind is that our fall enrollment studies somewhat underestimate the total number entering college: some students may enter at other times. Periodically we have collected full-year data but have not done so for the past few years. Despite these limitations, the figures on fall enrollment versus high school graduates are generally of interest.

First-time college enrollment spurted in an unprecedented way in 1964, increasing about 17 per cent from fall 1963. Another rise of about 18 per cent is expected for this fall (1965). This is quite different from what the colleges have been seeing--about 1 or 2 per cent rises from 1962 to 1963, and from 1961 to 1962. These figures closely parallel the changing figures on high school graduates which I cited earlier in this paper.

The college output in 1968 will consist heavily, although certainly not exclusively, of 1964 entrants. The forecasted total is for some 666,000 Bachelor's and first professional degrees. This figure is about half as large (54%) as the number entering four years earlier (1964). Like the 1964 entrants, the 1968 graduates will be substantially more numerous than the immediately preceding class; the rise in graduates is estimated as 19 per cent over 1967.

Other forecasted outputs that involve substantial numbers of 1964 high school graduates are those for Master's in 1969-70 and for Doctorates in 1972-73. The Master's degree group will then be about one-fourth the size of the Bachelor's group two years earlier. The 1972-73 Doctorate group, in turn, will be about one-sixth the Master's group of three years earlier. We are now dealing with very small fractions of our starting group of 1964 high school graduates: about 7 per cent for Master's and about 1 per cent for Doctorates.

Our 1964 college outputs, on which current data are available, may be worthy of somewhat more detailed mention. In 1963-64 there were about 460,000 Bachelor's degrees and about 42,000 1st professional degrees. At this point it might be appropriate to explain that in USOE statistics, 1st professional degrees (such as M.D. or Bachelor of Laws) have traditionally been grouped with 4-year Bachelor's degrees. In the past few years, however, data have been collected separately for 4-year degrees and for the 1st professional degrees requiring a total of 5 or more years of higher education.

For each level of degree the U.S. Office of Education reports a count by field of specialization. Perhaps a few of the findings of this field count for 1963-64 will be of interest. Starting with the 460,000 4-year degrees, we find that the largest of the 25 broad fields considered is education, accounting for virtually a quarter of degrees at this level. The second largest at the Bachelor's level is social sciences, with about one-sixth (17 per cent) of all degrees, and third is business and commerce with about one-eighth (12 per cent).

Among 1st professional degrees requiring 5 or more years (about 42,000 degrees) by far the largest number are found in health (33 per cent) and legal areas (26 per cent).

At the Master's level there were about 101,000 degrees in 1963-64. Education looms very large, accounting for 40 per cent of the total number of Master's degrees. Engineering is second with 11 per cent of all Masters, and social sciences is third with 9 per cent.

At the Doctorate level there were about 14,500 degrees in 1963-64. Here, the largest fields are physical sciences with 17 per cent of all Doctorates, education with 16 per cent, and social sciences and engineering with 12 per cent each.

In giving you an overview of USOE statistics on higher education, I have deliberately postponed the question "what _is_ higher education." The issue of definition arises in this paper because when I spoke of output of higher education I was speaking only of that segment on which the U.S. Office of Education collects data. Undoubtedly, data on other types of post-secondary output would be valuable for a variety of purposes.

For statistical purposes, the USOE includes in its universe of higher education those institutions which meet the following criteria: (1) Institutions accredited or approved by a nationally recognized agency, by a State Department of Education, or by a State university are included. (2) Institutions not meeting the first criterion are eligible for inclusion if their credits are accepted by not fewer than three accredited institutions. I want to emphasize that I am speaking of inclusion for statistical purposes and not of USOE approval of institutions; USOE does not accredit, rate, or otherwise approve institutions.

Clearly, other criteria of higher education are conceivable; in the realm of post-secondary education the borderline between in-out of higher education is a rather hard one to establish. There is probably general agreement that colleges of barbering, for example, are properly excluded. On the other hand, Major Seminaries of the Roman Catholic church or schools of nursing--many of which are not included in USOE statistics--may for some purposes belong in "higher education." In any case it seems clear that the 1964 USOE figures of about 1¼ million first-time students and about 5 million total enrollment represent only a modest portion of post-high school education.

Table 1. Actual and expected educational outputs at various levels.

	Date	N	Per cent based on: (17-18)/2 popul. * 1964	High school graduates 1963-64	1st time college 1964	B.A. and 1st prof. 1967-68	M.A. 1969-70
Population age (17+18)/2*	1964	3,340,000	100				
High school graduates	1963-64	2,302,000	69	100			
First-time college enrollments	Fall 1964	1,225,000	37	53	100		
B.A. and 1st professional degrees	1967-68	666,000	20	29	54	100	
Master's degrees	1969-70	168,000	5	7	14	25	100
Doctor's degrees	1972-73	28,000	0.8	1.2	2.3	4.1	16.4

* $\frac{1}{2}$ population aged 17 + 18

Source: Unpublished data prepared by Reference, Estimates, and Projections Branch, Division of Statistical Analysis, National Center for Educational Statistics, United States Office of Education.

MEASUREMENT OF EDUCATIONAL QUALITY: SUMMARY AND PROGNOSIS

Frederic D. Weinfeld, U.S. Office of Education

It is unfortunate that in a society which talks so much about the need for the best possible education for its youth we are forced by honesty to say that the evaluation of its educational processes and institutions is woefully behind the state of the art. "Good schools," "excellence," "quality education" and other loose terms are bandied about without definition. If we attempt to cut through the semantics we find that the terms are operationally defined by the criteria used in their evaluation. To some, therefore, a "good" school system is one which has a highly paid staff and a low pupil-teacher ratio, to others it is one which has an active PTA, or one which is building new and beautiful structures, or one which is expending large sums of money on modern technical learning aids, or one which is introducing innovations in techniques, materials, and curricula. In short, it is assumed that the school which does something more than other schools is a better school. There is an obvious hiatus here between school practices, policies, and facilities which are considered beneficial and their actual effect upon the students. Little focus is put on the real educational output of the school, the level of educational attainment of the students; instead inappropriate headcounts, such as the number of students going on to prestige colleges, the number of Merit Scholars, the amount of scholarship funds awarded to the students, or in some cases, the average test scores of the school are used as criteria of school quality. And so the logical criterion measure of output, the amount of change or growth in the student himself brought about by the school is neglected. Under the egalitarian views of some of our educators, it is assumed that our schools do equally well for all students and develop them to the full extent of their potential--whatever that misunderstood and troublesome cliché means.

The school system should be evaluated, as any other operating system would be, by the efficiency of the system, measured in output per unit input within classifications of schools grouped according to their size, funds available, type, etc. In other words, compared to other similar schools how much "bang" is a school getting for its "buck." This is indeed a stark way of looking at the problem. We must discard descriptive characteristics of a school system as the valid criterion of what it accomplishes. Though chrome ornaments on an automobile add to its decorative appeal they in no way add to the performance of the vehicle itself; so too in education the performance of the system is the proper criterion, not its "good looks." But if indeed we wish to use the beauty of a school plant as a criterion measure let us do it, but let us not call the resulting scale "quality of education" or impute it to be a measure of what goes on inside the school.

It has been assumed that the more money a school spends the better it is. Only insofar as higher expenditures do indeed contribute to the educational attainment of the students, is per pupil expenditure a valid criterion measure. In fact data has revealed that some schools with meager budgets are contributing more to the growth of the students than other less efficient, overrated schools with higher expenditures. The effects of other school characteristics, such as school policies and practices, upon student output must first be investigated before advocating school expenditure as valid measures of school quality.

Schools have many educational goals and outcomes, such as the ability to think and evaluate constructively and creatively, the appreciation of our democratic heritage, the acquisition of good habits and attitudes, etc., and specific criteria are needed for evaluating the schools effectiveness in meeting each of the defined objectives. For many of such goals we have not yet devised satisfactory measuring instruments. However, one major goal of the schools which is conceded by everyone is the acquisition of basic skills in the use of words and numbers. Since standardized achievement tests currently in use by the schools measure the level of student achievement in the various basic skills, they can be used as a criterion measure of the schools effectiveness in achieving this specific goal.

As hot as the "pursuit of excellence" has been in the past decade, the quarry has eluded us because the searching parties have been few and ill-equipped. The members of educational establishments have in the past been reluctant to evaluate themselves, today they are only slightly more receptive to the idea of finding out what is going on in the schools. They are concerned with the testing of students, and achievement test scores are, of course, proper criteria for the assessment of the output of the schools; however, most schools unfortunately, consider a perusal of the average test scores to constitute an evaluation of the school. Some school systems and States have even published local norm tables for various achievement tests with the bald statement that the norms were to be used to evaluate the quality of the schools without any recognition of the fact that the schools differed greatly in size, expenditure, staff, and facilities, and that they were educating students who differed greatly in their socio-economic status, family background, ability, motivation, and past preparation. Rarely have evaluations of schools been based on a sound research design which controlled in some way the variations in the characteristics of the school and the student body so that sensible comparisons between schools could be made.

It is evident that only when such controls are used should there be an attempt to proceed to investigate the effect of the treatment variables, those characteristics of the schools which are thought to enhance educational attainment.

Of all the studies of school quality done in the past few are worth reporting. One of the better research studies has been the New York Quality Measurement Project. This project, started back in 1957 by the Division of Research of the New York State Education Department, selected 103 schools throughout the State for longitudinal investigation. I.Q. data were collected and various grades were tested for three consecutive years with a battery of achievement tests. Some limited school data were also collected on school size, socio-economic level, type of community, staff, and teacher training. Mort's "The Growing Edge," an instrument for measuring the adaptability of a school, the degree to which a school employs modern instructional techniques, was also used as a criterion measure. Schools were then grouped by socio-economic level and community type for comparisons and separate norm tables were derived for these groups.

In this study a school's effectiveness in teaching the basic skills, measured by standardized achievement tests, was assessed by comparing its test results to those of other similar schools. After grouping the schools by communities which had similar characteristics there were, of course, still great differences in the educational attainment within each group. Also there were great differences in the expenditures of the schools. The amount of money spent by a school is only a rough measure of its quality level, and so it is easy to find schools with high expenditures which do not produce students with high achievement and similarly, we can find efficient low expenditure schools which turn out students with high achievement. In the New York study it was found, on closer examination, that the relative effectiveness of a school system varies with the I.Q. score of the student, the socio-economic level of the student, the subject matter content, and sex of the student. It is not simply a matter of one school being better than another, rather it is that one school does a better job with a particular type of student in a certain subject matter area. Schools have strengths and weaknesses in specific areas and are not simply universally good or bad.

One wishes, however, that more comprehensive input data on the students and school characteristics had been available in the New York study so that the analysis could be further refined. For then students could be grouped, or controlled, by home background, attitudes toward school, etc. After having controlled student input we could go on to investigate and try to discover the important school characteristics which influence the attainment of the students.

I might mention, as an aside, that such comprehensive data is available in the Project TALENT Data Bank from their large 1960 study. Also the Iowa Educational Information Center is now starting to collect interlocked comprehensive data on schools, including student home background data, test scores, teacher data and school data, which could be used in studies of school quality.

Two major studies of school quality underway at the present are worth mentioning. One is the Carnegie Assessment Project which is now in the stage of developing testing instruments to be used in a nationwide study to measure the wider outcomes of the schools. The other study is the Pennsylvania Quality Education Project which is now nearing completion. This is an exploratory attempt to develop criteria of school quality and to assess measures and indices of these criteria.

I would now like to turn to the effort of the U.S. Office of Education in the measurement of Educational Quality. The National Center for Educational Statistics is now in the midst of conducting the "Educational Opportunities Survey" as required by the Civil Rights Act of 1964. This fall almost a million school children, from Grades 1 through 16, will participate in a major study which will collect comprehensive data on the students, including their home backgrounds, attitudes, ability levels in several basic skills, their teachers, and their schools. The data from this study are intended to do triple duty. They will fulfill the primary purpose of surveying the lack of equal educational opportunity for children of minority groups, they will, we hope, be a source of invaluable data for other educational researchers, and they will provide another needed set of basic data to the Division of Operations Analysis. This Division, within the National Center, is now in the process of developing a mathematical model of educational attainment. The basic plan for this effort is to identify the pertinent variables, those school characteristics which affect educational attainment, and to determine their relative importances, controlling, of course, on student input variables and certain school characteristics. Having found these pertinent variables we intend to ascribe costs to them and finally to determine by linear programming the optimum allocation of available funds for the maximization of educational attainment.

Cost and quality are often discussed together. As I mentioned earlier, high expenditure does not insure quality in a school. It is evidently the judicious use of funds on factors contributing to the educational attainment of the student which is of import. It would be of great value to the Office of Education if we could determine the relative effects of various possible alternatives for improving school quality. (And here we intend to use the concept of school efficiency as a measure of school quality.) For then we could make our decisions on the basis of the most efficient alternative which would give us the maximum educational output for a given resource or dollar expenditure.

In summary, the path ahead in the measurement of school quality leads to the very simple realization that in order to do this properly we must first specify our goals and determine the appropriate criterion measures, then we must control on relevant student and school characteristics. Also, because of the wide organizational variations in the schools throughout the Nation it is advantageous to group the schools by geographic region, type of school, school size, tax base, or any other meaningful classification. By using such a procedure we can reasonably expect to improve our assessment of the schools.

OPERATIONS RESEARCH IN EDUCATION

Arnold A. Heyl, U.S. Office of Education

It is not my purpose today to survey the literature of operations research in education. I will not attempt to summarize such very large-scale operations research activities as the educational planning projects conducted primarily in Europe by the Organization for Economic Cooperation and Development, the general purpose of which is to plan for educational growth in a given country in relation to the total economic development of that country. Nor will I discuss the smaller-scale activities, more nearly resembling systems analysis, which are being conducted at a number of educational institutions in this country, the purpose of which is to rationalize the operations and decisions of those institutions or some other operating educational institution. There are among those projects a number of excellent examples of operations research in education, but at this time I would like to keep them in the background.

This afternoon I would like to take quite a parochial stance and discuss with you those operations analysis activities being undertaken and to be undertaken at the U.S. Office of Education. In January of this year the Division of Operations Analysis was established in the National Center for Educational Statistics in the U.S. Office of Education. I'd like to report to you on the "health and welfare" of that Division and on its future. Our broad mission is to develop and maintain a quantitative, analytical model of the educational system in the United States, utilizing the techniques of operations research to aid the educational decision-maker in formulating policies and charting future courses of action.

The words "quantitative" and "analytical" have mathematical connotations and the mathematical model, one of the hallmarks of operations research, is at the heart of our thinking. Such a model will include a representation of the flow of students through the system; will describe the growth and utilization of professional staff at all levels in the system; will take into account the characteristics of curricula; and will consider the administrative organization and financial resources of the system. It will also permit the educational system to be related to the total United States economic system so that impacts of education on the economy and of the total economy on education can be evaluated.

Much as we would like to do so, we are not in a good position to begin by formulating the model. We must grapple with more fundamental matters, and first among these is developing an operations research capability. In February the staff of the Division of Operations Analysis included three professionals. Now the number of senior research personnel has doubled and we hope for additional strength and experience in operations research as our group continues to grow. Present members of the staff are concentrating on selecting the relevant variables to describe the elements of the educational system and establishing their parameters and inter-relationships. Then we can proceed to structure the elements and develop mathematical models to represent that structure analytically. These separable activities cannot be carried out completely separately and sequentially and hence we have begun a number of studies which focus on specific segments of the total structure and will culminate in submodels of it.

The flow of students through the system is one of the segments under study, and Dr. Chandler summarized in her presentation some of the information already available in the Office which bears on this topic. With those data some fairly firm conclusions can be drawn concerning public school enrollment at the elementary and secondary levels. These enrollments have been charted and forecast at the national, and to some degree at State levels. Non-public school enrollments at these levels are less clearly defined, and we are currently trying to improve this complementary segment. Our immediate goal is to describe elementary and secondary student flow in both public and non-public schools by means of sets of curves and functions with their parameters.

Despite all the statistical knowledge we do have, gaps and voids continue to exist, and hence, the immediate goal is just the beginning. As Dr. Chandler also mentioned, we know all too little about what's inside a secondary education; what levels of training the 2 1/3 million high school graduates have had. We know even less, in any integrated fashion, about where those who leave secondary education go. In the immediate future efforts will be directed toward shedding light on these two unknowns. What kinds of students leave secondary school? What segments of the range of intellectual levels do they represent? What socioeconomic levels do they represent? To what kinds of occupations do they go? What other kinds of training do they get and by what means? Where do they ultimately take their place in the social and economic fabric of our Nation? More complete knowledge of the structure of potential our high school graduates possess, and a clearer picture of "dropouts" are already included among the stated requirements of educational decision-makers. The data which Dr. Chandler mentioned exist in other agencies are in general limited in scope either substantively or in area of applicability. We hope to broaden the view and sharpen the picture.

A similar set of educational problems and the research problems related to them can be stated concerning higher education. Their complexity is increased because of the increased diversity in programs of study at the undergraduate and graduate levels. And at these levels the additional dimension of student mobility enters the picture as a significant variable. We need not only to know what kinds of high school graduates enter what kinds of college programs, and how persistent they are; but also, what kinds of institutions enroll those graduates and where they are located. Data are available at several points along the way in higher education on both kinds and numbers of students. Our efforts will be directed toward selecting significant indicators, eliminating gaps in the data, and developing a meaningful structure.

The measurement of level of educational achievement, which comprises another major segment of our efforts, has taken on increased importance in recent months and we are taking steps to broaden our information on this basic indicator of the status and progress of education. Reliable measurement of the level of educational attainment at the national level is not yet available. Technical matters related to comparability of measuring instruments and philosophic differences related to national assessment have thus far militated against such measurement programs.

Dr. Weinfeld has already mentioned that this fall in an Educational Opportunities Survey required by the Civil Rights Act of 1964, some data on educational achievement will be collected, along with other data concerning students' backgrounds and their school and community environments, from a nationwide sample of elementary and secondary students and from a smaller sample of students enrolled in higher education. These data together with earlier data from project TALLENT will provide initial material for models that use achievement as a criterion. More broadly representative data may become available in 1967 as an outcome of the research project currently being sponsored by the Carnegie Corporation which Dr. Weinfeld also cited. The cornerstone of this project, the purpose of which is to develop a national assessment of educational achievement, is a group of measuring instruments which are specifically designed for program evaluation and specifically designed to preclude individual evaluation or any indication of a national curriculum.

Educational achievement is not only the indicator of the level to which we have educated but also plays a vital role as a primary criterion in the measurement of quality in education. Similar statements can be made concerning the importance of measurement of educational level in connection with evaluation of program objectives. How effectively are our curricula meeting our needs? Is content at one level well articulated with content at the next? How effective are our methods in fostering learning on the part of our students? Analysis of achievement data can provide at least partial answers which can be interpreted in the context of other aspects of the educational process.

In advance of the availability of these data, we have begun research in parameter and relationship estimation with variables which we feel with some certainty are relevant to the process of education. Toward that end, one of our current efforts involves some special analyses of data collected on some 400,000 secondary school students who participated in project TALENT in 1960. We plan to use a variety of regression analysis techniques in seeking structure among not only achievement variables, but also student, teacher, school, and community characteristics which are available in the TALENT data. As we progress in this area we plan to use other bodies of data which already exist and also to relate our findings from these efforts to current data. Resource allocation studies and studies of costs of education, which I'll mention again later, may also utilize these results.

You may have noticed that my remarks concerning educational achievement, as were my earlier remarks about student flow, have been heavily weighted toward elementary and secondary education, and measurement of college and university students had only the barest mention. At the lower levels curricula are less complex, measures are in greater variety and abundance and more easily developed where lacking (though some excellent general achievement measures which are well accepted exist in higher education), the volume to be measured is roughly ten-fold that of higher education; and I could go on ... But perhaps the most important reason for our beginning at the elementary and secondary levels is the recently enacted Public Law 89-10: The Elementary and Secondary Education Act, 1965. This Act specifically requires yearly program evaluation by techniques which include measurement of educational achievement. Nonetheless, as we progress and time and staff become available, it seems clear that our work will lead us into measuring achievement in higher education.

From what I have said it is apparent that the results of our efforts have implications for manpower, and we intend this to be so. The product of the educational process is manpower and the studies I have discussed are all bent toward finding out more about that product. As still another major area of endeavor, separate efforts must also be undertaken to relate that product to the demands in the total manpower structure. The detailed study of the feedback loop within the educational system from output as graduates to input as professional staff seems particularly appropriate for the Office of Education. How is level of preparation associated with level of teaching? What are the patterns of faculty mobility, and how do they relate to experience, salaries, personal and social goals, and other factors such as these? Over what ranges is the relationship between level of educational expenditure and quality of instructional staff significant? These are but a few examples of questions concerning which greater insight is needed. Our present work is focused on assessing the data base, bolstering where necessary, integrating within the area, and relating to other areas of data.

As we progress in our analysis of the products of education, we will assure that our work is related to other agencies' manpower research efforts and other sources of manpower data. Questions similar to those I raised concerning "dropouts" can also be asked concerning educational output generally.

The costs of education--in dollars, manpower, facilities, and other resources--represent still another area of investigation to be included as a part of the total model development effort. Estimation of the cost per student of education to specified levels will be developed not only in overall terms but also for broad fields within the total spectrum of education. In other studies the cost of educating various segments of the population to specified levels will be estimated, the value of such educational programs will be assessed, and the "flow of funds" through the educational system will be described.

Projected costs and projected effects must be related in cost-effectiveness studies so that alternative courses of action can be considered and evaluated and decisions concerning allocation of resources made. These studies, to be maximally useful, must take into account economic factors outside of education both of the kind which have an effect on education, and of the kind which are affected by education. For example, a significant decrease in the dropout rate at the secondary level forecasts an immediate "shortage" effect on certain segments of the labor supply, a later effect towards increasing college enrollment, and perhaps a still later "surplus" effect on other segments of the labor supply. The overall model of the educational system under development will enable consideration of such possible effects.

In summary, the analytical models of the educational system that are being developed will encompass quantitative relationships describing flow of students through the system and into the labor force, flow of faculty and staff into and out of the system, allocation of resources of all kinds to programs conducted by the system, and measures of the costs incurred for and gains derived from those programs. Our goal is to enable the educational decision-makers, through the use of these quantitative models, to evaluate the progress of current programs and to assess, prior to their adoption, proposed courses of action and proposed policy changes in the sphere of education.

Discussion

W. B. Schrader, Educational Testing Service

Rather than take issue with specific points in Dr. Chandler's paper, with which I am in substantial agreement, I would like to concern myself mainly with advocating a broader definition of output than she used.

First, it would be desirable if output were considered to include all those kinds of evidence which might serve as dependent variables or criteria in assessing the effects of the many influences currently at work in education. For example, information about the educational programs which high school graduates had followed could be regarded as describing output. The College Board is currently planning a study which will attempt to find out--at the topical level--what high school students are studying. The influence of the current ferment in curriculum may show up more clearly when fairly detailed data are collected. It seems possible, also, that topics may provide a better indicator than course titles for comparisons over a period of time.

Greater detail in existing reports should increase their usefulness as indicators of trends. For example, I regret that the category of "first-time full-time degree-credit students" is no longer reported. Of course, if the scope of output statistics is enlarged and if the various series are explored in greater detail, there is a danger that the sheer bulk of figures produced will overwhelm the users. Here, the developing use of computers in information retrieval may help to resolve the difficulty by making it possible to produce special purpose reports as needed from a comprehensive set of basic data.

Second, it would be highly desirable if the variables on which data were collected could be derived from a conceptual structure. Lazarsfeld[1] has pointed out that while Quetelet repeatedly suggested "that special data could be collected to form the empirical basis of a new concept, he never set a concrete example". Educational statistics needs new, conceptually-derived variables which, along with some currently used variables, might make up a system. This kind of approach seems essential if the kind of solid forecasts most useful for planning are to be produced. To describe the tides without considering the moon would undoubtedly be cumbersome. In somewhat the same way, the "bulges" to which Dr. Chandler referred were made plausible by Thompson and others who foresaw the effect of birth rates on future enrollments.

If a broader view is taken regarding the nature of output statistics, certain implications for data-collection seem clear. The compilation of data reported in tabular form by institutions has definite advantages. It seems important, however, that these statistics be augmented extensively by data collected directly from individuals, preferably using modern sampling methods. Collecting data directly from individuals would facilitate studies of relationships and the inclusion of more variables. Project Talent and other special studies are using this approach. It should be equally desirable for recurrent surveys. For such surveys, it should be possible to sample within a school or college to identify the individuals to be included. We have sampled in this way to develop test norms.

Finally, I would urge that many frankly methodological studies be undertaken both to clarify the interpretation of regularly collected data and to aid in relating data collected by different methods but bearing on the same question.

[1] In Woolf, H., ed. Quantification. Indianapolis, Ind.: Bobbs-Merrill, 1961.

IV

HUMAN RESOURCES AND ADVANCED EDUCATION

Chairman, JOHN K. FOLGER, National Academy of Sciences

THE SUPPLY AND DEMAND OF COLLEGE TEACHERS

A. M. Cartter, American Council on Education

Considering the importance of the problem to higher education, and the many hundreds of millions of dollars appropriated by the Federal government for the expansion of graduate education over the last few years, it is rather astonishing that we know so little about the present and probable supply and demand of college teachers. The general consensus today as expressed by several Federal agencies, the National Education Association, and many college and university Presidents and graduate deans, seems to be summed up in the following three propositions: (a) persons trained at the doctoral level are in increasingly short supply; (b) the quality of faculty (as measured by highest degree attained) in the nation's colleges and universities is deteriorating; and (c) the situation will worsen over the coming decade as a consequence of burgeoning undergraduate enrollments. Over the last few years various distinguished educational spokesmen have used such terms as "disastrous shortage", "serious crisis", the nation standing virtually paralyzed", "frightening figures", "a major national scandal" to describe the supply of college teachers, and have called for "heroic efforts", "crash programs", and new degrees short of the doctorate to stem the tide.

At the risk of flying in the face of commonly held opinion, I wish to argue the reverse of the above propositions; namely that: (a) the "sellers market" in academic personnel is likely to disappear over the coming decade, (b) the quality of faculty in the nation's colleges and universities has improved, not deteriorated, over the last ten years, and (c) the situation is moderately well in hand now, and will improve dramatically in the 1970's. In attempting to support these views the paper will first summarize events of the last ten years, then present a growth model helpful in projecting supply and demand conditions ahead to 1985.

The Last Decade

The belief that things are getting worse rather than better is largely attributable to the biennial research bulletins issued since 1955 by the National Education Association on "Teacher Supply and Demand in Universities, Colleges and Junior Colleges."[1/] The first report presented a distribution of total staff by highest degree for 637 reporting institutions in 1953-54. Successive reports, however, have only inquired as to highest degree of new teachers. The figures shown in Table I, taken from the various NEA reports, have led some readers to believe that a rapid deterioration in faculty quality was in fact occurring.

A few critics of these reports have noted that it is an improper procedure to compare average and incremental ratios, but no attempt has been made to estimate the magnitude of this distortion. Table II is an attempt to correct this procedure, using additional data from the NEA reports. Columns 1 and 2 are the data used by Maul to obtain the percentages in Table I. In addition, however, the biennial reports give the number of new doctorates each year who "continue in teaching", and thus do not show up in the "new teacher" series.[2/] These are shown in column 3 of Table II, and a ratio of new doctorates in teaching to new teachers is computed in column 4. Now a meaningful comparison can be made between the average ratio for 1953-54 and the incremental changes in both the number and degree level of college teachers. This series suggests a slight improvement in the proportion of senior college faculty with the doctorate.

One further factor should be considered which is also favorable to the view that the quality of faculty (as measured by highest degrees attained) has not deteriorated. _A priori_ one would assume that teachers with the doctorate are more likely to make a lifetime career out of teaching than those without a doctorate. It would be reasonable to assume that there is a differential net transfer rate for the two groups. A recent Office of Education study, to be published later this year,[3/] indicates that for 1962-63 the rate of those leaving college teaching for reasons other than death or retirement was 3.1% for doctorates and 7.1% for non-doctorates. Other data, discussed below, further indicate that the net transfer rate of doctorates into and out of teaching has been approximately zero in recent years--that is, that the in-transfer rate of doctorates from other employment was also about 3%. To illustrate the effect of a difference in the net transfer rate, assume that the rate is zero for Ph.D.'s and a 5% annual net loss for non-doctorates. For the 1963/64 class of new teachers, with an initial ratio of .484, five years later the ratio of doctorates to total continuing teachers would rise to .548. Unfortunately we have only one fragment of data from the COLFACS study to judge by, so this example is suggestive only; presumably the separations rate for non-doctorates is positive but not greater than seven percent.

If the data in Table II and its accompanying speculations, were the entire basis of the thesis that the percentage of college faculty with doctorate has been rising over the last decade, it would rest on a weak reed indeed. But this view is now supported by two new studies. One was recently presented by this author, drawn from data collected quadriennially by the American Council on Education.[4/] The findings are summarized in the first two columns of Table III. The other is a soon-to-be published study (COLFACS) by the Office of Education, whose findings are summarized in column 4 of Table III. Whether one views the comparison between total or full-time instructional staff, it seems clear that the percentage of doctorates has been rising for each type of institution. This conclusion is consistent with the N.E.A. data as presented in Table II above, although it is just the opposite of the conclusion which N.E.A. drew from its own material.

Faculty Forecasting Models

Projections of the demand for college teachers made over the last decade have varied widely, and most have been such poor predictors of actual developments that the basis on which the projections were made needs careful scrutiny. The best known model is that developed by Ray Maul in the 1959 NEA report,[5] and now used by the Office of Education.[6] The model consists of three ingredients: (a) an independent projection of future enrollment, (b) an assumed student/staff ratio, and (c) an assumed replacement rate for faculty deaths, retirements and shifts to other employment sectors. In the most recent presentation by the Office of Education the student/staff ratio is estimated to average 14:1 for the next decade,[7] and the replacement rate is assumed to be 6%. The choice of the latter percentage apparently derives from the earlier Maul model.[8] The result of this model when applied to Office of Education enrollment projections is to predict an aggregate need for some 556,000 new college teachers over the next ten years. Assuming constant quality of faculty, the Office of Education predicts a probable "deficit" of more than 120,000 doctorates by 1973/74.[9]

There are a number of aspects of the current OE model which I believe lead to a considerable exaggeration of future faculty needs. First, the projected student/staff ratio (18:1, based on total instructional staff) is lower than the experience of the last decade would indicate. Table IV, using Office of Education data, shows the increment of enrollment and increment in total instructional staff since 1953/54.[10] It has averaged 19.3:1 and there is no clear trend upward or downward. On reflection this does not seem an unusually high marginal ratio for a number of reasons. First, junior colleges, where the average ratio is 20:1 or greater, represent a larger portion of increments in enrollment than they do of the current total (nearly 30% of the annual increases as compared to less than 15% of the total). Second, enrollment in public institutions, where the ratio is moderately high, is expanding more rapidly than in private colleges and universities. Third, much of the expansion is occurring in already existing institutions, and one would expect there to be some manpower economies of scale associated with such growth. Finally, modest changes in technology (language laboratories, educational television, independent study, etc.) presumably work to increase the ratio despite enrollment expansion. A continuing <u>marginal</u> ratio of nearly 20:1 would mean that the <u>average</u> ratio will rise from 15.3:1 today to 17.3:1 by 1985. The Office of Education choice of an 18:1 ratio, therefore, appears to overstate the expansion needs by nearly 10%.

A second, and more major, criticism is the use of a 6% replacement rate for faculty, for I believe it overstates replacement needs by a factor of three. The reason for believing that this is such a major error is the following. If one applied this model to the last decade, beginning with 1953/54, then we should have experienced a decline in the percentage of doctorates on teaching faculties from about 40% to 30%; instead, as Table III indicated, it has risen by seven to ten percentage points for four year institutions. As I have indicated in another paper, the actual experience of the last ten years is consistent with a replacement rate of <u>slightly less than 2%</u>. Judging by the age distribution of present faculty (1962/63), and applying appropriate mortality rates and estimating retirements, my estimate of the actual replacement rate for the coming decade is shown in Table V.[11]

A third objection to the OE model is that included in full-time equivalent staff are personnel for administrative services (few of whom, below the level of academic deans, would be expected to have the doctorate), junior instructional staff (who by definition are teaching assistants without the doctorate) and a large number for "research." Since research personnel needs are determined by factors largely independent of the purely educational function, and doctorates are probably not a large fraction of the other two categories, it seems much more appropriate to concentrate just on the needs for teaching faculty. As a corollary, this requires counting only new doctorates who enter teaching as a component of supply, rather than the number who enter higher education in all of its various facets.

So much for the Office of Education forecasts; as a stone-thrower I should at least create my own glass house as a target for others. The starting point is a projection of college enrollments (E) and doctoral degrees (P) to 1985 in Table VI. The enrollment projection is similar to that of the Office of Education through 1974, and assumes that the ratio of undergraduate enrollment to the 18-21 age group rises to .55 by 1985.[12] (It is now approximately .40). The doctoral projection is that of the author, and while it is moderately higher than the most recent Office of Education projection, it is below that of the National Science Foundation.[13] Assuming that the projections turn out to be accurate--I believe they are about as good a guess as can be made--one can then attempt to analyze the staffing implications of such a future growth path.

If we are to assess the quality of instructional staff by highest degree obtained (a rough, but useful measure), we need to know the total size of faculty required and the likely number of teachers who will have the doctoral degree. Given the enrollment projection, the total faculty will expand as follows:

$$(1) \quad F_t = F_{t-1} + f(E_t - E_{t-1})$$

where F is faculty, E enrollment, and f the faculty coefficient (the inverse of the student/staff ratio). As indicated in Table IV, f has averaged .0517 over the last decade, and for the moment I will continue to assume that it remains constant.

Given the present number of doctorates on instructional staffs and the doctoral projection in Table V, the number of doctorates in teaching will grow in the following manner:

Table I

Percentage of Total Staff in 1953/54 and New Teachers in Successive Years Who Had the Doctorate

Total Staff in 1953/54	40.5%
New Teachers in 1953/54	31.4
1954/55	28.4
1955/56	26.7
1956/57	23.5
1957/58	25.3
1958/59	23.8
1959/60	25.9
1960/61	25.8
1961/62	27.3
1962/63	25.4
1963/64	28.3
1964/65	27.2

Source: "Teacher Supply and Demand in Universities, Colleges and Junior Colleges, 1963/64 and 1964/65, NEA Research Report 1965 R-4, Table 2.

Table II

Additions to College Teaching Staff and to Doctorates in Teaching 1953-1965

Year	New Teachers (1)	New Teachers With Doctorate (2)	Continuing Teachers Receiving Doctorate (3)	Ratio of New Doctorates in Teaching to New Teachers (4)
(Total Staff in 1953/54)	(58,719)	(23,768)	--	(.405)
1953/54	4,232	1,329	na	na
1954/55	4,694	1,333	822	.460
1955/56	6,337	1,695	856	.403
1956/57	8,308	1,953	1,528	.419
1957/58	9,293	2,354	1,529	.418
1958/59	9,100	2,254	1,825	.448
1959/60	10,221	2,650	1,894	.447
1960/61	11,184	2,886	1,987	.436
1961/62	10,439	2,851	2,115	.476
1962/63	12,186	3,092	2,334	.445
1963/64	13,562	3,833	2,732	.484
1964/65	16,059	4,361	3,084 est.	.463 est.

Source: Columns 1 and 2 from NEA Research Report 1965-R-4. Column 3 computed from Table Y of 1965 Report and comparable tables in earlier reports in the series. Column 4 is Col. 2+3 ÷ Col. 1.

$$(2)\ D_t = [1-(c+m+r-a)]D_t + bP_{t-1}$$

where the following new forms are introduced:
a = the accretion, or in-transfer, rate of persons with the doctorate who enter teaching from other employment;
c = the loss, or out-transfer, rate of doctorates leaving teaching for other employment;
m = mortality rate of the present teachers;
r = retirement rate of the present teachers;
b = the percentage of new doctorates who enter teaching; and
P = the number of doctoral degrees awarded.
(Using academic years as periods, the number of new Ph.D.'s entering teaching in year t depends upon doctoral output in year t-1.)

Given the values of the respective coefficients, equations (1) and (2) indicate how total faculty and doctorates in teaching will grow. In order to see the effect of various values for the coefficients, it may be useful to construct a supply and demand equation, as indicated in (3). Here one new coefficient, q, is added, defined as the percentage of new teachers with the doctorate.

$$(3)\ bP_t = (c+m+r-a)D_t + qf(E_t - E_{t-1})$$

This equation represents a supply-demand identity, the left hand term representing the quantity of new doctorates supplied, and the two right hand terms the replacement and expansion demand components.[14] From various studies we can estimate the approximate values of each of these variables for recent years.
b = .50 (the N.E.A. reports indicate an average of about .48 over the decade, closer to .50 for the last four years).
m = .0069 (calculated from the age distribution of faculty in the COLFACS study).
r = .0112 (calculated from COLFACS data, see note 11 above).
a = .0310 (the rate at which doctorates intended to leave higher education, from COLFACS data).
c = .0321 (estimated on the basis of Table V, above, which indicates that the net transfer rate (c-a) has been approximately .0011).
q = .33 (the estimated value over the past decade).
f = .0517 (from Table IV, above).
D = approximately 90,000 in 1963/64.
The combined factors which go to make up the replacement rate are equal to .0192, as indicated in Table V. These factors are small in magnitude and appear to have been relatively stable in recent years--although (c-a) obviously responds to changes in the relative salary level of academic personnel. The three significant coefficients are b, q, and f. Of these, f has remained relatively constant since 1958 at about .05, and is stable in the sense that it is the result of conscious decisions on the part of college and university administrators in determining the staff/sudent ratio. The percentage of new teachers hired with the doctorate, q, reflects the aspirations of the institutions, for most institutions equate a high q with excellence and a low q with deterioration of faculty quality. The percentage of doctorates who enter college teaching reflects the aspirations of graduate students for a majority enter graduate school with the intention of entering college teaching.[15] Both b and q are highly variable, depending upon yearly market conditions for new Ph.D.'s.

Using the projections for E and P for the next twenty years, assuming a constant faculty coefficient (f = .0517) and a constant replacement rate (rounded off at $[c+m+r-a]=.02$), what behavior might we predict for b and q? Column 1 of Table VII gives the predicted values for b if q remains constant--that is, the percentage of new Ph.D.'s who would have to enter college teaching in order to maintain a constant ratio of doctorates in teaching to total faculty. We might call this the "constant faculty quality" model.[16] Over the coming three years the required value of b is higher than we have experienced or can reasonably expect; therefore the quality of faculty is unlikely to be maintained. However, after 1968 b will steadily decline (to a low of less than 12% for the 1980-85 period), and will probably be lower than it has ever been in history.

Alternatively, column 2 of Table VII projects an "absorption" model, assuming that the percentage of new doctorates entering teaching (b) remains constant at 50%, and that all such available doctorates become employed in college teaching. The ratio of new teachers with the doctorate to annual additions to the instructional staff ($q = \frac{\Delta D}{\Delta F}$) dips for the 1965/66-1967/68 period, then steadily rises to new historical highs. Assuming a constant b, by 1977/78 <u>every</u> new college teacher would possess the doctorate; after that year the absolute number of non-doctorates would fall rapidly as doctorate-teachers displaced non-doctorates faster than the latter were reaching retirement age. Figure I illustrates dramatically the relationship between the available supply if b remains constant quality. After a temporary deficit in the 1965-68 years, the available supply begins to exceed demand by a rapidly growing amount, sharply altering the market conditions for college teachers.

The "absorption" and "constant quality" versions of the model seem to me to represent the outside limits; actual experience will probably lie somewhere between. Figure II shows the outer boundaries, and illustrates an intermediate case similar to the experience of the last decade discussed in the first section of the paper. In this example it is assumed that the overall ratio of doctorates to faculty continues to rise by one-half of one percentage point each year. The two intermediate lines indicate the values for both b and q (given the projections of E and P) for a steadily rising quality model. In this case q rises to about unity (its logical maximum) and b gradually declines to one-fifth.

Table III

Percentage of Doctorates Among 4-Year College and University Faculty

Type of Institution	All Instructional Staff 1950-51 (ACE)1/	1962-63 (ACE)1/	Full-time Instr. Staff 1953-54 (NEA)2/	1962-63 (OE)3/
	(1)	(2)	(3)	(4)
Public Universities	36.0	44.9	44.0	58.4
Private Universities	37.3	43.8	51.9	59.6
Public Colleges	28.2	33.5	30.7	42.6
Private Colleges	29.7	35.4	35.2	42.7
All Institutions	32.3	39.4	40.5	50.6

Sources: 1/ A.M. Cartter, "A New Look . . .", op. cit. p.270.
2/ "Teacher Supply and Demand in Degree Granting Institutions," NEA Research Bulletin XXXIII, 4 (December 1955), p. 138.
3/ "Doctorates Among Teaching Faculty," op. cit., Table III

Table IV

Average and Marginal Faculty Coefficients

1953/54 - 1963/64

	E	ΔE	F	ΔF	F/E	ΔF/ΔE	ΔE:ΔF
1953/54	2,236		182.0	--	.084		
1955/56	2,660	424	197.8	15.8	.076	.037	26.8:1
1957/58	3,047	387	226.5	28.7	.074	.074	13.5:1
1959/60	3,377	330	244.5	18.0	.072	.055	18.2:1
1961/62	3,861	484	266.6	22.1	.069	.046	21.7:1
1963/64	4,495	634	298.9	32.3	.066	.051	19.6:1
1953/54 to 1963/64		2,259		116.9		.0517	19.3:1

Source: "Projections of Educational Statistics to 1973/74" (OE-10030, 1964), pp. 8 and 24. Faculty considered here are members of the Instructional Staff at the level of Instructor or above. The extremely high and low ratios, for 1955/56 and 1957/58 may result from errors in reporting by institutions.

Table V

Estimated Annual Replacement Rate for College Faculty

Losses annually due to:	
Deaths	.69%
Retirement	1.12%
Net Transfer to Other Employment	.11%
Total Annual Losses	1.92%

Table VI

Total College Enrollment (E) and Doctoral Production (P)
Actual 1953-64, and Projected to 1985 (000's)

	E	ΔE	P
1953-54	2,207	--	9.0
1954-55	2,421	214	8.8
1955-56	2,627	206	8.9
1956-57	2,891	264	8.8
1957-58	3,009	118	8.9
1958-59	3,195	186	9.4
1959-60	3,344	149	9.8
1960-61	3,536	192	10.6
1961-62	3,804	268	11.6
1962-63	4,124	320	12.6
1963-64	4,433	309	14.5
1964-65	4,744	311	15.0
1965-66	5,185	441	16.1
1966-67	5,641	456	16.8
1967-68	6,064	423	18.0
1968-69	6,382	318	19.5
1969-70	6,676	294	21.4
1970-71	6,982	306	23.2
1971-72	7,315	333	26.0
1972-73	7,671	356	28.9
1973-74	8,027	356	31.5
1974-75	8,401	374	33.6
1975-76	8,750	349	35.7
1976-77	9,082	332	37.9
1977-78	9,369	287	40.4
1978-79	9,644	275	43.1
1979-80	9,936	292	46.0
1980-81	10,148	212	47.9
1981-82	10,288	140	49.7
1982-83	10,428	140	51.6
1983-84	10,487	59	53.2
1984-85	10,598	111	54.7

Source: Actual figures from Office of Education data; Projections by the author (See A. M. Cartter and R. Farrell, "Higher Education in the Last Third of the Century," The Educational Record, Spring 1965, pp. 119-128)

Table VII

Projected Percentages of New Doctorates Entering Teaching (b), and New Teachers with the Doctorate (q) for "Constant Quality" and "Absorption" Growth Medals: 1965-85

	Values of: b (q=.33)	Values of: q (b=.50)
1965-66	60%	24%
1966-67	62	26
1967-68	54	28
1968-69	41	41
1969-70	36	48
1970-71	34	51
1971-72	33	52
1972-73	31	55
1973-74	29	61
1974-75	28	64
1975-76	26	73
1976-77	24	81
1977-78	21	100
1978-79	19	111
1979-80	19	110
1980-81	15	163
1981-82	13	255
1982-83	12	264
1983-84	9	639
1984-85	11	340

Note: Based on assumed continuing values: f = .0517 and (c+m+r-a) = .02

Figure I

Doctorates Available, and Doctorates Required to Maintain Constant Quality of Faculty at 1963/64 Level

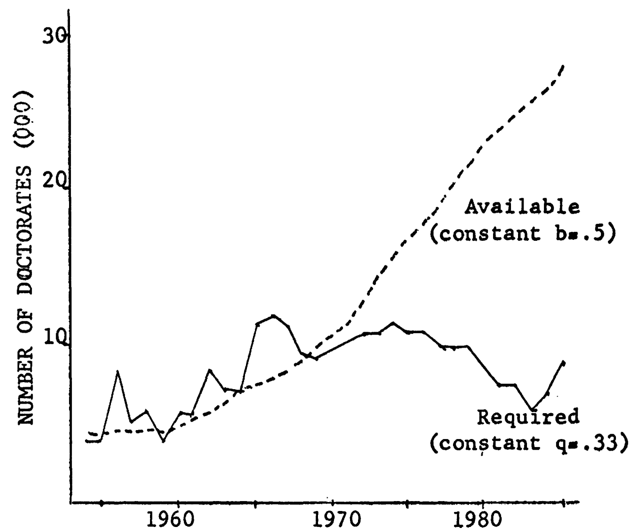
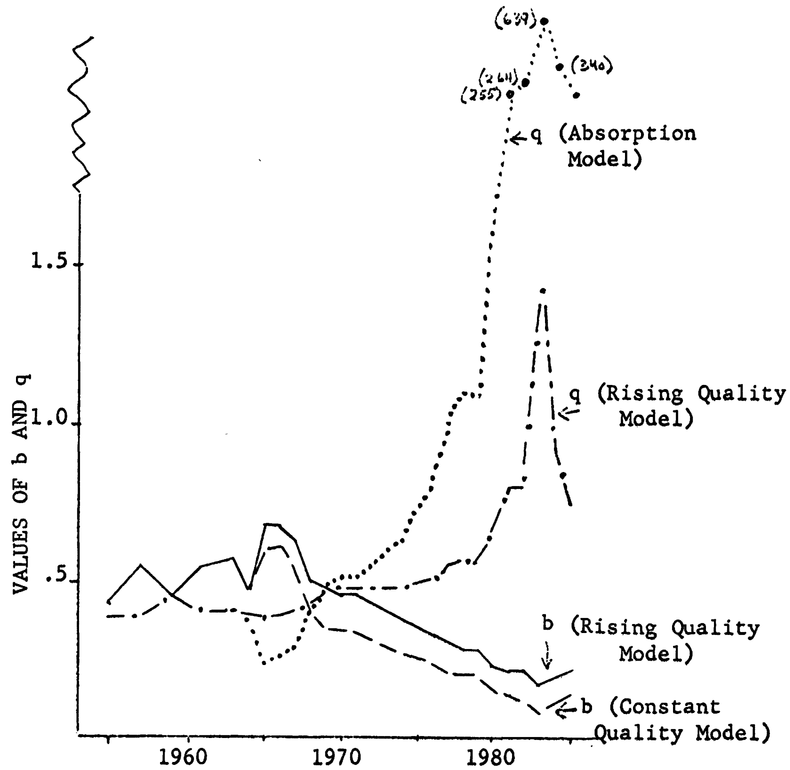


Figure II

Values of b and q for Absorption, Constant Quality, and Rising Quality Models



Conclusions

The preceding analysis suggests that educators have been much too pessimistic about the adequacy of both the present and future supply of college teachers. We seem to have learned little from the experience of the 1950's when the National Education Association and most public school officials were maintaining that there was a critical shortage of school teachers, only to find by the end of the decade that both the number and quality (as measured by formal preparation) of teachers had been steadily rising. Similarly, the despairing cries about the rapidly deteriorating situation on the college level have now proved to be in error, and the future looks bright beyond the next three to five years.

If the projections of total college enrollment and of doctorates to be awarded are even approximately correct, the sellers' market for college faculty will quickly disappear in the early 1970's. This has many implications for public policy and for the nation's colleges.

Given the time lag between entrance to graduate school and completion of doctorate, it is conceivable that graduate education facilities might be expanded too rapidly by basing decisions on degrees awarded in the recent past. The present faculty and facilities, at their current level of utilization, would turn out about 20,000 doctorates a year in a stable system. That is to say, because we are rapidly expanding we occasionally forget that the fifteen thousand doctorates awarded this year reflect the teaching capacity of the graduate schools about 1960. If, as the model suggests, the demand for new doctorates in teaching will stabilize or even decline after 1968, as a consequence of the declining rate of growth of the total system, then a serious question of public policy may be whether or not it is desirable to encourage many new institutions to enter the doctoral field. Four-fifths of the present nearly 250 universities awarding the Ph.D. are too small to be educationally or economically efficient. We might well ask whether public policy would be better served by consolidating and strengthening our existing graduate schools, rather than encouraging another ten or twelve new doctoral granting institutions to join the university ranks each year as is now occurring.[17]

The model also has serious implications for the future level of academic salaries. For the next three years the market will remain fairly tight, and the succeeding several years may be needed to regain temporarily lost ground. The 1970's, however, may usher in a "buyers' market," and academicians may experience again a decline in their relative income position. The model above assumed that the replacement rate remained constant over the next twenty years, but this is unlikely in a market where supply is relatively abundant. There may develop a trend for colleges to lower mandatory retirement ages (thus raising r), and the transfer rate of senior staff (c-a) will probably rise a few percentage points. For example, a tendency for b to fall as a result of a decrease in demand would tend to depress beginning academic salaries. As the upward pressure on salaries of new Ph.D.'s diminishes, colleges may let out-transfers increase and reduce in-transfers of older doctorates (i.e. c-a would rise) partly stemming the decline in b. Junior and senior faculty are relatively good substitutes from the point of view of performing the teaching function. Alternatively, the slack might be taken up by a rising faculty coefficient (a reduced marginal student/staff ratio).

If I were to hazard a guess fifteen or so years ahead, I would predict a fairly constant marginal faculty coefficient (f), a gradually diminishing percentage of new Ph.D.'s entering teaching (b) after 1970, a continuing modest improvement in the percentage of faculty with the doctorate $(q > \frac{D}{F})$, a positive net out-transfer rate ($c > a$ and gradually rising), and a slowing down in the upward drift of academic salaries becoming noticeable in the early 1970's. It may well be that the real challenge to Committee Z of AAUP will come in the 1970's when in all probability market forces will be an opponent rather than an ally in efforts to improve the relative income position of college teachers.

The discussion above has ignored field-by-field differences partly in the interests of brevity and partly because the aggregate data are better than that for individual disciplines. There are wide variations in the values of each of the coefficients from field to field,[18] but the demarcations between fields are too fuzzy to permit the application of such a model with any degree of precision to individual disciplines. Certainly shortages in many fields will continue beyond 1970, but the general outlook appears to be favorable for the continued expansion and improved quality of higher education in the United States.

Footnotes

1/The first report in the series had a slightly different title; see "Teacher Supply and Demand in Degree Granting Institutions, 1954-55", NEA Research Bulletin XXXIII, 4 (December 1955). The Series has been under the directorship of Ray C. Maul, and the most recent is NEA Research Report 1965-R4.

2/These data are not precisely comparable, but should be sufficient to illustrate the principle. Maul's data on new teachers is drawn from questionnaires to the colleges hiring new teachers, while his data on the employment of new doctorates is drawn from questionnaires to the graduate schools granting the doctorates.

3/"Teaching Faculty in Universities and 4-Year Colleges, Spring, 1962" by Dunham, R.E., Wright, P.S., and Chandler, N.O. (OE-53022-65). Preliminary data were presented in a paper "Doctorates Among Teaching Faculty" at the annual meeting of the American Educational Research Association, February 11, 1965. This study is commonly referred to as COLFACS.

4/See "A New Look at the Supply of College Teachers," The Educational Record, (Summer 1965), pp. 267-277.

5/Teacher Supply and Demand . . .", NEA Research Report 1959-R10, pp. 50-54. The same model was used also in the 1961 Report, but did not appear in the later reports.

6/"Projections of Educational Statistics to 1973-74" (OE-10030, 1964), p. 26.

7/The ratio, in terms of full-time staff equivalents, ranges from a low of 10:1 to a high of 16:1 in somewhat random fashion, but averages 14 for the decade to 1973-74. In terms of total instructional staff at the rank of Instructor or above, the OE projection ranges from 14:1 to 27:1, averaging 18:1. As Table IV indicates, this is lower than the average of the last decade. See "Projections of Educational Statistics . . ." op. cit., pp. 8 and 24.

8/The choice of the appropriate replacement rate is so critical to the model that it is surprising that no very serious attempts have been made to verify it. A difference of one percentage point makes a difference of about 40,000 teachers over a decade. Various assumptions have been used by different model builders -- e.g. 5% by the Fund for the Advancement of Education in Teachers for Tomorrow (1955) 5% by Brown in The Market for College Teachers (1965), 4% by Berelson in Graduate Education in the United States (1960), 3% by Wolozin in "How Serious Is the Faculty Shortage?", Challenge (June 1965).

9/Memoranda on "Estimates of Demand for and Supply of Higher Educational Staff," Higher Education Personnel Staff, Office of Education, October 26, 1964 and January 4, 1965.

10/Ideally one would like to use full-time equivalents for both measures, but national enrollment data is not available on this basis despite the fact that the Higher Education Facilities Bill of 1963 uses a full-time equivalent enrollment formula for the distribution of Title I funds.

11/Bolt, Kolton and Levine have recently published a model for scientific fields which is in close agreement with the above. Their estimate for scientists, based on a review of National Register data for recent years and an assumption that scientists retire at age 65, is:

Death rate	.009
Retirement	.006
Total	.015

See "Doctoral Feedback into Higher Education," Science (May 14, 1965), pp. 918-28. The retirement assumptions in my estimate are that 4% of faculty aged 60-64 voluntarily retire each year, and that from age 65 on teachers on the average retire from teaching one year after mandatory retirement age is reached. This is the equivalent of assuming that one-third retire at the mandatory age, one-third continue (probably at another college) for one year, and one-third for two years. An alternative assumption that 10% of teachers age 60 and above will retire each year would give a current rate of .0098.

12/See A. M. Cartter and R. Farrell, "Higher Education in the Last Third of the Century," The Educational Record, (Spring, 1965) for the development of this and alternative projections of enrollment.

13/See "Projections of Educational Statistics to 1973/74," op. cit., pp. 12-16 for OE Forecasts, and Comparisons of Earned Degrees Awarded 1901-62 -- With Projections to 2000 (NSF-1964), p. 54. The author's "A New Look at the Supply of College Teachers, " The Educational Record (Summer, 1965) compares these with other doctoral projections. For periods up to ten years ahead P may be taken as an exogenous variable, determined by the level of fellowship support, the capacity of graduate schools, etc. In projecting doctoral degrees, however, I have assumed after 1974 that P_t is a function of E_{t-7}, the value of the functional coefficient being .0058. From 1964 through 1974 the value of this coefficient is approximately (.0047 + .0001t). This model produced reliable estimates of doctorates for years before 1964, and a projection that falls reasonably between the low estimates of the Office of Education and the high estimates of the National Science Foundation. For the 1974-85 period it is very close to Lindsay Harmon's "Reference Series." See "Memorandum on Projected Doctorate Production", National Academy of Sciences (January 29, 1965).

[14]/If one were collecting data from colleges and universities, it would be more appropriate to express the supply-demand identity as follows:

(3A) $bP_t+aD_t+sD_t=sD_t+(c+m+r)D_t+qf(E_t-E_{t-1})$

This differs from equation (3) in that aD_t is shifted to the left hand side, since it is technically part of the supply of doctorate-teachers, and a new term sD_t appears on both sides of the equation (s being defined as the percentage of teachers who shift teaching positions from one college to another in any year); when aggregating sD_t cancels out. According to COLFACS data, s = .114 in 1962/63.

[15]/I would estimate that half of science students and at least 90% of non-science students would prefer--other things being equal--to enter college teaching. The fact that only about 20% of the former and 75% of the latter category do become teachers upon receiving their degrees is attributable to the economically attractive alternatives at the time of graduation. b is therefore assumed to be sensitive to relative salaries in academic and non-academic occupations. In the economist's terms I would assume that b is price (i.e. salary) elastic, and that q is relatively price inelastic.

[16]/This might be compared with Brown's "Quality-constant supply" function, which uses different (and I believe unlikely) assumptions. See <u>The Market for College Teachers</u>, op. cit., pp. 18-27.

[17]/One quick answer is that government and industry can absorb all the additional doctorates produced. This may turn out to be so, but if it does occur doctorates in non-educational employment will experience an increasing rate of growth. For example, if the educational system followed the path indicated by the constant-quality model, then doctorates entering non-teaching employment would grow from the present level of about 7,500 per year, to 26,000 in 1975 and to 54,000 by 1985. Over the last ten years the total number of employed non-teaching doctorates has grown about 4-5% per year; over the next twenty years it would expand at the rate of about 10% per year.

[18]/Take the retirement rate (r) as an example; over the next five to ten years it will probably average from a low of only .47% in Biochemistry to a high of 3.71% in Classics. Judging from the present age distribution of teachers, the combined mortality and retirement rate for the next several years will be about 2.5% in the Humanities, 1.7% in the Biological and Physical Sciences, 1.6% in Engineering, and 1.9% in the Social Sciences.

PATTERNS OF CHANGE IN THE LONG-RUN CAREER FIELDS OF JUNE, 1961, COLLEGE GRADUATES

Joe L. Spaeth, National Opinion Research Center

An important element in the process of allocating the current scarce supply of trained manpower to the wide range of positions that must be filled is the decisions of the individuals involved. Persons may change their minds regarding their field of study or the one they are preparing to work in during college and after graduation as well. Such decisions may have an impact on the distribution and quality of talent available to a given field. This paper is a preliminary analysis of changes in the prospective career fields of one college graduating class, covering a time span from the freshman year to three years after graduation.

In the spring of 1961 the National Opinion Research Center began a longitudinal survey of that year's June college graduating class. After a sample of 135 institutions was drawn, officials at each were asked for lists of prospective recipients of the collegiate bachelor's degree. All 135 co-operated, and from their lists a sample of about 41,000 names was drawn. Each of the students in the sample received a self-administered questionnaire to fill out and return to NORC. New questionnaires were sent in the spring of each year through 1964. In 1961, 85 per cent returned usable questionnaires. The figures for subsequent waves are: 1962, 76 per cent; 1963, 71 per cent; and 1964, 60 per cent. The last was in response to a very long mailed questionnaire of twenty-four pages for the men and forty-four for the women.[1]

Each of the questionnaire waves included questions on current career decisions, such as planned or actual enrollment in graduate school, employment, and plans for the future--expected career field, anticipated career activities, etc. The senior-year questionnaire also asked about college experiences, social backgrounds, and so on. All in all, we have a rather complete set of data on the aspirations and activities of our sample and how these have changed over time.

This paper deals with changes in respondents' long-range plans regarding the field--generally corresponding to an academic discipline--in which they plan to work during their career. The actual question is:

> Which field from the list. . .best describes your anticipated career field? Please enter the code number of the field you expect to be your long-run career and ignore any stop-gap job or temporary military service which might precede it.

The list contained about one hundred fields, ranging from classical languages to veterinary medicine. These detailed fields have been grouped for this analysis into the following twelve broad categories: physical sciences, biological sciences, social sciences, humanities, engineering, medicine, health fields other than medicine, education, business, law, the remaining fields listed--here called "other professions"--and NEC (not elsewhere classified), which for the men who are the subject of this paper refers to a code indicating that no field on the list closely corresponds with the person's career field intentions.

There are many approaches to the analysis and description of changes taking place in longitudinal data. Transition probability matrices can be compared to see if the change processes are the same over two or more different time spans. They can also be compared across categories of an independent variable to see whether different conditions imply different change processes. Another approach is to try to fit the observed process to a statistical model: Markov chains, or the stayer-mover model, for example.[2] Still another method is that adopted by James A. Davis for dealing with the first part of the data presented here. Using the same categories listed earlier, he analyzed the associations of a variety of independent variables with being in a field as a freshman, staying in it, or leaving it during college.

My concern today is different from those just enumerated, the enumeration being to specify what I am <u>not</u> going to do. In an effort to see whether there are patterns of change among fields regardless of the number of people entering and leaving them, independence values were computed just as with chi-square. These are compared with the observed values, with the cells markedly exceeding chance used as the basis for describing the patterns of change. A clearer idea of the procedure will come with discussion of the actual results. This discussion will be confined to men in order to avoid the complications that might arise concerning women's rather different career aims.

---

[1] The research reported herein was supported through the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education and Welfare, under Contract SAE-9102, and by the National Institutes of Health under Grants M5615, M5615-02, M5615-03, and M5615-04. The author wishes to thank Harold Levy, of NORC's Data Processing department, for writing and running the program for this analysis.

[2] Leo A. Goodman, "Statistical Methods for Analyzing Processes of Change," <u>American Journal of Sociology</u>, LXVIII (1962), 57-78, and "Statistical Methods for the Mover-Stayer Model," <u>Journal of the American Statistical Association</u>, LVI (1961), 841-68.

Before turning to the data on changes in career plans, one manpower-related question can be answered rather quickly. As Table 1 shows, the distribution of career fields for the men changed markedly between the freshman and senior years. Business, the social and biological sciences, education, and the humanities made marked relative gains; engineering, medicine, and other health fields suffered marked losses. After graduation, however, the picture is rather different. The distribution for the senior year is quite similar to that for the third year after graduation. With the exception of "NEC," which is too ambiguous to analyze, the largest absolute change is the 5 per cent gain for business, a 21 per cent relative gain. Other absolute changes do not exceed 2 per cent; relative ones rarely exceed 10 per cent. In the years immediately after college graduation none of the broad fields gains or loses much in comparison with the others, and the supply of manpower to various fields provided by this cohort, at least, does not shift markedly from one field to another.

Table 1 gives only the net balance among fields and says nothing about the number of individuals who may have changed fields in bringing that balance about. Table 2 gives the percentages of men who reported the same field for each of the two time periods. Overall, 58 per cent of all respondents reported the same career field in the senior year as in the freshman, and 70 per cent made the same report for the period spanning the senior year and the third year after graduation. (This is not a direct estimate of stability, since some of these men may have moved out of a field and then back into it again during the time period in question.) It seems apparent, however, that there was considerably less changing after graduation than before in all fields except education.

Such stability poses problems in the analysis of changes. If the analysis were to be done on all men in the sample, changers and nonchangers alike, patterns of change would tend to be swamped by the overriding presence of the nonchangers. The method adopted here--comparing observed with expected values as in chi-square computations--suffers particularly from this drawback. Accordingly, all cases which failed to show a change from one time to the next were eliminated before the computation of expected values. These are the cases in the top left to lower right diagonal of the tables.

Expected values were computed on the basis of statistical independence between time 1 and time 2. However, the statistic analyzed is not the traditional single-valued chi-square, but the ratio of observed to expected values. There is, of course, one of these ratios for every cell in the table, just as there is an expected value for every cell. This fact is at present the source of some difficulty, since there are expected values for the diagonal which was made empty by excluding nonchangers. This means that all off-diagonal expected values are slightly too low, on the average by a factor of one-twelfth. A gross adjustment to take this problem into account has been made and will be described shortly.

The matrix of O/E values contains 132 (144 - 12) cells of interest, the values of which range from zero up. Since scanning the matrix was done by the "eyeball" method, two criteria were established in order to simplify the search for structure: the higher was an O/E value of 2.000 or greater, the lower a value ranging from 1.500 to 1.999. The former characterizes all ratios that occur at least twice as often as chance, the latter ratios between one and one-half and twice as likely as chance. A crude adjustment for the presence of expected values in the diagonal was made by multiplying two, and one and one-half, by 13/12 to give the cutting points actually used, 2.167 and 1.625.

The O/E values for two time periods are shown in Charts 1 and 2--(1) from the freshman to senior years in college, and (2) from the senior year to three years after graduation. Chart 1, summarizing patterns of change during the earlier time period, shows a rather definite structure which will be clarified in the later one. Starting at the top and going down to the right, we see some sign of a relation between law and business, more definite links between engineering and the physical sciences, and a triplet composed of the life sciences-related fields of biology, medicine, and "other health." Finally, there is a rather amorphous cluster including humanities-education and humanities-social science doublets.[3]

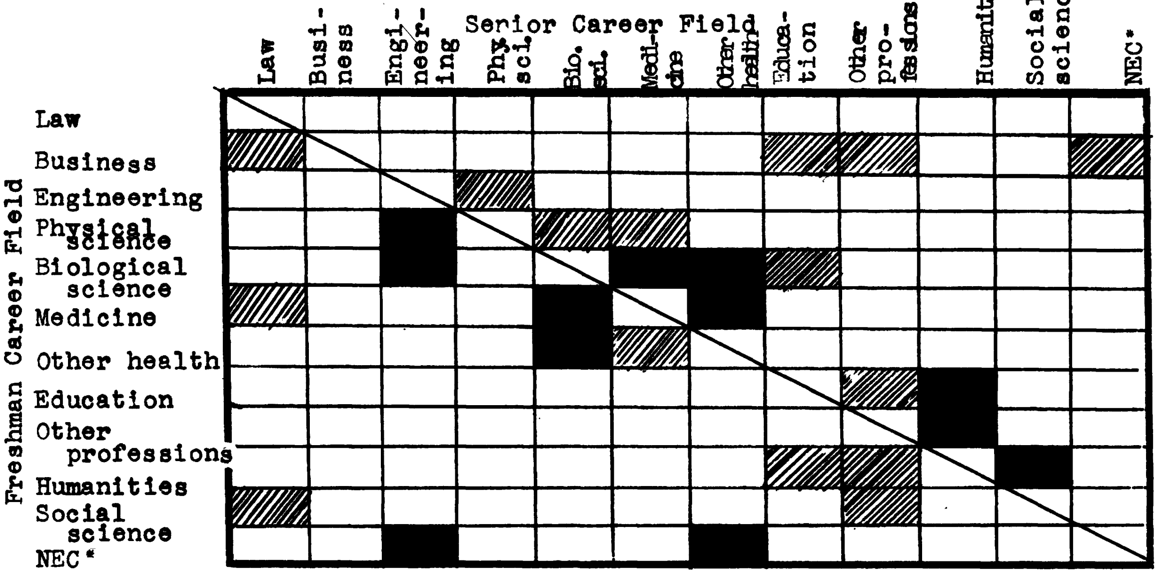
On the whole, results for the period <u>beginning</u> with the senior year are similar to those for the earlier period, as Chart 2 shows. The business-law cluster is joined by an engineering-business pair. Engineering and the physical sciences continue to be linked; the biological sciences are related to the physical sciences as well as to the two health-related categories. In fact, the fields ranging from law to "other health" form a kind of chain whose links are directional. That is, the flow of changers may exceed independence values both into and out of a given pair of fields, like engineering and the physical sciences. In the chain from Chart 2 all links are two-directional except the first; law does not supply a disproportionately large number of recruits to business, perhaps because prospective lawyers were still in school and lacked the employment experiences that might bring about such a change.

Continuing down the diagonal of Chart 2, we see that the amorphous cluster mentioned in connection with Chart 1 has split into two pairs--the social sciences with "other professions" and humanities with education. Even so,

---

[3]Since the charts merely summarize the data, two sets of tables are appended: the actual numbers of weighted cases involved and the O/E values on which this discussion is based.

1. SUMMARY OF CHANGES IN CAREER FIELD FROM FRESHMAN YEAR TO SENIOR YEAR (MALES)

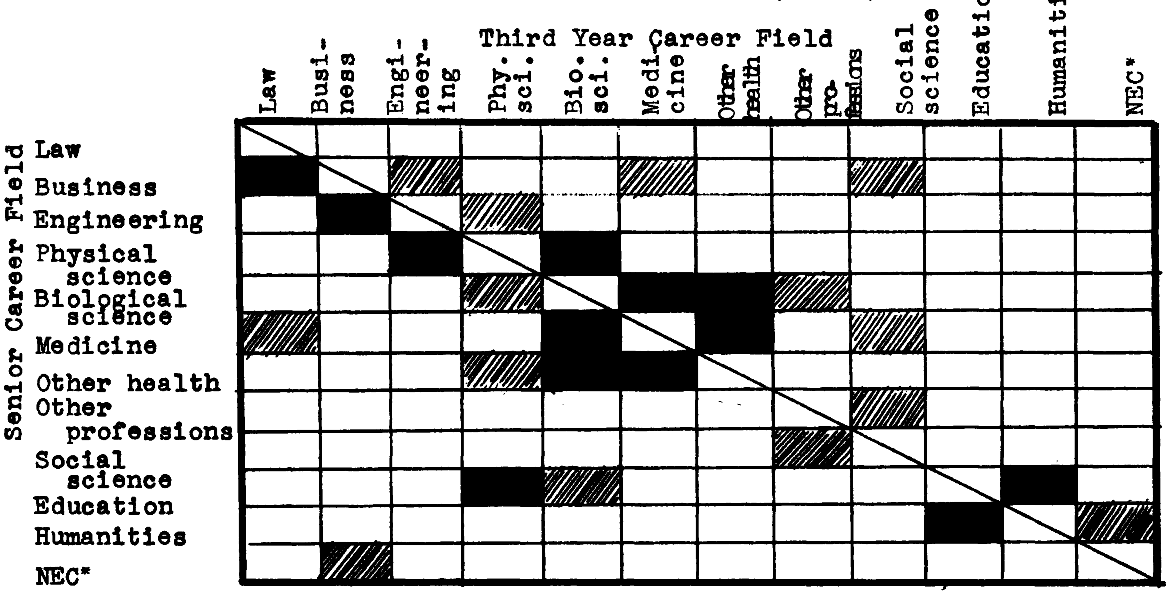


*NEC is "not elsewhere classified."

O/E > (13/12) 2.000

(13/12) 1,999 > O/E > (13/12) 1.50

2. SUMMARY OF CHANGES IN CAREER FIELD FROM SENIOR YEAR TO THIRD YEAR AFTER GRADUATION (MALES)



*NEC is "not elsewhere classified."

O/E > (13/12) 2.000

(13/12) 1,999 > O/E > (13/12) 1.50

the pattern of Chart 2 is on the whole quite similar to that for the college period.

But some of the differences are rather interesting. It is a well-known fact of life in the business world that some engineers will move into management; otherwise their career mobility is likely to be limited. Also, some engineers function as salesmen, drawing on their professional training to some extent. There is no evidence for such a switch in career plans during the college years, but it is quite clear for the period after graduation. There is also a weaker movement in the other direction. Perhaps some businessmen see a rosier future in the technical sphere.

Another difference between the two time spans involves the undergraduate linkage between the humanities and social sciences, which disappears after graduation. Instead the social sciences become consumers, as well as producers, of "other professionals." The actual fields involved in this interchange are probably sociology and clinical psychology on the one side and social work on the other.

The humanities-education link is a stable one, probably indicating movement of English and history people in terms of teaching those subjects to high school versus college students. Interestingly enough, the humanities are the only supplier of educators. In other words, physical, biological, and social scientists are not extraordinarily likely to add to the supply of trained elementary and secondary school teachers. These data apply to men only, of course. Perhaps this deficit is made up by the women.

One general configuration underlying the clustering seems to be a back-and-forth movement between substantively similar fields along a pure-applied dimension: engineering-physical sciences, biology-medicine, social sciences-"other professions," etc. Another reflects patterns of career development in the business world, engineering to business. That career field changes exceed chance among groupings of similar fields can be explained in part by the constraints imposed by training. If one is going to change fields, it is easier to go into a field with which one has some familiarity than to start all over again. An interesting problem raised by the existence of these clusters is the correlates of movement in one or the other direction. For example, do the abler men move from the applied to the pure fields or in the opposite direction; or is ability unrelated to such changes?

Up to this point, I have paid little attention to the order in which the fields have been presented except to note the chainlike relations among certain fields in the second time period. The order, which maximizes the number of cells adjacent to the diagonal, was partially dictated by the data--partially because more than one order is possible, even under the diagonal-maximizing criterion.

The order for the first time period is different from that for the second. Of the twenty-seven first-period ratios meeting the 1.625 minimum, sixteen, or 59 per cent, were as close to the diagonal as they could be.[4] For the second time period, eighteen of the twenty-seven values (67 per cent) were as close to the diagonal as possible. If the post-college order were imposed on the college data, the "fit" to the diagonal would be 13/27, or 48 per cent, a decline of 11 per cent.

The ordering of Chart 1 is interesting for another reason. Table 1 is arrayed according to the magnitude of relative change among the twelve fields over the undergraduate years, starting with the largest net gain and ending with the greatest net loss. When the freshman-senior net turnover data are arranged in that order, the fields with the greatest gains make those gains from nearly every field "under" them, and the fields with greatest losses lose to nearly every field "over" them.[5] Yet the order of Table 1 is very different from that of Chart 1. In other words, an adequate description of the net gains and losses between fields does not correspond with the ordering which describes the links between them.[6] This is not to say that one or the other of the two different descriptions is more desirable or accurate than the other, only that they refer to different aspects of the same data.

In addition to the table summarized in Chart 2, data are available for each of the three one-year time spans (including that beginning with the senior year) of the post-college period. Though the data are not presented here, they have a structure which closely resembles that of the four-year span reported here. The optimum order originally found was somewhat different from that given here, but when the present order is imposed the percentages along the diagonal are exactly the same for two of the time periods and less by one cell for the other, clustering around 70 per cent. Most of the doublets from the longer span occurred in the one-year ones, and the triplet was always present. The off-diagonal cases differed to a greater extent and were not nearly as stable as the ones next to the diagonal. In other words, the year-by-year patterns of change are reflected in Chart 2, which covers the entire post-college period.

[4] The biological sciences-"other health" links are one step removed from the diagonal, which is as close as they can be, since they are members of a triplet.

[5] Cf. James A. Davis, Undergraduate Career Decisions (Chicago: Aldine Publishing Company, 1965), 23-26.

[6] This difference stems, in part, from logical necessity. Of any two fields, only one may gain at the expense of the other. But, where this is the dominant pattern for the college years, it is much less important after graduation.

TABLE 1

DISTRIBUTION OF LONG-RUN CAREER FIELD OF MEN, FRESHMAN AND SENIOR YEAR IN COLLEGE, AND THIRD YEAR AFTER GRADUATION

Long-run Career Field	Freshman Year	Senior Year	Third Year after Graduation	Relative Change	
				Freshman-Senior (Per Cent)	Senior-Third Year (Per Cent)
NEC[a]	1.7	3.6	2.4	+112	-33
Social science	2.0	3.6	3.2	+ 80	-11
Business	15.7	23.8	28.7	+ 52	+21
Humanities	3.4	4.6	4.2	+ 35	- 9
Education	11.7	15.7	15.3	+ 34	- 3
Biological science	1.5	2.0	2.2	+ 33	+10
Law	5.1	6.0	6.0	+ 18	0
Other professions	10.9	11.5	12.2	+ 6	+ 6
Physical science	9.2	8.0	6.0	- 13	-14
Other health	3.5	2.2	2.2	- 37	0
Medicine	7.6	4.7	4.2	- 38	-11
Engineering	27.8	14.5	12.5	- 48	-14
Total	100.1	100.2	100.0		
Base N	17,545	19,324	19,180	$\lvert\Sigma\rvert$ 506	132
NA	2,148	369	513	$\lvert\bar{x}\rvert$ 42	11
Total weighted N[b]	19,693	19,693	19,693		

[a]NEC = "not elsewhere classified."

[b]Students in the larger undergraduate institutions were over-sampled, leaving some of the smaller schools underrepresented. Responses from the smaller colleges have been weighted so that the sample corresponds to a straight probability one.

TABLE 2

PERCENTAGE OF MEN WITH SAME CAREER FIELD AT AT TIME 2 AS AT TIME 1, BY CAREER FIELD

Long-run Career Field	Freshman-Senior	Senior-Third Year
Law	59.9 (877)	73.5 (1,142)
Business	76.2 (2,735)	81.9 (4,484)
Engineering	49.6 (4,825)	73.1 (2,743)
Physical science	54.2 (1,613)	69.9 (1,513)
Biological science	44.2 (258)	64.7 (374)
Medicine	50.7 (1,324)	82.5 (874)
Other health	41.6 (603)	71.1 (412)
Other professions	58.9 (1,883)	64.7 (2,175)
Social science	37.5 (344)	49.2 (666)
Education	74.4 (2,017)	72.6 (2,925)
Humanities	46.6 (583)	55.1 (856)
NEC[a]	66.8 (292)	17.4 (674)
Total	58.3 (17,354)	70.5 (18,838)
NA, Time 1 or Time 2	2,339	855
Total weighted N[b]	19,693	19,693

[a]See footnote, Table 1.

[b]See footnote, Table 1.

Two possibilities are opened up by these findings. The first is encompassed by the present discussion: the cells with O/E values of at least 1.625 can be treated as doublets and triplets which tend to form a chain. Taking the O/E values, or rather their reciprocals, as quantitative measures of the distance between fields, one could attempt to construct a higher dimensional model which took account of all 132 cells and not just the largest ones. The results would probably be rather complex, however, and might extend beyond the third dimension, in which case visualization would be impossible.

The second possibility for further analysis is the pinpointing of pairs and triplets. Movement between these groups can be analyzed along lines traditional in survey research. What are the characteristics of the changers, and do these characteristics differ when movement is in one direction rather than another, etc.?

In summary, then, the procedure of eliminating stable cases and examining the O/E values for men who changed their long-run career fields has shown that "affinities" exist between certain pairs or triads of fields. Since there are more changers in some fields than in others, these affinities do not necessarily describe those fields furnishing large numbers of recruits to other fields. However, these findings do describe, for the period immediately after college graduation, an important facet of the interchange between broad classes of fields.

TABLE A-1

FRESHMAN CAREER FIELD BY SENIOR CAREER FIELD

(N)

Freshman Career Field	Senior Career Field											
	Law	Business	Engineering	Physical Sciences	Biological Sciences	Medicine	Other Health	Other Professions	Social Sciences	Education	Humanities	NEC[a]
Law	525	140	3	7	2	8	3	58	27	40	38	26
Business	102	2,084	19	24	2	10	2	147	60	181	45	59
Engineering	162	903	2,392	388	35	56	36	241	98	289	77	148
Physical sciences . .	53	162	86	875	41	53	8	83	44	101	56	51
Biological sciences .	0	13	13	3	114	20	8	23	15	41	6	2
Medicine	82	137	28	52	77	671	62	46	54	55	36	24
Other health	19	104	13	10	30	19	251	50	14	84	4	5
Other professions . .	47	174	40	28	28	9	8	1,109	64	187	149	40
Social sciences . . .	26	49	6	4	1	5	2	50	129	42	24	6
Education	25	119	27	35	26	2	10	125	54	1,501	77	16
Humanities	10	43	2	15	0	12	5	67	48	97	272	12
NEC[a]	5	17	10	8	0	2	9	18	3	23	2	195

N	16,111
NA on either freshman or senior career field .	3,582
Total N	19,693

[a]See footnote, Table 1.

TABLE A-2

FRESHMAN CAREER FIELD BY SENIOR CAREER FIELD

(Observed/Expected Values)

Freshman Career Field	Senior Year Career Field											
	Law	Business	Engineering	Physical Sciences	Biological Sciences	Medicine	Other Health	Other Professions	Social Sciences	Education	Humanities	NEC[a]
Law	0.000	1.546	0.250	0.251	0.170	0.839	0.403	1.313	1.154	0.721	1.520	1.374
Business	2.135	0.000	0.855	0.465	0.092	0.567	0.145	1.799	1.387	1.765	0.973	1.686
Engineering	0.907	1.443	0.000	2.010	0.430	0.850	0.700	0.789	0.606	0.754	0.446	1.132
Physical sciences .	0.979	0.854	3.414	0.000	1.661	2.651	0.513	0.896	0.897	0.869	1.068	1.285
Biological sciences	0.000	0.351	2.645	0.263	0.000	5.128	2.627	1.273	1.567	1.807	0.587	0.258
Medicine	1.711	0.816	1.256	1.004	3.526	0.000	4.490	0.561	1.244	0.535	0.776	0.684
Other health . . .	0.736	1.149	1.082	0.358	2.548	1.993	0.000	1.132	0.598	1.515	0.160	0.264
Other professions .	0.827	0.874	1.514	0.456	1.082	0.429	0.489	0.000	1.244	1.534	2.710	0.961
Social sciences . .	1.648	0.886	0.818	0.235	0.139	0.859	0.440	1.853	0.000	1.240	1.571	0.519
Education	0.660	0.897	1.533	0.855	1.507	0.143	0.917	1.931	1.574	0.000	2.101	0.577
Humanities	0.438	0.538	0.188	0.608	0.000	1.425	0.760	1.717	2.322	1.980	0.000	0.718
NEC[a]	0.702	0.681	3.020	1.040	0.000	0.761	4.388	1.479	0.465	1.505	0.290	0.000

[a]See footnote, Table 1.

TABLE A-3

SENIOR YEAR CAREER FIELD BY THIRD YEAR CAREER FIELD

(N)

Senior Career Field	Third Year Career Field											
	Law	Business	Engineering	Physical Sciences	Biological Sciences	Medicine	Other Health	Other Professions	Social Sciences	Education	Humanities	NEC[a]
Law	839	137	15	10	0	4	2	45	21	39	22	8
Business	127	3,671	96	25	3	20	9	206	64	161	23	79
Engineering	29	480	2,005	65	3	5	4	80	3	24	1	44
Physical sciences . .	7	115	105	1,057	32	3	10	49	9	70	12	44
Biological sciences .	0	22	1	10	242	20	18	42	1	14	1	3
Medicine	18	18	1	3	31	721	36	22	12	7	4	1
Other health	3	38	2	9	14	10	293	14	4	23	0	2
Other professions . .	49	303	40	15	26	6	12	1,407	75	118	57	67
Social sciences . . .	20	75	4	14	1	2	8	109	328	70	26	9
Education	18	201	22	88	44	0	16	183	42	2,124	150	37
Humanities	13	51	10	3	0	5	0	66	18	179	472	39
NEC[a]	20	305	47	10	7	5	5	78	14	52	14	117

N	16,111
NA on either freshman or senior career field .	3,582
Total N	19,693

[a]See footnote, Table 1.

TABLE A-4

SENIOR YEAR CAREER FIELD BY THIRD YEAR CAREER FIELD

(Observed/Expected Values)

Senior Year Career Field	Third Year Career Field											
	Law	Business	Engineering	Physical Sciences	Biological Sciences	Medicine	Other Health	Other Professions	Social Sciences	Education	Humanities	NEC[a]
Law	0.000	1.441	0.803	0.728	0.000	0.918	0.306	0.924	1.466	0.946	1.303	0.441
Business	2.858	0.000	1.915	0.679	0.127	1.710	0.513	1.576	1.665	1.455	0.508	1.623
Engineering	0.719	2.073	0.000	1.944	0.140	0.471	0.251	0.674	0.086	0.239	0.024	0.996
Physical sciences .	0.281	0.804	3.734	0.000	2.424	0.457	1.016	0.669	0.417	1.128	0.472	1.612
Biological sciences	0.000	0.531	0.123	1.672	0.000	10.534	6.320	1.980	0.160	0.779	0.136	0.380
Medicine	2.152	0.375	0.106	0.433	7.000	0.000	10.906	0.895	1.659	0.336	0.469	0.109
Other health . . .	0.461	1.018	0.273	1.669	4.064	5.842	0.000	0.732	0.711	1.420	0.000	0.281
Other professions .	1.167	1.258	0.845	0.431	1.170	0.543	0.724	0.000	2.065	1.129	1.332	1.457
Social sciences . .	1.083	0.707	0.192	0.914	0.102	0.411	1.097	2.006	0.000	1.522	1.380	0.445
Education	0.411	0.800	0.445	2.425	1.898	0.000	0.926	1.421	1.109	0.000	3.360	0.772
Humanities	0.619	0.423	0.422	0.172	0.000	0.905	0.000	1.069	0.991	3.425	0.000	1.696
NEC[a]	0.657	1.745	1.368	0.396	0.434	0.624	0.416	0.871	0.532	0.686	0.451	0.000

[a]See footnote, Table 1.

METHODS OF PROJECTING SUPPLY AND DEMAND IN HIGH-LEVEL OCCUPATIONS

Harold Goldstein, U. S. Bureau of Labor Statistics

The use of statistics as a guide to action often requires that they be drawn upon to make long-term projections. This is one of their most challenging uses because it implies the analysis of causation of economic and social events through the statistics. Not only do we gain insight into the underlying laws operating, but we also find out the inadequacies of the existing bodies of statistical data themselves. Thus, research on projections contributes to the never-ending reevaluation and improvement of the Nation's statistical programs. It is in this context that I want to review the research methodology in manpower projections.

Projections of supply and demand in high-level occupations (which I take to mean occupations for which higher education is the usual means of preparation) are used for several purposes: (1) planning educational programs and estimating what expansions in enrollments must be provided for in order to meet the Nation's needs for trained workers; (2) evaluating the feasibility of launching new programs requiring high-level personnel (studies were made, for example, before the United States launched the space program and the national communities mental health program); (3) evaluating the feasibility or the implications of attaining certain standards of services--such as reducing the pupil-teacher ratio in secondary schools or attaining a desirable ratio of engineers to technicians in industry; and (4) the vocational guidance of individuals.

Questions have been raised as to the value of such projections, in view of the substantial technical difficulties in making accurate projections and the considerable amount of flexibility that individuals have shown in adapting themselves to occupations for which they had not been trained, and that institutions have shown in adapting their operations to shortages of certain occupations. However, this flexibility, while it should be encouraged, has its limitations: we cannot easily make physicians and physicists out of lawyers and social workers, and the process of retraining or reeducating can be costly in time and resources. Prudence requires that an educational establishment on which billions of dollars are spent annually be shaped with some regard to the best appraisal that can be made of future manpower requirements. The same is true with respect to attempts to launch immense new programs requiring specialized personnel or to achieve substantial changes in standards of service. In view of the vast frictions, chaos, and costs, to individuals and to society, which could result if these expensive programs were launched simultaneously with no regard to their manpower implications, a modest investment in research on projections will clearly more than pay for itself.

Kinds of Projections

Different kinds of projections are required to meet the different needs outlined above. In planning educational programs, we need to know how many people should be provided with each kind of training. In this context we should assume that the demand is independently determined, reflecting society's needs. The unknown for which this equation is being solved is the inflow to the supply from training programs. It is sometimes useful in such studies to illustrate the implications for our ability to meet future demand of a continuation of the present trend in the number receiving training; this would answer the question, What would happen if we did nothing to change our educational effort? Such a computation would be useful in showing how much of a change in enrollments in each field of study would have to be accomplished in order to meet the projected demand.

In another context, it may be the demand that is projected illustratively and the supply that is projected independently. For example, when the government is considering a large new program or a great expansion in an existing one, the question arises, Will we have the necessary manpower supply without taking any special steps to assure it? In this case the supply should be projected on the basis of current trends, and the total demand projected on the basis of the addition of demand resulting from the proposed program to the demand already existing in the economy projected in the absence of such a program.

One must recognize that in reality demand and supply are not independent, but rather interact on each other; that a short supply of workers in an occupation forces industry to adapt work processes and manning arrangements, and thus affects the demand for labor; and that, conversely, demand creates supply by drawing in workers from other occupations or from outside of the labor force who can qualify in the shortage occupation.

For vocational guidance, a student needs a realistic appraisal of the extent of employment opportunity. Both supply and demand should, therefore, be projected on the basis of the most probably course of events in the judgment of the analyst. This type of projection is most nearly like a "forecast."

For all of these purposes, projections are needed for years in advance, usually at least five, and often 10 to 20 years. For educational planning they are also needed at a State or local level, although this need is considerably less with respect to the high-level occupations considered in this paper, since they seek jobs in a nationwide market.

The same general body of techniques and data are required for all these types of projections. They are, therefore, best made by a staff which is working continuously in the field and which develops the necessary expertise. At the same time it is essential that whoever makes projections keeps clearly in mind the questions each projection is designed to answer and the kind of projection that is needed.

Methods of Projecting Demand

The following discussion will be based on the experience of the Bureau of Labor Statistics in this area of work over the past two decades, and on studies by other government agencies and international organizations. There has been a great interest in this subject in recent years, arising in part out of the greater investments in education being made in all countries. Work has gone on in many different countries, and the OECD, UNESCO, and ILO have attempted to develop and encourage improvements in techniques through international conferences, technical assistance, and publications. 1/, 2/, 3/*

If one may define a general analytical approach to projections of manpower demand, it follows a logic something like this:

1. Ascertain the factors affecting demand for the occupation, and study how they have operated. Identify, if possible, any economic variables that measure or serve as a proxy for each factor and that are capable of being projected independently, and study their past relationship to demand for the occupation.

2. Project the factors or the variables, and their relationship to demand for the occupation.

3. Project demand for the occupation on the basis of these relationships.

This procedure has been followed over the years for many occupations. In most cases it is found that the factors affecting demand for the occupation are the level of demand for the goods and services produced by the occupation (or the institutions in which the occupation is employed), and the way in which the work of the occupation is organized and combined with that of other occupations to produce the goods or services. This leads us to project (a) the demand for the products of each sector of the economy in which the occupation is employed; (b) total employment in each sector (on the basis of a study of the relationship of output to employment); and (c) the changing patterns of use of the various occupations by each industry sector (affected by technological change and reflected in the changing occupational composition of that sector's work force). 4/, 5/, 6/ In order to project the demand for the products of each sector, we need a system of general projections of the patterns of economic growth and of the relationships between levels of income and general economic activity and patterns of expenditure of each type of goods and services.

If I have seemed to be leading the college president, who simply wants to know whether he should expand his engineering school, down the garden path of an elaborate system of economic analysis that takes in the Nation's whole economy in all its complexity, I make no apologies. No occupation can be considered in and of itself; its demand arises out of the complex fabric of our society, out of the growth or decline of industries, the dramatic changes in technology, the availability of workers in other occupations related in the work process. Its supply, too, is not independent, but depends upon the total supply of educated workers, and the numbers entering other occupations.

Having described above the logical steps leading us to making broad-scope economic analyses as a basis for projections, let me briefly outline the analytical steps currently followed in the work of the Bureau of Labor Statistics, and in that of many other agencies engaged in similar projections

1. The general level of economic activity is projected.--This way involves a projection of the population, a projection of the labor force, a computation of the number of persons who would be employed if this labor force is fully utilized (allowing for some minimum, inescapable level of frictional unemployment), and a projection of the gross national product that would be turned out by this number of employed workers, given an assumed growth in output per man-hour and an assumed change in the number of hours worked by each person.

2. The general character of the economy is projected.--This involves establishing a reasonable relationship between such strategic variables as investment, consumption, government expenditure, net foreign trade, net inventory change, and income payments to the various factors of production.

3. Patterns of consumption are projected.--This involves development of information on the expenditure patterns of families of different income levels, and projections of the numbers of such families.

---

* References are at the end of the paper.

4. Industry production levels are projected.--The production generated in each industry by final consumption of each of certain quantities of each product or service is estimated by means of regression analysis or input-output analysis.

5. Employment levels in each industry are projected.--This requires a projection of the change in output per man-hour and of the change in annual hours of work per employee.

6. Employment by occupation in each industry is projected.--This requires development of information on the past and present occupational composition pattern of each industry and an appraisal of how this will change under the impact of changing technology. The way in which the occupational composition of each industry is changing cannot always be projected by means of a study of the industry itself. Patterns of occupational use, particularly in relation to new technology, are introduced from one industry to another, and one can learn much from the experience of other industries. Thus, the managerial occupations and the way they are used in industry have certain elements common in all industries. Similarly, the introduction of a new technology, such as numerical control of machine tools, spreads from industry to industry, and these phenomena have to be examined independently of the study of changing occupational patterns of individual industries. Also, the relative employment of members of different occupations whose work is related (such as engineers and technicians, or nurses and hospital attendants) is affected by considerations of supply of the other occupations involved, and these factors have to be considered across the board and outside of the data we can accumulate on the occupational composition of each industry.

There are cases in which a projection for an occupation can be made with considerably less than the elaborate analysis described above. Occupations directly serving people, for example, can be projected on the basis of the growth and changing characteristics of the population. Thus, demand for teachers has been projected on the basis of projections of population of school age, trends in the proportion of this population enrolled in school, and trends in the ratios of teachers to pupils, which have to be analyzed in terms of the many institutional factors affecting these ratios. 7/ In the same way the typical method followed in projecting demand for medical personnel has been to project past ratios of personnel to population, allowing for the effects of such factors as changing age composition and urbanization. But for the very large number of occupations whose employment is in industry, particularly those widely scattered in different industries, such as engineers, chemists, accountants, administrative workers, etc., a more elaborate economic analysis is needed.

In addition to the projections that can be made of population-serving occupations, there are other individual projection approaches that can be made independently of the general system described above. For example, the demand for automobile mechanics has been projected on the basis of the potential number of cars and trucks in use, which is related to the number of families and the average number of cars per family. Employment of scientists and engineers has been projected on the basis of expenditures on research and development. 12/ However, these projections are not completely independent of a general economic analysis for many of the ingredients that enter them derive from considerations of income levels, corporation expenditure levels, etc. As a general principle, any means of alternative independent projection is worth pursuing as a check against the systematic projections and to get some sense of the range of error in the latter. Many cross-checks should be built into the system itself; e.g., employment or production in the building materials manufacturing industries ought to have some reasonable relationship with employment in construction or the volume of construction put in place.

Recent developments in data collection and analytical techniques make possible some improvements in the projection of demand. An extensive research project on economic growth, in which the Bureau of Labor Statistics, the Office of Business Economics of the Department of Commerce, and several other government agencies and university groups are engaged, applies analyses of consumer expenditures, investment expenditures, and input-output relationships among industries in a systematic approach to economic projections. 8/ New research by BLS and by some university groups on the interdependence of the size of the labor force and the level of employment opportunity has led to improved methods in this area. 9/ A major remaining gap is the lack of statistics on employment by occupation. An outline for a system to provide this is being developed by BLS; this would provide data for successive years on the occupational composition of individual industries, and aid in the projection of changing patterns.

New data are also becoming available that will help in the projection of demand in certain occupations. This may be illustrated by a single example. The traditional approach to projections of demand for medical personnel outlined above has some obvious weaknesses, but has been followed for many years, lacking a better one. Ratios of physicians to population need not be constant. On the one hand, rising standards of medical care and rising ability to purchase it (as a result of per capita income growth, prepaid medical plans, etc.) make for an increase in the ratio of physicians to population. On the other hand, changing medical practices and improvements in method may reduce reduce the need for physicians in relation to population. An example of the latter is the rapid cure of many diseases by antibiotics, reducing the number of doctors' visits for each illness. In view of these considerations, when one makes projections of medical manpower requirements on the basis of ratios to population he is on exceedingly marshy ground. We may get some new light on this problem from the experience of prepaid full-service medical plans, such as that of the various Group Health Associations which have developed through experience the necessary ratios of doctors required to serve their populations.

These new sources of data and techniques of analysis will enhance the accuracy of projections of demand.

It should be clear from the above description that the techniques for projection of demand are analytical in their orientation: they depend on the understanding of causes of economic change, not on any mechanical projection of past trends in employment in any occupation.

In view of all the difficulties I have described, I should not leave the subject of projections of demand without the reassuring note that it is indeed possible; that it has been done as part of a continuing research program and these projections are reviewed and revised repeatedly. The most recent general projection was published early this year in the Manpower Report of the President, and spelled out in more detail in a later publication. 11/

In this study requirements for professional, technical, and kindred workers, of whom over 8.5 million were employed in 1964, were projected to rise by more than two-fifths by 1975, almost twice as fast as the one-quarter increase projected for total employment. Demand in the technical fields was found to be rising most rapidly: needs for scientists and engineers (of whom 1.3 million were employed in 1963) were expected to rise by 50 percent by 1975; for science and engineering technicians (825,000 employed in 1963) demand was projected to rise by two-thirds by 1975. The same rate of increase was projected for college teachers; but demand for elementary and secondary school teachers and most health professions was expected to rise more slowly.

The growth in demand for managerial personnel was projected at about one-fourth from 1964 to 1975, with more rapid increase in needs for salaried managers and officials; slower growth for self-employed businessmen.

Methods of Projections of Supply

Although not all the problems of projected demand have been adequately solved, the art of projection demand is far more developed than that of projecting supply. The latter is the great neglected field in the area of manpower projections. This arises in part from the preoccupation of economics with the various kinds of studies and the development of the kinds of data that lend themselves to demand analysis. It also arises because of some of the severe conceptual and measurement problems in the area of labor supply. One group of these problems centers around the inherent flexibility and adaptability of human beings. People have multiple skills, and are capable of working in many occupations other than those for which they have had specific training or in which they have been previously employed. Most people actually do move, during the course of their working lives, among a number of occupations, and this mobility is found not only among the less-skilled but even among the most-skilled and specialized occupations. 10/ From this it follows that the labor supply in any occupation is difficult to quantify. If the analysis of the supply of labor is viewed as a system of stocks and flows, not only is the stock impossible to measure accurately, but also the flows into and out of each occupation are difficult to predict.

For high-level personnel, the major inflows are persons who complete training (e.g., receive a degree), those who enter without having received the formal training (a significant proportion of persons employed as engineers, for example, have never received an engineering degree), or come into the country as immigrants. The major outflows are deaths, retirements, withdrawals of women from the labor force for family reasons, movement to other occupations, and emigrants from the country.

For the inflows we have good measures of the number of persons receiving degrees at various levels. On the other hand, there is very little information on interoccupational mobility--the number of people who enter each occupation without having received formal training. Such studies in broad terms have shown that the quantity of such mobility, even among professional occupations, is considerable. 13/, 14/

For the outflows we have reasonably accurate measurements of the impact of deaths and retirements by using tables of working life for men and women developed by the Bureau of Labor Statistics, which can be applied to the present members of each occupation separately by age to make estimates of the prospective losses resulting from deaths and retirements over the next 10 or 20 years. There are greater difficulties in developing techniques for estimating losses from occupations resulting from occupational mobility. A series of two-year followup surveys of persons reported in certain occupations in the 1960 population census has provided some insights, but misclassification of these persons by occupation in the census makes the results difficult to interpret. 16/

Followup studies of people who receive degrees in each field have found a substantial movement out of many occupations, even in the period immediately after graduation. 15/

One of the hopeful new developments in this area of measuring occupational mobility is the National Education Association's annual surveys of public education systems, in which they get data on the number of teachers who leave the system (classified by whether they went to another teaching job, retired, or withdrew from the labor force, etc.), and the number of teachers who entered the system (classified by whether they were new graduates, reentrants, or persons who moved to other teaching jobs). These surveys help to sort out the great amount of mobility from place to place within the teaching profession from the inflows and outflows from the profession, and also help to deal with the question so important in occupations in which large numbers of women are employed: What is the pattern of inflows and outflows associated with women's family responsibilities?

The whole subject of projections of supply may be summarized by saying that, with the exception of deaths and retirements, our methods of projection are exceptionally weak and in need of considerable research. A good deal more information is available on the supply in some occupations than in others; the scientific and technical, medical and teaching fields are among the better-documented occupations. 17/

A general final comment might be made about all the elements of manpower requirements and supply projections: these cannot be done effectively on an ad hoc basis; they require instead the development of a staff responsible for continuing work in this area. This is true not only because the techniques are complex, but also because a well-thought-out-program of research to fill major gaps in our knowledge and techniques is needed, and also because the very nature of projections requires that they be reviewed at frequent intervals in the light of changing developments, and that the necessary lessons be learned from past mistakes and applied to future work.

References

1. "Employment Forecasting." Final report on the International Seminar on Employment Forecasting Techniques, Organization for Economic Cooperation and Development, Brussels, 4th-7th June, 1962.

2. "Forecasting Manpower Needs for the Age of Science," Office for Scientific and Technical Personnel, Organization for European Economic Cooperation, Paris, September 1960.

3. "Mediterranean Regional Project: Forecasting Educational Needs for Economic and Social Development," OECD, Paris, October 1962.

4. U. S. Department of Labor, Bureau of Labor Statistics, "Changing Manpower Requirements in Banking," by Rose Wiener, Monthly Labor Review, September 1962.

5. _______ "Employment Requirements and Changing Occupational Structure in Civil Aviation." BLS Bulletin No. 1367, June 1964.

6. _______ " Employment Outlook and Changing Occupational Structure in Electronics Manufacturing." BLS Bulletin No. 1363, October 1963.

7. _______ "A New Look at Manpower Needs in Teaching," by Maxine G. Stewart, Occupational Outlook Quarterly, May 1964.

8. _______ "The Federal Government's Program of Economic Growth Studies," by Jack Alterman, Monthly Labor Review, August 1965.

9. _______ "Labor Force Projections for 1970-80," by Sophia Cooper and Denis F. Johnston, Monthly Labor Review, February 1965.

10. U. S. Department of Labor, Bureau of Labor Statistics, "Occupational Mobility of Scientists," BLS Bulletin No. 1121, February 1953.

11. _______ "Manpower Needs in 1975," by Howard Stambler, Monthly Labor Review, April 1965.

12. National Science Foundation, "Scientists, Engineers and Technicians in the 1960's--Requirements and Supply," NSF 63-34, 1964.

13. "Labor Mobility in Six Cities," by Gladys Palmer. Social Science Research Council, New York, 1954.

14. U. S. Department of Labor, Manpower Administration, Office of Manpower, Automation and Training, "Formal Occupational Training of Adult Workers," Monograph No. 2, December 1964.

15. "Two Years After the College Degree," by Laure M. Sharp, Bureau of Social Science Research, Inc. National Science Foundation Bulletin 63-26, 1963.

16. "Education and Training of Technicians," by Howard V. Stambler and Annie Lefkowitz, Monthly Labor Review, November 1964.

17. National Science Foundation, "Scientific and Technical Manpower Resources," NSF 64-28, 1965.

DISCUSSION

Abbott L. Ferriss, National Science Foundation

1. Dr. Cartter first establishes that the percentage of doctorates among teaching faculty has increased during the past decade. The evidence from biennial NEA data, as Dr. Cartter says, is less than convincing. But he also has accumulated evidence from the American Council on Education quadriennial publication, American Universities and Colleges, which shows an increase of 7 percentage points AY 1951 to AY 1963 in 781 accredited institutions for all instructional staff. This evidence inadequately represents the universe of institutions. The 781 institutions in the volume for academic year 1963 were only 53% (781/1477) of degree-granting institutions in that year, and only 37% (781/2100) of all institutions beyond the high school.

To evaluate the quality of faculty in higher education, one must assess the quality of the full-time equivalent instructional staff for degree credit courses in all or as much of education beyond secondary school as can be assembled. This would include two-year as well as four-year institutions, the full-time equivalent of part-time instructional staff, the professional schools as well as the regular colleges, and the junior instructional staff as well as others. The COLFACS estimate of 50.6% doctorate among full-time employed teaching faculty in degree-granting institutions reflects only 73% of the full-time equivalent faculty in higher education. The ACE 1964 publication The American Junior College yields an estimate of 9.3% doctorates among teachers in junior colleges, which may be compared with 10% which Berelson estimated in his assessment of faculty supply and demand several years ago. The junior college must be considered, for its share of enrollments have increased from 11% to 18% during the past ten years.

The full-time equivalent of part-time, also must be considered, but for this segment no estimate of doctorate-holding is available. To estimate doctorate-holding among these categories of teachers we may use .5 (from COLFACS) for the full-time teaching staff of degree-granting institutions, .09 for junior college full-time teachers, and .3 for the full-time equivalent staff of part-time teachers, for AY 1963; the result is 41.2% with doctorates of FTE teaching staff. Using .4 for the part-time yields 42.5%, and using .5 for the part-time yields 43.8%. These values are 9 to 11 percentage points above Cartter's AY 1951 estimate (32.2%). If we assume the latter to be an upper limit, which I would take it to be, the twelve-year period has indeed witnessed an improvement in the quality of the teaching cadre in higher education.

A better measure of quality of instruction is ratio of students to teachers with the doctorate. Taking a few liberties with Ray Maul's data for AY 1955 provides an estimate of 39.7 full-time students per doctorate-teacher. The comparable AY 1963 estimate (from COLFAC and the USOE Faculty and Other Professional Staff Survey) is 36.0 students. By this index the quality of instruction has improved about 10% (1 - 36/39.7 = 9.3%).

In his conclusions, Dr. Cartter points to the inconsistency between past estimates of shortages of doctorates in higher education and the situation 10 years later of an increase in the percent doctorates. Without cries of alarm, help would not have come. Help did come, as Orlans has shown, and the educational establishment today is better because of it. One important function of a prediction is to make possible an evaluation of a future situation.

2. The second important feature of Dr. Cartter's paper is the presentation of a model for the prediction of future teacher requirements for faculty and for doctorate teachers. He introduces a few elaborations which previous model-builders have overlooked, but in the main the variables are the same as others. His overview points to the additional statistics needed to adequately activate a reliable model.

The Bolt-Koltun-Levine model (Science, May 14, 1965) for evaluating the consequences of various levels of feedback of doctorates into higher education, was a distinct improvement over previous models because it applied to particular disciplinary fields. Dr. Cartter's model adds no new variables and does not attempt to control for field of study. Separate consideration of fields is important, since the market exogenous to the educational establishment most certainly is not uniform among fields, as David Brown has pointed out. To use this approach would require much additional processing of available information, but such is needed for sound educational planning and the formation of national policy.

In addition to separate consideration of homogeneous groups of fields, independent consideration of types of institution will make possible much more careful control of another important source of variance in estimates of future teacher requirements: the student-teacher ratio.

A third important element would consist of classifying enrollment both by institutional type and full or part-time status.

For those educators who may enter a catatonic state when anyone suggests that they supply new data, let me hasten to add that the above requires no new data. It only requires the appropriate ordering of data already collected.

I now want to review the values Dr. Cartter assumes for elements in his prediction equations, examining each factor separately.

m - mortality rate of present teachers. This is of minor importance relative to some other values. The most recent mortality rates by occupation are based upon 1950 data. This is not complementary to a statistical system which prides itself upon the advanced state of its technology. Not only can death registration data coupled with the Census be employed for this purpose, but matching of death records against the Doctorate Record File and/or the National Register File could produce more accurate mortality rates by field than now are available.

r - retirement rate. Also of minor importance, this factor, nevertheless, should be estimated more accurately. Cartter used the age distribution from the COLFACS survey, and assumed a schedule of retirement 0, 1, and 2 years after 65. This is a refinement over previous procedures. In applauding him, I also point out that an accumulation of information on retirement experience, perhaps through a more extensive retirement survey, through T.I.A.A., or through other means, is needed.

a, c - The rate of transfer out of higher education of doctorates and the in-transfer rate of doctorates to higher education from other employment. /Note: in Eq. (2) a is in-transfer, but below Eq. (3) a is out-transfer. The latter apparently is in error./ The net (loss or gain) is the significant statistic. Dr. Cartter estimates that .0011 of the doctorates are lost annually by transfer out to other employment. In the Bolt-Koltun-Levine model the net transfer rate is estimated for scientists, from 1960-1962 matched cases in the National Register of Scientific and Technical Personnel, at -.001. The negative sign is quite significant, for it connotes a net gain of doctorates to the educational establishment in the interchange, rather than a net loss. Cartter's assumption (re Eq. 3) that this statistic is stable is quite questionable. Doctorates employed outside of higher education are evidently responding to salary increases in higher education. For the 1960 cohort of scientists, the experience between 1962 and 1964 revealed a net gain to the educational establishment of 2.89% per year, rather than 0.1% per year. One defect in this statistic is that it includes transfers to and from educational institutions, irrespective of level and function of the doctorates. Dr. Cartter's estimate -- the basis for it is not adequately documented -- could be in the proper direction for the decade over which he makes it, but the National Register data on scientists, who, with the professions, undoubtedly have the highest employability outside the educational establishment, for the 1960-64 period, unmistakably record a net inflow to the educational establishment. I conclude that more adequate data is needed on in- and out-transfer of doctorates. If doctorate holders are sensitive to salary changes, annual or biennial data are needed to make an accurate assessment of the flow, and this net interchange in Dr. Cartter's model should be more precisely scheduled over the future decades, perhaps upon the basis of assumptions on comparative salary levels.

b and q - Equation 2 estimates the number of doctorates in teaching. /In the first member of the right side of the equation, the subscript of D is t. Evidently, from the text above, this should be (t-1)./ Omitted is the percent of new doctorates who already are teaching. Dr. Cartter may have intended that b include new doctorates continuing as well as those newly entering teaching, as the value of b (Equation 3) would indicate. However, allowance also should be made for new doctorates continuing in higher education in q, the percent of new teachers with the doctorate, in equation 3. The percent of non-doctorate teachers at (t-1) who achieve the doctorate by t should be introduced as a third member of the equation. Data of the Doctorate Record File, NEA, and USOE provide a basis for estimating that this may be 2% or 3% of the full-time equivalent instructional staff in higher education. Four to five thousand teachers may be so "upgraded" annually.

f - The ratio of the increment of faculty to the increment of students (the inverse of the incremental student-faculty ratio) is set at slightly less than 20 to 1. The greatest source of error in projections of teacher requirements, as may be shown by ex post facto examination of Ray Maul's 1959 projections is the student-faculty ratio. To hold it constant, as Dr. Cartter does, is to deny the trend during the past decade* as well as his own arguments (1st criticism of the "OE model"). A more advisable approach is to assume a continuation of present trends, an increase of approximately 0.25 annually in the student-teacher ratio, or, to provide schedules of alternative assumptions. (See item 4, below.)

3. Several factors affecting higher education, the quality of teaching, and future teacher supply and demand, are not elements of Dr. Cartter's formula. To enumerate them briefly:

1. The education of Americans abroad, estimated at 17,200 persons in AY 1964.

2. The provision of Americans as faculty and scholars to foreign institutions, estimated at 3,400 persons in AY 1964.

3. Foreign scholars in the U. S., a supply source, consisted of 8,400 in AY 1964 (Open Doors, 1964).

* Total enrollment per FTE instructional staff has increased from 16.6 in Fall 1958 to 18.3 in Fall 1965 (estimated).

4. The education of foreign students in the U. S., estimated at 36,000 undergraduate and 39,000 other students in 1964.

The latter educational endeavor is quite significant in promoting the diffusion of science and technology, and in developing the scientific manpower for resource and institutional cultivation among underdeveloped nations. Within this context the demand for teachers for the educational systems of our own and foreign countries must be viewed.

4. The student-teacher ratio in the model considers the total staff, instructor or above, and does not make allowance for part-time teachers nor include junior staff. The text says that this ratio, 15.3:1 today, may be expected to rise to 17.3:1 in 1985. It says, "The Office of Education choice of an 18:1 ratio, therefore, appears to overstate the expansion needs by nearly 10%." Cartter's ratio of 17.3:1 would set our sights at approximately 80,000 more instructional staff in 1985 than would result from using the USOE factor [(18.0-17.3) 10,600,000]. I believe he should have said that the USOE estimate understates rather than overstates expansion needs. Perhaps Dr. Cartter's criticism of the USOE choice of a student-teacher ratio should be reconsidered.

My preference is to use full-time equivalent instructional staff, as did Ray Maul. This ratio was 16.6:1 in 1958 and has risen to 18.3:1 today. In the interests of conservatism and accepting Cartter's four reasons for expecting an increase in the number of students per teacher, I am inclined to project the ratio at a 0.25 incremental addition annually for about 10 years. Such a procedure would provide an increasingly conservative statement of teacher requirements.

5. An implication Dr. Cartter derives from the results is to question the wisdom of expanding the higher education system through new institutions entering the doctorate-granting field. Capacity of the higher educational system to produce doctorates is not a component of the model, although he estimates that 20,000 doctorates can be produced annually. I do not interpret the model as suggesting that the demand for new doctorates in teaching will decline or stabilize after 1968. There will be a continuing need to increase the percentage of teachers with the doctorate, and a continuing demand for doctorates in non-teaching positions.

6. The "deficit" of 120,000 which Dr. Cartter attributes to a USOE estimate is actually not an official USOE document, but rather a working memorandum, which cannot properly be attributed to the Office. To call this the current USOE model is not accurate.

7. Dr. Cartter considers his projections of doctor's degrees to be below those of Dr. Karel, who prepared projections published by the National Science Foundation. Karel's projections are confusing since they mix professional medical degrees with doctorates, but if the medical professional degrees are removed, Cartter's projections for AY 1970 are only 600 more than Karel's.

DISCUSSION

Laure M. Sharp, Bureau of Social Science Research

For purposes of this discussion, I would like to distinguish between the two topics which are the subject of Dr. Spaeth's paper. The first topic is change in choice of career fields between the freshman and senior year in college. The second deals with change in choice of career field between the senior year and the third year following graduation. To summarize my conclusions: the ingenious technique developed by the author seems to be much more productive in connection with the latter time period than with the earlier one. For the former, it strikes me as a statistical exercise which adds relatively little to the existing body of knowledge in this area; for the latter, I think it is a useful tool which --with some further refinements or adaptations--can make a valuable contribution in an area where new and better data are sorely needed.

First of all, there is a slight technical bottleneck in connection with the use of the method for change patterns between the freshman and senior years in college. As Dr. Spaeth indicated in his presentation, data for both years were actually collected during the senior year, in 1961. Furthermore the wording of the retrospective question differs slightly from the question dealing with current choice. The recall item reads as follows:

> Career preference when you started college. Give your single strongest preference even if it was vague or if there were several alternatives.

Therefore, one may question the use of a method using independence values and a time 1 versus time 2 comparison when the 2 items were asked at time 2 and in different ways.

But more important than the technical question is the substantive one. In terms of new findings, the approach yields little that was not previously known, and has less explanatory power than earlier analyses using traditional statistical techniques. There has been a great deal of research in the past 10-15 years in the area of motivations, behavior and values of college students, including occupational choice. It has been well established that clear, early occupational choices occur most often at the two extremes--among students oriented toward the traditional professions, medicine and to a lesser extent law (and in these fields a family tradition is often present) and among lower-class students whose sights are fixed on education or business, the upper limits of their aspiration spectrum. Except at these two extremes, freshmen occupational goals are vague for a great many students, perhaps the majority. It is precisely one of the basic functions of the 4-year college stay to provide for students a clarification of their own interests, abilities, and suitable career commitments. Depending on family background and the type of high-school attended, freshmen often have little knowledge of occupational alternatives (for example, underexposure to the social sciences is common) or come to college with an inappropriate evaluation of their competitive standing (for example engineering is often initially selected by students who do not have the necessary background and ability in mathematics and science). On this topic--of what happens during the college years--we are fortunate in having a voluminous and carefully researched literature which has been building up over the past two decades, partly through small-scale psychological and sociological studies done on many campuses with captive student populations, and more recently, through foundation or government-supported large-scale research efforts, such as those of the Cornell group, conducted by Rosenberg and his associates, the work of Ann Roe and her colleagues, at Harvard, the recent studies conducted by NORC and analyzed by Jim Davies, which Dr. Spaeth has mentioned, and many others. While there is of course always room for more and better data, and innovation in methodology, I do not see any "pay-off" in terms of new insights or a better model through the use of the methods presented by Dr. Spaeth.

Let me now turn to the second topic, change in career fields between the senior year and 3 years after graduation. Here we are indeed gaining much new and useful information from the data presented by Dr. Spaeth. Not only do these data consist of genuine time 2 versus time 1 responses, but the time 2 responses were given after reality-testing of academic preferences and tentative choices in the present opportunity structure of the occupational world. Concerning the dynamics of early career choices and changes in the period following college graduation we do not have the wealth of data available for "captive" college populations. We have neither the small insight-providing studies which would give us hypotheses to test with larger samples, nor the relevant basic statistical data which would provide the needed parameters. To the best of my knowledge, the only recent information about the transition from educational institutions to the labor market stems from the work done by Dr. Spaeth and his colleagues at NORC, and from the studies we are involved with at the Bureau of Social Science Research, where under sponsorship of the National Science Foundation we have so far conducted two elaborate nation wide follow-up surveys of the class of 1958. These types of studies are beginning to provide the parameters, although I feel that much of the depth needs to be filled in through more intensive studies in the future.

What Dr. Spaeth's findings point to most overwhelmingly is of course stability in choice of career fields, rather than change. He is the first to say so, since he has developed the particular technique he presented here to overcome the handicap to the analysis of change patterns caused by the stubbornness of the data. Obviously, he has a right to concentrate his attention on the small minority of cases which were indeed subject to change, but I can't help being more intrigued by the very fact that the great majority of his respondents displayed consistency and stability over the 3-year period with respect to career fields. This is all the more impressive because the wording of the question--which in effect uses academic fields and careers fields interchangeably--might at times be conducive to inconsistency (for example, men seeking administrative and managerial careers might easily have selected two different answer categories in 1961 and 1964 without having made an actual career change). This stability is shown not only by the large number of identical choices in the 2 years, but also by the clustering of practically all change patterns around the diagonale, which means that observed changes involved closely related fields. Whatever dispersion there is seems to be largely caused by very small numbers of actual cases, and in one particular case, seems to be hard to understand on the face of it, so that one suspects a possible error: twenty men who in their senior year planned a business career had switched to medicine 3 years later (a switch which would normally require substantial additional undergraduate preparation). Furthermore, as pointed out by Dr. Spaeth, the ordering of fields leaves some groups off the diagonale, although the "drastic" change which is thereby shown is artificial. And last but not least, the unavoidable necessity of making some arbitrary classification decisions may also produce some shifts which are more apparent than real. Thus, 88 respondents who chose an "education" career in their senior year and are shown as having switched to "physical science" careers may have merely shifted from high-school teaching to teaching at the jr. college or 4-year college level--a shift which under the NORC classification system removed them from education and into physical science. So, all in all and despite some deliberate--and justified--rigging to emphasize change, we see very little movement and very few drastic switches--and this is an extremely important finding which Dr. Spaeth's method illustrates elegantly. If I may speculate for a moment, I think this represents an important and basic social trend which has been greatly accelerated since World War II. As you may notice, I do not happen to agree with Harold Goldstein and some of the members of this audience that our labor force even at the professional level is characterized by great occupational flexibility and that shifts from one field to another will continue to be common. I feel that we have moved much more decisively to early specialization, usually determined at the time the bachelor's degree is received and that subsequent field shifts will be the exception rather than the rule.

In order to gain a fuller understanding of the mechanisms involved here, two efforts are needed, both rather cumbersome but I think indispensable for a meaningful analysis of the current occupational shifts. In the first place, broad categories, like the ones used in the paper under discussion--while much more manageable for tabular presentation--are conceptually inadequate. Nowadays, a shift from one physical science to another--even from one subspecialty to another--is a significant departure, and probably the preferred mechanism on the one hand for adjusting the allocation of available professional manpower to the needs of the economy, and on the other for enabling college graduates to correct early career decisions. Data from our own studies dealing with switches from undergraduate to graduate fields of study show considerable shifts from one social science to another, or from one engineering specialty to another, but relatively little switching across fields. If Dr. Spaeth had shown us the more complex matrices based on detailed, rather than broad fields, we would have had more evidence of change. Furthermore, such matrices would be extremely useful because they might provide some clues as to what linkages between fields are common and how much elasticity there is between related fields. Switches from electrical to nuclear engineering, from clinical to experimental psychology, from French to Swahili have meaningful manpower implications and could be assessed very efficiently through the method developed by Dr. Spaeth.

The second, related point which we need to keep in mind is the birth of new occupational fields, many of which do not fit neatly within the established standard categories into which we customarily organize our data. Here again, we may both understate or overstate occupational shifts, depending on rather arbitrary classification decisions. A political science student who becomes a systems analyst may or may not continue to work in the discipline in which he was trained. An engineer who becomes a contract negotiator for complex research programs may be classified as a business executive although his engineering training continues to be of paramount importance in his work. This is a cumbersome and tedious problem which all researchers handling educational and occupational data have struggled with in the past and will continue to struggle with in the future. However, classification decisions turn out to influence very significantly the judgments we make about trends in the relation between education and occupation, and about the extent to which the present generation of college graduates engage in highly specialized or broadly convertible careers.

V

MATCHING OF CENSUS AND VITAL RECORDS IN SOCIAL AND HEALTH RESEARCH: PROBLEMS AND RESULTS

Chairman, MONROE G. SIRKEN, National Center for Health Statistics

THE METHODOLOGY OF COMPUTER LINKAGE OF HEALTH AND VITAL RECORDS

David M. Nitzberg, Harvard School of Public Health
Hyman Sardy, Brooklyn College

Introduction

Record linkage is a process whereby records pertaining to the same individual from two or more files are brought together to form a combined record. In order to link records successfully they must be matched on identifying information that is common to the files to be linked and that has (a) high discriminating power, (b) low probability of change during an individual's lifetime, and (c) low likelihood of being recorded erroneously. Such information, together with rules for matching the records, as well as criteria for deciding when a pair of records match sufficiently to be declared a linkage, determine a linkage procedure. Its success is measured by

(1) its speed and cost,
(2) the number of valid linkages produced,
(3) the number of false linkages produced, and
(4) the number of valid linkages missed.

There is, of course, no a priori knowledge of which individuals have records in more than one of the files. Furthermore, there is often only a limited amount of common identifying information available. Also complicating the problem is the presence of noise, that is to say errors, such as surname misspellings and age discrepancies in the information which is available.

This paper briefly discusses our research to develop efficient computer linkage techniques. We are using the IBM 7010 data processing computer at the National Center for Health Statistics to perform actual linkage operations between cohort and death records. However, since our work is still in progress we can only indicate some of the goals we have set for ourselves and present some preliminary results.

Background

Densen and Shapiro (1) have pointed out that the limitations inherent in using existing data (i.e., information collected routinely, not for specific research purposes) can often be overcome by combining the information contained in records from several sources. Moriyama (2) has documented the value of vital records in health research, especially when such records are linked with others.

Computer linkage research has already been carried out (and is continuing) in Canada, England, and the United States (3-11). This research has demonstrated that when there is an abundance of identifying information common to the files to be linked, large-scale computer linkage is possible -- even though noise exists and people do not have unique identity numbers. The research we are conducting, however, is geared to studying the nature of the problem when only limited amounts of identifying information are available. We hope this will lead to the development of computer techniques which will permit linkage prior to the time when widespread use of identity numbers will simplify the methodological problems existing today.

Death Clearance

We chose to do our research in the area of death clearance since it is frequently performed in health research and is methodologically typical of most linkage operations. Death clearance is a process whereby a file of records pertaining to a group of individuals, a cohort file, is linked with a file of death certificates to determine which individuals have died and to extract information from their death certificates. Death clearance by computer would facilitate long-term follow-up studies since large study cohorts could then be placed under automatic surveillance for mortality at frequent intervals, with high precision and at relatively low cost. A contract in 1963 with the New York City Department of Health* by the National Center for Health Statistics (NCHS)** to develop computer death clearance techniques has made the research possible.

The death file we are using is composed of magnetic tapes made from the death index cards routinely keypunched by the New York City Health Department from death certificates. All deaths occurring in the city, as well as deaths of city residents reported to New York City as occurring outside the city, are included. Our file covers the years 1961-1963 (inclusive); Appendix I shows the format of the 1963 cards.

On the tapes used in our computer runs we added the Social Security numbers

---

*Dr. Paul M. Densen, Deputy Commissioner of Health, New York City, is project director for this contract.
**We are grateful to Mr. S. Binder, Chief of the Data Processing Division of the NCHS for his help.

of the deceased when these were given on the actual death certificates (they are not now keypunched on the index cards) in order to study the value of this unique identity number. One of the several cohorts we are linking has this information recorded. Even though at present many of the deceased do not have Social Security numbers -- 59%* of recent New York City death certificates do not have this item recorded -- an ever increasing proportion of our population will. Furthermore, widespread use of these numbers is being fostered by requirements of the Internal Revenue Service. All of this indicates that even at present the Social Security number is an item of identifying information which may be worth recording and using in linkage operations.

We will limit our discussion here to what we are doing with the largest cohort we are linking, the coronary heart disease (CHD) population of the Health Insurance Plan of Greater New York, better known as HIP. HIP has placed about 120,000 people under observation for specific manifestations of coronary heart disease in order to study its incidence and prognosis. Their study cohort includes all persons 25-64 years of age enrolled in 12 of their medical groups. The prepaid group medical practice setup of HIP with its central record system makes such an undertaking possible (13). Nevertheless, mortality for so large a group over a number of years remains a problem since it is not known what percentage of its members' deaths are reported to them. Obviously, death clearance by manual procedures for such a large group for a number of years is out of the question. With HIP's 1961 enrollment cards (Appendix II shows the format of these records) as our cohort file, we are performing a death clearance operation using computer techniques. This is being done for all members (176,481), not just those between 25 and 64 years old, of the medical groups in the HIP-CHD study for 1961, 1962, and 1963.

Methodology

At present, we have in operational form IBM 7010 programs** with which to explore death clearance techniques. Names are coded phonetically by a Soundex program according to the Russell-Soundex system developed by the Library Bureau of Remington Rand and based on the work, early in this century, of Mr. R. C. Russell (14).

A Matching program brings together records from the HIP and death files having the same Soundex code. The program then compares each of these HIP-death pairs to see whether there is agreement on items of identification which are common to both. For numerical information, it is possible to specify matching rules which permit slight discrepancies (e.g., requiring that $|\text{HIP age - Death age}| \leq k$ for some value of k)* to be tolerated when deciding whether agreement exists between the fields. For alphabetic information, however, exact agreement is necessary, except in the case of surname prefixes such as Von, Mc, D', etc., which are often keypunched in various ways with respect to the rest of the name. Our program treats as identical such variations as: MC COY and MCCOY, VON MEER and VONMEER, D'AMOS and DAMOS, etc. This class of alphabetic discrepancies became apparent after our initial runs, but were overcome easily by making simple changes to our original program.

Noise suppression techniques such as allowing $\pm$ k year differences when matching ages and ignoring blanks and apostrophes appearing within surnames are straightforward. Other sources of noise, as, for example, name misspellings and address spelling variations, are not so easily suppressed. The value of the Soundex name code is that it enables one to bring together records containing similar names on the assumption that many misspellings lead to similar sounding variations of the true spelling. The problem remains, however, of defining how "dissimilar" names can be and still be declared matching -- how can we overcome the loss of potential linkages because of misspellings without thereby creating many false linkages? The results we are beginning to accumulate indicate that, although Soundex is a gross noise filter, e.g., such different and commonly occurring surnames as Jones and James have the same code of J520, it does help overcome many name misspellings. As a result of our experience, we hope to be able to estimate just how many are overcome validly in death clearance operations. We also hope

*Estimated from a sample of 4196 death index cards, and agreeing with a sample we collected of 816 New York City death certificates, for 1961-1963.
**The programs were written by Dr. H. S. Levine (of HIP) and Mr. J. Hayden, especially for this project under the NCHS contract.

*We will get a distribution of age differences for the valid linkages so that an optimum k for death clearance can be estimated.

our results will enable us to modify Soundex so that it can be made a finer, and hence more efficient, noise filter for names.

The Matching program produces a listing of record pairs meeting and surpassing a set of minimum matching criteria specified at the beginning of the computer run. In the case of the HIP-CHD cohort we compared 176,481 people against a total of 281,208 death records in a four-hour computer run. This resulted in 89,306 pairs representing 37,777 different HIP members* meeting our minimum matching criteria, which were exact agreement on Soundex code of first and last names and age agreement within five years. Table 1 gives the number of record pairs matching on the fields specified by asterisks. It should be noted that the numbers given in the table are not cumulative -- that is, the absence of an asterisk means the pairs do not match on that field. Also, the heading "age ± 5 years" means that age agrees within five years but not within one year.

The numbers presented in the table are preliminary. They are based on the initial production run of the programs during which invalid characters due to keypunching errors, blanks and apostrophes in surnames, and several other minor difficulties prevented the matching of a small number of records. A supplementary computer run for these records is planned and our figures will be adjusted accordingly. The changes, however, will not alter the table significantly since the number of records involved is relatively small. Since it is possible to partition the HIP file and do death clearance on each part separately without affecting the accuracy of the operation (although efficiency might be degraded thereby), we have delayed doing this "mopping up" process until we could do it for all of our cohorts at once.

In all, there were 700,738 pairs matching on Soundex, of which 111,252 pairs matched exactly on first and last names as well but whose ages were not within at least five years of one another (and so were not put out by the Matching program). This means there were a total of 132,123 pairs matching exactly on first and last names. This clearly shows that name alone is a very poor identifier of people.

The 7,036 pairs matching exactly on name and age within one year are most interesting since they indicate very clearly the problem of trying to link with insufficient identifying information common to both files. Since only about 3,000 deaths are expected -- of which only about 15% will not be among these 7,036, mainly because of noise -- it is apparent that name and age alone lack the discriminating power upon which to decide which pairs constitute valid linkages.

Looking further at the figures we see that there were 26,075 pairs matching on Soundex of first and last names plus age within one year, and that 7,036 of these matched exactly on first and last names also. This indicates two things: first, that the Soundex code increased the number of possible pairs by a factor of 3.7 (the 26,075 pairs are reduced to 7,036 when exact full name matching is required); and second, that items of identifying information used in conjunction with one another very quickly narrowed down the field of possible pairs (e.g., from 176,481 x 281,208 possible pairs to 132,123 pairs by use of first and last names, and then to 7,036 pairs by use of age to within one year in addition). This seems to indicate that if there were no noise in our data, then one other good variable (e.g., the maiden name of an individual's mother) would probably reduce the 7,036 pairs to the correct 3,000 or so pairs with few false positives. Since noise exists however, schemes such as Soundex must be used to overcome it. This raises the number of possible pairs, which in turn might be reduced if yet another identifying variable were available.

---

*A given HIP or death record may appear in more than one pair. It is not uncommon for different people with the same common names, like Mary Smith for example, to appear in the HIP file and be within five years of age of several different deceased persons also with the same name. This gives rise to more pairs than the number of different people involved.

Table 1

Number of HIP-Death Pairs Matching
on the Fields and Criteria Specified

Exact Soundex Code of First and Last Names	Exact First Name	Exact Surname	Age ± 5 Years	Age ± 1 Year	Number of Pairs
*			*		14,426
*	*		*		30,818
*		*	*		4,152
*	*	*	*		13,835
*				*	5,615
*	*			*	11,772
*		*		*	1,652
*	*	*		*	7,036
					89,306

Conclusion

By clerically eliminating the pairs that do not seem to constitute valid linkages* and validating the remaining pairs by using information which could not be used by our present Matching program (e.g., address, since only the death file contained this field), we hope to find most of the deaths which did, in fact, occur.

Since this work is exploratory in nature, we wish to find as many of the deaths as possible even though we cannot do it by completely automated means now. HIP has been notified of a large proportion of the deaths of its members, so they will be able to tell us about a number of the deaths our computer-manual death clearance operation missed. Likewise, we will be able to give them a list of deaths about which they had no knowledge and which they will verify through their normal channels of contact with their members. Since the two procedures, HIP's and ours, for finding deaths are independent, we will then be able to estimate the number of deaths we both missed. If this number is small (less than 5% of the total deaths), as we expect it will be, then we will be able to analyze why our death clearance techniques did or did not link the deaths we know about, knowing that the unfound deaths cannot greatly bias these results.

This analysis, together with our other work, should lead to useful estimates of the parameters of death clearance procedures which, in turn, should lead to a better understanding of the methodology of record linkage and to computer techniques for automating the linkage operation. Whether this is actually so, remains to be seen. We feel confident, however, that our attempt to follow a cohort of 176,000 people for deaths for three years with the aid of computer techniques currently available will prove successful.

---

*This is being done by: (a) Human judgment to decide whether two names having the same Soundex code, but not agreeing exactly, were different names or variants of one another that could reasonably be expected to arise from misspelling or recording error; (b) the use of such other information as date of birth, spouse's name, address, etc., that appears on only one record of the pair and on HIP records other than enrollment cards or the death certificates themselves. Time does not permit fuller discussion of this procedure here.

REFERENCES

1. Densen, P.M. and Shapiro, S. Research needs for record matching. 1963 Proceedings of the Social Statistics Section of the American Statistical Association, pp. 20-24.

2. Moriyama, I.M. Uses of vital records for epidemiological research. J. Chron. Dis. 17:889-897, 1964.

3. Newcombe, H.B., Kennedy, J.M., Axford, S.J. and James, A.P. Automatic linkage of vital records. Science 130(3381):954-959, Oct. 16, 1959.

4. Newcombe, H.B. and Kennedy, J.M. Record linkage. Communications of the Association for Computing Machinery 5(11):563-566, Nov. 1962.

5. Newcombe, H.B. and Rhynas, P.O.W. Family linkage of population records. The use of vital and health statistics for genetic and radiation studies. Proceedings of the Seminar sponsored by the U.N. and the W.H.O. (held in Geneva, 5-9 Sept. 1960), U.N. Publication, Sales No. 61 XVII.8:135-154, 1962.

6. Kennedy, J.M. The use of a digital computer for record linkage. Ibid., pp. 155-159.

7. Binder, S. Present possibilities and future potentialities. Ibid., pp. 161-170.

8. Phillips, W., Bahn, A.K. and Miyasaki, M. Person-matching by electronic methods. Comm. ACM 5(7):404-407, July 1962.

9. Phillips, W. and Bahn, A.K. Experience with computer matching of names. 1963 Proc. Social Statistics Section of the A.S.A., pp. 26-38.

10. Acheson, E.D. Oxford record linkage study: A central file of morbidity and mortality records for a pilot population. Brit. J. Prev. Soc. Med. 18:8-13, 1964.

11. DuBois, N.S.D. A document linkage program for digital computers. Behavioral Science 10(3):312-319, July 1965.

12. Shapiro, S., Weinblatt, E., Frank, C., Sager, R. and Densen, P. The HIP study of incidence and prognosis of coronary heart disease: Methodology. J. Chron. Dis. 16:1281-1292, 1963.

13. Shapiro, S. Research in prepaid group practice programs. Amer. J. Public Health 54(12):2040-2050, Dec. 1964.

14. Waugh, W.C. Russell-Soundex (An address given by Mr. Waugh). Remington Rand, Inc., Library Bureau, SP-V3012 (undated).

APPENDIX I

Format of the 1963 New York City Death Index Cards

Columns	Field	
1-5	Death Certificate number	
6-9	Month, day, and year of death	
10	Borough where death occurred	
11	Type of institution where death occurred	
12-14	Institution number	
15	Borough of residence	If non-resident of N.Y.C., state and county of residence
16-19	Health area of residence	
20	District of residence	
21	Blank	
22	Sex	
23	Color or race	
24	Marital status	
25-27	Age at death (2 columns for age and one for units -- hours, day, months, years -- of age)	
28	Nativity of deceased (U.S., Puerto Rico foreign)	
29-32	Cause of death	
33	Operation?	
34	Attendant at autopsy	
35	Religious affiliation of cemetery where buried	
36-40	If death due to accident, type and line number of special accident report, and borough and district of occurrence of accident	
41-44	Medical examiner case number, if any	
45-63	Address of deceased	
64-80	Name of deceased (surname, first name)	

APPENDIX II

Format of the HIP Enrollment Cards

Columns	Field
1-19	Name (surname, first name)
20-22	Blank
23	R = Renewal
24-26	Original Entry Date (month, year) into HIP
27-28	Blank
29-36	Certificate Number (policy number)
37-38	Borough of Residence
39-42	Medical Group
43-46	Blank
47	Associated Hospital Service Class
48	Blank
49	HIP Class
50-51	Number of Persons Covered on Certificate (or Policy)
52-56	Contract Number
57-61	Effective Date of Contract (month, day, year)
62-63	Sex and Family Status
64	Blank
65-67	Month and Year of Birth
68-80	Blank

USE OF CENSUS MATCHING FOR STUDY OF PSYCHIATRIC ADMISSION RATES

Earl S. Pollack, National Institute of Mental Health

Studies or analyses designed to measure the rate of occurrence of a particular event in specific population groups are extremely common. In most of those concerned with illness or mortality, the numerators are obtained from interviews, vital records, or hospital or agency case records and are related to published population data. The assumptions implicit in such a procedure are : (1) that each individual counted in the numerator has been enumerated in the population and (2) that each individual is classified identically in both numerator and population denominator with respect to the characteristics under study.

An alternative procedure involves identifying the individuals to whom the event of interest has occurred and locating for each of these persons the Census document used for tabulating population data. If this procedure is successful in locating the Census records for all of the persons in the study, both of the above assumptions will be fulfilled. It is the purpose of this paper (1) to describe a study using this procedure, (2) to present data indicating the relative success of the Census matching procedure for various groups and (3) to discuss the implications of failure to find matching Census schedules for the analysis of rates.

The Psychiatric Admission Rate Study

In 1960, a study was begun to determine the rates at which persons come under psychiatric care in specific population groups defined primarily by a number of socioeconomic and family relationship variables. Data of this type were needed to help plan for the development of programs of psychiatric care and to provide a set of hypotheses for further study into the etiology of mental disorders. Further impetus to the formulation of the study was provided by the impending availability of a large volume of population data to be prepared from the 1960 Census.

Two states, Maryland and Louisiana, were selected as the locale for the study on the basis of their extensive programs for central reporting of data on persons coming under care in inpatient and outpatient psychiatric facilities and because of the interest in the study expressed by key persons in the agencies responsible for the mental health programs in these states. With the cooperation of the Louisiana State Department of Hospitals and the Maryland Departments of Mental Hygiene and of Health, the Office of Biometry of the National Institute of Mental Health collected basic identifying information on each person admitted to the public and private inpatient and outpatient psychiatric facilities in the two states during the year following the Census. This information was given to the Bureau of the Census, where 1960 Census schedules for these individuals were located, and detailed tabulations of the demographic, socioeconomic and family relationship characteristics of these patients were tabulated.

Matching Method

Since the major focus of this paper is on the results of the matching procedure, the procedure itself will be described only briefly. For each person admitted to a psychiatric facility in the two states during the study period, a transcription sheet was submitted to the Bureau of the Census containing the following information: name, sex, color, date of birth, psychiatric diagnosis, facility to which admitted, history of previous admissions, residence as of time of admission and as of April 1, 1960, and name of head of household on that date.

It should be emphasized that the matching was carried out by hand at the Bureau of the Census, not by computer. The Census schedules are filed by enumeration district (ED), a small geographic subdivision assigned to a single census enumerator and consisting of an average of about 250 housing units. Therefore, the success of the matching operation depended heavily on the accuracy of the address for each person admitted to a psychiatric facility. The transcription sheets were sent first to the geography unit where the appropriate ED number was assigned to each address. If an address was given in rather vague terms, it could have been assigned with equal justification to more than one ED. Therefore, on the transcription sheet, space was allotted for the assignment of a maximum of seven ED's for a given address.

The transcription sheets were next sent to a processing unit where an attempt was made to locate the Census schedule corresponding to each individual in the study. The schedules are filed in books according to ED. The search was carried out in two stages: (1) finding the page in an ED book which contained the same address as that given for the patient on the transcription sheet and (2) identifying the patient on that page. A set of rules was provided for each of these stages. On transcription sheets where several possible ED's were indicated, each was searched in turn until the address was found. If the address could not be found in any of the ED books indicated and if a search for the patient's name in each of those books also proved unsuccessful, the patient was considered a "non-match." Further procedures were involved for those individuals who were included in the census 25 percent sample. Since this is not important for this presentation, the discussion which follows will pertain primarily to the census 100 percent data.

Results of Matching

Transcription sheets containing basic information on 13,036 Louisiana patients and 14,450 Maryland patients were submitted to the Bureau of the Census for purposes of locating the corresponding 1960 Census schedules. Of these, matching schedules were found for 67 percent of the Louisiana patients and 64 percent of the Maryland patients. The fact that the matching procedures failed to locate census schedules for approximately one-third of the patients raises a serious question about the validity of admission rates based on matched cases only. An analysis of match rates according to specific characteristics and in relation to existing knowledge about completeness of census enumeration will help to place in perspective the effect of these match rates on the analysis of admission rates.

The extent of matching varied considerably from one category to another. Match rates according to specific characteristics are presented in Tables 1 through 7. The following are a few of the highlights:

1. Matching was most successful among those under 18 years of age; least successful among those 18 to 24; higher among males than females; higher among whites than non-whites.

2. Match rates were far lower for alcoholics than for any other diagnostic group.

3. For household heads and members of their immediate families, three-fourths of the matching schedules were found, whereas the match rates for other relatives were 63 percent in Louisiana and 57 percent in Maryland and for non-relatives only 40 percent and 49 percent, respectively.

4. Among married persons the match rates were 77 percent for Louisiana and 76 percent for Maryland.

These rates have been presented in some detail in the tables with the hope of illuminating reasons for failure to find matching Census schedules. Possible reasons for this failure are:

1. Inadequate or poorly defined addresses.

2. Differences between census and admission records in name and age.

3. Clerical errors.

4. Persons not enumerated in the census.

No specific studies have been conducted which would permit a classification of non-matched cases into these four categories. Comparison of the above results with those of other census matching studies reveals consistently lower match rates for the present study (1,2,3). The study most comparable to the present one was that conducted by the University of Chicago in cooperation with the National Vital Statistics Division in which deaths during the four-month period following the 1960 Census were matched against Census schedules. Preliminary analysis indicates that the overall match rate obtained in that study was approximately 80 percent (4) Since the Census matching for this study was carried out immediately prior to that for the present study using the same clerical staff and almost all of the same procedures, one would expect the quality of the search to be comparable between the two studies.

It is likely, therefore, that differences in match rates between the two studies are due to differences in quality of the addresses used as a basis for matching, differences in the extent to which those under study had been enumerated in the population, or some combination of these two factors.

Unfortunately, based on information presently available, there is no way of determining the relative influence of each of these two factors. Evaluation of the 1960 Census is still being carried out, but some preliminary data providing estimates of the net census undercount according to sex, color and age are presented in Table 8 (5). These undercounts, particularly among the whites, may seem too low to have any appreciable effect on the match rates. It should be pointed out, however, that these are estimated average counts over the total population, and furthermore, that they pertain to the population of the United States rather than to those of the two states under study.

As a result of the studies evaluating the 1950 Census, the rates of omission from the census by household composition were estimated as follows: Head or wife, 2.0 percent; child of head, 1.7 percent; other relative of head, 4.1 percent; unrelated individuals, 8.2 percent (6). The investigators who carried out these studies stated that a sizeable proportion of persons in enumerated households who were omitted from the Census may be persons with no regular place of residence.

What, then, can be said about the extent of Census under-enumeration among patients included in the present study? Based on the data presented above the following statements can be made: (1) In categories in which match rates tend to be high (Table 1) census undercounts tend to be low (Table 8). (2) Table 4 indicates high match rates for household heads and immediate members of their families, lower rates for other relatives and very low match rates for unrelated individuals. As indicated above the rate of omission from the census was lowest among household heads and members of their families; next lowest among other relatives; and highest among unrelated individuals. (3) It seems reasonable to expect that Census enumeration among alcoholics and persons living alone would be more difficult than for the population in general. If this is indeed true one would expect low match rates among these groups due to underenumeration alone.

Implications of Nonmatching and Underenumeration for Analysis of Admission Rates.

Suppose we wish to compare rates of admission to psychiatric facilities between two subgroups of the population, say, single and ever married. Ideally, we would divide the total number of single persons admitted to psychiatric facilities by the total number of single persons in the population and compare the result with a similar ratio for married persons. In this study two factors complicate this comparison: (1) the numerator of this ratio is incomplete due to the inability to find all of the Census schedules, and (2) the denominator is also understated because of underenumeration of the population in the census. To complicate this problem further, the extent to which Census schedules were found is unknown for many of the categories to be considered in the analysis and the proportion of underenumeration is unknown for every category of the population.

How, then, can we obtain a valid comparison of the admission rates when only incomplete counts of both numerator and denominator are available? How can we make use of the available data to approximate the results that we would obtain if complete counts of both numerator and denominator were available? These questions can be answered more readily if we consider first the following formulation of the problem.

Consider, again, a comparison of rates of admission to psychiatric facilities between two categories of the population, single and ever married.

Let Y_i = number of admissions in the i^{th} marital status category

(i = 1, 2 for single and married, respectively)

P_i = total population in the i^{th} category

y_i = number of matched admissions

p_i = enumerated population

m_i = proportion of admissions matched to Census schedules

e_i = proportion of the population in the i^{th} category which was enumerated in the census.

$\hat{Y}_i = y_i$ + an estimate of $(Y_i - y_i)$

$R_i = {}^{Y}i/P_i$

$r_i = {}^{y}i/p_i$

$r'_i = {}^{\hat{Y}}i/p_i$

Ideally, we would like to compute the "true" admission rate, ${}^{Y}i/P_i$, but complete data on neither the numerator nor the denominator are available. There are two alternatives open to us:

(1) compute ${}^{y}i/p_i$

(2) compute ${}^{\hat{Y}}i/p_i$ by adding an estimate of the number of nonmatched cases in the i^{th} category to the known number of matched cases.

But ${}^{y}i/p_i$ is a "good" estimate of ${}^{Y}i/P_i$ only if ${}^{m}i/e_i$ is close to unity, and ${}^{\hat{Y}}i/p_i$ will, on the average, overestimate ${}^{Y}i/P_i$, because the expected value of $\hat{Y}_i = Y_i$, but p_i is almost always less than P_i.

In comparing rates between two marital status categories, however, the problem is not necessarily that of obtaining "good" estimates of $R_i = {}^{Y}i/P_i$, but rather to obtain "good" estimates of $R_1 - R_2$, if we are interested in the excess risk of admission among single persons or ${}^{R}1/R_2$ if we are interested in the risk of admission among single persons relative to that among married.

For purposes of this presentation only the relative risk, ${}^{R}1/R_2$, will be considered. Using data on matched admissions and enumerated population, the relative risk can be written.

$$\frac{r_1}{r_2} = \frac{m_1Y_1}{e_1P_1} \cdot \frac{e_2P_2}{m_2Y_2} = \frac{Y_1}{P_1} \cdot \frac{P_2}{Y_2} \cdot \frac{m_1e_2}{m_2e_1} = \frac{R_1}{R_2} \cdot \frac{m_1e_2}{m_2e_1},$$

where ${}^{R}1/R_2$ is the "true" relative risk. If the ratio ${}^{m}i/e_i$ is relatively constant, i.e., if ${}^{m}1{}^{e}2/m_2e_1 = 1$, the observed relative risk is approximately equal to the "true" relative risk.

If we add to the numerator the estimated number of nonmatched admissions, the resulting relative risk is

$$\frac{r'_1}{r'_2} = \frac{\hat{Y}_1}{e_1P_1} \cdot \frac{e_2P_2}{\hat{Y}_2} = \frac{Y_1}{P_1} \cdot \frac{P_2}{Y_2} \cdot \frac{e_2}{e_1}$$

If $e_1 = e_2$, the observed relative risk is equal to the true relative risk.

This formulation suggests that to simplify the interpretation of ratios of admission rates, two conditions must be fulfilled:

(1) $\frac{m_1e_2}{m_2e_1} = 1$ when using rates of the form $\frac{y_i}{p_i}$

(2) $\frac{e_2}{e_1} = 1$ when using rates of the form $\frac{\hat{Y}_i}{p_i}$

When these conditions hold, ratios of rates provide consistent estimates of the "true" relative risks.

To obtain some idea of the extent to which these ratios might deviate from unity, examples of each of these two types of ratios were computed based on the data presented in Tables 1 and 8 and are shown in Table 9. It should be reiterated, however, that the data on census undercounts, presented in Table 8, are merely crude estimates for the entire United States.

It will be noted that the greatest deviation from unity for the ratio m_1e_2/m_2e_1, occurred in the age group 65 and over in each state and resulted from low match rates among non-white males. These were not offset by the corresponding enumeration rate which, far from being low, was estimated as a 7.9 percent overcount. Aside from those categories, two-thirds of the ratios differed from unity by less than 10 percent. Similarly, more than two-thirds of the ratios e_2/e_1 deviated from unity by less than 10 percent.

This provides some indication that if similar ratios could be obtained, based entirely on Louisiana and Maryland data, the assumptions made in assessing relative risks of admission to psychiatric facilities will be fulfilled with relatively small error. Except for the meager pieces of evidence provided thus far by the Census post-enumeration surveys, the e_i are unknown. Match rates, on the other hand, can be estimated for some variables, but are unknown for others. Therefore, since only part of the information required to make a choice between the two alternatives is available, both rates, y_i/p_i and $\hat{Y}_i/p_i$, will be computed for each category, where possible. If the results of a given comparison are consistent for the two sets of rates, they can be interpreted with greater confidence, perhaps, than results based on only one set of rates. If, on the other hand, conclusions differ between the two sets of rates, they will be viewed as inconclusive. In such an event, however, some interpretation will be made based on the knowledge available on match rates and enumeration rates for the categories involved.

The findings of this investigation could have rather far-reaching implications for studies in which numerators of rates are not obtained from matching Census schedules. Such numerators are essentially complete, but the corresponding denominators are incomplete due to underenumeration of the population. Marked variation in the extent of underenumeration among the categories being compared could result in substantial differences in rates due to differences in Census coverage alone. A careful comparison of the results obtained from the two sets of rates, y_i/p_i and $\hat{Y}_i/p_i$, may shed some light on this problem. In addition, more detailed data on the extent of Census coverage in segments of the population defined by a number of variables would be extremely helpful.

REFERENCES

1. Kaplan, D.L., Parkhurst, E. and Whelpton, P.K.: The comparability of reports on occupation from vital records and the 1950 Census. Vital Statistics - Special Reports, U.S. Dept. of H.E.W., Public Health Service, 53, June 1961, 1.

2. General Register Office: 1951 Census of England and Wales, General Report. London, H.M. Stationery Office, 1958.

3. Guralnick, L. and Nam, C.B.: Census - NOVS study of death certificates matched to census records. Milbank Mem. Fund. Quart., April 1959, 144.

4. Hauser, P.M. and Lauriat, P.: Record marching -- theory and practice. Proceedings of the Social Statistics Section, Am. Stat. Assn., 1963.

5. Taeuber, C. and Hansen, M.H.: A preliminary evaluation of the 1960 censuses of population and housing. Unpublished paper, Sept. 1963.

6. U.S. Dept. of Commerce, Bureau of the Census: The Post-enumeration Survey, 1950, Technical Paper No. 4. Washington, D.C., U.S. Bureau of the Census, 1960.

TABLE 1

Percent of Matching 1960 Census Schedules Found for Psychiatric Admissions, Louisiana and Maryland, by Age, Sex and Color

Age, Sex and Color	Louisiana: Total persons admitted	Louisiana: Percent matched	Maryland: Total persons admitted	Maryland: Percent matched
All classes	13,036	66.8	14,450	63.9
Under 18 years	2,643	76.5	3,428	70.8
18-24	1,235	55.9	1,232	55.0
25-44	5,432	64.4	5,795	59.9
45-64	2,982	67.2	2,942	66.1
65 years and over	744	66.7	1,053	67.4
White males	5,258	66.3	6,167	64.3
Under 18	1,412	78.0	1,850	71.7
18-24	381	54.3	430	57.2
25-44	1,919	60.7	2,112	58.8
45-64	1,278	65.4	1,370	65.0
65 and over	268	66.0	405	65.2
White females	4,549	71.1	5,255	69.5
Under 18	765	76.5	955	72.3
18-24	456	59.0	456	56.8
25-44	2,013	70.8	2,235	68.7
45-64	1,032	72.7	1,083	73.0
65 and over	283	72.1	526	71.7
Non-white males	1,533	59.2	1,710	51.1
Under 18	241	76.3	393	65.1
18-24	183	53.6	183	47.5
25-44	674	56.5	782	45.4
45-64	331	58.3	281	49.8
65 and over	104	50.0	71	49.3
Non-white females	1,696	64.0	1,318	56.1
Under 18	225	67.6	230	67.0
18-24	215	54.0	163	52.8
25-44	826	64.0	666	51.1
45-64	341	66.3	208	60.1
65 and over	89	70.8	51	66.7

TABLE 2

Percent of Matching 1960 Census Schedules Found for Psychiatric Admissions, Louisiana and Maryland, by Selected Mental Disorder

	Louisiana		Maryland	
	Total persons admitted*	Percent matched	Total persons admitted*	Percent matched
All diagnoses	13,036	66.8	14,450	63.9
Alcoholism	1,255	54.7	1,612	50.9
Diseases of the Senium	510	62.5	755	63.7
Schizophrenia	3,382	62.4	3,322	58.7
Psychoneurotic reactions	2,449	74.2	2,280	70.2
Transient situational personality disorders	940	76.5	1,263	74.0
All other diagnoses	3,626	66.5	4,054	66.1
Undiagnosed	874	74.7	1,164	65.8

*Persons admitted more than once were counted only once and if more than one diagnosis was given for a person, the first one was used for this table.

TABLE 3

Percent of Matching 1960 Census Schedules Found for Psychiatric Admissions, Louisiana and Maryland by Type of Psychiatric Facility*

	Louisiana		Maryland	
	Total persons admitted	Percent matched	Total persons admitted	Percent matched
Public mental hospitals	4,592	60.7	4,807	60.6
VA hospitals	609	60.8	310	55.8
Private mental hospitals	**	**	1,255	74.7
General hospitals	2,849	65.4	938	78.3
Outpatients clinics	6,172	74.8	8,212	64.7

*Counts of persons are unduplicated within each type of facility, but some duplication of individuals exists between types of facility.

**Data on private mental hospital patients in Louisiana were not available for matching.

TABLE 4

Percent of Matching Census Schedules Found for
Persons in the 10 Percent Sample*, Louisiana and Maryland,
by Relationship to Household Head

Relationship To Household Head	Louisiana		Maryland	
	Total in Sample	Percent Matched	Total in Sample	Percent Matched
Total	1,228	69.9	1,025	68.3
Head of Household	287	77.0	239	72.8
Wife of head	279	76.0	182	78.6
Child of head	382	74.9	296	72.6
Total immediate family of head	948	75.8	717	74.2
Other relatives of head	86	62.8	70	57.1
Non-relatives and persons living alone	161	39.8	165	49.1
Inmates	32	62.5	73	64.4

TABLE 5

Percent of Matching Census Schedules Found for
Persons Aged 25 and Over in the 10 Percent Sample,*
Louisiana and Maryland, by Marital Status

Marital Status	Louisiana		Maryland	
	No. in Sample	Percent Matched	No. in Sample	Percent Matched
Total	851	68.5	695	67.1
Married	491	77.4	376	76.1
Widowed	59	57.6	59	61.0
Divorced	83	51.8	50	46.0
Separated	103	56.3	89	57.3
Never married	106	55.7	113	55.8

TABLE 6

Percent of Matching Census Schedules Found for
Persons Aged 25 and Over in the 10 Percent Sample*
Louisiana and Maryland, by Educational Level

Education	Louisiana		Maryland	
	No. in Sample	Percent Matched	No. in Sample	Percent Matched
Total	851	68.5	694	67.1
None	35	68.6	14	71.4
Elementary	387	70.0	265	63.0
High School	294	65.0	279	72.4
College	97	70.1	87	64.4
Unknown	38	76.3	49	63.3

* Random sample of admissions

TABLE 7

Percent of Matching Census Schedules Found for Persons
Aged 25 and Over in the 10 Percent Sample,*
Louisiana and Maryland, by Sex and Employment Status

Sex and Employment Status	Louisiana		Maryland	
	No. in Sample	Percent Matched	No. in Sample	Percent Matched
Males	426	66.2	372	63.4
Working	185	69.7	159	69.8
Looking for work	92	53.3	64	48.4
Unable to work	92	68.5	77	63.6
Inmate	10	60.0	24	62.5
Other	34	85.0	25	64.0
Unknown	13	46.2	23	60.9
Females	425	70.8	322	71.4
Working	73	67.1	57	71.9
Looking for work	15	53.3	11	72.7
Keeping house	276	75.4	154	77.3
Unable to work	36	55.6	42	54.8
Inmate	8	37.5	26	57.7
Other	8	75.0	16	81.3
Unknown	9	77.8	16	68.8

*Random sample of admissions

TABLE 8

Estimated Census Net Undercount, by Sex, Color and Age, United States, 1960**

Age	Total	Percent Undercount			
		White Male	White Female	Non-white Male	Non-white Female
Total	2.3	1.1	1.7	10.3	7.1
Under 5	2.6	2.1	1.4	7.9	6.4
5-14	2.1	2.3	1.3	4.9	3.8
15-24	4.0	3.3	2.3	13.9	9.5
25-44	2.6	2.2	0.7	16.0	6.2
45-64	2.3	0.2	1.8	13.0	12.8
65 and over	0.9	(8.1)*	4.5	(7.9)*	2.6

*Overcount

**Source: Taeuber, C. and Hansen, M.H.: A preliminary evaluation of the 1960 Censuses of population and housing, Bureau of the Census. U.S. Department of Commerce, September, 1963, unpublished.

TABLE 9

A: Ratio $\frac{m_1e_2}{m_2e_1}$ for White-Non-white and Male-Female Comparisons by State

	Louisiana				Maryland			
	Nw/W		M/F		Nw/W		M/F	
	Male	Female	White	Non-white	Male	Female	White	Non-white
Total	.985	.953	.927	.958	.877	.854	.919	.944
Under 18	1.027	.920	1.029	1.148	.953	.965	1.001	.988
18-24	1.109	.988	.930	1.044	.932	1.003	1.017	.947
25-44	1.084	.957	.871	.987	.899	.788	.868	.992
45-64	1.023	1.027	.885	.882	.879	.927	.876	.830
65+	.758	.963	.809	.637	.839	.912	.803	.667

B: Ratio of Enumeration Rates e_2/e_1 for White-Non-white and Male-Female Comparisons, United States

	Nw/W		M/F	
	Male	Female	White	Non-white
Total	1.103	1.058	.994	1.036
Under 18	1.049	1.041	1.009	1.017
18-24	1.123	1.080	1.010	1.051
25-44	1.164	1.059	1.015	1.117
45-64	1.147	1.126	.984	1.002
65+	1.002	.980	.883	.903

THE MATCHING OF CENSUS AND PROBATION DEPARTMENT RECORDS SYSTEMS*

Jon E. Simpson and Maurice D. Van Arsdol, Jr.

University of Southern California

Records linkage as a new methodological technique has generally been neglected by social scientists. Also, with a few notable exceptions, little attention has been given to the linkage of Federal and community data sources on an individual case basis to maximize the efficiency of data collection and analysis.[1]

The present research, known as the Southern California Records Matching Project, is concerned with the matching of two disparate systems - a metropolitan probation department and the United States Bureau of the Census - which are not designed to obtain comparable information, and for which there has been little *a priori* concern with possible records linkage. By merging the two systems, it is possible to obtain a better quality of data for delinquents on a broader range of variables than is usually provided by local data systems. Furthermore, in the past, delinquency rate analyses have been inhibited by differences in the recording of information between adjudicating agencies and the Bureau of the Census, the usual general population data source on which rates are constructed.[2] More explicitly, police and probation departments provide information for the numerators of delinquency rates (usually consisting of population and housing characteristics of the delinquents and their families), while the U.S. Censuses of Population and Housing furnish data for delinquency rate denominators (i.e., information on the presumed corresponding characteristics of population aggregates). If, as is often the case, census data pertain to populations or variables that differ in an unknown way from those defined by delinquency adjudicating agencies, the data may not be comparable and the rates will be in error. Subsequent analyses of delinquent and nondelinquent populations may be invalid.

An alternative to correct for these limitations is records linkage. Direct delinquency rates and comparisons can be derived by matching, on an individual case basis, the probation department records for juveniles (which form the basis of the rate numerators) with comparable individual 1960 census returns (which provide data for rate denominators) for these juveniles and their households. With this technique, numerators and denominators of delinquency rates as well as comparisons between subsamples of the delinquent and general populations refer to the same phenomena - census variables for which comparable data are available to pursue theory building endeavors, to enhance flexible statistical analysis, as well as to evaluate some of the contemporary conceptions of delinquency. In addition to internal comparisons involving subcategories of the delinquent population, the delinquents and/or their households can be examined in terms of: (1) the total population delimited by the same gross criteria which pertain to the delinquent, e.g., in this study, only those households in which there are one or more children between the ages of 10 and 17; (2) the "nondelinquent" siblings of the delinquent; and (3) computer "matched" nondelinquent households within the same or other communities selected on the basis of the presence of a youth with comparable characteristics. Finally, it is possible to analyze the type and incidence of delinquency from one neighborhood (enumeration districts) to the next, contrasting similar as well as dissimilar areas.

This report discusses both the techniques and results of the records matching approach as applied in the present study. Also consideration is given to the attendant problems which have special implications for the tabulation program.

The Records Matching Procedure

In order to obtain a universe of "delinquent" cases, summary information on case identification, name, age, sex, race, and offense was collected for the 23,543 juvenile cases referred to the Probation Department from July 1, 1959 through December 31, 1960 - an 18 month period centered on April 1, 1960. The study population was limited to 13,351 cases comprised of juveniles age 10 to 17 inclusive, who were defined as official cases through the filing of formal juvenile court petitions as a result of alleged delinquent acts.[3]

Procedures were instituted for the collection of intake data from the case folders, and an editing system was devised to furnish standardized identification information for the Bureau of the Census. Thus, information on age, sex, race, and offense was coded, a specification was made of the address of the adult members of the household with whom the juvenile would most likely be matched, and the cases were allocated to census tracts. To achieve maximum efficiency in retriving the required information from the case folders, only the intake "face sheets" or "book sheets" were examined.

Work by the Bureau of the Census. Rules had been previously established for determining whether a match was obtained between the Probation Department and Bureau of the Census records. The key criteria for establishing match status were relationship of juvenile to head, and age, sex, and race of juvenile. Actually two different matching procedures were implemented by the Bureau of the Census.

In the "feasibility" phase, designed to evaluate the practicality of merging the two records systems, the face sheets for a systematic random sample of 2,316 cases comprising one-sixth of the total projected study sample were matched by hand (visually) with the enumeration schedules in storage at Jeffersonville, Indiana. The Bureau allocated the tracted cases to enumeration districts and searched the district files for the records of the juveniles and/or adults in the juvenile's household. Each match failure on the first (primary) address was reallocated to new enumeration districts and/or a search and reallocation of second and third addresses, when available, was completed. Institutional lists were reviewed for cases not located in housing units. The cases for which a complete match - juvenile and head of household - was not effected were returned to the project staff for estimates of new addresses which could be rematched with the census returns. Although the matching rates for this stage will be discussed in the following section on results, it should be noted that the manual procedures located both the juvenile and adult in 84.0 percent of the cases.

Subsequent to the feasibility study and before arrangements could be completed to forward the remaining five samples of the study universe for the matching cycle, the enumeration schedules were destroyed. The relevant matching information was transferred to microfilm reels which could not be utilized without prohibitive increases in financial and time commitments. An alternative method combining the Census listing books and the computer tapes which contain the 25 percent sample information proved almost as effective as the first method with a substantial reduction in cost.[4] The face sheets were matched with the listing books which provide for address, sample status, surname of the head of household, number of household persons, Fosdic page number, ED (enumeration district) number and administrative material. For the cases in the 25 percent sample, the ED and Fosdic page designations as well as the information pertaining to relationship to head of household, age, race, and sex of the juvenile from the face sheets were punched on cards. The cards were matched with the 25 percent sample tapes for Los Angeles County which were specifically constructed for this project from the Los Angeles City and California files. The match failures were processed by techniques similar to those described for the manual method with one additional step. The 25 percent sample microfilm records were used as a final verification source for unmatched and "marginal" cases.

As each case was located, all the population and housing characteristics of the household were added to a tape file for the delinquent population. The general population tape file, which contains the data for the delinquency rate denominators and which is comprised of all families and housing units with one or more children 10 to 17 years of age, was derived from the Los Angeles County tape file constructed for matching purposes.

Finally, comparative information on the relative effectiveness of the two methods suggests that the computer alternative produces substantially similar matching rates and allows the same degree of confidence in the accuracy of the match as the visual procedures. The determining advantages in the choice of alternatives are the lower costs and time requirements to match the 25 percent sample cases by the computer method as contrasted with the availability of complete count data in the microfilm record or enumeration schedules for matching all cases.

Matching results. Table I summarizes the Bureau of the Census search results for the two basic matching operations which have been described in this report.[5] The outcome of the more recent and complete study universe search is reported under the headings: "Listing Book Search" and "Computer Matching Rates." The "Visual Matching Rates" section presents the findings for the earlier feasibility study. the 2,919 sample cases are shown in

TABLE I

SUMMARY OF BUREAU OF CENSUS SEARCH OF LOS ANGELES COUNTY PROBATION DEPARTMENT JUVENILE CASES

	Outcome	Number of Cases	Percent
	Listing Book Search		
(1)	Total Cases	13,351	100.0
(2)	Total Cases Within Scope	12,597	94.4
	(3) Case out of 25 Percent Sample	9,678	76.8
	(4) Case in 25 Percent Sample	2,919	23.2
(5)	Total Cases out of Scope	754	5.6
	Case in Group Quarters	321	42.6
	Case out of Los Angeles County	364	48.3
	Duplicate	22	2.9
	Status Undetermined	47	6.2
	Computer Matching Rates		
(6)	Total 25 Percent Sample Cases	2,919	100.0
(7)	Juvenile and Adult Found	2,270	77.8
(8)	Juvenile Found - Adult Not Found	59	2.0
(9)	Juvenile Not Found - Adult Found	196	6.7
(10)	Juvenile and Adult Not Found	394	13.5
	Visual Matching Rates		
(11)	Total Cases (Feasibility Study)	2,125	100.0
(12)	Juvenile and Adult Found	1,785	84.0
(13)	Juvenile Found - Adult Not Found	57	2.7
(14)	Juvenile Not Found - Adult Found	80	3.8
(15)	Juvenile and Adult Not Found	203	9.5

A total of 13,351 cases (Row 1), which comprise the study universe, were submitted to the Bureau of the Census. As may be seen in Rows 2-4, a total of 12,597 cases were within the scope of the study design, of which 23.2 percent had been allocated to the 25 percent sample enumeration. The listing book and collateral search techniques established that 754 juveniles (Row 5) were out of scope because of residence in group quarters or out of Los Angeles County, duplication, or lack of information.

The computer matching results for Rows 6-10. Both the juvenile and adult were found in 77.8 percent of the cases; the juvenile, but not the presumed adult, was located in an additional 2.0 percent of the cases; the adult only was found in 6.7 percent of the referrals; and neither the juvenile nor the adult was matched in the remaining 13.5 percent of the cases.[6]

The outcome of the earlier feasibility study to test the practicality of merging Probation Department and Census records is reported in Rows 11-15. The somewhat higher rates are probably a function of the more elaborate efforts to obtain a match and the less conservative requirements for accepting a match. Thus, in order to acquire new addresses for rematching, an extensive field follow up of unmatched cases involving a review of school, public assistance, vital statistics, Youth Authority, and Juvenile Index records, and a detailed study of the Probation Department case folders was completed for the feasibility sample, but only the examination of probation files was undertaken for the unmatched cases of the entire study population. Furthermore, location of the juvenile and appropriate adult in the enumeration schedules did not assure that the 25 percent sample information would be obtained without case attrition as occurred, by definition, with the computer matches.

Matching rates by date of referral and social categories. Although the results for the total study population are not yet available, the findings for the feasibility sample provide data for tentative conclusions. An analysis of the matching percentages and numbers of cases located for all possible combinations of months evenly bracketing the census date (i.e., two to eighteen) indicated no substantial decrease in matching rates as a result of using the full eighteen month period, rather than the usual maximum of two to four months. To illustrate the matching rates for the 18 month range of referrals, the highest percentage of juveniles matched for any single month was 94.3 for December 1959, and the lowest percentage matched, 73.5, occurred for July 1959. However, only three months were characterized by rates less than 80 percent. These figures may be compared with the total matching percentage of 86.7 (Rows 12 and 13) for the entire 18 month period.

The findings by sex, race, and offense classifications suggest that the matched cases are representative of the Probation Department universe of "official" cases. With offense controlled no statistically significant sex or race variations were found in comparisons of the matched and unmatched cases. Although differences by offense were not statistically significant, there was a tendency to achieve more success in locating the juveniles brought to the attention of the court for auto theft, major traffic, and property violations than for those youth processed for sex delinquencies and offenses against the persons (e.g., robbery and forcible rape).

To determine whether the matching criteria were systematically relaxed with respect to the matched 25 percent sample cases and at the same time to assess the degree of relationship between the Probation Department and the Bureau of the Census recording of key information, the correspondence between the respective entries for relationship to head, sex, race, and age was reviewed. The two systems were in one to one agreement on more than 96 percent of the relationship to head, sex, and race comparisons and for over 92 percent of the juveniles by chronological age $\pm$ one year to account for the 18 month entry span in the probation referrals bracketing the April 1, 1960 census date.

In summary of the matching results, it is clear that the "delinquent" cases in Los Angeles County can be traced and identified in the returns of the 1960 U.S. Census of Population. This objective can apparently be accomplished for a time span of 18 months without major attrition of cases by month and without eliminating subcategories of the universe based on sex, race, or offense.

Implications for the Tabulation Program

Any matching study must consider the possible biases introduced as a consequence of the limitations associated with the data from the systems that are merged and with the matching process. This final section will focus on selected matching problems currently under examination in connection with the tabulation program; a description of which follows for purposes of orientation. The Bureau of the Census has tabulated group data on an individual case basis for over 100 population and housing variables. In addition, two variable cross-tabulations for 27 key variables and specially constructed indices were completed. The tabulation format provides for 36 sort groups and all possible combinations based on 2 family, 2 sex, 3 age, and 3 ethnic categorizations of the delinquent population.[7] The printouts for the delinquent and general populations include rates and reflect the weighting program used in preparing the 1960 Census sample results.

Underenumeration and sampling errors may be introduced by both the Probation and Census records systems. Certainly, the probation cases are not representative of all youths who have committed delinquent acts or even the "official" delinquents known to adjudicating agencies. However, the total universe of court referrals meeting the project specifications was included and less than one percent was deleted due to administrative difficulties. The underenumeration of census returns for the Probation Department base population can only be estimated with considerable subjective judgment, since there are no survey results that are directly applicable to this specific base population. Nevertheless, an evaluation of this issue resulted in gross underenumeration estimates of approximately 4 and 6 percent as the low and high parameters, respectively. As noted earlier, 23.2 percent of the "in scope" cases were also within the 25 percent sample. At present, estimates have not been developed to indicate the extent to which underenumeration and sampling errors effected the ratio actually obtained.

Match failures and mismatches represent the two basic sources of potential bias resulting from the search process. Since our investigation was not initiated until three years after the 1960 census enumeration, special procedures, such as a sample survey, to acquire information that would permit an analysis of the characteristics of the unmatched cases were not possible. The representativeness of the matched cases can only be tested on the basis of Probation Department data on age, sex, race, date of referral, and offense, as described for the feasibility study. The findings reported on the close correspondence of the Probation and Census entries for the matched cases lends support to the assumption that few cases were mismatched.

In conclusion, it must be noted that the matching approach facilitates the analysis of possible sources of bias to a degree not possible, or usually not explored, with other types of data collection methods. Records linkeage is a technique that, with continued refinement, should stimulate the type of interest that has been manifested in the procedural matters pertaining to censuses, surveys, and sampling. Although its utility is being explored with reference to juvenile delinquency in this project, applications of the method should be consistent with a wide variety of substantive interests.

Footnotes

1. Illustrative exceptions include: David L. Kaplan, Elizabeth Parkhurst, and Pascal K. Whelpton, "The Comparability of Reports on Occupation from Vital Records and the 1950 Census," Vital Statistics Special Reports, LIII (June 1961); Lillian Guralnick and Charles B. Nam, "Census NOVS Study of Death Certificates Matched to Census Records," The Milbank Memorial Fund Quarterly, XXXVII (April 1951), pp. 144-153; Monroe G. Sirken, "Research Uses of Vital Records in Vital Statistics Survey," The Milbank Memorial Fund Quarterly, XLI (July 1963), pp. 309-316; Evelyn M. Kitagawa and Philip M. Hauser, "Methods Used in a Current Study of Social and Economic Differentials in Mortality," in Emerging Techniques in Population Research (New York: Milbank Memorial Fund, 1963), pp. 250-266; and the NIMH Matching Study under the direction of Earl S. Pollack, which is matching census records with mental patients' records to obtain tabulations of patients' characteristics by family characteristics.

2. See, for example, the studies summarized by: Terrence Morris, The Criminal Area (London: Kegan Paul, 1958); Bernard Lander, Toward an Understanding of Juvenile Delinquency (New York: Columbia University Press, 1954); Henry D. McKay, Rate of Delinquents by Communities in Chicago, 1953-1957 (Chicago: Institute for Juvenile Research, 1959), mimeographed; Clifford R. Shaw, et al, Delinquency Areas (Chicago: The University of Chicago Press, 1929); and Clifford R. Shaw and Henry D. McKay, Juvenile Delinquency and Urban Areas (Chicago: The University of Chicago Press, 1942).

3. A total of 9,378 cases were eliminated as a consequence of the following criteria which were implemented in the order listed: under age 10 (3,427 cases - primarily unofficial and/or dependency cases), unofficial (4,089 cases - petition not filed, closed at intake or non-court investigation), dependency (1,842 cases - guardianship, parental neglect, transient, and court consents), and over 18 years of age (20 cases). The final universe of 13,351 cases reflects the additional loss of 814 cases because of duplication or residence outside of Los Angeles County at the time of census enumeration based on the records reviewed by the Population Laboratory.

Males represented about four-fifths of the universe. Anglo-Americans constitute approximately two-thirds of the cases, and Negroes and Spanish surname cases each comprised about one-sixth of the sample. Individual offense categories accounted for 8.5 to 26.0 percent of the total cases with the following percentage distribution by offense type: auto theft (15.2), offense against property (26.0), offenses against persons (8.5), sex delinquencies (9.9), major traffic offenses (14.1), narcotics (4.1), and delinquent tendencies (22.2). Neglect and other dependency referrals and minor traffic violations were previously eliminated from the universe.

4. The 25 percent sample is the most effective source of information from the point of view of the study objectives since complete count data is available for only a few variables.

5. Results described in this report relating to the comparison of Bureau of the Census records with other records are based on work done by the Bureau of the Census and transmitted to the authors in a manner such that the identification of individuals was not possible and the confidentiality of reports to the Bureau of the Census was maintained.

6. The data tabulations which do not involve the youth characteristics include the 6.7 percent of the cases (row 9) in which the adult only was found as well as the 77.7 percent of the cases (row 7) in which both the juvenile and adult were found for a total of 84.5 percent. The former cases were retained for the analyses in view of the very definite evidence that these households were correctly matched even though the youth was not enumerated.

7. The elaborate and detailed tabulations are deemed appropriate in view of the exploratory objectives of the study and the time delays that would occur with a series of runs by the Bureau of the Census.

*This investigation was supported in whole by Public Health Service Research Grant 06729 from the National Institute of Mental Health. The data are drawn from the Southern California Records Matching Project. Agencies conducting the research are the University of Southern California Department of Sociology and Anthropology Population Research Laboratory and the Youth Studies Center. Patricia A. Nelson, Muriel Schad, and James Taggart assumed the primary responsibility for the collection of data from the Probation Department files. John C. Beresford, Staff Assistant, Population Division, United States Bureau of the Census has coordinated the work of the Bureau of the Census in connection with this work. Dr. Georges Sabagh participated in the formulation of the original project design and assisted in the development of the tabulation program. The authors are indebted to Dr. Calvin Goldscheider for a critical reading of this paper.

MATCHING FOR CENSUS COVERAGE CHECKS

Walter M. Perkins and Charles D. Jones, Bureau of the Census

Perhaps it is true that "a rose is a rose is a rose" - this is a little outside our field. We do, however, disagree with anyone who makes the same claim for a match. It has been our job to try to answer the same question for many different individuals, "Was this person counted in the 1960 Census?" In this process, we have frequently met the creature called a "match" - and sometimes its opposite the "nonmatch". These are no problem - it is the gradations and variants that are troublesome. For example, we do find a creature that can only be described as "probably a match - but maybe not" and that other category "seemingly not a match - but possibly is." Then there is that rather mysterious category - "the impute;" the census interviewer has been told that a given address is occupied, but has been unable to find anyone home on repeated visits. The electronic computer, in its infinite wisdom, has imputed at this address a young couple plus a five year old daughter. On a revisit, however, we find the occupant at the time of the census was an elderly man. Was he counted - or wasn't he? It would seem clear that he wasn't; yet, from another point of view, it appears that he has been counted - not wisely, but too well!!

Speaking more seriously, we believe that the gradations in matching - and the variants - rule out the possibility of developing a single optimum matching procedure. The effects of a false match on the one hand, and the failure to make a valid match on the other hand, can each have a drastically different effect on the reliability of one survey as compared with another; and such value differences may make a matching procedure efficient in one survey, yet entirely inappropriate for another. What this group has been doing is developing and improving a variety of matching procedures, where each procedure is designed for a defined set of survey needs and objectives.

With this in mind, we want to take most of our time here to consider the needs and objectives of the coverage evaluation study, as related to matching. We will first give a summary of our matching methods - enough to make it clear that they differ significantly from others discussed at this session. For any that are interested, we have prepared a handout that describes in greater detail the matching procedures used in the matching coverage check studies.

Summary of Matching Procedures

The bases for the coverage evaluation were samples of persons who should have been counted in the census, selected from independent sources. We then check to see if these persons could be found in the census enumerations.

The matching was done by clerical coding, and professional review on a selective basis - the computer was not used. The information available for the match was quite variable. At the minimum, it consisted of the sample person's name and address; at the maximum, it included not only the sample person whose enumeration status was being checked, but the name of every person in his household at the time of the census, and the race, sex, age, marital status, and relationship of each person to the head of the household.

When any match - even vaguely possible - was found in census enumeration records, a two-part code was assigned as a measure of the degree of match. The first part of the code is based on the minimum information only - name and address. Here the definitions could be clear-cut, and there was relatively little chance for marginal cases. Independent verification indicated a very low error rate in these codings.

The second part of the code considers all information other than the name and address of the sample person. The question was "What additional evidence, if any, is provided by such information that the given census enumeration is, or is not, an enumeration of the sample person?" The code allowed for five categories of decision, ranging from "very strong additional evidence that the sample person was enumerated" to "very strong additional evidence that the census household being checked did not include the sample person." Clearly, the categories of decision were subjective. They could be illustrated, but not precisely defined, since it simply was not feasible to write into instructions the great variety of data combinations that might be encountered. What we did in the training of coders was not only to provide a variety of examples drawn from an early sample of actual match situations, but also to justify in specific detail each illustrative code. Some of these illustrative examples and justifications have been included in the handout materials. The names and addresses in the examples have been changed to protect the confidentiality of Census records; yet the variety of matching situations presented is authentic. We were satisfied with our training results for this coding, inasmuch as independent verification of the codes showed a high degree of consistency.

Match-Related Requirements of Coverage Evaluation

With what has just been said as a background, we should like to consider for a moment what the needs and objectives of a coverage evaluation study are - and how these affect matching:

(1) In our record check studies, we start with a sample of persons obtained from a source other than the 1960 Census. As we said earlier, in matching these persons against the 1960 Census, we seek to determine which ones have been missed in the census enumeration. Right here, we have an important difference between matching for coverage evaluation and many other matching studies. For most matching projects, the matched cases are the useful product, i.e., it is the matched cases that are studied, comparing information from the two matched sources. The focus

of a coverage study, on the other hand, is on persons that are unmatched. (The problem in these coverage studies, of course, is to distinguish between genuinely-missed persons and persons who are unmatched for other reasons.)

In most matching projects, by far the majority of cases turn out to be matches. Consequently, when it is the unmatched rather than the matched that is the base for the results, matching errors automatically become more serious percentagewise. Thus, if the "true" missed rate in a population is about 3 percent, the error of failing to match even 1 percent of the sample, when the persons involved are actually enumerated, biases the missed rate about a third!

(2) From one point of view, all errors made in the matching process can be put in two categories - positive errors, i.e., matches made between the sources when the persons involved are actually different, and negative errors, i.e., matches not made for sample persons who are actually listed in the source being searched. In coverage evaluation, these two types of match error are equally serious, since each distorts the missed rate to an equal degree - even though in opposite directions. The estimate of the missed rate is biased precisely to the degree that the expected number of positive match errors differs from the expected number of negative match errors.

In matching projects that are not used for coverage evaluation, positive and negative match errors are not necessarily of equal importance. Typically, the positive match error has the worse effect, since the false match puts a case in sample that does not belong, thus introducing a false association between the data from the two sources being matched. The negative match error, on the other hand, results merely in a loss of sample data - i.e., an increase in the nonresponse rate. Clearly, under these conditions, the two types of matching error are in no sense compensating.

(3) The two characteristics so far discussed - (a) the focus on unmatched cases rather than matched cases, and (b) the equal importance of positive and negative match errors - are basic to coverage evaluation studies. The 1960 record check studies had, in actual operation, a third characteristic that affected the matching results - a great range in the amount of information available for the match. For the most part this gradation was a reflection of the different sources from which our samples of names were drawn, and of the great variation in how current those sample lists were. One sample, for example, provided for the census search a name, in some cases an age, and a mailing address current as of January 1960. At the other extreme, one sample originally provided a great deal of information about the sample person, his personal characteristics, residence, other members of his household, their characteristics, and even neighbors and their personal characteristics - all this with one slight drawback.....the list was 10 years old!! In this case, of course, we did not put the information directly into a census matching operation, but first went through a location procedure in which we attempted to learn whether the person was still alive in April 1960, his current residence, and details about the household in which he currently lived. When the location procedure was successful, we typically had a good deal of detailed information upon which a census enumeration identification could be based.

Supporting Evaluation Procedures

Once the match-related requirements of a survey are clearly recognized, it becomes easier to choose between alternative matching procedures - it is also easier to see when no matching procedure, by itself, can possibly do the job required. This was the case in the coverage evaluation program. The sources of our lists of sample persons provided us with addresses as well, but in a significant number of cases these addresses were not where the sample persons were living at the time of the 1960 Census. Under these circumstances, any matching procedure whatever would return the sample persons as unmatched, not because they were missed in the census, but because they were enumerated at addresses that were not searched. Such negative matching errors would substantially overstate the estimate of the missed rate.

A field reconciliation operation was set up to reduce the number of erroneous decisions that would otherwise result from the matching operation. Field reconciliation was carried out by letter, by telephone, and by personal visit. All unmatched persons plus all problem cases were sent for reconciliation. Only the "sure" matches were treated as final decisions. Broadly speaking, three types of reconciliation were done, as follows:

(a) In all cases, careful questioning was done to determine all other addresses where the sample persons should or could have been enumerated. (This probing, obviously, was designed to reduce the number of negative matching errors we spoke of a moment ago.)

(b) In some cases, especially where a rural address was involved, we had difficulty in conducting a satisfactory census search because it was difficult to determine the exact geographic location of the address, and consequently, which Census Enumeration District should be examined. Here, the goal of the reconciliation was to pinpoint the location of the address on a map; this geographic location was then readily translated into an equivalent Census Enumeration District.

(c) Cases in which the matching operation found in the census a "probable" or, perhaps, just a "possible" misspelling of the sample person's name, were also included in reconciliation. Here, the goal of reconciliation was to search for the persons with the name as spelled in the census. If no such named person was found, or that named person was identified as the sample person, this was considered as confirming evidence for a match.

Summary

As stated earlier, we are happy to supply any interested person with a detailed description of the matching procedures used in the record check coverage studies of the 1960 Census. We hope it is clear that we are not recommending these procedures for indiscriminate use in surveys that require matching. In general, we suspect the procedures we followed entail costs that are higher per person matched or searched for than the costs of other systems of matching. In developing or selecting a system of matching, an essential criteria is the function that the matching system must perform in the given survey. In the paper, we have discussed the three requirements of coverage evaluation that are particularly critical to the matching of the three, undoubtedly the most important is the fact that coverage evaluation matching focuses on unmatched rather than matched cases.

BIBLIOGRAPHY

(1) U.S. Bureau of the Census. Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960: Record Check Studies of Population Coverage. Series ER60, No. 2. Washington, D.C., 1964.

(2) U.S. Bureau of the Census. Procedural Studies of the 1950 Censuses, No. 1, Infant Enumeration Study: 1950.

(3) Census-NOVS Study of Death Certificates Matched to Census Records, Milbank Memorial Fund Quarterly, April 1959, Vol. XXXVII, No. 2, pp. 144-153.

(4) U.S. Bureau of the Census. Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960: Accuracy of Data on Housing Characteristics. Series ER60 No. 3. Washington,D.C., 1964.

(5) U.S. Bureau of the Census. Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960: Accuracy of Data on Population Characteristics as Measured by Reinterviews. Series ER60, No. 4. Washington,D.C., 1964.

(6) U.S. Bureau of the Census. Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960: Accuracy of Data on Population Characteristics as Measured by CPS-Census Match. Series ER60, No. 5. U.S. Government Printing Office, Washington,D.C., 1964.

(7) U.S. Bureau of the Census. Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960: The Employer Record Check. Series ER60, No. 6. U.S. Government Printing Office, Washington, D.C., 1965.

HANDOUT

MATCHING FOR CENSUS COVERAGE CHECKS

In a number of census coverage checks, records for persons selected from record sources independent of the 1960 Census were compared to the census records for those persons, their households, and their addresses. For each comparison, a two-part code was clerically assigned to indicate the degree of information agreement between the two sources. The first part of that code was for address-name agreement while the second part of the code was for agreement of all information between the sources not used in the address-name coding. A large number of examples was used to train the clerks on how to assign composite codes - especially for the assignment of the second part of that code. The definitions given for the second part of the code were in relative terms and required some subjective interpretation by the coders. The examples served to guide their subjective interpretation.

Once the codes had been assigned, a sample of cases were reviewed in order to categorize the cases by MATCHED or NON-MATCHED by composite code. the MATCHED category consisted of those cases where the evidence was clear that the sample persons had been enumerated in the census. The NON-MATCHED included, in addition to census missed persons, all the matching problem cases. All NON-MATCHED cases were subjected to a reconciliation process carried out by office review, further field work, and additional census search for possible matches.

In this handout are the address-name codes with their descriptions, definitions of terms used in those descriptions, the supplemental information codes with their descriptions, some examples used for training the coders how to make composite code assignments (Illustrations 1-10), how these composite codes were used to classify each case as MATCHED or NON-MATCHED (Illustration 11), distribution of MATCHED and NON-MATCHED cases for some of the studies (Illustration 12), and the work sheet used to record all information pertinent to the matching operation used for these matching coverage studies (Illustration 13).

ADDRESS-NAME CODES

CODE	DESCRIPTION OF COMPARABILITY BETWEEN THE TWO SOURCES
A	Same address and same name
B	Same address and similar name or similar address and same name
C	Similar address and similar name or noncontradictory address and same name
D	Same or similar address and noncontradictory name or different address and same name
HA HB HC HD	Same criteria as for Codes A-D above respectively except that the name part of the comparison was made on some person in the household other than the sample person (these codes were frequently used when the names for the sample person, as recorded in the two sources, were quite different, or there was no possible match for the sample person in the census household being examined).
X	All other address-name comparisons than those described above.

DEFINITIONS OF TERMS USED IN DESCRIPTIONS

SAME NAME

Two names are same if:

(a) They are identical with at least one given name present;

(b) The last and one given name are identical with

(1) no disagreement with respect to the other name or initial or

(2) only minor variation with respect to the other given name;

(c) An accepted contraction or nickname is given in one source for a first name; last name being identical.

SIMILAR NAME

Two names are similar if they differ too much to be considered "same" and the difference can be attributed to:

(a) Careless spelling or an error in interpreting handwriting;

(b) Phonetic spelling of the name

NONCONTRADICTORY NAME

Two names are noncontradictory if the last names are same or similar and:

(a) There is no first name in one of the sources; or

(b) Given names and initials between the two sources agree on initial basis.

SAME ADDRESS

Two addresses are same if:

(a) They are identical (this includes rural addresses if such addresses are numerically specific);

(b) The difference between the sources can be attributed to omission of street type, compass point, or apartment designation.

SIMILAR ADDRESS

URBAN

There are two parts of an urban address (a) street number and (b) street name. Two urban addresses are similar if:

(a) street number is identical while street name is similar where similar street name is described in the same terms as for a similar name for persons;

(b) street name is identical while street number is similar where similar street number is described as a number where the difference could arise from:

(1) a mistake in writing a digit in the number,

(2) an omission of a digit in a number, or

(3) the transposition of digits in a number.

RURAL

Two rural addresses are considered to be similar if:

(a) They are the same but are not numerically specific;

(b) They differ in description with some identical descriptive terms in both sources but the complete descriptions between the two sources are not contradictory.

NONCONTRADICTORY ADDRESS

Two addresses are noncontradictory if they differ in description but are not in opposition in the meaning (this term is used only for rural addresses where addresses often are described using different civil level descriptions for the road name).

DIFFERENT ADDRESS

Two addresses are different if they differ too greatly to be considered either same, similar, or noncontradictory.

SUPPLEMENTAL INFORMATION CODE

CODE	DESCRIPTION OF COMPARABILITY BETWEEN THE TWO SOURCES
++	Excluding information used in the address-name coding, the remaining evidence is very strong that the sample person is included in the Census household being compared. Minimum requirements for this code is that there be other household members in both sources and that the personal characteristics (age, race, sex, relationship to head, marital status) of the sample person and the household information essentially agree between the sources.
+	Excluding information used in address-name coding, there remains significant additional information that the sample person is included in the Census household being compared. Some items to consider are: (a) Sample person's age as given between the two sources is within 1 year of agreement; (b) Household composition, though not in complete agreement between the two sources, has more positive agreements than negative disagreements; (c) "Race" and "marital status" are not not sufficient for significance but add positive evidence if race is other than white or Negro or marital status is other than married or never married and these agree between the sources.
0	Excluding information used in address-name coding, the remaining evidence, or balance, provides no significant amount of additional evidence that the sample person is or is not included in the household being compared. Some instances of use of this code are: (a) When only address and name is given between the sources; (b) When information on personal characteristics is available but is not sufficient to be considered significant evidence; (c) When additional information is available but there are enough contradictions in the available information to offset any positive significant evidence in both sources.
?	Excluding information used in address-name coding, there is a significant amount of remaining evidence that the sample person is not included in the household being compared. This contradictory evidence can be of the form "probable different household" or some disparity in personal characteristics for the possible sample person match between the two sources.
??	Excluding information used in address-name coding, there is very strong evidence that the sample person is not included in the Census household being compared. The contradictory evidence here can take the form "different household" or different personal characteristics--especially large differences in age where the possible matching sample person has an age that would place him in another generation from the sample person in the independent record source.

MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 1

The examples given in these Illustrations were used for training coders on how to make the composite code assignments. The examples come from actual cases encountered in the match. For this handout, however, the addresses and names have been altered to protect the confidentiality of census records.

	ADDRESS AND NAMES	RELATIONSHIP TO HEAD	AGE
INDEPENDENT RECORD SOURCE	1830 W. Flat St.		
	*Cavallo, Sue P.	Hd.	-
	Russo, Jack L.	Father	51
	_____, May B.	Mother	47
	_____, Charles J.	Brother	23
	_____, Jill A.	Sister	14
	_____, Margaret	Sister	7
	_____, Doris H.	Daughter	20 Months
CENSUS RECORD	1830 Flat 3rd		
	Russo, Jack L.	Hd.	51
	_____, May B.	Wife	47
	_____, Charles T.	Son	23
	**_____, Sue T.	Daughter	21
	_____, Jill A.	Daughter	15
	_____, Margaret T.	Daughter	7

*Sample Person **Possible Sample Person

Code "HA++"

Justification

"HA" - "1830 W. Flat St. vs. 1830 Flat, 3rd." is same address.
- "Russo, Jack L." given in both sources is same name for a household member.

"++" - Both sources show the same name for the following members of the Russo family; May B. Charles (J.vs.T), Jull A. and Margaret (T). The ages and relationships are the same for the respective members of the two households. The sample person"s name, "Cavallo, Sue P." (Russo) might reasonably be the married name of Russo, Sue T.

MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 2

The examples given in these Illustrations were used for training coders on how to make the composite code assignments. The examples come from actual cases encountered in the match. For this handout, however, the addresses and names have been altered to protect the confidentiality of census records.

	ADDRESS AND NAMES	RELATIONSHIP TO HEAD	AGE
INDEPENDENT RECORD SOURCE	1318 Steeple St.		
	Holster, James K.	Hd.	42
	_______, Edna	Wife	39
	_______, Allen	Son	18
	_______, Keven	Son	15
	Smith, Dave	Step-Son	12
	*_____, Louise	Step-Dau.	10
CENSUS RECORD	1314 Steeple		
	Holster, James	Hd.	42
	_______, Edna	Daughter	39
	_______, Allen	Son	18
	_______, Keven	Son	15
	_______, John	Son	12
	**_______, Jill	Daughter	10

*Sample Person **Possible Sample Person

Code "HB++"

Justification

"HB" - "1318 Steeple St. vs. 1314 Steeple" are considered <u>similar addresses</u>.

- "James K. Holster vs. James Holster" are considered same name for a household member.

"++" - Note that we have exactly 6 persons in the two households being compared, with ages that can be exactly matched - year for year. Four names (including the one already coded in the HB) are the same. True, the other two names (including the "possible" sample person) are entirely different, but these might represent changed names - before and after adoption.

The relationships shown are different in 3 cases, and appear to be unreliably recorded in the Census. Certainly is is unreasonable to show "Edna" as a 39-year old daughter of the 42-year old head.

MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 3

The examples given in these Illustrations were used for training coders on how to make the composite code assignments. The examples come from actual cases encountered in the match. For this handout, however, the addresses and names have been altered to protect the confidentiality of census records.

	ADDRESS AND NAMES	RELATIONSHIP TO HEAD	AGE
INDEPENDENT RECORD SOURCE	1601 Wayne		
	*Hayes, Hattie	Hd.	63
	_____, Sue Ann	Grand Daughter	7
	Henson, Kari Lynn	Grand Daughter	6
CENSUS RECORDS	1601 Wayne		
	**Haze, Hattie	Hd.	63
	Henson, Carolyn	Rel.	8
	Haze, Suzan	Daughter	6

*Sample Person **Possible Sample Person

Code "B++"

Justification

"B" - "1601 Wayne" is shown in both sources

- "Hayes, Hattie" vs. "Haze, Hattie" are considered <u>similar names</u>.

"++" The same 3-person household composition is shown on both sources (a head and two children) with the "possible" sample person listed as head and 63 years old.

The 7 year old granddaughter shown in the source (Sue Ann Hayes) has a phonetically-similar name and the same age as the 6 year old daughter shown in the census (Suzan Haze). The relationship, granddaughter vs. daughter, might be considered noncontradictory.

The second child is shown by both sources as having the same last name, Henson, (Which is other than the last names of the other household members) and phonetically-similar first names. The age and relationship comparison are "noncontradictory" as defined.

MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 4

The examples given in these Illustrations were used for training coders on how to make the composite code assignments. The examples come from actual cases encountered in the match. For this handout, however, the addresses and names have been altered to protect the confidentiality of census records.

	ADDRESS AND NAMES	RELATIONSHIP TO HEAD	AGE	MARITAL STATUS
INDEPENDENT RECORD SOURCE	Apple Tree Road (Rural)			
	Burr, Daag K.	Hd.	70	
	*____, Betty	Wife	70	
CENSUS RECORD	Route 52, 1st. house after State 25			
	Burr, Daag K.	Hd.	70	Married
	**____, Betty T.	Wife	70	Married

*Sample Person **Possible Sample Person

Code "C++"

Justification

"C" - "Apple Tree Road (Rural) "vs. "Route 52, 1st. house after State 25" are considered noncontradictory addresses. "Burr, Betty" vs. "Burr, Betty T" are considered same name.

"++" - Both sources list two persons in the household - Burr, Daag K., age 70, as head, and the "possible" sample person as his wife, age 70. The rather uncommon name "Daag" should be considered to have a particularly high weight as evidence.

MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 5

The examples given in these Illustrations were used for training coders on how to make the composite code assignments. The examples come from actual cases encountered in the match. For this handout, however, the addresses and names have been altered to protect the confidentiality of census records.

	ADDRESS AND NAMES	RELATIONSHIP TO HEAD	AGE	MARITAL STATUS
INDEPEN-DENT RECORD SOURCE	3932 Steel St. *Grant, Troy Gene	Hd.	65	
CENSUS RECORD	3932 Steel **Grant, Gene T.	Hd.	65	Married

*Sample Person **Possible Sample Person

Code "A+"

Justification

"A" - "3932 Steel St. vs. 3932 Steel" are considered same address.

"Grant, Troy Gene" vs. "Grant Gene T." are considered same name.

"+" - Both census and source show household consisting of one person of the same age.

Note that, for matches involving only the personal characteristic of the sample person, this is about as good a match as can be expected. Knowledge of race and sex would not contribute appreciably to the match.

MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 6

The examples given in these Illustrations were used for training coders on how to make the composite code assignments. The examples come from actual cases encountered in the match. For this handout, however, the addresses and names have been altered to protect the confidentiality of census records

	ADDRESS AND NAMES	RELATIONSHIP TO HEAD	AGE	MARITAL STATUS
INDEPENDENT RECORD SOURCE	128 Dallas St.			
	Milano, Eugene S.	Hd.	22	
	*______, Lucille G.	Wife	21	
CENSUS RECORD	128 Dallas - Basement			
	Milano	Hd.	-	Married
	**______, ?	Wife	-	Married

*Sample Person **Possible Sample Person

Code "D+"

Justification

"D" - Note that "128 Dallas St." vs. "128 Dallas - Basement" are considered same address.

"Milano, Lucille G." vs. "Milano,?" are considered noncontradictory names.

"+" The source and the Census both show households consisting of exactly 2 persons, a married couple.

This is a very week "+", which might reasonably have been considered as "0". If the households being compared had had more than 2 people (with agreement in number and relationship), or if the Census had reported age for at least one of the couple that agreed with the source within one year, then the code would be definitely "+".

MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 7

The examples given in these Illustrations were used for training coders on how to make the composite code assignments. The examples come from actual cases encountered in the match. For this handout, however, the addresses and names have been altered to protect the confidentiality of census records.

	ADDRESS AND NAMES	RELATIONSHIP TO HEAD	AGE	MARITAL STATUS
INDEPENDENT RECORD SOURCE	Middlebrook State School for the Mentally retarted			
	Michael, Carl B., M.D.	Superintendent		
	*Hanes, John	Inmate		
CENSUS RECORD	Middlebrook State School			
	**Hanes, John M.	Inmate	45	Never

*Sample Person **Possible Sample Person

Code "BO"

Justification

"B" -"Middlebrook State School for the Mentally Retarded vs. Middlebrook State School" are considered <u>similar address</u>.

"Hanes, John vs. Hanes, John M." are considered <u>same name</u>.

"O" -The additional information consists only of the fact that the "possible" sample person is an inmate. This is certainly not very distinctive for an institutionaladdress. (The agreement of address is of course, already included in code "B"). If the source had shown an age for the sample person which was within one year of the age given in the census household, a code "+" would have been assigned.

Note: This case illustrates a possible deficiency in what was transcribed from the census records. In this case, the records would be reexamined to see if "Michael, Carl B." was included in the census as being the superintendent, and, if not, to transcribe the name of the person recorded as being in charge.

MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 8

The examples given in these Illustrations were used for training coders on how to make the composite code assignments. The examples come from actual cases encountered in the match. For this handout, however, the addresses and names have been altered to protect the confidentiality of census records.

	ADDRESS AND NAMES	RELATIONSHIP TO HEAD	AGE	MARITAL STATUS
INDEPENDENT RECORD SOURCE	1498 West 168th St. Apt. 2B			
	Spurlock, Stanley	Hd.		
	________, Babs	Wife		
	Stile, Agnes	Inmate		
	*Jackson, Carl	Lodger	23	Married
	________, Alice	Lodger's Wife	23	Married
CENSUS RECORD	1417 W. 166th St. Apt. 3D			
	**Jackson,	Hd.	-	Married
	Jackson,	Wife	-	-

*Sample Person **Possible Sample Person

Code - "XO"

Justification

"X" - "1498 West 168th St., Apt. 2B vs. 1417 W. 166 St., Apt. 3D" are considered <u>different addresses</u>. Note that the difference arises <u>from three</u> "similarities" - the house number, the street number and the apartment number. Had any two of these been the same, the two addresses would have been considered similar.

"Jackson, Carl vs. Jackson" are considered <u>noncontradictory names</u>.

"O" The only evidence that "Jackson, Carl" is "Jackson, ___" beyond the poor name-address comparison is the fact that in both sources the "possible" is shown as married, wife present. This fact is <u>not</u> sufficient in itself to merit a "+", and, in any case, the fact that the couple is shown as living by themselves in one source and as living as lodgers in someone else's household in the other source is offsetting evidence in the "?" direction.

MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 9

The examples given in these Illustrations were used for training coders on how to make the composite code assignments. The examples come from actual cases encountered in the match. For this handout, however, the addresses and names have been altered to protect the confidentiality of census records.

	ADDRESS AND NAMES	RELATIONSHIP TO HEAD	AGE	MARITAL STATUS
INDEPENDENT RECORD SOURCE	2845 Lee St.			
	*Spencer, Samuel G.	Hd.	34	Married
	________, Geraldine B.	Wife	34	Married
CENSUS RECORD	2845 Lee St.			
	Spencer, Geraldine B.	Hd.		
	**________, Samuel	Son		

*Sample Person **Possible Sample Person

Code "A?"

Justification

"A" - "2845 Lee St." is present in both sources.

- "Spencer, Samuel G." vs. "Spencer, Samuel" are considered same name.

"?" - Geraldine B. Spencer is listed as 34 year old wife of Samuel G. Spencer in the source and and as his mother in the Census. This is significant evidence that the two "Samuel Spencers" may be different people - possible father and son.

MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 10

The examples given in these Illustrations were used for training coders on how to make the composite code assignments. The examples come from actual cases encountered in the match. For this handout, however, the addresses and names have been altered to protect the confidentiality of census records.

	ADDRESS AND NAMES	RELATIONSHIP TO HEAD	AGE	MARITAL STATUS
INDEPENDENT RECORD SOURCE	1354 Main			
	*Smith, Charles M.	Hd.	20	
	Myers, Eva	Grandmother	87	
CENSUS RECORD	1354 Main, Apt. 4			
	Meyers, Eva.	Hd.	88	Never

*Sample Person

Code "HB??"

Justification

"HB" -"1354 Main vs. 1354 Main, Apt. 4" are considered same address.

"Myers, Eva vs. Meyers, Eva" are considered similar names (Sample person not involved)

"??" -The sample person, Smith, Charles M., clearly does not appear in the Census household.

MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 11

After the composite codes were assigned, a sample of cases were reviewed in order to categorize the cases as MATCHED or NON MATCHED by composite code. The table below shows how the codes were divided into these categories.

Address-Name Code

SUPPLEMENTAL INFORMATION CODE	A	B	C	D	HA	HB	HC	HD	X
++	MATCHED - Requires only clerical review and verification; Sample person enumerated in Census.								
+									
0	NON-MATCHED - Requires professional review to conclude about match status or whether further field and office work is required before concluding about enumeration in the census.								
?									
??									

MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 12

Percentage Distribution of "MATCH" and "NON-MATCH" cases for 4 record check studies reported in ER 60 No. 2 a/

	STUDY			
	TOTAL 4 Studies reported in ER 60 No.2 a/	Sample of Birth Records for Children born during inter-censal period 4/1/50 to 4/1/60	Sample of persons selected from 1950 Census and 1950 PES b/ records	Sample of Persons selected from alien registration records; January 1960
TOTAL	100.0	100.0	100.0	100.0
MATCHED (only clerical review and verification required)	84.0	86.6	82.7	62.8
NON-MATCH (professional review required to conclude about match status or whether further search or field work is needed)	16.0 c/	13.4 c/	17.3 c/	37.2 c/
Approximate number of cases for which census matching was attempted	6200	3700	2300	200

a/ U.S. Bureau of the Census, Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960: Record Check Studies of Population Coverage. Series ER 60 No. 2, Washington, D.C. 1964.

b/ U.S. Bureau of the Census, the POST-Enumeration Survey: 1950. Bureau of the Census Technical Paper No. 4. Washington, D.C. 1960.

c/ These percentages are not missed rates. Rather, they reflect the total group of matching problems from which the missed cases are finally culled by the process of professional review, further Census search, and field follow-up.

MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 13

Form 60-28-1.16 U.S. DEPARTMENT OF COMMERCE Bureau of the Census WORKSHEET TO SEARCH CENSUS RECORDS FOR SAMPLE PERSONS	1. Control Number	2. Best Match Code
	3. Source	4. Date on Source

5. INFORMATION ON INITIAL SOURCE

a. Address	City	State

Line No. b	Name c	Relation to Head d	Sex e	Race f	Date of birth Month g	Year h	Age i	Marital Status j
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								

6. FIRST POSSIBLE MATCH - CENSUS

a. Address (Listing Book)	b. Address (FOSDIC Book) Same as Listing Book

c. ED Number	d. L.B. Page	e. L.B. Line	f. FOSDIC Page	g. Key	h.	i.	j.	k.	l.

Line No. m	Name n	Relation to Head o	Sex p	Race q	Date of birth Month r	Year s	Age t	Marital Status u
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								

DISCUSSION

Jack Silver, Bureau of the Census

In comparing a set of study records with an independent set, the number of successful matches depends not only on the specifications for a match and the "true" proportion of cases in the independent file which actually match, but also on the reliability of the mechanism, whether computer or clerical. Many of us have experience with clerical matching operations in which the number of successful matches increases each time the operation is repeated. It would seem to me that a useful part of any clerical matching operation would be estimates of the "failure to match" rate (when match is possible) and the "successful match rate" when match is not possible under the specifications.

Computers <u>must</u> follow specifications. Clerks performing visual matching, however, have been known to modify specifications to increase the number of matches by taking advantage of information in ways not permitted by the specifications. On the other hand, they may fail to make permissible matches for any number of reasons. Random samples of the study records could be used to determine the extent to which specifications are not followed strictly. (It is quite conceivable that one may want to modify the original specifications to take advantage of clerical ingenuity used in increasing the number of matched cases.)

While listening to the Simpson and Van Arsdol experience in matching juvenile delinquency records against the 1960 Census, I wondered whether part of the failure to match rate could be attributed to some undercoverage of a population which may be inherently more difficult to enumerate. I think it would be interesting to know whether the failure to match rates would have been statistically different had the same specifications been employed in matching the records for a more representative group of persons in this age group against the 1960 Census.

Discussion

J. William Lloyd - National Cancer Institute

The four papers presented this morning have given us a comprehensive view of record linkage procedures and some of the problems encountered in their use.

Rather than review the papers individually, I would prefer to offer some general comments on record linkage procedures with reference to what our speakers have presented.

My first job in the field of statistics was with a local health department where the compilation of vital statistics had for many years consisted of single entry of events in a set of ledgers. Each of these ledgers being maintained by geographic area and finely subdivided by the demographic characteristics of interest. Before my arrival on the job, it had been decided that the volume of events to be recorded was too much for the ledger system and should be replaced by a mechanized system of counting--namely, a key punch machine and a 250-card-a-minute sorter.

As you may have anticipated, what should have been a much more rapid and efficient system of enumeration was at least for a short while an uncontrollable Frankenstein. Coding schemes had to be developed, the clerk who had made the Journal entries had to be trained to key punch--a task he never did master. He was convinced that the new-fangled system wouldn't work. And the sorter, in addition to throwing cards in the wrong pocket, would without warning chew up cards by the handful.

Now we are in the day of the "black box"--the computer that can pair 200,000 records of one system with 300,000 records of another, and quite rapidly select the 38,000 pairs that meet some specification. Our old ledger clerk would be amazed! While the problems generated by our new equipment, and our new technique of record linkage are not quite analogous to these encountered with our earlier mechanized procedure, we have responded in quite the same manner as we did to the 250-word-a-minute sorter. In our eagerness to make use of the high speed computer, we sometimes forget that we are still doing exactly the same thing as the ledger clerk. Each of the decisions made by the ledger clerk when he visually reviewed a pair of records is a decision that we must make. The problem that we encounter in working with the high-speed computer is that we have to anticipate the kinds of situation which might have arisen in a manual review, and specify decisions for each situation.

Before cataloging decisions of this sort, however, we should first ask ourselves if it is reasonable in terms of expected productivity to even consider matching of the two systems. While no detailed set of rules can be devised to cover every set of records which could be matched, there are some general criteria which should be considered before any record linkage is attempted.

The first of these, which may seem obvious, is that each record in one group should have a chance to appear in the other group. For example, in the HIP study we might expect some attrition in the death file over time among persons retiring to Florida and therefore not reported as New York residents at the time of death. There will be times, of course, when this criteria is not strictly met but can be corrected for by estimating the occurrence of an event among the nonmatched cases. In the HIP study such estimates might be based on out of state death claims filed with the insurance company. The same sort of a procedure was suggested in the paper on psychiatric admissions. The bias resulting from failure to meet this criteria can be considerable. In one of the studies being conducted by our agency we have noted quite different patterns of mortality between those who have remained in a local community and those who have left.

As other criteria for successful record linkage, I would suggest those pointed out by the authors of the first paper. They bear repeating. The common identifying information of the two systems should have

(1) high discriminating power
(2) low probability of change during an individuals lifetime and
(3) low likelihood of being recorded erroneously.

The unexpected low rate of matching in the psychiatric admissions study may be due in part to the dependence on information which does not fit these criteria, i.e. address at admission. Dr. Pollack, it should be noted, also requested the patient's address at the time of the census. Some of the nonmatches may therefore be due to faulty recall. The probation records, in contrast, showed a much higher rate of matching for an eighteen month period around the census with no apparent tendency to decrease over time.

One other factor which might account for the different match rates of these two studies is that there may be considerable differences in levels of enumeration and/or identification by enumeration district for large metropolitan areas as compared to complete states (which may have a large rural component). The authors of the last paper have pointed out that problems exist in the classification of rural addresses by enumeration district. Finally, I would raise the question as to whether matching of survey records to those of the decennial census, using name and address as the primary information, appears to be sufficiently fruitful for estimating vital statistics rates.

VI

NEEDED DEVELOPMENTS IN SOCIAL STATISTICS

Chairman, FREDERICK F. STEPHAN, Princeton University

MONITORING SOCIAL CHANGE: A CONCEPTUAL AND PROGRAMMATIC STATEMENT

Wilbert E. Moore, Russell Sage Foundation
Eleanor Bernert Sheldon, Russell Sage Foundation

Lists of needed research and of needed data constitute a dismal catalogue of our identifiable ignorance, and leave untouched the greater ignorance beyond. Aside from their possibly chastening effect as a deflator of arrogant self-satisfaction, the utility of such lists may not be high. We have no authenticated record of the consultation of such lists by researchers "at liberty" and seeking new topics. Perhaps if the compendia of ignorance were presented as opportunities for those seeking to make their mark rather than as failures of their predecessors to hit the mark, the positive effect on cumulative knowledge would be greater.

Despite some unease at presenting yet another list of things that someone else ought to do, we venture to do so, for we hope to make some contribution to the state of thinking about large-scale structural change in American society, and to follow this up with some small steps in getting new and better data collected and new and better analyses undertaken.

With a pedantry that we hope is forgivable, we want to start with some explication of this paper's title, though not in the precise order of the critical words. By social change we mean, as just indicated, large-scale structural transformations, and to avoid tedium in positive and negative definition of this set of qualifiers, we ask the indulgence of identification by systemic explication of such changes as we go along. At the moment, we may indicate some of the things we are not concerned about: the formation of individual personality in the transition from infancy to a possibly eventful adulthood; the emergence of varying leadership styles in small groups; the dynamic processes in voluntary associations or formally constituted administrative organizations. These do constitute part of the spectrum of social change, but they may have little bearing on the major shape of contemporary society, and it is to the latter that we wish to attend.

Monitoring social change we mean in the full, ambiguous sense of the term. We are concerned with "tuning in," with recording and verifying the messages we may get or produce relating to structural alterations. But we are also concerned with the use of such information for entry into the system, to alter the magnitudes, speed, or even direction of change in terms of explicit, normative criteria. Many people are attempting to manipulate the system, and we think we have some obligation to unite such sophisticated skills as science affords us with such practical policies as explicit value-orientations validate.

And that brings us to the programmatic part of our proposal. For the major aspects of structural change, to which we turn presently, we hope to secure expert appraisals of the current state of knowledge, and of what knowledge, within our readily potential means, would be appropriate. Specifically, we shall ask an expert on each of our major topics to summarize: (a) what we know retrospectively concerning trends in the area under review, with a note on whether that knowledge is quantitative or possible quantifiable; (b) what we know about current state, and, in combination with (a), about prospective state; (c) what additional trend data are needed, and why? For this third query, we seek both an intellectual rationale, such as potential utility in the explanation and prediction of social phenomena outside the particular topic under inspection, and, where appropriate, a policy rationale, a focus on the use additional and hopefully sophisticated trend data would have for monitoring social change in our second sense, that is, for use in entering the system. We also hope to elicit some additional methodological contributions, of presumed interest to statisticians, but which we can only suggest here: problems of uniformity in definition and modes of reporting, determining the proper periodicity for serial observation, suggesting aggregative indexes or other summary measures and finding the common mensuration units to make such synthesis tenable and meaningful for analysis and policy.

Our principal task so far has been that of conceptualization: how do we most usefully think about, that is, analyze, the major structural features of American society for the purposes we have just indicated. For a time it appeared that we might end up rediscovering introductory sociology, much of which relates precisely to this problem of parts or features or analytically distinct functions of societies. However, we have had to depart somewhat from well-established and authenticated conventions, for our focus differs in at least two significant respects: we are concerned with trends of change, and we are particularly alert to the way deliberate policy does and could affect those trends.

Let us outline first the major headings that we think will conveniently serve to organize the structural changes that merit attention. We shall then turn to some particular topics, which we hope will indicate more explicitly, and with brief illustrative detail, what the grand design entails. We propose five major rubrics for examining structural changes in American society and its constituent features: (1) the demographic base, giving an indication of aggregative population trends, shifting differential contributions to those trends, and the distribution of people across the geographic surface; (2) major structural components of the society, where we come closest to the rediscovery of introductory sociology by looking at the functionally distinct ways in which a society produces goods, organizes its knowledge and technology, reproduces itself and regulates adult sexuality, and maintains order; (3) distributive features of the society, meaning the way various goods, benefits, and services get allocated through the several sectors of the population; (4) aggregative features of the society, where we attempt to look at what the society as a whole is like, by at least implicit comparison with other societies and by explicit comparison with our own historic past; (5) finally, the meaning of welfare is singled out as a conception of public and private benefits that has changed both in its significance and in the degree to which the society approximates the achievement of its own changing standards. We recognize that synthetic statistical indicators might be readily attainable in some areas under inspection while others may call for considerable analytic research. We also realize that even these major rubrics maybe subject to change pending further review.

Space and time are notoriously limiting parameters, and we shall have to be highly selective and taxonomic in filling in some greater substance under these headings. We hope only to be illustrative, and perhaps a little provocative.

The Demographic Base

Population Magnitudes and Geographic Distribution

With respect to the demographic base of American social structure, we hope to start from a constructive summary of historic trends in such magnitudes as total population, age-sex composition, total and differential fertility and mortality, and the regional and residential redistribution of the population. Despite the rather extensive, quantitative documentation of most of these trends, various critical questions remain moot: for example, trends in the social selectivity of migrants, that is, the attributes of migrants additional to their "race," age, and sex; also the analysis of new variables affecting fertility, which have destroyed the long-established patterns of differential fertility. The notable failures of past population predictions and the visible scars of those failures scarcely counsel an abdication of continuing attempts to improve our predictive understanding of demographic behavior, especially since these magnitudes affect almost every conceivable aspect of social functioning and welfare policy.

Major Structural Components

We now turn to the ways a modern society organizes its basic social functions.

Production of Goods and Services

Though other social scientists envy economists the excellence of their aggregative measures, and some envy the influence of these measurements on private and public policy, the envious ones also note that our institutional structure rather than the analysts produced the pricing mechanism as a way of adding up otherwise diverse quantities. This is not the place to enter into an extensive critique of economic measurement, but it should be noted that some of the monetary quantities are essentially arbitrary, and some other quantities--such as the equivalence of man-hour units in computing the Gross National Product--are simply wrong. What we should like to see, aside from a critical examination of existing indexes, is an analysis of "relative shares"in the production of goods and services, not only as between the private market and the state, but also among other non-governmental and non-market mechanisms of production and distribution--for example, family and kinship production, mutual aid, private charity and philanthropy.

Labor Force and Occupations

Labor force and occupational trends provide a principal way of linking economic production with other structural features of the society. Changes in age-sex participation rates reflect such other changes as the rising educational standards for labor force entry, the growth of public and private pensions that permit formal retirement, and the changing pattern of child-bearing in families, permitting labor-force reentry by mature women. Occupational structures reflect not only the major shifts between economic sectors (agriculture, industry, services), but also such trends as upgrading in terms of distributions by skill levels, continuing specialization--the fertility rate of new occupations--and bureaucratization, that is, the relative decline of self-employment. We think that considerable ingenuity is still needed in examining trends in these and other dimensions of economic activity, including the much-discussed but little-measured competition between men and machines.

Knowledge and Technology

In any society it cannot be assumed that knowledge is uniformly shared by the adult population, as technical and esoteric information may be in highly specialized custody. But especially in modern industrialized societies, there is a kind of "knowledge establishment" which functions to preserve, retrieve, and distribute varieties of specialized heritages and, even more importantly, to produce new knowledge. We should like to know how to define and measure the store of knowledge and its changes, at least by conventional divisions. We should also like to know the shifts among the agencies that perform research--the government, industry, universities, and foundations, and shifts among sources of support. We wish among other things to get beyond the rather silly notion that technology is a kind of prime mover to which other structures must adjust. On the whole, we get the technical change that we deserve or at least that we support, so that we are better off in national defense and a myriad of consumer gadgets than we are in potable water supplies and traffic safety.

Family and Kinship

One standard doctrine of structural change is that the colonial or pre-industrial American family represented a different kind of kinship system than that of the contemporary family: in particular that it was a "large-family system" in the sense of strong ties between generations and among collateral kinsmen. whereas the modern family is small in the special sense that it consists solely of parents and their immature children. (Neither of these conceptions has any direct connection with the number of children produced in a family.) Reappraisal of both sides of this contrast, the old large-family system and the contemporary small-family system, put it in considerable doubt. Quantitative historical materials may not permit us to get much beyond the composition of households as distinct from other types of kinship patterns, but even there cross-sectional differentials by region, occupation, and so on, at successive periods of time, should prove instructive. Similarly, trends in rates of marriage, separation, divorce, and illegitimacy can be taken as indicative of family functioning and malfunctioning. The internal structure of the family as an operating unit, including husband-wife relations and child-rearing practices, also no doubt differs cross-sectionally and through time, and measurement of these differences and changes will take ingenuity proportional to the importance of doing so.

Religion

For alleged reasons of national policy--reasons that are at least debatable--the American population is not officially enumerated in terms of religious affiliation, to say nothing of depth of religious conviction or forms and frequency of religious participation. We are thus not in a position to answer very effectively the changing position of religion in American society, or to appraise the truth or interpret the possible significance of the supposed upsurge of religion after World War II. Although empirical studies indicate the importance of religious differences in fertility behavior, we have little comparable knowledge about, say, religion as a factor in occupational choice or in political participation. Local studies have been made, but our focus is broader, and insistently temporal. Clearly, in terms of American institutional principles, trends in religion are not of direct concern to public policy as represented by government, though this line too is repeatedly blurred-- as in tax exemption and aid to education. Yet our proposals are not exclusively policy-oriented, and here especially is a place where new research under private auspices will be needed.

The Polity

All modern countries are "welfare states" in some form or degree, and we should like to see an appraisal of the absolute and relative importance of government in American society. Is it true that the Federal establishment has grown "at the expense of" state and local government, to say nothing of the market and other private mechanisms, or has it grown in addition to these others--that is, performing many services not previously performed at all? This kind of question cannot be answered by use of aggregated budgets and the proportions of the governmental sector in the Gross National Product. A much more analytical approach is required, for we must attend both to the changing size of the social universe as well as to the relative shares of different structures through time. In this regard, we feel the need for a reappraisal of the changing relationship of government to other social organizations. A simple confrontation between business and government, for example, was never exactly the situation, but now appears dangerously naive in view of restrictive, competitive, cooperative, and mutually dependent relationships that operate simultaneously.

How do we conceptualize and measure the functions of the state? Measures of "political effectiveness" are an inviting possibility. Voting behavior and other indexes of political participation are clearly relevant to the operation of democracy. Whether meaningful indexes of international effectiveness can be developed may be debatable, but presumably some such calculus enters into budgetary and other policies relating to diplomacy and defense. Crime and civic disorder presumably represent debits against internal political effectiveness, as perhaps does delay in the courts or the unequal administration of justice according to social position. Many of the appropriate statistical data are either poor in quality or non-existent, but this is the type of situation which calls for the sort of renewed effort and ingenuity that we hope to encourage.

Distributive Features of American Society

Our next major category relates to the way various benefits of economic, political, and voluntary activities are distributed through the population. Though the benefits are often allocated in a highly organized way, we believe there is a useful distinction between the major structural features, just discussed, and the distributive features which can be viewed as attributes or shares of individuals or families.

Consumption

It is of course too simple to say that the purpose of economic production is consumption, for this leaves in some doubt production for capital expansion to say nothing of national defense, maintaining order, or a host of other collective activities. Nevertheless, consumption levels clearly do represent an important index of economic effectiveness. Here we should continue and expand our examination of trends in income levels and income distribution, noting also changes in relative shares of various goods and services in the budgetary behavior of consumers. Changes in non-market sources of income such as charity, direct relief, public insurance, and pensions should be examined, along with attempts to develop minimum standards regardless of market contribution. Since that portion of income that is "discretionary" is not predictive of consumer behavior, an appropriate question is, what are the determinants of differentials and trends in "styles of life"? For example, what are the identifiable characteristics of families that accumulate goods and those who buy experiences, of those who save at a high rate for their own deferred consumption or that of their children or for establishing charitable trusts as compared with those who currently spend or overspend their income? The commerical utility of such predictors is apparent, but so is their analytic utility and possibly their utility for various public policies such as taxation rules.

Health

Good health is scarcely debatable as a welfare goal nearly universally shared, but its measurement is difficult in a prosperous society that enjoys high life expectancies and a low incidence of killing or crippling communicable diseases. Thus we believe it appropriate not only to examine trends in the traditional health indicators (life expectancies, infant mortality, incidence of morbidity and disability) but also to appraise new attempts to define health status in terms of role performance (can the breadwinner do his job, the child attend school and do his lessons, the housewife perform her tasks?) We should note that trends in mental health also merit examination, and here there is clearly a need for greatly increased sophistication in definition, identification, and measurement.

This is perhaps a suitable point for interjecting the reminder that we are concerned with the methodological problems of how data are produced as well as with the end products for purposes of analysis and policy. Health statistics, like crime statistics, are typically the product of large numbers of individual officials and local agencies. One way of achieving standardization is the sample survey, which for health status presumably encounters fewer causes for concealment than would a survey of the incidence of crimes or criminals.

Education

We have already referred to the "knowledge establishment," but here we are concerned with education in its distributive features. Trends in educational attainment, with suitable cross-sectional differentials, are relatively easy to pick up from census sources. But these data are relatively crude, for their use depends upon some notion of equivalence of "school years." Since schools are notably different in standards and individuals are at least equally different in performance, the school-year unit is not very homogeneous. Moreover, some individuals are genuinely self-educated, in whole or in part. Hence it has been suggested, here and there, that we should develop measures of educational status rather than school completion, measuring current level by some sort of achievement test. This sort of approach is interesting, even if providing no retrospective trends and thus only useful for replication in the future. But the conceptualization and tooling up for this kind of a survey should be approached with exquisite caution, as there may be little or no correlation between the human mind as a memory-storage-and-data-retrieval system and the human mind as a problem-solving mechanism.

There is another educational problem which calls for considerable statistical finesse. So far, no one has disentangled the relative historical importance of education as an investment for future economic growth and education as a consumer benefit deriving from economic prosperity. The advances in education, as standardly measured, are undoubtedly a combination of these and possibly other factors, but it would be analytically important to sort them out. The policy implications, particularly for newly developing countries, leap out at us.

Though we are dealing here with education in its distributive aspects, there is no need to be too precious about distinctions adopted for mere convenience. Accordingly, we suggest that education be viewed broadly from still another point of view, that of the apparently growing importance of adult retraining for persons faced with technical displacement, and of continuing education for technicians and professionals who are constantly faced with incipient obsolescence. Again, retrospective trend data may not be available on these changes, seemingly recent in their cogency, but there is every reason to expect that the phenomena and the problems will be of future significance. If we cannot reconstruct the past, we may be able to monitor the future.

Recreational and Expressive Activities

Attention to the businesses associated with recreation as constituting economic opportunities and concern for the constructive use of leisure offer reinforcing testimony to the apparently growing importance of leisure time and its discretionary use in modern American society. We say "apparently," for we must be cautious about what we are measuring or comparing. No known society, however impoverished, is without art forms, periodic festivals, and more frequent indulgence in fun and games. What we should like to know are the trends in both the magnitude and the form of recreational activities. If it was indeed early industrialism that introduced the practice of ceaseless toil, not even known in most slave systems, it would be interesting to trace the rise and fall of that grim condition of man.

The forms of recreational activities also warrant examination in terms of trends, though here one must tread cautiously indeed. On quick inspection, this is a very mixed bag of goodies: visiting, active or creative hobbies, reading, travel, active and passive sports and entertainment, and participation in varieties of voluntary associations. Cross-sectional data on time-budgets of behavior are just beginning to appear, and temporal trends are almost impossible to establish. Those who are involuntarily idle and those who can afford to be may have little in common, as do those who have minimal work commitments and those who are wedded to their jobs. But never before was the saying so true, or so possible, that "Man does not live by bread alone," and it would be foolish not to try to establish the changeful significance of that biblical aphorism.

Aggregative Features

We next turn to a partial attempt to characterize the over-all features of American society as a total operation that may distinguish it from others and certainly distinguish it from its own past. We have just been examining the prospects for getting distributive answers to the question, How are we doing? We now want to attend briefly to a pair of aggregative features of the society, and thus to get aggregative or collective answers to the same question.

Social Stratification and Mobility

The ideal of American society as an equal equalitarian system was never entertained seriously, if by equalitarian one means a social order without distinctions. Certainly notions of relative talent, merit, and "worth" have always prevailed, and no society is without such distinctions. Still, certain areas of absolute equality have ancient roots--for example, in the operation of civil and criminal law, and, later, in the franchise. Men more sophisticated than the Founding Fathers had the good sense to translate the equality of men into the equal opportunity of men--an ideal we have yet to attain.

What we should like to see is a thoroughgoing reappraisal of evidence concerning the relative degrees of absolute inequality (say, in property, power, and income) in American society through time. This is a view of stratification in one of its strict senses. We should also like to see an examination of relative degrees of rigidity and openness of sectors and strata according to merit. This is the converse of stratification in terms of impermeable boundaries. Thus trends in intergenerational and career mobility warrant analysis. Just as an example, it is almost certainly true that sociology and social history texts that emphasize the greater mobility opportunities in the nineteenth than in the twentieth century are plainly wrong; but it would be helpful to know more precisely not only the total probability distribution by type of origin but also the relative importance of various avenues or channels of status changes.

"Cultural" Homogeneity and Diversity

There is a class of social scientists and literary critics that bemoan the standardization and degradation of tastes in a "mass culture." Why standardization and degradation form an equation is not at all clear, but the more measurable component of the two, standardization, is also dubious. Clearly, some ideals and levels of economic consumption are becoming more standardized, as are, perhaps, rather ephemeral fads and fashions. Equally clearly, pluralism in political and religious conviction, in some elements of ethnic tradition, and simple eclecticism in preferences ranging from art forms to cuisine about in American society. The questions are: in which respects are we becoming homogeneous, standardized, and even "massified"--if that, doubtfully, is a word; and in what respects do we exhibit hardy survivals and new manifestations of pluralism? The answers may not have grand policy implications at the federal political level, but they would have some considerable analytic value.

The Meaning of Welfare

We return, finally, to an explicit consideration of deliberate social change, which we choose to limit to conceptions of welfare. This is the second sense of "monitoring" social change, which we identified in the beginning. (Welfare is of course also a consequence of unplanned change, and monitoring that is a major basis for deliberate intervention.)

Welfare and Its Measurement

An interesting historical essay could trace conceptions of welfare in various societies at different times. Such a general task would not be easy, for it would approximate the anthropologist's or sociologist's conceptions of comparative value systems. But for the Western world or American society one might set an aim at once more modest and more precise: how well were the distributive features of society previously noted performed, and by what mechanisms?

A consideration of welfare conceptions leads to certain distinctions and certain notions about measurement. If the avowedly highest goal of human welfare is the achievement of individual immortality, there is no pragmatic way in which we can measure the quality of performance. But we take a more mundane view, and assert that this view is never irrelevant in any substantial human aggregate: how are we doing, here and now?

We should still, however, think that in the measurement of welfare we must deal with several analytically distinct conceptions: (a) collective welfare, such as national independence and defense, international power and influence, the preservation of cultural integrity, the preservation of valued political principles and forms such as democracy and an independent judiciary, and the protection and increase of public wealth as represented in monuments, parks, and areas left in their natural state; (b) many of these welfare considerations also have a distinct dimension of concern for the future: the preservation of these values for one's progeny and for generations yet unborn; (c) distributive welfare, ranging from safety and the amenities of the work place through material well-being as provided by the market and its substitutes, to the preservation of pluralism, privacy, and individual liberties.

We think that measurements of these welfare functions are imaginable, starting with such simplicities as monetary values and per capita participation rates, and proceeding wherever analytic ingenuity leads us.

We did not say that the tasks of monitoring social change are simple. We did imply, and now say explicitly, that we hope to continue thinking about them and then set about doing them.

DISCUSSION

Ida C. Merriam, Social Security Administration

If I correctly understand the underlying thrust and purpose of the work on which the authors of this paper are engaged, it is an attempt first to find or to select organizing concepts around which can be developed measures of social change, and then to encourage analysis and research in the selected areas.

It is of interest--and depending on one's point of view, might be regarded either as encouraging or discouraging--that the concepts which the authors have found most useful are the conventional categories of demography, labor force, production, consumption, family and kinship, and so on. Within each of these areas, they have identified analytic tasks still to be done. With this there can be no quarrel, except on details, and it would be fruitless to pick out details for discussion from the very broad range of topics mentioned in the paper.

I would like to comment on two aspects of the paper. First, on what I believe is a suppressed goal. In spite of, or because of, their critical remarks about the economists aggregate indicators, I have a notion that what the authors would really like to do is to develop as useful an overall index as the GNP but one which would apply to the entire social order, including its social and political as well as its economic dimensions.

The goal is one which should not be lightly dismissed. A large part of the advance in the social sciences has come through improved techniques of indexing information. Kenneth Boulding 1/ has suggested that ". . . it is probably fundamental to all knowledge processes that we gain knowledge by the orderly loss of information. . . . Indexing is a process of filtering out irrelevant information . . ." in such a way as to make clear the broad outlines of highly complex social systems. The economists began developing price indexes in the last quarter of the 19th century; the concept of the GNP stems from the 30's. The technique is now beginning to be applied in other disciplines as well. Data collection is still basic and essential for social science research, but we may have reached a time when the important seminal work will center around use of existing data to create crude indexes--which can later be improved and provide the basis for more sophisticated data collection.

I find it difficult, however, to conceive of any single measure of social order that could sum up all the components of the social system. I think the authors are well advised to deal with segments of the social structure. I suspect that as they move on, however, they will focus more sharply on problems of measurement and of index construction within sectors.

My second comment relates to the brief section of the paper on the meaning of welfare. This is the one section where the idea of monitoring is stressed. The concept of welfare could be the beginning of a significant synthesis of separate measures and indexes. The distinction between collective welfare and distributive welfare is useful. I would question, however, the idea that welfare as here defined can be regarded as the result of deliberate social change. Some welfare activities are, of course, deliberately decided on through the political process. Many others represent the consequences of structural change in the economic and social order not directed to this end.

If I may, I will illustrate this point by reference to some of our own recent work. Mr. Stephan suggested that I describe briefly some of the current activities of the Office of Research and Statistics in the field of social statistics. The Social Security Administration started many years ago to develop a series of trend indicators based on aggregate data. For example, estimates of the total income loss from sickness or in another series the total expenditures for health and medical care, and the percent covered by private insurance or public programs. The most inclusive of these macro-indicators is the social welfare expenditure series in which we bring together estimates for the public and private sector and relate them to trend data for the GNP, total government spending and other measures.

Last year, we developed a new indicator which measures distributive rather than aggregative performance. I am referring to the SSA poverty index, which has attracted some attention and has now been adopted by the Office of Economic Opportunity for use until such time as a more refined index--on which we are working--can be developed. I will not describe the index, which has been fully explained in two recent articles in the Social Security Bulletin. 2/ Its major feature is that it establishes different poverty cut-off points for families of differing size and composition, based on a concept of equivalent levels of living.

Now for my illustration of the importance of generalized and nonspecific welfare trends. Using our poverty index as a classifier for income data from the CPS, there were 35.3 million persons or 19 percent of the population, living in poverty in 1963. In 1964, before specific new anti-poverty measures got under way, the number had dropped to 34.1 million or 18 percent of the population. This happy result of general economic growth was, of course, differentially distributed among various subgroups in the population. Next year an effort will have to be made to disentangle the effects of general conditions from the effects of the war on poverty. This will not be easy, but at least we have some basis for measurement.

If as social scientists we are really to monitor social change, we will have to give increasing attention to the development of indexes that are not only descriptive but also predictive. The beginning research program which Wilbert Moore has described is surely to be welcomed and encouraged for its potential contributions to this end.

1/ In The Meaning of the 20th Century, Harper and Row, 1964.

2/ See Orshansky, M., "Counting the Poor" and "Who's Who Among the Poor," in the Social Security Bulletin, January and July 1965.

MONITORING SOCIAL CHANGE

Discussion: Robert B. Voight, Bureau of the Census

I consider this paper a most timely assessment of the many issues with which social scientists, in the broadest sense of the term, should concern themselves in the present atmosphere of interest in social issues. Not since the early 1930s, perhaps, has there been the need for as well as the opportunity to monitor social change.

The authors have identified the areas of most pressing concern at a point where technological advances in methods of communication, and the statistical manipulation and display of the total inventory of data permit an open dialogue between scientists in many disciplines. In fact the opportunities presented by data banks, computerized information systems, and the ability to collect and store information appear so glamorous at the moment that we are most susceptible to the ingenious possibilities of the systems and may easily lose sight of the end purposes to be served. Dr. Moore and Dr. Sheldon remind us of the real responsibilities which these conveniences make possible to achieve. A recent example of a well balanced approach to the collection, storage, and treatment of social data is the Inter-University Consortium for Political Behavior where some 50 major universities across the country have joined forces to exploit the advantages of a central information system.

In scanning the broad spectrum of social changes for which the authors point out the need for monitorship, I suggest there is a need for anticipating future potential development and ensuring that such phenomena be subjected to constructive continuing evaluation as early as possible. Effective monitorship, I believe, calls for the refinement of measures, the need to establish priorities among the various phenomena being observed to narrow the consideration of the investigators to the more crucial issues, and greater concern with the attitudes and intentions of people which induce social change. Here it would be appropriate to pay particular attention to the risk takers and innovators in our society for here is the opportunity to anticipate and forecast future change.

While the paper was not explicit in indicating the group or groups that logically should take the lead in monitoring social change, it would appear most logical that the direction and impetus come from those outside the Government. It would be unfortunate for those closely involved in the administration of programs directly effecting change to also act as monitors.

The authors are to be commended for issuing a provocative challenge to the social science disciplines and I urge that the topic be considered for further discussion at next year's meetings of this association.

VII

STATISTICAL COVERAGE IN THE SOCIAL SCIENCE ENCYCLOPEDIA

Chairman, W. ALLEN WALLIS, University of Rochester

BAYESIAN INFERENCE IN THE INTERNATIONAL ENCYCLOPEDIA OF THE SOCIAL SCIENCES (Abstract)

Harry V. Roberts, University of Chicago

While the origins of Bayesian inference are old, the approach is still novel and controversial. Theoretical developments are rapid. The ultimate relationship to the "sampling theory" approach is not yet clear. For these reasons, the problems of what to say in an encyclopedia article on Bayesian inference overshadow the problems of how to say it. But two offsets have greatly eased the difficulties of writing the article.

First, at least in published literature, the usual misunderstandings of scientific controversy have scarcely arisen. It seems to be widely recognized that the major practical difference between Bayesian inference and sampling theory is the extent to which judgment is introduced into formal stage of a statistical analysis.

Second, research in Bayesian methods has brought out many interesting connections between Bayesian theory and sampling theory. Most, if not all, sampling theory techniques seem to have useful Bayesian interpretations. When prior distributions are "informationless," Bayesian inferences often agree, or nearly agree, with sampling-theory inferences; when this agreement does not occur, the existence of disharmony can be a stimulus to further theoretical development, especially since the Bayesian procedures often have desirable sampling properties.

STATISTICS IN THE INTERNATIONAL ENCYCLOPEDIA OF THE SOCIAL SCIENCES

William Kruskal, University of Chicago

A number of interesting problems have been faced in planning the sixty-odd statistical articles for the forthcoming International Encyclopedia of the Social Sciences. An encyclopedia has many functions, and users approach it in many ways. Should articles be primarily essays about parts of statistics, without attempt at full coverage? Should articles be handbook-like terse summaries? How should controversial areas, and widely differing viewpoints towards statistics be treated? What about level of exposition? Following discussion of such issues, brief descriptions were given of the intensive work on accuracy, clarity of exposition, etc., and of the structure of the statistical topics to be treated.

PRELIMINARY LIST OF STATISTICS AND STATISTICS-RELATED ARTICLES IN THE INTERNATIONAL ENCYCLOPEDIA OF THE SOCIAL SCIENCES

Topical Articles

BAYESIAN INFERENCE (Harry V. Roberts)
CAUSATION (Herbert A. Simon)
CENSUS (Conrad Taeuber)
CLUSTERING (David Wallace)
COHORT ANALYSIS (Norman B. Ryder)
COMPUTATION (Robert L. Ashenhurst)
CONCENTRATION CURVES (P. C. Mahalanobis)
CONTENT ANALYSIS (Matilda W. Riley & Clarice S. Stoll)
COUNTED DATA (Douglas G. Chapman)
CROSS-SECTION ANALYSIS (Yair Mundlak)
CYBERNETICS (M. E. Maron)
DECISION THEORY (Herman Chernoff)
DISTRIBUTIONS, STATISTICAL
- I Special discrete distributions (Frank A. Haight)
- II Special continuous distributions (Donald B. Owen)
- III Approximations to distributions (N. L. Johnson)
- IV Mixtures of distributions (Wallace R. Blischke)

ERRORS
- I Nonsampling errors (Frederick Mosteller)
- II Effects of errors in statistical assumptions (Robert M. Elashoff)

ESTIMATION
- I Point estimation (D. L. Burkholder)
- II Confidence intervals and regions (J. Pfanzagl)

EVIDENCE, STATISTICS AS LEGAL (Hans Zeisel)
EXPERIMENTAL DESIGN
- I The design of experiments (William G. Cochran)
- II Response surfaces (George E. P. Box)
- III Quasi-experimental design (Donald T. Campbell)
- IV Social experiments (Fred L. Strodtbeck)

EXPLANATION (Sidney Morgenbesser)
FACTOR ANALYSIS
- I Statistical aspects (A. E. Maxwell)
- II Psychological applications (Lloyd G. Humphreys)

FALLACIES, STATISTICAL (Irving John Good)
FIDUCIAL INFERENCE (R. A. Fisher)
GEOGRAPHY: Statistical Geography (Brian J. L. Berry)
GOODNESS OF FIT (Herbert T. David)
GOVERNMENT STATISTICS (Nathan Keyfitz)
GRAPHICS (Calvin F. Schmid)
HYPOTHESIS TESTING (E. L. Lehmann)
IDENTIFIABILITY, STATISTICAL (Olav Reiersøl)
INDEX NUMBERS
- I Theoretical aspects (Erik Ruist)
- II Practical applications (Ethel D. Hoover)
- III Sampling (Philip J. McCarthy)

LATENT STRUCTURE (Albert Madansky)
LIFE TABLES (Mortimer Spiegelman)
LIKELIHOOD (Allan Birnbaum)
LINEAR HYPOTHESES
- I Regression (E. J. Williams)
- II Analysis of variance (Julian C. Stanley)
- III Multiple comparisons (Peter Nemenyi)

MARKOV CHAINS (Patrick Billingsley)
MATHEMATICS (R. Duncan Luce & Patrick Suppes)
MULTIVARIATE ANALYSIS
- I Overview (Ralph A. Bradley)
- II Correlation (Harold Hotelling)
- III Discriminant functions (T. W. Anderson)

Biographical Articles

Lexis, Wilhelm, 1837-1914 (R. K. Bauer)
Lotka, Alfred J., 1880-1949
(Joseph J. Spengler)
Pearson, Karl, 1857-1936 (Helen M. Walker)
Petty, William, 1623-1687
(Phyllis Deane)
Poisson, Siméon Denis, 1781-1840
(Robert Féron)
Quetelet, Adolphe, 1796-1874
(Paul F. Lazarsfeld)
Spearman, C. E., 1863-1945 (Raymond B. Cattell)
Stouffer, Samuel A., 1900-1960
(M. Brewster Smith)
Süssmilch, Johann Peter, 1707-1767
(I. Esenwein-Rothe)
Von Mises, Richard, 1883-1953
(Hilda Geiringer)
Von Neumann, John, 1903-1957
(Oskar Morgenstern)
Wald, Abraham, 1902-1950
(Harold A. Freeman)
Wiener, Norbert, 1894-1964 (M. E. Maron)
Wilks, S. S., 1907-1964
(Frederick Mosteller)
Willcox, Walter F., 1861-1964
(Frank Notestein)
Yule, G. Udny, 1871-1951
(M. G. Kendall)

QUASI-EXPERIMENTAL DESIGN[1]

Donald T. Campbell

Northwestern University

This phrase refers to the application of an experimental mode of analysis and interpretation to bodies of data not meeting the full requirements of experimental control because experimental units are not assigned at random to at least two "treatment" conditions. The settings to which it is appropriate are those of experimentation in social settings, including planned interventions such as specific communications, persuasive efforts, changes in conditions and policies, efforts at social remediation, etc. Unplanned conditions and events may also be analyzed in this way where an exogeneous variable has such discreteness and abruptness as to make appropriate its consideration as an experimental treatment applied at a specific point in time to a specific population. When properly done, when attention is given to the specific implications of the specific weaknesses of the design in question, quasi-experimental analysis can provide a valuable extension of the experimental method.

While efforts to interpret field data as experiments go back much farther, the first prominent methodology of this kind in the social sciences was Chapin's Ex Post Facto Experiment (Chapin & Queen, 1937; Chapin, 1955; Greenwood, 1945), although it should be noted that due to the failure to control regression artifacts, this mode of analysis is no longer regarded as acceptable. The American Soldier volumes (Stouffer et al., 1949) provide prominent analyses of the effects of specific military experiences, where it is implausible that differences in selection explain the results. Thorndike's efforts to demonstrate the effects of specific course work upon other intellectual achievements provide an excellent early model (e.g., Thorndike & Woodworth, 1901; Thorndike & Ruger, 1923). Extensive analysis and review of this literature are provided elsewhere (Campbell, 1957; 1963; Campbell & Stanley, 1963) and serve as the basis for the present abbreviated presentation.

The core requirement of a "true" experiment lies in the experimenter's ability to apply at least two experimental treatments in complete independence of the prior states of the materials (persons, etc.) under study. This independence makes resulting differences interpretable as effects of the differences in treatment. In the social sciences this independence of prior status is assured by randomization in assignments to treatments. Experiments meeting these requirements, and thus representing "true" experiments, are much more possible in the social sciences than is generally realized. Wherever, for example, the treatments can be applied to individuals or small units (such as precincts or classrooms) without the respondents' being aware of experimentation or that other units are getting different treatments, very elegant experimental control can be achieved. An increased acceptance by administrators of randomization as the democratic method of allocating scarce resources (be these new housing, therapy, or fellowships) will make possible field experimentation in many settings. Where innovations are to be introduced throughout a social system, and where the introduction cannot in any event be simultaneous, a use of randomization in the staging can provide an experimental comparison of the new and the old, using the groups receiving the delayed introduction as controls. Nothing in this article should be interpreted as minimizing the importance of increasing the use of true experimentation. However, where true experimental design with random assignment of persons to treatments is not possible, due to ethical considerations, lack of power, or in feasibility, application of quasi-experimental analysis has much to offer.

The social sciences must do the best they can with the possibilities open to them. Inferences must frequently be made from data lacking complete control. Too often a scientist trained in experimental method rejects out of hand any research in which complete control is lacking. Yet in practice no experiment is perfectly executed, and the practicing scientist overlooks those imperfections which seem to him to offer no plausible rival explanation of the results. In the light of modern philosophies of science, no experiment ever proves a theory, it merely probes it. Seeming proof results from that condition in which there is no available plausible rival hypothesis to explain the data. The general program of quasi-experimental analysis is to specify and examine those plausible rival explanations of the results which are provided by the uncontrolled variables. A failure of control which does not in fact provide a plausible rival interpretation is not regarded as invalidating.

It is well to remember that we do make assured causal inferences in many settings not involving randomization: (The earthquake caused the brick building to crumble; the automobile crashing into it caused the telephone pole to break; the language patterns of the older models

[1]The preparation of this review was supported in part by Project C-998, Contract 3-20-001, with the Media Research Branch, Office of Education, U.S. Department of Health, Education, and Welfare, under provisions of Title VII of the National Defense Education Act. This symposium presentation is essentially the same as the current draft of my article for the International Encyclopedia of the Social Sciences.

and mentors caused this child to speak English rather than Kwakiutl; etc.) While these are all potentially erroneous inferences, they are of the same type as experimental inferences. We are confident that were we to intrude experimentally, we could confirm the causal laws involved. Yet they have been made assuredly by a nonexperimenting observer. This assurance is due to the effective absence of other plausible causes. Consider the inference as to crashing auto and the telephone pole: we rule out combinations of termites and wind because the other implications of these theories (e.g., termite tunnels and debris in the wood, wind records at nearby weather stations) do not occur. Spontaneous splintering of the pole by happenstance coincident with the auto's onset does not impress us as a rival, nor would it explain the damage to the car, etc. Analogously in quasi-experimental analysis, tentative causal interpretation of data may be made where the interpretation in question squares with the data and where other rival interpretations have been rendered implausible.

For the evaluation of data series as quasi-experiments, a set of twelve frequent threats to validity have been developed. These may be regarded as the important classes of frequently plausible rival hypotheses which good research design seeks to rule out. All will be presented briefly even though not all are employed in the evaluation of the designs used illustratively here.

Fundamental to this listing is a distinction between internal validity and external validity. Internal validity is the basic minimum without which any experiment is uninterpretable: did in fact the experimental treatments make a difference in this specific experimental instance? External validity asks the question of generalizability: to what populations, settings, treatment variables, and measurement variables can this effect be generalized? Both types of criteria are obviously important, even though they are frequently at odds, in that features increasing one may jeopardize the other. While internal validity is the sine qua non, and while the question of external validity, like the question of inductive inference, is never completely answerable, the selection of designs strong in both types of validity is obviously our ideal.

Relevant to internal validity are eight different classes of extraneous variables which, if not controlled in the experimental design, might produce effects mistaken for the effect of the experimental treatment. These are: 1. History: the other specific events occurring between a first and second measurement in addition to the experimental variable. 2. Maturation: processes within the respondents operating as a function of the passage of time per se (not specific to the particular events), including growing older, growing hungrier, growing tireder, and the like. 3. Testing: the effects of taking a test upon the scores of a second testing. 4. Instrumentation: in which changes in the calibration of a measuring instrument or changes in the observers or scorers used may produce changes in the obtained measurements. 5. Statistical regression: operating where groups have been selected on the basis of their extreme scores. 6. Selection: biases resulting in differential recruitment of respondents for the comparison groups. 7. Experimental mortality: the differential loss of respondents from the comparison groups. 8. Selection-maturation interaction: In certain of the multiple-group quasi-experimental designs, such as the nonequivalent control group design, such interaction is confounded with, i.e., might be mistaken for, the effect of the experimental variable.

Factors jeopardizing external validity or representativeness are: 9. The reactive or interaction effect of testing, in which a pretest might increase or decrease the respondent's sensitivity or responsiveness to the experimental variable and thus make the results obtained for a pretested population unrepresentative of the effects of the experimental variable for the unpretested universe from which the experimental respondents were selected. 10. Interaction effects between selection bias and the experimental variable. 11. Reactive effects of experimental arrangements, which would preclude generalization about the effect of the experimental variable for persons being exposed to it in nonexperimental settings. 12. Multiple-treatment inference, a problem wherever multiple treatments are applied to the same respondents, and a particular problem for one-group designs involving equivalent time-samples or equivalent materials samples.

Perhaps the simplest quasi-experimental design is the One-Group Pretest-Posttest Design, 0_1 X 0_2 (where 0 represents measurement or observation, and X represents the experimental treatment). This common design patently leaves uncontrolled the internal validity threats of History, Maturation, Testing, Instrumentation, and, if selected as extreme on 0_1, Regression. There may be situations in which the analyst could decide that none of these represented plausible rival hypotheses in his setting: A log of other possible change-agents might provide no plausible ones, the measurement in question might be nonreactive (Campbell, 1957), the time span too short for maturation, too spaced for fatigue, etc. However, the sources of invalidity are so numerous that a more powerful quasi-experimental design would be preferred. Several of these can be constructed by adding features to this simple one. The Interrupted Time-Series Experiment utilizes a series of measurements providing multiple pretests and posttests, e.g.: 0_1 0_2 0_3 0_4 X 0_5 0_6 0_7 0_8. If in this series, $0_4 - 0_5$ shows a rise greater than found elsewhere, then Maturation, Testing, and Regression are no longer plausible, in that they would predict equal or greater rises for $0_1 - 0_2$, etc. Instrumentation may well be controlled too, although in institutional settings a change of administration policy is often accompanied by a change in record-keeping standards. Observers and participants may be focused on the occurrence of X, and may fail to take into consideration changes in rating standards, etc. History remains the major threat, although in many settings it would not offer a plausible rival interpretation. If one had available a parallel time series from a group not receiving the experimental treatment, but exposed to the same extraneous sources of influence, and if this control time series failed to show the exceptional jump from $\underline{0}_4$ to $\underline{0}_5$, then the plausibility of History as a rival interpretation would be greatly reduced. We may call this the Multiple Time-Series Design.

Another way of improving the One-Group Pretest-Posttest Design is to add a "Nonequivalent Control Group." (Were the control group to be randomly assigned from the same population as the experimental group, we would, of course, have a true, not quasi, experimental design.) Depending on the similarities of setting and attributes, if the nonequivalent control group fails to show a gain manifest in the experimental group, then History, Maturation, Testing, and Instrumentation are controlled. In this popular design, the frequent effort to "correct" for the lack of perfect equivalence by matching on pretest scores is absolutely wrong (e.g., Thorndike, 1942; Hovland et al., 1949; Campbell & Clayton, 1961), as it introduces a regression artifact. Instead, one should live with any initial pretest differences, using analysis of covariance, or graphic presentation. Remaining uncontrolled is the Selection-Maturation Interaction, i.e., the possibility that the experimental group differed from the control group not only in initial level, but also in its autonomous maturation rate. In experiments on psychotherapy and on the effects of specific coursework this is a very serious rival. Note that it can be rendered implausible by use of a time series of pretests for both groups, thus moving again to the Multiple Time-Series Design.

There is not space here to present adequately even these four quasi-experimental designs, but perhaps the strategy of adding specific observations and analyses to check on specific threats to validity has been illustrated. This is carried to an extreme in the Recurrent Institutional Cycle Design (Campbell & McCormack, 1957; Campbell & Stanley, 1963), in which longitudinal and cross-sectional measurements are combined with still other analyses to assess the impact of indoctrination procedures, etc., through exploiting the fact that essentially similar treatments are being given to new entrants year after year or cycle after cycle. Other quasi-experimental designs covered in Campbell & Stanley (1963) include two more single-group designs (the Equivalent Time-Samples Design and the Equivalent Materials Design), Counterbalanced or Rotational Designs, Separate Sample Pretest-Posttest Designs, Regression-Discontinuity Analysis, the Panel Impact Design (see also Campbell & Clayton, 1961), and the Cross-Lagged Panel Correlation, which is related to Lazarsfeld's Sixteen-Fold Table (see especially Campbell, 1963).

Related to the program of quasi-experimental analysis are those efforts to achieve causal inference from correlational data. Note that while correlation does not prove causation, most causal hypotheses imply specific correlations, and thus examination of these probes, tests, or edits the causal hypothesis. Further, as Simon and Blalock have emphasized (e.g., Blalock, 1961), certain causal models specify uneven patterns of correlation. Thus the $A \rightarrow B \rightarrow C$ model implies that r_{AC} be smaller than r_{AB} or r_{BC}. However, the use of partial correlations or the use of Wright's (1920) path analysis are rejected by the present writer as tests of the model because of the requirement that the "cause" be totally represented in the "effect." In the social sciences it will never be plausible that the "cause" has been measured without unique error and that it also totally lacks unique systematic variance not shared with the "effect." More appropriate would be Lawley's (1940) test of the hypothesis of single-factoredness. Only if single-factoredness can be rejected would the causal model as represented by its predicted uneven correlations pattern be the preferred interpretation.

A word needs to be said about tests of significance for quasi-experimental designs. There has come from several competent social scientists the argument that since randomization has not been used, tests of significance assuming randomization are not relevant. The attitude of the present writer is on the whole in disagreement. However, some aspects of the protest are endorsed: Good experimental design is needed for any comparison inferring change, whether or not tests of significance are used, even if only photographs, graphs, or essays are being compared. In this sense, experimental design is independent of tests of significance. More importantly, tests of significance have come to be taken as thoroughgoing proof. In vulgar social science usage, finding a "significant difference" is apt to be taken as proving the author's basis for predicting the difference, forgetting the many other plausible rival hypotheses explaining a significant difference which quasi-experimental designs leave uncontrolled. Certainly the valuation of tests of significance in some quarters needs demoting. Further, the use of tests of significance designed for the evaluation of a single comparison becomes much too lenient when dozens, hundreds, or thousands of comparisons have been sifted, and this is still common usage. And in a similar manner, the author's decision as to which of his studies is publishable, and the editor's decision as to which of the manuscripts is acceptable, further biases the sampling basis. In all of these ways, reform is needed.

However, when a quasi-experimenter has compared the results from two intact classrooms employed in a sampling of convenience, sample size, small-sample instability, a chance difference, is certainly one of the many plausible rival hypotheses which must be considered, even if only one. If each class had but five students we would interpret the fact that 20% more in the experimental class showed increases in favorable-ness with much less interest than if each class had 500 students. In this case there is available an elaborate formal theory for the plausible rival hypothesis of chance fluctuation. This theory involves assumptions of randomness, which are quite appropriately present when we reject the null model of random association in favor of a hypothesis of systematic difference between the two classes. If we find a "significant difference," the test of significance will not, of course, tell us whether the two classes differed because one saw the experimental movie, or for some selection reason associated with class topic, time of day, etc., which might have interacted with rate of autonomous change, pretest instigated changes, reactions to commonly experienced events, etc. But such a test of significance will help us rule out this 13th plausible rival hypothesis, that there is no difference here at all that a model of purely chance assignment could not account for as a vagary of sampling. Note that our statement of probability level is in this light a statement of the plausibility of this rival hypothesis, which always has some plausibility, however faint. In this orientation, a practice of stating the probability in descriptive detail seems preferable to using but a single apriori decision criterion.

REFERENCES

Blalock, H.M. 1964 Causal inferences in nonexperimental research. Chapel Hill: The University of North Carolina Press, 1964.

Campbell, D.T. 1957 Factors relevant to validity of experiments in social settings. Psychological Bulletin 54:297-312.

Campbell, D.T. 1963 From description to experimentation: Interpreting trends as quasi-experiments. Pages 212-242 in Harris, C.W. (editor), Problems in measuring change. Madison, Wis.: University of Wisconsin Press.

Campbell, D.T.; and Clayton, K.N. 1961 Avoiding regression effects in panel studies of communication impact. Studies in Public Communication No. 3, 99-118.

Campbell, D.T.; and McCormack, Thelma H. 1957 Military experience and attitudes toward authority. American Journal of Sociology 62:482-490.

Campbell, D.T.; and Stanley, J.C. 1963 Experimental and quasi-experimental designs for research on teaching. Pages 171-246 in Gage, N.L. (editor), Handbook of research on teaching. Chicago: Rand McNally.

Chapin, F.S. 1955 Experimental designs in sociological research. New York: Harper. (Rev. ed.)

Chapin, F.S.; and Queen, S.A. 1937 Research memorandum on social work in the depression. New York: Social Science Research Council, Bulletin 39, 1937.

Greenwood, E. 1945 Experimental sociology: A study in method. New York: King's Crown Press.

Hovland, C.I.; Lumsdaine, A.A.; and Sheffield, F.C. 1949 Experiments on mass communication. Princeton, N.J.: Princeton University Press.

Lawley, C.N. 1940 The estimation of factor loadings by the method of maximum likelihood. Proceedings of the Royal Society of Edinburgh 60:64-82.

Stouffer, S.S. (editor) 1949 The American Soldier. Princeton, N.J.: Princeton University Press. Vols. I and II.

Thorndike, R.L. 1942 Regression fallacies in the matched groups experiment. Psychometrika 7:85-102.

Thorndike, E.L.; and Ruger, G.J. 1923 The effect of first-year Latin upon knowledge of English words of Latin derivation. School and Society 81:260-270, 417-418.

Thorndike, E.L.; and Woodworth, R.S. 1901 The influence of improvement in one mental function upon the efficiency of other functions. Psychological Review 8:247-261, 384-395, 553-564.

Wright, S. 1920 Correlation and causation. Journal of Agricultural Research 20:557-585.

ON WRITING STATISTICAL ARTICLES FOR THE INTERNATIONAL ENCYCLOPEDIA OF THE SOCIAL SCIENCES

F. J. Anscombe, Yale University

1. Statistics was originally a social science. Today the statistical method finds many applications in the social sciences. (For example, a college senior undertaking a research project as part of an honors program is, I believe, much more likely to find himself unexpectedly faced with a statistical problem if his field is one of the social sciences than if it is in the natural sciences or humanities.) But most of the impetus for development of statistical methodology in this century seems to have come from outside of the social sciences. Statistical methodology is basically not oriented towards any particular subject matter; it is as neutral as mathematics. Although most of the methods and concepts are fully as appropriate in the social sciences as anywhere else, the examples of statistical phenomena that come to mind most readily are not (for me) drawn from the social sciences. I have found it a bit embarrassing, in writing for the Encyclopedia, to try to impart a social science flavor, and no doubt some other contributors have had the same experience. When working on an article on outliers, I found that all the good examples I could recall concerning outliers were from the natural sciences. The last example of an outlier phenomenon that I have seen in a social science (it was in political science) was remarkable because nearly every observation seemed to be an outlier! All this is a matter of the accidents of one's personal experience.

2. What should the articles on statistics be like? Presumably they are addressed, not to professional statisticians, but primarily to social scientists who already have some acquaintance with statistics. There is surely no point in trying to do what is already done well in many textbooks, namely (i) explain how to make statistical calculations (of correlation coefficients, standard errors, analyses of variance, factor analyses, etc.) and (ii) give the mathematical theory underlying these methods. There is probably little point in trying to interest novices; it would be hard to compete with, say, Wallis and Roberts's paperback "The Nature of Statistics". Our best target would presumably be to try to do what is not well done in most books -- to address an adult reader, impart wisdom and insight, try to bridge the gap between the too glib textbooks and reality. Some target, yes!

3. Harry Roberts's very interesting talk raises the problem of controversy within statistical science. How much should appear in material addressed to non-professional statisticians and general readers? Surely it is unwise to suppress controversy so that a united front should be falsely presented to the outside world. But there are more appropriate places than the Encyclopedia for an author to participate actively in such controversy. We do not expect a physician to be as argumentative with his patients about fundamentals of medicine as he might be with some of his colleagues. Most of the controversies in statistics seem to have had a by-product that is both valuable and not in itself controversial, namely, an increased understanding of the diversity of statistical problems and of the many factors that enter into them. Ideally, the Encyclopedia articles on statistics should not hide the existence of controversy, but as far as possible divert attention to this by-product rather than attempt to persuade the reader to any one side. For example, the controversy round about 1950 over a randomized test for association in contingency tables brought out a distinction between decision making and inference, and that distinction can be expressed in dispassionate terms unlikely to raise anyone's ire. The present controversy over Bayesian methods has made us think more about the various kinds of uncertainty and vagueness in statistical problems, and these can usefully be described and discussed without partisanship.

4. Whether there is any point in making the above remarks now, I do not know. They would have been more appropriate several years ago in a discussion of policy, instead of now when most of us contributors have done our bit. But only now, having blundered through an assignment, can I (for one) see such policy questions clearly enough to venture an opinion on them.

VIII

SOME STATISTICS ON THE OLDER POPULATION

Chairman, MARGARET G. REID, University of Chicago

PREVIEW OF THE SSA's STATISTICAL PROGRAM ON HEALTH CARE OF THE AGED

Ida C. Merriam, Social Security Administration

Both statisticians and health care specialists can look forward with anticipation to the fund of data that will come from the new program of health insurance for the aged established by the 1965 Amendments to the Social Security Act. For many years, we have used what scraps of information were available to estimate the medical care needs, the hospital utilization, the medical expenditures and the aggregate resources devoted to health care of the aged population. By this time next year, we shall have not only a surer basis for estimate and a new situation that will make us struggle to interpret trend lines, but also the beginnings of an unprecedented volume and kind of information on the patterns of health care of aged individuals. I would stress the phrase patterns of health care, for the new program gives us a unique opportunity for measurement of individual by individual receipt of health services over a year or many years, from record data.

What I shall give you today is definitely a preview of the statistical program that will become possible as a result of HIB. It is a preview in that this is the first public summary of our research and statistical plans; a preview, also, in that we are in the early stages of planning. Furthermore, the SSA is still in the process of discussion and negotiation with potential administrative agents and carriers. The system or systems through which data will be generated are not yet fixed. We do, however, have sufficiently clear notions as to the kinds of data that will become available to sketch in a general picture.

First it might be well for me to describe briefly the major features of the health insurance program. As I am sure you all know, the so-called Medicare program establishes two related health insurance programs for aged persons. A basic plan (Part A) provides protection for all persons aged 65 and over against the costs of inpatient hospital care, post-hospital extended care services, post-hospital home health services, and outpatient hospital diagnostic services. These benefits are automatically available to OASDI and railroad retirement beneficiaries. The cost will be paid through a separate payroll tax and trust fund. Uninsured persons now aged 65 or over are also eligible for these benefits, with the cost paid from general revenues. An intensive effort is being made to register all such persons. This activity has already gotten under way and a major informational campaign will be carried on throughout the fall in an attempt to reach all old people.

I might say parenthetically that as statisticians we are looking forward with some curiosity to see what is the total count of aged under the program at the end of this campaign as compared with population estimates. We in SSA are also planning to use the opportunity to survey the characteristics of the uninsured group--particularly those under 70 or 75--to settle some questions about the nature of the gaps in OASDI coverage.

The second part of the health insurance program (Part B) is a voluntary plan providing payments for physicians and for other medical and health services not covered by the basic plan and financed through monthly premiums of $3 (until 1968, when the amount may go up) matched by an equal amount from the general revenues of the Federal Government. All aged persons are eligible to enroll in the plan and it is anticipated that between 80 and 95 percent will do so. The way was smoothed for social security beneficiaries by an increase in cash benefits amounting to at least $4 a month for a retired worker and $6 for a couple.

Benefits under both parts will become payable beginning July 1, 1966, except for post-hospital extended care benefits which start on January 1, 1967.

Not all medical services are covered under the program. Among the important omissions are drugs and dentistry. We shall be studying the effect of such omissions and the problems and costs involved in their coverage. But protection is also limited for types of services that are covered. The patient (or someone on his behalf) must pay the first $40 of hospital costs, and for days of care beyond 60 and up to the maximum of 90 in a spell of illness, he pays $10 a day. There is also a deductible of $20 for outpatient hospital diagnostic services received in each 20-day period--after which the program pays 80 percent of the charge. This deductible, however, counts as an expense under the supplementary medical insurance program, which pays 80 percent of the reasonable costs or charges for covered services above the first $50 in a calendar year. There is a lifetime limit of 190 days for inpatient psychiatric hospital services. And there is a limit of 100 days of post-hospital extended care services during any spell of illness, with the patient paying $5 a day after the 20th day.

If you are thinking that this a complicated program, you are right; but there is one more detail I want to add. A spell of illness, which sets the bounds on eligibility for hospital and post-hospital care, begins with the first day of hospitalization. It ends when the individual has been out of a hospital or extended care facility for 60 consecutive days. A person may be discharged and readmitted several times during a spell until he uses up his 90 days, but a new spell does not begin until he has been out for 60 days. The reason I mention this detail is that it means that for administrative reasons--to determine individual eligibility--we must know the total period of hospital or institutional care, not just the covered period. As statisticians we are going to make use of this circumstance--that is to say we plan to get information at the time of hospital discharge even though this comes well after 90 days.

In the case of skilled nursing home or other extended care facilities the problem of collecting meaningful data is a bit more complicated. The average stay in such institutions is about two years and it may well be that we should be satisfied with current reports on covered days only and get information on uncovered days by special sample studies or daily census type inquiries. These latter studies might be carried out by some group or groups other than SSA.

This may be a good point at which to note that the DHEW has announced the assignment of specific operating functions under the 1965 Amendments to constituent units. The SSA is responsible for the general management and operation of the two health insurance programs. The Welfare Administration is responsible for standards for the State programs of medical assistance and for administration of the new project grants for health programs for school and pre-school children that were established by the 1965 Amendments. Both will work closely with the Public Health Service which is assigned principal responsibility for the professional aspects of the hospital and medical insurance programs.

Under the law, the Secretary is required to use State public health agencies or other appropriate State agencies in determining which hospitals and other institutional providers of service meet the standards and conditions for participation. The State agencies will resurvey all providers under Part A periodically. They may also provide consultation to providers of service to help them meet established standards.

In the administration of Part A benefits, the bill provides that associations or groups of providers of service (hospitals, extended care facilities, and home health agencies) may nominate certain organizations, public or private, to serve as intermediaries between them and the Federal Government. Individual providers may, however, elect to deal directly with the Secretary. Under the supplemental voluntary insurance plan, public agencies and private organizations will also act as administrative agents, but they will not be selected by the nomination process. Instead, the Secretary is required, to the extent possible, to enter into agreements with interested and qualified agencies and organizations which he believes are capable of doing the job.

Hospitals and other institutional providers of service are to be paid on a reasonable cost basis. The formula for determining reasonable cost will be worked out in consultation with providers and the Health Insurance Benefits Advisory Council. Since this Council has not yet been appointed, you will understand why I cannot now tell you in any detail just what information relating to hospital costs will become available. I might observe that the first day deductible of \$40, like the premium for the Part B benefits, is fixed only through 1968. Thereafter the amount is determined each year by the Secretary in relation to changes in average per diem hospital costs. We shall certainly be watching closely the components of costs and variations among hospitals in this respect.

Payments to physicians will be made by carriers. The carrier is obligated to see that the charges of physicians (and other noninstitutional providers) are reasonable and not higher than charges for comparable services to the carrier's other policy holders or subscribers. In determining reasonable charges, the carriers will consider the customary charges for similar services generally made by the physician or other person or organization furnishing the covered services and also the prevailing charges in the locality for similar services. Payment by a carrier for physicians' services will be made on the basis of a receipted bill or an assignment under which the reasonable charge will be full charge for the service.

As I said at the outset, final decisions have not yet been reached as to the precise role to be played by the various administrative agents. It is clear, however, that there are strong administrative--as well as research and statistical--considerations pointing toward central record keeping. For Part A the most efficient system would appear to be one in which hospital admission and discharge reports for all aged persons (and similar reports from extended care facilities, hospital outpatient departments and home health agencies) flow directly to the SSA central record keeping and central computer facilities. No matter how frequently aged persons move or where they get their hospital care, it would thus be possible to maintain a current record of eligibility for additional services. For Part B benefits the problems are more complicated, but we are hopeful that information on covered services and payments can be incorporated in the master beneficiary utilization tape. Since the deductible for hospital outpatient services counts towards the deductible or coinsurance amounts under Part B, there is an administrative tie-in. More importantly, we are stressing the value for research purposes of the linkage of information on the hospital and medical services received by aged individuals. This value is recognized by those primarily concerned with the administration of the program, and by potential carriers and administrative agents with whom we have talked.

We anticipate, therefore, that we shall have the basis not only for cross-sectional analyses of hospital and health service utilization, but also for longitudinal studies of the patterns of covered services received by individuals from age 65 or the start of the program until death. In the case of hospital care, we should have almost complete reporting of all episodes. For the other services, we shall have in the records only those services for which the program pays at least in part. For example, there will be aged persons who spend less than the deductible for physicians or other services covered under Part B and probably others who spend more than $50 in the year but fail to realize that they could get reimbursement. Neither the carriers nor SSA will know about these services and expenditures. It may seem somewhat niggling to mention the gaps when the data potentially available are so much more extensive than we have ever had before. I do it to keep some balance.

In addition to utilization data, the basic statistics will include data on costs and on charges--total and covered--and of course on characteristics of all covered persons and providers. The potentialities for combining these several kinds of information open new vistas for analysis and research.

We are currently in the midst of specifying the essential statistical input of the system, developing detailed tabulation plans for data that should be currently available, and beginning to outline some of the special studies that should be undertaken at an early date. For this audience it might be of interest if I mention some of the statistical problems we are now debating.

What population base should we use to derive annual utilization rates? The usual procedure is to use mid-year population figures, but in this age group where the death rate is high the use of a mid-year figure will distort the rates. An alternative we are exploring is to add monthly eligible population data and divide by 12 to obtain average person-years of exposure. This method may have particular advantages when one is dealing with subgroups of changing composition, such as public assistance recipients, or the population of geographic areas.

We are still discussing the desirable geographic detail. We have also recognized that for some purposes we shall want data by residence of the beneficiary and for others by location of the provider. Utilization rates, which represent the probability of persons in a given area being hospitalized (or receiving some other service) should use the population resident in an area as the base, with a breakdown between those receiving services in and outside the area. Studies concerned with the adequacy of existing facilities or the organization of health care will call for counts of the characteristics, including usual place of residence, of all aged persons served in a specified time period.

One troublesome problem has to do with information as to race. For analysis of utilization and patterns of care this is an item of obvious importance. Unfortunately in the past few years, SSA records have developed rather sizable gaps as a result of the fact that such information was not obtained for persons assigned social security account numbers under the Internal Revenue Service registration project. In the years up to 1961, about one-half percent of the social security numbers issued did not have the race item filled in on the social security account number application form. From April 1962 through March 1964, 29 percent of the 14 million numbers issued had race unreported. As a result we lack this information for at least 1.2 million persons now aged 65 and over. The SSA does not want to ask a question as to race in connection with applications for benefits, since the purpose could easily be misunderstood. The special Internal Revenue Service form is no longer in use, and the SSA is again asking for and tightening up on procedures to get the information on new account number applications. But the problem of correcting for past omissions is more difficult and we shall probably be able to solve it only in part.

There are obviously a great variety of special studies that will become possible and desirable once the Medicare program has been in operation for a year or two. Indeed, I am sure, we will be trying to answer certain kinds of questions after the first months, or even weeks. There will be special analyses relating to the program itself--utilization experience of different groups, where people receive services, various aspects of costs and financing under the program. There will be important methodological studies both before and after the program goes into operation. We are, for example, presently looking into the form which a system of classification of medical procedures might take. We shall be concerned with such problems as improved methods of reporting of deaths, improved classification systems for services, alternative sampling and analytical techniques for longitudinal studies, studies of the lag between receipt of services and the receipt of the bill. There will be a whole range of studies of the impact of the program on hospitals and the organization of medical services, on medical manpower and medical prices, on voluntary insurance, on the remaining medical expenditures of aged persons and on their levels of living and--tentatively, since no one really knows how to measure this--on the quality of medical care.

We shall be giving particular attention to the studies called for by Congress, specifically studies and recommendations concerning the adequacy of existing health personnel and facilities, forms of health care alternative to inpatient hospital care, and the effects of the deductibles and coinsurance provisions upon beneficiaries, persons who provide health services, and the financing of the program.

Many of these studies will be carried out by the SSA. For others, particularly those relating to medical manpower and facilities or the organization of medical services, primary responsibility will rest with the PHS. Studies of the services received by public assistance recipients will be of joint concern to SSA and the WA. And we hope to encourage and support special studies by research groups outside the government. There are many aspects of the impact of the program that can best be evaluated in local communities, and many special problems or issues that should receive critical attention from statisticians and researchers in a variety of agencies.

What are we doing now to make possible valid before and after studies? We seriously considered the possibility of mounting a special survey of the medical expenditures of aged persons this coming January or February; but it was simply not possible to marshall the necessary manpower. We shall have to make do, therefore, with special tabulations and analyses of data collected in earlier surveys by the Health Information Foundation and by the SSA in its 1963 Survey of the Aged. One study we are planning involves tabulation of the first year's experience under the HIB program of the aged persons in our 1963 sample who survive through June 1967. Whether we will decide to resurvey this group to get additional current information on such matters as living arrangements, continuation of voluntary health insurance, etc., or to undertake a new cross-sectional sample survey after the HIB program has been in operation for a few years we have not yet decided.

There are, of course, other ongoing statistical series that will reflect the impact of the program. We have, for example, been talking with staff of the National Center for Health Statistics about the possibility of their expanding as early as next January their collection of information relating to health service utilization of aged persons, particularly physicians' services, and possibly also medical expenditures, in order to provide the basis for more adequate interpretation of later trend data.

As I warned you at the outset, this has necessarily been a very general preview of what statisticians can expect and look forward to having in the way of new data. In conclusion I would say simply that we who are directly involved are excited about the potentialities for statistical analysis and research in this new health insurance program. We are overwhelmed by the size of the job to be done between now and next July 1 and thereafter, but still hopeful that what proves feasible and realizable will not be too far from the ideal.

HEALTH CHARACTERISTICS OF PATIENTS IN NURSING AND PERSONAL CARE HOMES

E. Earl Bryant, National Center for Health Statistics

Arne B. Nelson, National Center for Health Statistics

Carl A. Taube, National Center for Health Statistics

Introduction

In 1961 the National Center for Health Statistics began planning a nationwide study of establishments which provide nursing or personal care to the aged and chronically ill, known as the "Resident Places Survey - 1 and 2." These surveys were the first of a series of ad hoc institutional population surveys to be conducted by the Center.

To make the study as meaningful as possible, a number of people in the Public Health Service who had experience in nursing home administration and statistics were asked to serve on a working group to make recommendations regarding solutions to several outstanding problems. These included the types of information needed, the scope of the survey, and a procedure for classifying the various types of establishments which are known as nursing homes, homes for the aged, etc. After several meetings of the group the following recommendations were made:

1. Because of the large volume and nature of the data needed, two surveys should be conducted. The first should be limited to the types of data that could be readily obtained by mail with reasonable reliability. This hopefully would include information about establishments such as their admission policies, size, etc., and certain personal and health characteristics of residents or patients. Information about health should be cast in very general terms similar to those used in a survey developed in 1953 by the Commission on Chronic Illness and the Public Health Service. 1/ The second survey should be conducted by personal visits to obtain information that could not be readily collected by mail. This would include more detailed information on the residents, including data on chronic conditions and impairments, and information on the characteristics of the employees.

2. The scope of the first survey should include not only nursing homes, homes for the aged, etc., but also mental, chronic disease and geriatric hospitals and long-term units of general hospitals caring for geriatric and chronic disease patients. With the exception of mental hospitals, these types of institutions and units serve predominatly the aged population. Mental hospitals should be included because of the presence of a sizable proportion of aged patients suspected to require primarily geriatric rather than psychiatric care.

3. Establishments should be classified on an a priori basis according to the type of service provided and the availability of nursing staff to provide care. This would impose a standard procedure of classification on the heterogeneous systems used by the State licensing and regulatory agencies.

All of these recommendations were adopted and both surveys have been conducted, the first in the spring of 1963 and the second in the spring of 1964. This paper, however, will be limited to a description of the design and methodology of the non-mental part of the first Resident Places Survey and some of its findings on the health and demographic characteristics of the residents in the nursing and personal care care homes.

Development of the Questionnaire and Procedures

The questionnaire was designed to be as self-explanatory as possible, consistent with the need to keep it simple. While general instructions were given, definitions and explanatory notes accompanied each question as necessary. The form was composed of three parts. Part I was concerned with certain establishment statistics such as admission policy, number of beds, residents, admissions, discharges, and charges for care of residents. Part II was used as a listing sheet to establish a sampling frame of residents and to record the date of admission, date of birth, race, and sex of each resident listed. For a systematic sample of the residents listed in Part II, health data were recorded in Part III. This sample was composed of residents whose names fell on predesignated lines of the questionnaire. Part III of the questionnaire contained 6 health related items, each subclassified into broad groups. For example, a person would be in one of the three Bed Status categories: "In bed hardly ever," "In bed part of the time," or "In bed all or most of the time." Other items pertained to walking, hearing, vision, continence, and mental status.

In developing the procedures and questionnaires, two pretests were conducted. The first involved the mailing of questionnaires to 38 homes in the Atlanta, Georgia, and Washington, D.C. metropolitan areas. The second pretest of 18 homes was conducted in Cincinnati, Ohio. Follow-up visits were made for each pretest to evaluate how well the respondents understood the questions, whether available records were consulted, and whether they were able to provide objective and consistent answers.

The pretest experience confirmed that much of the information about the establishments and residents was available either in records or could be reliably reported by the respondent. Prior to the pretests, there was skepticism as to whether health data could be reliably collected in a mail survey. It was almost certain that such information would not be consistently recorded in medical records. Thus, it would be necessary to rely heavily on some employee, such as a nurse or other responsible person, to know the facts through personal observation. In an attempt to evaluate the answers provided on the mail questionnaire regarding the patient's health, the original respondent was asked in the follow-up visit to answer these questions again for each sample resident without benefit of the answers provided a week or two previously. Upon comparing the two sets of information for more than 300 sample residents, it was found that consistent answers had been given for better than 95 percent of each of the health items. In several instances in which answers differed, a change in the patient's condition was stated to have taken place since completion of the mail questionnaire. It is, of course, not possible to determine by this type of check whether or not the information provided was valid. It is believed, however, for the type of data sought that there is a high correlation between reliability and validity; the respondent was usually in close contact with the residents and should have known whether they were confined to bed, their walking ability, etc.

Another question of concern in developing procedures for the survey was the upper size limit for inclusion of a home in the mail survey. For what size home would the amount of work required in listing the residents discourage respondents from participating in the survey? A number of very large nursing homes were visited in the pretest to discuss this question. As a result, it was decided that only homes with less than 300 beds would be in the mail survey; personal visits would be made to the larger homes to select a sample of residents and to help complete the questionnaire. The wisdom of this choice can be evaluated on the basis of the response and lack of complaints in the national survey. The response rate was the highest among the largest establishments surveyed by mail (i.e., 100-299 beds) and it took less follow-up effort to obtain a response than for the smaller homes.

The Sampling Frame

The sampling frame for the survey was the Master Facility Inventory (MFI). 2/ The MFI was developed by the National Center for Health Statistics by merging a number of listings of all types of hospitals and resident institutions in the United States. The most current lists of nursing and personal care homes used were those collected from State licensure agencies in 1961 by the Division of Hospital and Medical Facilities of the Public Health Service. To collect information needed for classifying establishments by type of service, size, and ownership, questionnaires were mailed to all places listed in the MFI. The coverage of the MFI was improved by the addition of places reported in the survey in answer to the question: "Does the owner of this establishment own or operate any other related or similar establishments which are not included in this report?" On the basis of preliminary research to evaluate coverage of the MFI, the sampling frame for RPS-1 is estimated to be about 85-90 percent complete in terms of places and 90-95 percent in terms of beds.

In addition to nursing and personal care homes, the scope of RPS-1 included mental hospitals, long-stay geriatric and chronic disease hospitals, and long-stay units of general hospitals which provided care to the aged and chronically ill. Excluded were homes with less than 3 beds, homes which did not routinely provide some level of nursing or personal care, i.e., provided room and board only, and homes providing care to children only.

The group of establishments in the sampling frame providing nursing or personal care were further classified into four subclasses which were defined as follows:

1. Nursing care home. An establishment which provided nursing care to more than half of its residents during the week prior to the MFI survey and which employed either a registered nurse or a licensed practical nurse 15 hours or more per week.

2. Personal-care-with-nursing home. An establishment which provided some nursing care but less than that provided by a nursing care home.

3. Personal care home. An establishment which did not provide nursing care, but routinely provided personal care.

4. Domiciliary care home. An establishment which routinely provided only minimal personal care.

Sample Design

The sampling for the Resident Places Survey-1 was based on a stratified multistage probability design. The establishments in the sampling frame were sorted into 16 primary strata consisting of four size groups subclassified into four type-of-service groups. Further stratification within each primary stratum was accomplished by sorting on geographic area and type of ownership. The first-stage sample was a systematic selection of establishments within each stratum. The sampling fractions varied by size strata from 1 in 15 for establishments with less than 30 beds to unity for establishments with 300 or more beds. The second-stage sample was a systematic selection of the residents or patients who were on the register of the sample establishments on the day that the questionnaire was completed. The second-stage sampling fraction was of such size to obtain a self-weighting sample of 1 in 15 residents.

Survey Procedure

The survey was completed by personal visit in the 134 places which maintained 300 or more beds.* One hundred percent response was obtained from these larger places. Questionnaires were sent by first class mail to 3042 sample places maintaining less than 300 beds. Three waves of follow-up were used to obtain the final mail survey response rate of 93 percent. Forty-two percent of the establishments replied to the initial mail inquiry. Two mail follow-ups, the first by regular mail three weeks after the initial mailing, and the second by certified mail 6 weeks after the initial mailing, raised the response rate to 66 percent and 84 percent respectively. The final follow-up was a combination of telephone reminders and personal visits, undertaken two weeks after the last mail follow-up. For establishments with 100 or more beds, appointments were made to complete the survey by personal visit. For places with less than 100 beds, a plea was made to the respondent by telephone to return the questionnaire. If the respondent indicated that he had some problem in completing the form, an offer was made to visit the place to aid in completing the questionnaire.

As mentioned previously, the larger homes in the mail survey, those with 100-299 beds, responded more readily than those with under 100 beds, even though the task of completing the questionnaire was more time consuming for these larger places. At the end of the mail follow-up, 90 percent of the establishments with 100 beds or more had responded as opposed 82 percent of those with less than 100 beds.

About two-thirds of the schedules returned by mail were acceptable without need for further query. The remaining third did not pass the editing criteria, and a fail edit query was mailed to obtain the missing information. A response was obtained from 82 percent of those queried.

* RPS-1 was a cooperative effort by the NCHS and the U.S. Bureau of the Census. The field operations as well as certain parts of data processing were carried out by Census personnel.

The need for fail edit follow-up is one indicator of the quality of the response. Used as such, it shows that in this survey the adequacy of response was highest for those places responding to the initial mailing. Seventy-three percent of these questionnaires were acceptable without further query. By comparison, only 62 percent of the questionnaires returned after mail follow-up passed the editing criteria. For questionnaires returned as a result of the final telephone and personal visit follow-up, the pass-edit rate increased to 69 percent, which undoubtedly was influenced by the personal visits. The completeness rate was very low, however, considering the fact that the interviewers were instructed to check the questionnaires for completeness before leaving the home. It is interesting to note that the larger homes, in addition to having a higher overall response rate, also seem to be better respondents in terms of the quality of response. For the homes with less than 30 beds, 41 percent of the returned questionnaires required fail edit queries, as opposed to around 30 percent for the homes with 30-299 beds.

Personal and Health Characteristics of Residents

In 1963, an estimated 505,000 persons were receiving some type of care as residents or patients in 16,370 nursing and personal care homes in the United States. Slightly more than half of the persons resided in nursing care homes; about a third were in personal-care-with-nursing homes, and a tenth were in homes which provided personal care but not nursing. The latter category includes establishments defined previously in this paper as "Domiciliary Care Homes" and "Personal Care Homes."

Most of the persons were in care homes because of advanced age and the various problems associated with aging. Nearly a third of the residents were 85 years of age or more. Their average age was 78 years. Although more than a third of the homes reported a policy of accepting adults of any age, only 2 percent were under 45, and 12 percent were under 65. There was little variation in the age pattern by type of home; the residents in nursing care homes were the oldest by a slight margin, and in personal care homes, the youngest, with average ages of 78 years and 76 years respectively.

The elderly female greatly outnumbers the elderly male in nursing and personal care homes. This is due entirely to the larger frequency of females at ages 65 and over where the ratio was almost two to one (Table A). Men were in a slight majority at ages under 65. The predominance of women reflects in part the sex differences in older ages of the U.S. population. However, when comparing the number of females per 100 males at specific ages with those of the U.S. civilian population it is obvious that a higher proportion of aged women than of aged men were in care homes.

Nonwhites composed a relatively small proportion of the residents. Only 4 percent was nonwhite, a rate of less than 2 per 1,000 nonwhite persons 20 years of age and over in the U.S. population. This compares with a rate of around five for whites. Half of the nonwhite residents were in the South region, but the highest rate was observed for the Northeast at 2.4 per 1,000 population.

The health of residents was studied in terms of their walking status, bed status, mental awareness, continence, hearing, and vision. For each of these health related categories, except hearing, residents were classified into one of three groups depending upon the extent to which they were disabled. Hearing status included only two classes: "No serious problems with hearing" and "serious problems or deaf."

On the basis of information reported in the survey by proxy respondents such as nurses and other personnel of the homes, it is estimated that 17 percent of the residents were confined to bed all or most of the time, and another fourth were in bed part of the time over and above that required for ordinary rest or sleep (Table B). Also, a fourth of the residents never walked or were completely dependent on others to get about. In all, only 58 percent of the residents could walk unassisted or with a cane or crutch.

A fifth of the residents were totally incontinent; that is they normally could not control either their bowels or bladder. An additional 8 percent were incontinent in one respect or another but not in both.

More than half of the residents were disoriented; about a fifth were confused all or most of the time, and a third were confused part of the time.

Only a small proportion of residents were indicated to be blind, but a considerable number had serious problems with seeing, 3 percent and 16 percent respectively. A similar proportion of the residents had a serious problem with hearing or were deaf.

Relatively more women than men were restricted in some way, especially in terms of their ambulation. Almost half of the women were confined to bed at least part of the time and a like proportion had a problem with walking. This compares with a ratio of less than 4 out of 10 men with bed confinement or walking problems. It is also apparent from Table B that a larger proportion of women than men had severe restrictions in terms of all health-related categories studied except hearing and vision.

One might expect these sex differences to be largely due to age differentials between sexes. This is not true, however. As shown in Table B, the proportion of women with severe restrictions in terms of bed, walking,

Table A. Resident population of nursing and personal care homes and the U.S. civilian population 20 years +, the number of females per 100 males in each population, by age and sex: United States, 1963

Age	Resident Population			U.S. Civilian Population*(in 000's)		
	Males	Females	Females per 100 males	Males	Females	Females per 100 males
Total	173,063	332,179	192	53,374	59,365	111
20-64	32,021	27,657	86	45,597	49,575	109
65-74	35,147	54,472	155	5,150	6,185	120
75-84	65,233	142,010	218	2,242	2,988	133
85+	40,662	108,040	266	385	617	160

*Source: Current Population Reports, Population Estimates, Series P-25, No. 276, July 1, 1963.

Table B: Percent distribution of residents in nursing and personal care homes by extent of disability in certain health-related categories according to sex and age: United States, April-June, 1963

Sex and Age	Total: Number	Total: Percent	Bed Status: In bed hardly ever	Bed Status: In bed part of time	Bed Status: In bed all or most of time	Walking Status: Walks unassisted	Walking Status: Walks with some help	Walking Status: Never walks	Continence Status: Continent	Continence Status: Partly continent	Continence Status: Totally incontinent	Mental Status: Always aware of surroundings	Mental Status: Confused part of time	Mental Status: Confused all or most of time	Hearing Status: No serious problem	Hearing Status: Serious problem or deaf	Vision Status: No serious problem	Vision Status: Serious problem	Vision Status: Blind
			Percent Distribution																
Both Sexes																			
All ages	505242	100	57	26	17	58	18	24	73	8	19	50	32	18	84	16	81	16	3
Under 65	59678	100	70	17	14	66	14	20	82	5	13	62	27	11	94	6	90	8	3
65-74	89619	100	61	24	15	61	17	22	77	7	16	55	31	14	91	9	87	11	3
75-84	207243	100	58	26	16	60	18	23	73	8	19	50	32	18	86	14	82	15	3
85 +	148702	100	49	30	21	50	21	28	68	9	23	43	36	21	74	26	72	24	5
Males																			
All ages	173063	100	61	24	14	65	16	19	75	8	16	54	31	15	84	16	83	14	3
Under 65	32021	100	75	14	11	73	13	14	85	5	10	66	24	10	95	6	91	7	2
65-74	35147	100	65	23	13	67	16	17	78	8	14	57	30	13	90	10	87	10	3
75-84	65233	100	58	27	15	64	17	20	72	9	19	51	33	17	84	16	82	15	3
85 +	40662	100	52	30	18	60	19	21	70	10	20	46	35	19	71	29	73	22	5
Females																			
All ages	332179	100	55	27	19	54	19	27	72	8	20	48	33	19	84	16	80	17	3
Under 65	27657	100	63	20	17	58	16	26	77	6	17	57	31	13	94	6	88	9	3
65-74	54472	100	58	25	17	57	18	25	76	7	17	54	31	15	91	9	86	12	2
75-84	142010	100	58	26	17	58	18	24	73	7	19	49	32	19	87	13	82	15	3
85 +	108040	100	48	30	22	47	22	21	67	9	24	42	36	22	75	27	71	24	5

continence, and mental status was almost without exception higher in each age group than for men of corresponding ages, and the magnitudes of the age-sex differences were similar to the difference when sex alone is considered. For vision status, the pattern by age was similar for both men and women. Hearing impairments, however, were more prevalent among men in the upper age groups than among women.

Even though the differential age distribution does not account for observed sex differences in disability, it is a significant factor affecting health of both men and women. For each health category there is increasing disability with increasing age. The disability patterns are marked by 3 broad age groups. For example, consider the distribution of residents by bed status and age. About 3 out of 10 of those under 65 had some degree of bed restriction. The proportion increases to about 4 out of 10 for age groups 65-74 and 75-84, and to 5 out of 10 for residents 85 and over. This same general pattern holds for walking, continence and mental disability.

The picture is somewhat different if only extreme disability is related to age. Considering bed and walking status, the proportions are quite similar for each age group through age 84. At ages 85 and over a significantly larger proportion of residents were bed ridden or could not walk. For continence and mental status, there appears to be a more gradual increase in the proportion of residents with total incontinence or with severe mental confusion as the age of residents increases.

When one health characteristic is related to another, such as bed status to walking, continence, and mental awareness,it is seen that increased restriction in one category is associated with increased restriction in the others, as illustrated in Table C. As one would expect, for residents who were in bed all or most of the time, relatively few of them were able to walk; more than four-fifths never walked and a tenth were able to walk only if helped. Nearly three-fifths of these bed ridden patients were totally incontinent and a third were confused all or most of the time. In contrast, only a fourth of the residents with partial bed restriction never walked, a fourth were totally incontinent, and a fifth were confused all or most of the time. Correspondingly, residents in bed hardly ever were much less restricted in these other health-related items than those with bed restriction. It should be noted however, that residents ordinarily out of bed were not free of infirmities. For some 289,000 such people, about 4 out of 10 were confused at least part of the time, a sixth required help with walking and a tenth were incontinent.

A question of concern relates to the level of care being provided to these people. No definitive answers can be given since the emphasis of the Resident Places Survey was on the characteristics of people rather than on their care. However, a crude indicator of the level of care is the primary type of service provided in the homes. As shown in Table D, the proportion of residents with extreme disabilities were related to the service category of the home; the proportion was the highest for nursing care homes and the lowest for personal care homes. Recall that places were classified as "nursing care homes" if they employed a registered nurse or a licensed practical nurse 15 hours or more per week and provided nursing care to the majority of their residents. Survey results also indicated that more than 90 percent of the nursing care homes had a nurse or nurse's aide on duty 24 hours a day and that 55 percent of them employed a full-time registered professional nurse to supervise nursing care. The remainder of the nursing care homes, with few exceptions, employed licensed practical nurses and part-time registered nurses to supervise nursing care. Personal-care-with-nursing homes, although not a requirement of the classification criteria, frequently employed professional and practical nurses. About two-thirds of these homes had a registered nurse or a licensed practical nurse in charge of nursing care, and about 8 out of 10 homes provided round-the-clock nursing service.

The survey also indicated that a few of the establishments classified as "personal care homes" employed nursing personnel and had a nurse or nurse's aide on duty 24 hours per day. Thus it is possible that the small proportion of disabled and infirm residents in personal care homes were receiving some level of nursing care.

<u>Summary and Conclusions</u>

A sample survey of establishments providing nursing and personal care to the aged and chronically ill was conducted during the spring of 1963, by the National Center for Health Statistics in cooperation with the U.S. Bureau of the Census. Results indicated a total resident population of about 505,000 persons.

Data collected in the survey provide a general picture of the health and personal characteristics of residents. Although a few were relatively young, the vast majority were quite aged as indicated by an overall average age of 78 years. Women outnumbered men by a ratio of 2 to 1. Almost all of the residents were white.

Table C: Percent distribution of residents in nursing and personal care homes by extent of disability in selected health-related categories according to sex and bed disatbility status: United States, April-June 1963

Sex and Bed Status	Total		Walking Status			Continence Status			Mental Status		
	Number	Percent	Walks Unassisted	Walks with some help	Never Walks	Continent	Partly Continent	Totally Incontinent	Always aware of Surroundings	Confused part of time	Confused all or most of time
		Percent Distribution									
Total	505,242	100	58	18	24	73	8	19	50	32	18
Both Sexes											
In bed hardly ever	288,675	100	83	13	4	90	5	5	63	26	11
In bed part of time	130,381	100	37	35	28	63	13	24	37	42	21
In bed all or most of time	86,186	100	6	11	83	32	11	57	26	39	35
Females											
In bed hardly ever	182,603	100	81	14	5	90	4	6	62	27	11
In bed part of time	88,252	100	33	36	31	63	12	25	36	42	22
In bed all or most of time	61,324	100	5	10	85	32	10	58	25	38	37
Males											
In bed hardly ever	106,072	100	87	10	3	90	5	5	66	24	10
In bed part of time	42,129	100	44	34	22	63	14	23	39	42	19
In bed all or most of time	24,862	100	8	14	78	35	11	54	27	41	32

Table D: Percent distribution of residents in nursing and personal care homes by extent of disability in selected health-related categories according to primary type of service provided in the homes: United States, April-June, 1963

Primary type of service	Total		Bed Status			Walking Status			Continence Status			Mental Status			Hearing Status		Vision Status		
	Number	Percent	In bed hardly ever	In bed part of time	In bed all or most of time	Walks unassisted	Walks with some help	Never Walks	Continent	Partly continent	Totally incontinent	Always aware of surroundings	Confused part of time	Confused all or most of time	No serious problem	Serious problem or deaf	No serious problem	Serious problem	Blind
		Percent Distribution																	
All types of service	505242	100	57	26	17	58	18	24	73	8	19	50	32	18	84	16	81	16	3
Nursing care	286373	100	47	31	22	47	22	31	66	9	25	43	36	21	84	16	79	17	4
Personal care with nursing	170678	100	67	21	12	70	14	16	81	6	13	58	28	14	85	15	84	13	3
Personal care	48191	100	80	14	6	82	10	8	88	5	7	67	25	9	84	16	82	15	3

More than half of the residents had some level of disability or infirmity. About 4 out of 10 were confined to bed at least part of the time, and a large proportion of those who were usually out of bed except for rest or sleep had impairments. Measured in terms of frequency, the largest health problem was not physical, but mental; half of the residents were unaware of their surroundings either part or all of the time compared with two-fifths who were unable to walk, three-tenths who were incontinent, and a fifth who had problems with seeing or hearing.

Thus it is clear that many residents of nursing and personal care homes are often in ill health, as indicated by data collected in RPS-1. For many purposes, however, more specific information is needed on the conditions associated with ill health, the availability and provision of medical and nursing services, and on other factors which relate to health. The second Resident Places Survey should go far in filling this need. A number of reports will be published periodically over the next year or two about such topics as the prevalence of chronic conditions and impairments, types of services provided, the use of special aids such as walkers and wheelchairs, charges for care of residents and source of payment, and the availability of staff to provide care. Meanwhile, several reports are being published on the results of RPS-1 which include, in addition to statistics on the health of residents, characteristics of the institutions themselves, and the utilization of facilities. 3/, 4/, 5/, 6/.

In time, the Center plans to study the health of all segments of the institutional population. The data collected will augment that being obtained through household interviews, health examinations, hospital records, etc., about the noninstitutional population to provide a comprehensive description of the health of the nation's people and of related matters. 7/ The *ad hoc* institutional population surveys will be repeated or new ones conducted on the basis of current needs and national interest.

References

1/ Solon, J. A., Roberts, D. W., Krueger, D. E., and Baney, A. M.: *Nursing Homes, their Patients and their Care*, Public Health Monograph No. 46, PHS Publication No. 503, August 1956

2/ National Center for Health Statistics: Development and Maintenance of a National Inventory of Hospitals and Institutions, *Vital and Health Statistics*, PHS Publication No. 1000, Series 1 - No. 3, Public Health Service, Washington, D.C., Feb. 1965

3/ National Center for Health Statistics: Institutions for the Aged and Chronically Ill, *Vital and Health Statistics*, PHS Publication No. 1000, Series 12 - No. 1, August 1965

4/ National Center for Health Statistics: Characteristics of Residents in Institutions for the Aged and Chronically Ill, *Vital and Health Statistics*, PHS Publication No. 1000, Series 12 - No. 2, September 1965

5/ National Center for Health Statistics: Characteristics of Residents in Mental Hospitals, *Vital and Health Statistics*, PHS Publication No. 1000, Series 12- No. 3, (To be published)

6/ National Center for Health Statistics: Utilization of Institutions for the Aged and Chronically Ill, *Vital and Health Statistics*, PHS Publication No. 1000, Series 12 - No. 4, (To be published)

7/ National Center for Health Statistics: Origin, Program, and Operation of the U.S. National Health Survey, *Vital and Health Statistics*, PHS Publication No. 1000, Series 1 - No. 1, August 1963

Appendix Table: Approximate standard errors of estimated percentages shown in Tables B - D.

Base of the percent (number of residents)	Estimated Percent				
	2 or 98	5 or 95	10 or 90	25 or 75	50
	Standard Error (expressed in percentage points)				
25,000	0.6	0.9	1.2	1.7	2.0
50,000	0.4	0.6	0.9	1.2	1.4
100,000	0.3	0.4	0.6	0.9	1.0
250,000	0.2	0.3	0.4	0.6	0.7
500,000	0.1	0.2	0.3	0.4	0.5

Discussion

Jerry A. Solon, University of Pittsburgh Graduate School of Public Health

The Merriam and Bryant papers, to their credit, do not leave me with the discussant's impulse to "pick at" the presentations. I'm inclined, rather, to urge our two presenters on, along the lines of activity that they have opened up.

Both presenters have shown, with their papers, admirable finesse in reducing a large and intricate subject area to a neat nutshell. But for this gift of selectivity and balance, they might have been overwhelmed by the massive and complex data which faced them -- as was the writer of a recent newspaper feature story on the prospects of caring for the aged, who dourly intoned that "The statistics tilt ominously toward the future like an avalanche pouring over the slopes toward the helpless village below." It is to our benefit that Merriam views with benevolence rather than pessimism the potential "avalanche" of data from the emerging program of health insurance for the aged. Yet while portraying the potential to us, she has not denied us an appreciation of the difficulties faced in developing a national data system on health services utilization under the new Medicare program. Within the confines of available time, she has merely spared us some of the cruel challenges imposed by the multitude of intricately varying provisions in the Medicare legislation.

Similarly, Bryant has presented a clearcut summary of a vast amount of data. These data are beginning to appear in a series of monographs from the National Center for Health Statistics dealing with its surveys of nursing and personal care establishments.

Both reports stem from activities which we need to recognize as being revolutionary in our time. This perspective may be a bit difficult to grasp and appreciate, as close as we are to the situation. Let us acknowledge, however, that the kind of national studies which Bryant et al. are reporting are breakthroughs. There was not until this decade the extent of interest in nursing and personal care homes, nor the thrust of resources in that direction, to permit such national survey data to be gathered. The best we were able to do 10 years ago was to mount a 13-state study, largely on a self-selecting basis. That study of nursing homes, in which I was fortunate to have a part, was itself a pioneering foray made possible by the combined impetus of the Commission on Chronic Illness and the Public Health Service. But not until now have we seen a representative national study accomplished.

In parallel fashion, health insurance in the Social Security framework has been long proposed and opposed until the historic passage which we have witnessed this year. In yesterday's session on "Needed Developments in Social Statistics," Frederick Stephan referred to the United States as an underdeveloped country in some areas of social statistics. Merriam and Bryant have made it clear that in the areas which they are covering, we are no longer so much an underdeveloped country as a developing country.

Both papers have evidenced a grasp toward important underlying concepts upon which to build data. Thus, Bryant's data are constructed upon a basic classification which is itself a worthy contribution. The categorization of nursing and personal care homes, devised with thoughtful concepts and skillful method, provides a framework which gives coherence to what would otherwise be an amorphous mass of slippery data.

Merriam penetrates to essentials by stressing the objective of portraying patterns of health service utilization with the individual as a unit, as distinguished from mere cumulation of service units on a mass basis. The elusiveness of some of the potential for organizing utilization data in such illuminating ways can be illustrated here. The legislation provides for inpatient care in hospitals and extended care facilities within the framework of "spells of illness." Utilization data can be especially meaningful in such a framework; unfortunately, the statutory definition of what constitutes a "spell of illness" defeats this potential by specifying that a "spell of illness" ends 60 days after discharge from the inpatient institution. In order to make health service utilization data more meaningful, we have in a recent study been formulating the concept of "episodes of medical care"; we proceed then to make sense of detailed units of service by clustering them within episodes of care gravitating around particular health problems and objectives. The Act's specification of a "spell of illness" in the terms mentioned imposes an administrative necessity which, while geared to the limitation of benefits, disengages the utilization data from the real phenomenon to be represented -- patterns of health care. Whether some conceptual and methodological devises may yet be devised to permit casting the utilization data into the holistic terms of episodes of medical care remains to be seen.

The methodological schemes reported by the two papers invite admiration. Our confidence is engaged by Bryant's description of the sampling design and survey procedure. The follow-up techniques used to gather in the responses were thorough, and the planned combination of mail and personal visits was effective. Merriam's proposal for the employment of "average person-years of exposure" as a base upon which to calculate utilization is a promising refinement. In fact, a variety of features in the plans which she has outlined for us bespeak a fine sense for capitalizing richly on the opportunities -- for example, to study the characteristics of the uninsured as well as the insured in the aged population, to study the impact of the program upon medical care facilities and providers, to link the program-produced data with supplementing surveys which might be conducted by the National Center for Health Statistics.

It is significant that in undertaking its institutional population surveys, the National Center for Health Statistics gave precedence in its sequence of survey priorities to nursing and personal care homes rather than general hospitals. It is understandable that having accomplished as much as it has in the nursing and personal care institutional area, the Center would now swing its attention away from the nonhospital establishments. The appearance of the Medicare program upon the scene, however, invites the Center's reconsideration. With the baseline data which its survey activity has already established, the Center has in effect set the stage for taking account of the impact Medicare will have on use of "extended care facilities." Relevant here is an observation which Merriam made in an earlier paper to the National Industrial Conference Board, in which she anticipated that "The problem of increased utilization is not as serious with respect to hospitals as with respect to skilled nursing home and other extended care facilities." We may hope that the Center will not rest on its laurels in its already accomplished nursing home surveys, but will rather proceed to plan appropriately timed resurveys in ways which will reflect the influences of the new health insurance benefits.

Both developments reported in these papers -- planning a statistical program on health care of the aged, and surveying nursing homes -- occur in organizational settings which are reported to have had the cooperative participation of numbers of agencies and units in addition to the centrally-involved agency. The potential for cooperative planning and activity achieved in these instances is not always realized successfully. A potential exists in such situations for a paralyzing competitiveness for role among organizational units concerned with the emerging activity. The dynamics of interplay among organizational units which leads to enhancing or inhibiting the activity is another whole story, not undertaken in these papers, which could be instructive for research in large-scale organizations.

DISCUSSION

John A. Sonquist, The University of Michigan

In view of the fact that it was impossible for Mr. Haber to present his paper, I should like to make a few general remarks about problems in dealing with statistics on the older population and then comment briefly on the papers by Mrs. Merriam and by Mr. Bryant.

Lenore Epstein[1] has pointed out that any study of economic and social characteristics of the older population should deal adequately with five problems; skewed distributions, the proper unit of analysis, inadequate measures of income and resources, differing family sizes and structure, and non-current resources such as assets, insurance, and potential help from relatives.

Averages can be quite deceptive, since distributions of variables collected from members of the older population are often skewed. The Gini index, as a measure of inequality based on the Lorenz curve, is a useful measure of the inequalities in the income, or any other distribution. It has the value 0.0 if every unit has the same income; it rises to 1.0 if one unit has all the income. For a recent national cross section of spending units the index is .48 for the whole population (considering wage income) and .86 for units headed by someone aged 65 or over.[2]

The unit of analysis also constitutes a problem in dealing with the characteristics of the aged. There are obvious problems in using families (a group of persons related by blood, marriage or adoption and living together in the same household). The poverty of the parents may be hidden in the combined family income unless detailed income figures for each adult in the family are obtained. Moreover families combine and separate over time. This tends to make estimates on a family basis unstable and inadequate. Since the vast bulk of aid to elderly people takes on the form of housing provided by younger family members, there is a strong argument for separating out related adults and couples into more basic units. Why not use individuals? Husbands and wives form a close economic and psychological unit. The data on income, for example would be misleading because many wives, would show up with no income.

Measures of the income and resources of the aged are subject to criticism. Perhaps one that has not yet been made (perhaps Mr. Haber would have) is related to the selection of the family in many studies as the basic unit of analysis. The respondent in the interview may be a younger family member. If the respondent is not the person who gets the income serious reporting errors may occur as to both the amount and sources of the income accruing to elderly family members. Insofar as there exist concentrations of elderly people living in low-cost housing areas in the large metropolitan areas, the low response rates in such areas now being noted by research organizations may bias aggregate figures downward.

But these are probably not the most serious of the problems associated with the development of adequate measures of income and resources. Economic resources are not only cash income. They include paid-for homes lived in rent-free, smaller families and lower food and clothing requirements, home grown food, making repairs on one's own home, and even variations in the climate requiring lower heating bills.

Perhaps the most central problem in assessing the income and resources of the aged has to do with the income value of housing either owned outright by the aged or provided "free" in the homes of younger family members. Data obtained in the 1962 Survey of Consumer Finances[2] showed that among spending units where the head was aged 65 or over 58 percent owned their own home and almost half owned it mortgage free. Almost 30 percent had $10,000 or more equity in their house. About one fourth paid rent and 12 percent lived with relatives. Spending units headed by an elderly person tended to have more rooms per person and their houses tended to be somewhat more run down than among younger families.

After developing really adequate units of analysis and a more comprehensive measure of level of living that takes account of non-money income to some standard of needs, three additional factors need to be considered.

Housing is a source of difficulty and rigidity as well as income. Do older people want all those rooms? Or is there no alternative form of housing meeting their needs? What maintains this disproportion in the consumption patterns of the aged?

Methods need to be developed for providing answers to the question of the adequacy of current income (after providing for housing) with respect to needs other than housing.

Additional attention needs to be given to the problems of emergencies (of which sickness is, to be sure, the most important, but nevertheless only one of many types) and the resources (assets which are perceived as useable, and insurance) to meet these needs. The real need is to study consumption patterns together with money income and other goods and services received.

But the collection of statistics on the objective situation of the aged is not enough.

Health and income are only means to a complex and interrelated series of personal goals. The consumption of durable and non-durable goods and services are goals, but are also means toward the living of a satisfying life as a member of family, community and society. Therefore it is not enough to collect data on hard "facts," such as health status and income. These need to be related to data on attitudes and values with respect to present and past income and to the perceived needs and desires of the aged themselves. Economic statistics need to be coupled with measures of the aged person's integration into his family, his participation in primary groups other than the family and in formal organizations, and his contacts with other persons his age and with those younger.

I have two comments to make on the papers by Mrs. Merriam and Mr. Bryant. Inadequate statistics on the race of those now covered by the Social Security program can be improved by putting an appropriate box on the medical service utilization form filled out by the doctor. Since the social security number is also on the form, master files can be up-dated. I would hope that the data-collection program of the Social Security Administration will include attitude data as well as data on utilization of services. This would have to be done by the SSA, since the files will be confidential.

Mr. Bryant's paper represents a significant contribution in an area about which all to little has been known. He is to be complimented especially for presenting distributions rather than means. I have only one question to raise, which concerns possible biases in his respondents, the employees of the nursing homes. In many institutions the lower level staff are somewhat underpaid and tend to view their job as one of running their ward or section with as little "trouble" for themselves as possible. This is particularly true in State-operated institutions. In many cases the informally established rules for patients are excessively oppressive. Consequently there is a tendency for patients who have a certain amount of self-respect and independence to be classified as "disturbed" or "senile" when in fact they are just the opposite. The extent to which this biases Mr. Bryant's statistics is difficult to ascertain. But I would predict that there is a negative correlation between the income of the informant and the proportion of disturbed or senile patients he reports.

1. Lenore Epstein, paper delivered at the 5th Congress of the International Association of Gerontology, San Francisco, 1960. See also Janet Fisher in Aging and the Economy, by Orbach and Tibbetts, University of Michigan Press, 1963.

2. Data reported by J. N. Morgan, Measuring the Economic Status of the Aged, International Economic Review, January, 1965. See also The 1962 Survey of Consumer Finances, Katona, Lininger and Kosobud, University of Michigan, 1963.

IX

RECENT RESEARCH ON RESPONSE ERRORS

Chairman, ROBERT FERBER, University of Illinois

ON SOME ASPECTS OF RESPONSE ERROR MEASUREMENT[1,2]

William G. Madow, Stanford Research Institute

1. Introduction. Both response error and response bias[3] occur in almost every survey, whether a survey of a sample or a survey of an entire population.

The understanding and measurement of response error are essential for four main purposes:

(1) To improve the assessments of the accuracy of data provided by a particular survey, and to determine how much weight can be given those data in procedures for making conclusions, decisions, or actions in which the data are inputs,

(2) To determine how to improve the making of the survey and the estimates based on it,

(3) To improve the allocation of resources to the different parts of a survey, viewed as an economic enterprise, in order to maximize the information provided by a survey for a given cost or to minimize the risks of conclusions, decisions, or actions based in whole or part on the survey, and

(4) To contribute to the accumulation of information on response error for use in other surveys and in research on survey making.

In many surveys response errors are large; usually, they are measured or reduced only with considerable effort and cost. In some surveys, the main component of response error is response variance, which can be reduced by technical improvements in the size or method of making the survey. In other surveys, the response bias is large. To reduce response bias is often far more costly than to reduce response variance; sometimes the response cannot be sufficiently reduced to regard it as small. Sometimes, the knowledge and understanding of the bias are adequate for taking it into account when conclusions, decisions or actions are taken; sometimes other evidence than the survey itself will keep the response bias from leading to error; sometimes, in time series, the changes in the data are sufficiently greater than the changes in the biases for estimates of change to be relatively unaffected by response bias. To know whether the response bias will produce errors in analyses requires knowing something of the size and nature of the response bias.

To measure response variance and response bias, to determine how they depend on various parts of the survey, to determine how they are related to respondent and interviewer characteristics, to determine how they are related to other variables on which information is obtained in the surveys, to determine how they change over time, to determine how and at what cost they can be reduced, and to determine how much they increase the risks of conclusions, decisions and actions utilizing the survey; all these seem essential to the improvement of survey making and use.

In this paper we emphasize:

(1) The possibility of using double sampling methods for reducing response bias,

(2) The contribution of explanatory variables to the understanding of response error, and to the reduction of response error,

(3) The importance of studying as behavioral science models the choice mechanism used by the respondent in selecting one of the possible replies that he can give when a question is put to him, not only in order to understand response error but to reduce it, and to lower the cost of double sampling methods for reducing response error.

We also discuss the concept of the exploratory survey, a type of survey that we distinguish from the analytic, general purpose or special purpose surveys. Since the treatment of response error may be quite different in surveys for time series or in panel surveys than in one-time surveys, we also briefly consider that classification.

A suggestion that we make, although we do not implement, is an attempt to resolve an issue that often arises between the survey designer and the survey user. This issue is the contradiction between the user's knowing he cannot state precisely how the data will be used, and his consequent wish that he be provided with generally good data; and the designers knowing he cannot provide data adequate for all uses and his consequent wish that he be provided with exact statements of how accurate specified data must be for the uses to be made of them.

The suggestion is that the designer and user, instead of discussing accuracy begin by considering some perhaps oversimplified models of using the data, discuss the requirements on the data resulting from the desire to achieve certain risk levels through the use of these models, and conclude by adjusting the accuracy levels to account for other possible uses not so easily specifiable.

[1] This paper owes much to the work, published and unpublished, of the staff of the Bureau of the Census. Of the many papers they have published in this area only one is mentioned: Measurement Errors in Censuses and Surveys by Morris H. Hansen, William N. Hurwitz and Max A. Bershad, Bull. Int. Stat. Inst. Vol. 38 (1961) pp. 359-374.

[2] This research was supported by National Science Foundation Grant GS-83.

[3] We shall refer to response variance and response bias collectively as response error.

2. Classifying surveys by their uses.

We first discuss two classifications of surveys that are related to the uses of surveys. The word "surveys" as used here includes not only sample surveys, but also surveys of an entire population or universe.

One classification of surveys is by whether the surveys are "one-time" surveys, or whether the surveys are made to provide data for a time series, and in the latter case, whether the survey is a panel survey made at a fixed number of time periods with complete or partial renumeration of the members of the sample at the time periods.

Another classification of surveys is by whether the survey is intended to provide data for general use, such as data provided by the Bureau of the Census or the Bureau of Labor Statistics or various other government agencies, or whether data are to be provided for special purposes, such as the data provided by marketing surveys made for a specific company in order for it to decide what action it wishes to take with respect to its products, or whether the survey is an analytical survey made in order to test various specified social science hypotheses such as might occur in studies intended to test hypotheses concerning the nature of voting behavior or whether the survey is an exploratory survey, i.e. a survey made in order to obtain information concerning which variables are important in characterizing one or another kind of behavior.

The classifications of surveys made above are not classifications into mutually exclusive classes. Many surveys partake of several of these aspects. From the point of view of survey design however, these classifications lead to different designs, and consequently the consideration of these classifications is relevant to a discussion of survey design.

Let us discuss the differences among these various types of surveys in relation to a common body of information. For this purpose the topography of a geographical area provides a useful illustration.

In a general purpose survey, the topographical features such as mountains, rivers, etc. would have been identified; for each of these features specific items such as heights, lengths, etc. would be measured by some procedure. The list of topographical features and the measurements together with indications of how accurate these measurements would be would then be published as a general purpose set of data available for any user who could himself decide the extent to which he wished to rely or make use of these data. In making the survey, certain users and their needs might have been considered but, in addition, the data be for general use. The data may make it possible for some decisions to be made but might not be adequate for others - perhaps because of the wide focus rather than a narrow focus.

In a special purpose survey, attention would usually be put on one or more of the topographical features; the objective might be to learn whether one mountain is larger than another without it being terribly important to know the exact size of either of the two mountains, or it might be necessary to learn something of the slope of the mountain that would go beyond the detail published in a general purpose study, perhaps, in order to estimate the cost of moving material from one level to another of the mountain. A frequent characteristic of the special purpose survey is that it is intended to provide data required to make a decision or take some action.

An analytical survey might be concerned with testing hypotheses concerning the ages of the mountains, or a hypothesis concerning the nature of the phenomena that led to the existence of some of the existing topographical features, and the prediction of what these topographical characteristics might be at some future time period because of the nature of the fundamental processes creating change in the topographical characteristics.

The exploratory survey would be concerned with attempting to determine what were the real topographical features themselves, which of these features best characterized the area in question, and develop hypotheses concerning the underlying processes.

The main reason for discussing this classification at this point is that different kinds of samples and different types of analysis would be appropriate for the different types of surveys identified in the classification. For example, a special purpose survey with a very specific objective can often have far smaller size and be less rigorously made than a general purpose survey of the same area intended to provide information suitable for many uses, uses which may not have even been specified at the time that the survey was made. While this does not mean that all general purpose surveys must be of exceptionally high quality, it does imply that if a survey can be done for specific purposes the costs can be reduced.

The distinction between analytical surveys and either general or special purpose surveys is not a distinction between whether one wishes to estimate relations or only to estimate specific numbers. Both general and special purpose surveys may well have as their major objectives the estimation of both numbers and relations, or even only relations. Similarly general purpose studies, special purpose studies and analytical studies all may be intended to provide data to serve as a basis for decisions. The main difference between analytic surveys and either general or special purpose surveys is that analytical surveys are concerned with the processes that led to the data that are obtained or to the relationships that are obtained, whereas special purpose and general purpose surveys are primarily concerned with the data or relationships themselves and their uses in decision and action.

It is almost inevitable that analytical surveys would contain what are sometimes called "soft" data as well as the relatively "hard" data that are often obtained in special or general purpose surveys, although this is not meant to imply that only hard data are obtained in such surveys.

From a design point of view, the main difference between exploratory surveys and either general or special or analytical surveys is that in the case of exploratory surveys it is so much more important to sample all the possible sources of variability of the data to be obtained whereas in any of the other three the sources of variability can be limited by setting forth some more restrictive objectives that can be stated for an exploratory survey.

3. Steps in survey making related to response error.

The steps of the survey that include the selection of a respondent, the design of a questionnaire, the selection of an interviewer, and the bringing of these three entities together may be viewed as the planning of a choice experiment in which the respondent presented with a combined stimulus of the question or questionnaire and interviewer in the environment in which the interview takes place chooses a response either from a pre-assigned list of possible responses or from a set of alternatives among which the respondent decides to make the choice.

Not all elements of this experiment need be present. For example, it may be that instead of the respondent there is a file of information concerning people or businesses, and the interviewer, who in this case is the person who records the information, may abstract the information from the record onto the form. This procedure need not be error free. For example, if the forms from which the information is being extracted are the notes of physicians in medical records the abstraction of the information will be subject to many sources of error.

The interview may be conducted by mail or by telephone. In the former case the respondent may have an opportunity to read the entire questionnaire before he responds to individual questions, whereas either in a personal or telephone interview he is asked the questions sequentially and his responses may be affected by that fact. There may be no written questionnaire; the interview may be unstructured, and may be essentially non-directive as in the case of psychiatric interviews or similar types of interviews that occur in surveys.

The structure of the interviewing situation is that for each question, the respondent makes a choice from a more or less specified set of alternatives. The choices, (responses) are usually made sequentially although, as in the case of a mail survey, sometimes they are made collectively. The set of alternatives may be spelled out as in the case of a forced choice question, or to some extent defined by the respondent himself - usually without stating the set to the respondent.

The information provided by the respondent may be categorical as for example in the case of sex or it might be more nearly continuous as in the case of age or income. No distinction is being made according to the type of information provided by the respondent.

4. The accuracy required of estimates. Small analytic models.

The requirements of the analysis of the information provided by the survey must indicate the accuracy demanded of that information. Yet, statisticians and survey designers have long had the experience that it is difficult, if not impossible to expect the users of data to indicate the uses sufficiently precisely for the accuracy requirements on the survey to be possible to determine. Most frequently, such accuracy requirements have been arrived at by a process of negotiation in which the user will first suggest it would be nice if the data were completely accurate, the designer will inform the user of the cost of such an achievement, even assuming it to be possible which usually it is not, the user will then reduce the accuracy requirements he would like to place on the data, the designer will indicate the costs and feasibility of the reduced level of accuracy, etc. until either an agreement is reached or the decision is made either not to obtain the data or to obtain it by other means. One reason why it is so often difficult to state the required accuracy of the data is that the uses are themselves not too specifiable especially at the time of designing the survey. For example if one is talking about television ratings then it is well known that many decisions are made just by looking at the television ratings, but that many other aspects of information are used than those provided by the survey itself. Similarly if a government policy is to be based on the results of a survey, that policy will take into account far more then the numerical results of the survey; political aspects and fundamental economic considerations may also play a very large part. These considerations are even more important when one considers analytical and exploratory surveys in which the nature of the analysis will be affected by the results obtained in the survey itself. Between the date on which the design is fixed and the date on which the data begin to be available, several months may pass, and the needs may evolve and change. Furthermore, as the analysis evolves the needs and accuracy requirements may both change.

Even if the needs do not change, it often occurs that when decisions on the accuracy required are obtained by the process of negotiation between the survey user and the survey designer, it turns out that the negotiations either did not cover all the actual uses to be made of the data, or that the user did not fully understand to what he was agreeing.

Very often, once data have been collected then despite all the limitations placed upon the use of the data by the survey designer or statistician, the user often feels that since the data are the best he has available, he must use them and rely primarily on the consistency of the conclusions obtained from the data with other sources of information, or the internal consistency of the conclusions as the basic support for his analysis. It is true that if the data are highly variable then there will tend to be a sufficiently large number of inconsistencies for the user to become concerned and to limit the conclusions that he was prepared to draw from the data. But, when the data are biased, there may be few inconsistencies, and yet the survey as a whole is wide of the mark. A sufficiently sophisticated user who makes comparisons with relatively unbiased outside sources of data, and evaluates these data themselves will often use the survey in a useful way. However, it must be emphasized that it is the user whose judgment becomes critical rather than the objective properties of the survey data themselves.

Sometimes when a time series is being analysed it is felt that biased data may be used because the effects of the bias will disappear or be greatly reduced when investigating the changes over time. The biases may be due to response, as in income estimates, or to the use of a study population that differs from the population concerning which conclusions are to be drawn, as in the case of TV ratings. But often the biases will change over time, and may even increase in relative importance. Just as we attempt to measure the accuracy of estimates at a given time so must such estimates be attempted over time; in particular the changes in response bias should be estimated even if it is believed that response variance is stable.

There is an intermediate level that might help determine the required accuracy of the data without committing the user to use methods of analysis that are too simple or inappropriate for his purposes; this method is the construction of small models. These models may be analysed mathematically or by simulation. But they will lead to conclusions on the accuracy required of the data so far as their requirements are concerned, and such analyses will often provide an adequate basis for deciding the accuracy to be required of the data.

5. Questionnaire and Interviewer Effects on Response Error.

We have suggested how it may be possible more often than now is done, to arrive at conclusions concerning the accuracy of the estimates desired from a survey. Let us now turn to the ways in which accuracy may be measured and to the factors on which the accuracy depends.

One of the most important factors to consider, and one of the most difficult properly to assess, is the choice of questionnaire. Most discussions of questionnaire construction come down to telling the questionnaire constructer to be wise and to be careful, and point out that unless one is wise and careful variance and biases will result. Experience indicates however, that despite all the wisdom and judgment of the past, every new questionnaire presents a new challenge, and the ability of respondents to find weaknesses in the questionnaire continues to be effective even after serious and able attempts of questionnaire constructors to find and remove biases before the respondents show him the weaknesses. Pretests or pilot studies of questionnaires are known to be indispensable, and often these pretests and pilot studies include alternative questionnaires. Improvements of questionnaires are usually costly to demonstrate; and often it is difficult to justify the improvement in information in terms of the extra cost and time and, perhaps, the more competent interviewers and greater training that are required. Even when two or more versions of a questionnaire are used in the same study so that differences primarily due to the questionnaires can be isolated, and the versions are distributed among the sample of respondents so that valid comparisons can be made, the versions may have the same essential biases. Biases because of unwillingness to report, for example, may not be reduced sufficiently by an improvement in a questionnaire; similarly inability to report because of forgetfulness may be only partially reduced by the use of questions intended to stimulate the memory; the different versions of the questionnaire may yield data more similar to each other than any is to the true data that one would like to elicit.

It is rare that one-time surveys can adequately develop questionnaires, especially since so few tests of questionnaires are conducted in a way that provides objective evidence.

Once the questionnaire or questionnaires have been selected, the issue becomes that of the accuracy of the information requested on the questionnaires.

Other important topics in survey making that require further research in spite of the large and costly efforts that have been made are the training of interviewers and the manner of conducting interviews. Interviewer training has been difficult to evaluate because of the very high cost of experimental studies that are generalizable. Whether the interviewer should ask questions exactly as they appear on the questionnaire or be allowed to vary from the questionnaire seems to depend on the survey and the organization conducting the survey. It is well known that even slight deviation from questions as worded may produce unexpected effects in the responses obtained. Yet, whatever may be the emphasis placed on the importance of the interviewer not adjusting the questionnaire, changes are often made by the interviewer even when this goes counter to the instructions. Even when no change is explicitly made, the interviewer's attitude toward the question communicates itself sufficiently to the respondent to alter the meaning of the question. For example, the interviewer may feel that questions on income should not be asked, and have difficulty in asking such questions; even when asking the questions the interviewer may ask them sufficiently tentatively or concernedly so that the respondent becomes aware without any words being spoken that make the interviewer's attitude explicit. Similarly an interviewer with long experience in using a particular questionnaire may stop asking all the questions or using all the aides provided because of sufficient familiarity with the questionnaire to believe that these practices are unnecessary. For example, the interviewer may have been given a list to show to the respondent at a certain point but instead of showing the list to the respondent, the interviewer may eventually memorize the list and repeat it - perhaps incorrectly.

Similarly the problems that the interviewer faces in attempting to communicate with the respondent could be the subject of further study. Many respondents will not refuse, but will lower the quality of information they provide by the way in which they participate. The interviewer faced with a resentful or hurried or troubled respondent will often be affected by the respondent and the quality of information will suffer thereby. Similarly, the interviewer, in obtaining information from a proxy respondent, will find that it is more difficult to ask probing questions concerning the desired respondent in any fruitful fashion.

Under these conditions, the measurement of response variance and response bias become almost indispensable in a survey aimed at more than the establishment of very large differences.

6. Response Variance, Response Bias and Response Error.

It will be helpful at this point to introduce some definitions and symbols in order to make precise the analysis with which we will be concerned.

Suppose that we are dealing with the population of N elements, denoted by 1,2,...,N.

Let μ_i, i=1,...N be the "true value" of a variable for the ith element of the population.

Let x_i, i=1,...N be the random variable that is the choice of the respondent i, if respondent i is asked the question for which the true value for that respondent is μ_i.

Let a_i, i=1,...N be the expected value of x_i, i.e., $E\,x_i = a_i$.

Then, for the ith respondent, the variance of response, RV_i, is the expected value of the square of the difference $x_i - a_i$, i.e.

$$RV_i = E\,(x_i - a_i)^2$$

and the response bias, RB_i, is the difference between the expected response and the true value i.e.

$$RB_i = a_i - \mu_i$$

Finally, the mean square error of response, M_i, is the expected value of the square of the difference $x_i - \mu_i$, i.e.

$$M_i = E(x_i - \mu_i)^2 = E(x_i - a_i)^2 + (a_i - \mu_i)^2.$$

We shall not assume that the errors of response are uncorrelated. Also, we shall be assuming the existence of a "true value" even though, in practice a "true value" may not exist. For example, in cases of illness, the differences among physicians on whether a person is ill often are sufficiently great to cause us to remember that illness is itself a continuous variable and that what illness is will often be a matter of opinion. To define "illness" so that except for measurement errors among doctors, the same diagnoses would be reached would be most difficult. This problem is serious enough when it comes to ordinary medical conditions but in dealing with psychiatric problems the differences are much greater. Similarly when one is dealing with opinion and attitude questions the true value may well be a true probability distribution of opinions or attitudes of the person. Even in deciding whether a person is unemployed, or is a member of the labor force, large elements of opinion are present; whether there is really a true value for these variables is often doubtful. Sometimes, as in the case of unemployment, there may be knowable true values for many elements of a population without there being true values for all the elements of a population. Many are certainly employed, many are certainly unemployed, many are certainly not in the labor force, but many also are in the fringe groups where the true values are uncertain. To change definitions to eliminate fringe groups is not necessarily to increase the information provided by the survey.

Let us suppose now that the objective of a survey is to estimate a function,

$$f = f\,(\mu_1, \mu_2, \ldots, \mu_N)\ ,$$

of the "true values", $\mu_1, \mu_2, \ldots, \mu_N$.

For this purpose, an estimator, g', based on sample values $x'_1, x'_2, \ldots, x'_n$, i.e.

$$g' = g(x'_1, x'_2, \ldots, x'_n)\ ,$$

is defined.

Let $E\,(g'\ \mathcal{S})$

be the expected value of g', given the selected sample. Then, the response variance of g', $RV_{g'}$, is defined to be

$$RV_{g'} = E\sigma^2_{g'|\mathcal{S}} = E\left\{ E\left[(g' - E(g'|\mathcal{S}))^2 \middle| \mathcal{S} \right] \right\}$$

The response bias, $RB_{g'}$ is by definition

$$RB_{g'} = Eg' - f$$

When $E(g'|\mathcal{S})$ is evaluated it will be some function $h'_a = h(a'_1,\ldots,a'_n)$, i.e.

$$E(g'|\mathcal{S}) = h(a'_1,\ldots,a'_n) = h'_a$$

Define h'_μ to be the same function of $\mu'_1, \mu'_2,\ldots, \mu'_n$ that h'_a is of $a'_1,\ldots, a'_n$.

Often, $h(\mu'_1,\mu'_2,\ldots,\mu'_n)$ will be the estimate of f that would have been used, had there been no response error. Sometimes, another function would have been used, possibly $g'_\mu = g(\mu'_1, \mu'_2,\ldots, \mu'_n)$. Let us denote by $k'=k(\mu'_1, \mu'_2,\ldots, \mu'_n)$ the function that would be used to estimate f if there were no response error.

Then,

$$* \quad E\left[E(g'|\mathcal{S}) - Ek'\right]^2 = E\left[E(g'|\mathcal{S}) - k'\right]^2 + 2E\left[E(g'|\mathcal{S}) - k'\right]\left[k' - Ek'\right] + E(k'-Ek')^2$$

and we shall define the expected square sample bias, $SB^2_{g',k'}$ by the equation

$$SB^2_{g',k'} = E\left[E(g'|\mathcal{S}) - k'\right]^2,$$

the regression coefficient of $E(g'|\mathcal{S}) - k'$ on k', $\beta_{E(g'|\mathcal{S}) - k', k'}$, by

$$\beta = \beta_{E(g'|\mathcal{S}) - k', k'} = \frac{E\left[E(g'|\mathcal{S})\ k'\right]\left[k' - Ek'\right]}{E(k' - Ek')^2}$$

and the mean squared error had there been no response error, $MS_{k', f}$, by

$$MS_{k', f} = E(k' - f)^2 = E(k' - Ek')^2 + (Ek' - f)^2$$

As mentioned above k' will often be either g'_μ or h'_μ.

Let $SB^2_{g',k'}$ be partitioned into

$$** \quad SB^2_{g',k'} = E\left[E(g'|\mathcal{S}) - k') - (Eg' - Ek') - \beta(k' - Ek')\right]^2 + (E g' - E k')^2 + \beta^2 E(k'-Ek')^2$$

$$= S^2_{E(g'|\mathcal{S}) - k',k'} + (E g' - E k')^2 + \beta^2 E(k'-Ek')^2$$

Since

$$E(g' - f)^2 = \sigma^2_{E(g'|\mathcal{S})} + E\,\sigma^2_{g'\mathcal{S}} + (Eg' - f)^2$$

$$= E\,\sigma^2_{g'\mathcal{S}} + E\left[E(g'|\mathcal{S}) - f\right]^2$$

and since

$$E\left[E(g'|\mathcal{S}) - f\right]^2 = E\left[E(g'|\mathcal{S}) - E(k')\right]^2 + 2E\left[E(g'|\mathcal{S}) - Ek'\right]\left[E k' - f\right] + (E k' - f)^2$$

it follows that from * and ** that

$$E\left[E(g'|\mathcal{S}) - f\right]^2 = S^2_{E(g'|\mathcal{S}) -k',k'} + (\beta+1)^2\sigma^2_{k'} + (Eg' - f)^2$$

Hence we arrive at the fundamental formula for the mean square error of the estimator g' about the value estimated, f:

$$E(g'-f)^2 = E\sigma^2_{g'|\mathcal{S}} + S^2_{E(g'|\mathcal{S}) -k',k'} + (\beta+1)^2\sigma^2_{k'} + (Eg' - f)^2$$

Of the four terms, the first is the response variance and the last is the square of the bias. We have chosen to express the two middle terms using the regression of $E(g'|\mathcal{S}) - k'$ on k' since the within sample bias will so often be a function of k'.

It is easy to express these two middle terms using the regression of $E(g'|\mathcal{S})$ on k'. It will be noted that, especially if k' is an unbiased estimate of f, the occurrence of a negative regression coefficient of $E(g'|\mathcal{S}) - k'$ on k' can yield values of $E(g'-f)^2$ that are smaller than $E(k' - f)^2$.

The dominant term in $E(g' - f)^2$ may well be $(Eg' - f)^2$, the overall bias, squared. For this reason we turn in Section 7 to the use of double sampling methods of eliminating or reducing the bias.

7. <u>The reduction or elimination of response bias by double sampling.</u>

In order to avoid complications unnecessary for the discussion of the ideas we shall deal in this section with the simplest case, that in which a simple random sample has been selected from which a simple random subsample is selected for which estimates of the bias are obtained, and a difference estimate is used.

The true values for the population are $\mu_1, \mu_2,\ldots, \mu_N$. The objective of the survey is to estimate $\bar{\mu}$, the arithmetic mean of $\mu_1,\ldots, \mu_N$.

The responses of the elements of the sample are designated by $x'_1, x'_2,\ldots, x'_n$ and

$$E(x'_i\,|\,i) = a'_i,$$

where a'_i is also a random variable because of the use of simple random sampling.

Let

$$\bar{x}' = \frac{x'_1 + \ldots + x'_n}{n},$$

$$\bar{a}' = \frac{a_1' + \dots + a'_n}{n}$$

$$\bar{\mu}_1' = \frac{\mu_1' + \dots + \mu'_n}{n}$$

where the μ_i''s are the true values for the elements of the sample

For the sample random subsample of $n_1 < n$ elements, let

$$\bar{x}'' = \frac{x_1' + \dots + x_{n_1}'}{n_1}$$

$$\bar{a}'' = \frac{a_1' + \dots + a_{n_1}'}{n_1}$$

$$\bar{\mu}_1'' = \frac{\mu_1' + \dots + \mu_{n_1}'}{n_1}$$

We suppose that values of x_1', x_2', ..., x_{n_1}' are obtained from the respondents and that values of $\mu_1', \dots, \mu_{n_1}'$ are obtained from records. (Sometimes, when more intensive reinterviewing is done, instead of the true values, $\mu_1', \dots, \mu_{n_1}'$, values $y_1', \dots, y_{n_1}'$ believed to have smaller biases than $x_1', \dots, x_{n_1}'$ will be used. This type of analysis will be reported on later)

Let us define $\bar{z}'$ by the equation

$$\bar{z}' = \bar{x}' - (\bar{x}'' - \mu'')$$

Then $\bar{z}'$ is an unbiased estimate of $\bar{\mu}$, i.e.

$$E\ \bar{z}' = \bar{\mu}$$

Let $E\ x_i = a_i$; $\sigma^2_{x_i} = \sigma^2_i$; $\sigma_{x_i x_j} = \sigma_{ij}$

where

$$\sigma_{x_i x_j} = E\ (x_i - a_i)\ (x_j - a_j)\ .$$

Hence we assume that the response deviations of different respondents may be correlated.

Then

$$E\ (\bar{z}' - \bar{\mu})^2 = E\Big[(\bar{x}' - \bar{a}') - (\bar{x}'' - \bar{a}'')\Big]^2 + E\Big[(\bar{a}' - \bar{a}'') - (\bar{\mu}'' - \bar{\mu})\Big]^2$$

Also

$$E(\bar{x}' - \bar{a}')^2 = \frac{\sigma^2_w}{n}\ \{1 + (n-1)\ \rho\ \}$$

where

$$\sigma^2_w = \frac{1}{N} \sum_{i=1}^{N} \sigma^2_i$$

and

$$\rho\sigma^2_w = \frac{1}{N(N-1)} \sum_{i \neq j} \sigma_{ij}\ .$$

The response variance contribution is:

$$E\Big[(\bar{x}' - \bar{a}') - (\bar{x}'' - \bar{a}'')\Big]^2 = \frac{n-n_1}{nn_1}\ \sigma^2_w\ (1-\rho)$$

Also

$$E\Big[(\bar{a}' - \bar{a}'') + (\bar{\mu}'' - \mu)\Big]^2 = \frac{n-n_1}{nn_1}\ S^2_{a-\mu} + \sigma^2_{\bar{\mu}'},$$

where

$$S^2_{a-\mu} = \frac{1}{N-1} \sum_{i=1}^{N} \Big[(a_i - \mu_i) - (\bar{a} - \bar{\mu})\Big]^2\ .$$

Thus, finally,

$$\sigma^2_{\bar{z}'} = \sigma^2_{\bar{\mu}'} + \frac{n-n_1}{nn_1}\ \sigma^2_w\ \ (1-\rho) + \frac{n-n_1}{nn_1}\ S^2_{a-\mu}$$

where the first term, $\sigma^2_{\bar{\mu}'}$, is the term that would occur if there were no response error, the second term results from response variance, and the third term results from the sample response bias.

If no subsample is selected, then

$$E(\bar{x}' - \bar{\mu})^2 = \sigma^2_{\bar{a}'} + \frac{\sigma^2_w}{n}\Big[1 + (n-1)\ \rho\Big] + (\bar{a} - \bar{\mu})^2$$

where $\sigma^2_{\bar{a}'} = \frac{N-n}{Nn}\ S^2_a$ and $S^2_a = \frac{1}{N-1} \sum_{i=1}^{N} (a_i - \bar{a})^2$.

If, to simplify we put

$$\rho = 0$$

$$\sigma^2_{\bar{a}'} = \sigma^2_{\bar{\mu}'}$$

then

$$\Delta = E\ (\bar{x}' - \bar{\mu})^2 - E(\bar{z}' - \bar{\mu})^2 = (\bar{a} - \bar{\mu})^2 - \frac{n-n_1}{nn_1}\ S^2_{a-\mu} - \frac{\sigma^2_w}{n_1} \simeq (\bar{a} - \bar{\mu})^2 (1 + \frac{n-n_1}{nn_1}) - \frac{n-n_1}{nn_1}\ T_{a-\mu} - \frac{\sigma^2_w}{n_1}$$

where

$$T_{a-\mu} = \frac{1}{N} \sum_{i=1}^{N} (a_i - \mu_i)^2\ .$$

If $\bar{a}-\bar{\mu}$ is small then Δ is small or negative, and clearly there is little or nothing to gain from the subsampling technique. Indeed, if $\bar{a}-\bar{\mu}$ were not frequently large relative to $\bar{\mu}$, there would be no point in discussing response bias.

When $\bar{a}-\bar{\mu}$ is large, then unless $S^2_{a-\mu}$ is large relative to $(\bar{a}-\bar{\mu})^2$, it should not require a very large value of n_1 to make $\frac{n-n_1}{nn_1} S^2_{a-\mu} < .1\ (\bar{a}-\bar{\mu})^2$; even a smaller multiplier than .1 should not be difficult to attain. Thus the costs of a subsample interview may be quite large compared to that of a first stage interview and still the double sampling technique would be sufficient.

Obviously, these results also indicate that the use of design methods such as stratification and regression estimates to reduce $S^2_{a-\mu}$ will greatly increase the efficiency of the double sampling approach to reduce response bias. These will be discussed in the larger paper but they demonstrate the importance of discussing explanatory variables - a subject to which we now turn.

8. Explanatory Variables. Response error will often vary with other variables. People with low incomes tend to overreport income and people with high incomes tend to underreport i.e., the response error in reporting income is a function of the "true" income. People who see a physician frequently and recently, tend to report their illnesses better than people with the same illnesses who have seen a physician less frequently and not recently, i.e. the response error in reporting illness is a function of how often and how recently a physician was seen for the illness.

The study of the relationships of response error to the characteristics of the respondent, the interviewer, the questionnaire and the environment in which the interviewing occurred provides information of importance in

a. Understanding the data that the survey provides
b. increasing the ability to design future surveys, and,
c. improving the estimates resulting from the current survey (See Section 7.)

We shall refer to the variables to which response error is related as explanatory variables.

Explanatory variables can be classified, as in the preceding paragraph according to whether they are characteristics of the respondent, the interviewer, the questionnaire or the environment. When records or other sources outside the respondent or interviewer are used, then the explanatory variables may also be classified according to whether they are available:

a. For each member of the population

b. For various strata of the population

c. For each member of the sample from the interview

d. For members of a subsample from sources outside the interview

e. For members of a separate sample from sources outside the interview, the members of the separate sample also being interviewed

When the explanatory variables are available for each member of the population, or for various strata of the population, then the usual conditions of stratification, or stratification after initial sample selection occur.

When the explanatory variables are available from the survey itself then the double sampling procedures with stratification and regression become available. However, in such cases the explanatory variables may themselves be subject to response error; sometimes the explanatory variables may not be available from the interview. For example, the number of visits to a physician stated in medical records may be quite different from the number stated by the respondent, and better related to the error of response. There will also be information only available from the physician or medical records, and not available from the respondent. Explanatory variables not available from the interviews may be more closely related to response error than those available from the interview, but sometimes their usefulness is limited because the cost of obtaining data on them is so great that they are obtained only for the subsample (See Section 7).

When a separate survey is made to obtain information to compute an estimate relatively free of response error the variance of the estimate will be larger than would the estimate based on a subsample of equal size from the interview sample, and if made at a different time or under different conditions the response error may be quite different.

Let us now consider the explanatory variable in connection with double sampling and let us consider only the use of the explanatory variable for stratification purposes, leaving regression models for the fuller study.

The population now is assumed to consist of elements (i,j), $i=1,2,\ldots, H$, $j=1,2,\ldots, N_i$ and the response random variable for element (i,j) is x_{ij}.

Let $E(x_{ij}\ i,j) = a_{ij}$, $\sigma^2_{x_{ij}\ i,j} = \sigma^2_{ij}$, and let the "true value" for (i,j) be μ_{ij}.

To estimate

$$\bar{\mu} = \frac{1}{H} \sum_{i=1}^{H} \bar{\mu}_i$$

where

$$\bar{\mu} = \frac{1}{N_i} \sum_{j=1}^{N} \bar{\mu}_{ij} \quad ,$$

we can use

$$\bar{x}' = \frac{1}{n} \sum_{i,j \in \mathcal{S}} x'_{ij}$$

where $\mathcal{S}$ is a simple random sample of n elements selected from the N ($N = \sum_{i=1}^{H} N_i$) elements of the population.

The sample, $\mathcal{S}$, is stratified and we denote the number of elements selected from the ith stratum by n_i, and assume $n_i > 0$, $i=1,2,\dots, H$.

From the n_i elements in $\mathcal{S}_i$, where $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_H$ is a partition of $\mathcal{S}$, we select m_i elements by simple random sampling and for these m_i elements we obtain the μ_{ij}.

Let

$$\bar{z}' = \sum_{i=1}^{H} \frac{n_i}{n} \left[\bar{x}'_i - (\bar{x}''_i - \bar{\mu}''_i) \right]$$

where $\bar{x}'_i$ is the mean of x for $\mathcal{S}_i$; $\bar{x}''_i$ is the mean of x for the subsample of $\mathcal{S}_i$ and $\bar{\mu}''_i$ is the mean of μ for the subsample of $\mathcal{S}_i$, $i=1,2,\dots,H$

Then $\bar{z}'$ is unbiased, and

$$\sigma^2_{\bar{z}} = \sigma^2_{E(\bar{z}' \ (n))} + E\, \sigma^2_{\bar{z}'|(n)}$$

where $(n) = (n_1, n_2, \dots, n_H)$

It can be shown that

$$\sigma^2_{\bar{z}'} = \sigma^2_{\bar{\mu}'} + \sum_{i=1}^{H} E\left(\frac{n_i}{n}\right)^2 \left(\frac{n_i - m_i}{n_i m_i}\right)\left(\sigma^2_{w_i}(1-\rho_i) + S^2_{a_i - \mu_i}\right)$$

where

$$\bar{a}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} a_{ij}, \quad \bar{\mu}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \mu_{ij},$$

$$\sigma^2_{\bar{\mu}'} = \frac{N-n}{N\,n} S^2_\mu, \quad S^2_\mu = \frac{1}{N-1} \sum_{i=1}^{H} \sum_{j=1}^{N_i} (\mu_{ij} - \bar{\mu})^2$$

$$\sigma^2_{w_i} = \frac{1}{N_i} \sum_{j=1}^{N_i} \sigma^2_{ij}$$

$$\rho_i\, \sigma^2_{w_i} = \frac{1}{N_i(N_i-1)} \sum_{j \neq k} E\left[(x_{ij} - a_{ij})(x_{ik} - a_{ik}) \middle| i,j,k \right]$$

and

$$S^2_{a_i - \mu_i} = \frac{1}{N_i - 1} \sum_{j=1}^{N_i} \left[(a_{ij} - \mu_{ij}) - (\bar{a}_i - \bar{\mu}_i) \right]^2$$

Then, if the m_i are fixed, it follows that

$$\sigma^2_{\bar{z}'} \sim \sigma^2_{\bar{\mu}'} + \sum_{i=1}^{H} \frac{N_i}{N} \left[\frac{N_i}{m_i N} - \frac{1}{n} \right] \left(\sigma^2_{w_i}(1-\rho_i) + S^2_{a_i - \mu_i} \right)$$

while if $m_i = p\, n_i$ then

$$\sigma^2_{\bar{z}'} = \sigma^2_{\bar{\mu}'} + \frac{1-p}{pn} \sum_{i=1}^{H} \frac{N_i}{N} \left[\sigma^2_{w_i}(1-\rho_i) + S^2_{a_i - \mu_i} \right]$$

If no subsample is selected, then as in Section 7,

$$E(\bar{x}' - \bar{\mu})^2 = \sigma^2_{\bar{a}'} + \frac{\sigma^2_w}{n} \left[1 + (n-1)\rho \right] + (\bar{a} - \bar{\mu})^2$$

Now the decision on whether to use the subsampling method depends largely on the relative sizes of $(\bar{a} - \bar{\mu})^2$ and

$$\frac{1-p}{pn} \sum_{i=1}^{H} \frac{N_i}{N} S^2_{a_i - \mu_i}$$

where we are omitting the terms in σ^2_w and $\sigma^2_{w_i}$

To measure the effectiveness of the explanatory variables in general, not only if the double sampling method is used, we suggest the ratio

$$\frac{\sum_{i=1}^{H} \frac{N_i}{N} (\bar{a}_i - \bar{\mu}_i)^2}{S^2_{a-\mu}}$$

i.e. the variance between strata divided by the total variance of the individual bias, or the equivalent ratio

$$\frac{\sum_{i=1}^{H} \frac{N_i}{N} S^2_{a_i-\mu_i}}{S^2_{a-\mu}}$$

(We ignore the ratios $(N_i-1)\ N_i$.)

Similarly the effectiveness of explanatory variables in reducing bias when regression estimates are used should be measured by the ratio of the variance of $a-\mu$ about its regression line in terms of the explanatory variables.

Similar remarks hold for ratio estimates and for multiple explanatory variables using difference, regression or ratio estimates.

The use of double sampling with stratification will often lead to the desirability of using optimum strata. Considerable steps in that direction may be made from a general knowledge of the relation of the bias and explanatory variables, such as would result from earlier studies. The high cost of obtaining the μ_i will often justify multistage or sequential obtaining of the μ_i with improvements in the stratification at various stages of selection.

9. Response models. As has been mentioned earlier, the response to an interviewer's question is a choice made by the respondent. The mechanism governing that choice is relevant to the understanding of the data and the possibilities for reducing response bias.

We shall consider several cases.

Suppose that a question has q possible replies $y_1, y_2, \ldots, y_q$. Suppose that in the population there are N_i persons for whom the correct response is y_i, $i=1,2,\ldots,q$, $N_1+N_2+\ldots+N_q=N$.

Suppose that the response model has the following characteristics. Of the N_i persons for whom the correct response is y_i, M_i will give that response if selected, and the remaining N_i-M_i would choose one of all the possible responses with equal probability, i.e. with probability $1/q$.

Let u_i' be the proportion in the sample whose response is y_i . Then

$$u_i' = \frac{m_i + \frac{n-m}{q}}{n} = \frac{m_i}{n} + \frac{n-m}{nq}$$

where m_i is the number selected from the M_i whose correct response is y_i and give it, $i=1, 2, \ldots, q$ and $m=m_1 + \ldots + m_q$. Now

$$* \quad E\, u_i' = \frac{M_i}{N} + \frac{N-M}{Nq}$$

where $M = M_1 + \ldots + M_q$ and hence u_i' is a biased estimate of $\frac{N_i}{N}$

$$E\, u_i' = \frac{N_i}{N} + \frac{\left(M_i - \frac{M}{q}\right) - \left(N_i - \frac{N}{q}\right)}{N}$$

If the same proportion of guessing takes place for each response then $M_i = t\, N_i$, say, and

$$E\, u_i' = \frac{N_i}{N} - \frac{(1-t)\left(N_i - \frac{N}{q}\right)}{N}$$

which shows that if $\frac{N_i}{N} > \frac{1}{q}$ then u'_i has a downward bias and that if $\frac{N_i}{N} < \frac{1}{q}$ then u_i' has an upward bias. From * it follows that unbiased estimates of $M_1, \ldots, M_q$ can be obtained if an unbiased estimate of the total proportion guessing, $1-\frac{M}{N}$, can be obtained in a supplementary interview or through other means. If the proportion, t, guessing is constant and an estimate of t can be obtained then, the ratio estimate

$$r_i' = \frac{1}{t}\left(u_i' - \frac{1}{q}\right) + \frac{1}{q}$$

will approximate $\frac{N_i}{N}$

To compute the variance of r_i' and to compute estimates of the mean

$$\sum_{i=1}^{q} \frac{N_i}{N} y_i$$

and the variance of the estimate under the restrictive assumption is not difficult but is left to later publication.

Estimates of N_i/N without the assumption that the proportions guessing are equal for all i, appear to require supplementary estimates of the M_i/N rather than only M/N.

Clearly many variations of this "guessing model" are possible. All are subsets of a more general model which can be formulated as follows:

The elements of a population may be classified into H classes. In the h^{th} of the H classes, there are M_h elements. Designate by $y_1, y_2, \ldots, y_q$ the possible responses to a question. Then there are H response matrices $P_1, P_2, \ldots, P_H$,

$$P_h = (p_{ijh}) , \; i,j=1,\ldots,q; \; h=1,\ldots,H .$$

If an element is in class h, and his "true response" is y_i, then p_{ijh} is the probability that he will actually respond y_j.

In the preceding case, H=2; the matrix P_1 is the identity matrix of q rows and q columns and the matrix P_2 is also a square matrix all of whose elements equal 1/q. M_1, the number of elements in the first class is M , the number of persons that do not guess, and, M_2 the number of elements in the second class is N-M the number who do guess.

Obviously, the general model need not include square matrices, or can handle them by putting certain of the p_{ijh} equal to 0. Furthermore, many variants on the models can be studied; for example, the main diagonal of P_2 might be 0 and the remaining elements all equal to 1/q-1.

A second class of models may be associated to the name "learning theory". The possibilities are many and only the simplest case can be discussed here because of the length of the paper.

Perhaps it will be useful to begin by considering the following case:

A person is asked to report his illnesses. It is well known that, in general, the greater the number of times he has seen a physician for an illness the higher the probability that he will report the illness.

If the person has seen the physician h times for an illness, denote by p_h the probability that he will report the illness, and hence $1-p_h$ is the probability that he will not report the illness.

If access to medical records is available then, by use of the double sampling procedure previously discussed, it is possible to obtain estimates $p''_1, p''_2, \ldots, p''_H$.

If x'_h is the number of cases of the illness reported by persons who made h visits in the larger sample and if on the basis of the subsample it is found that the proportion in the subsample, falsely reporting the condition is t''_h, then

$$x'' = \sum_{h=1}^{H} x'_h \cdot \frac{t''_h}{p''_h}$$

will approximate the "true number" having the illness. The mean square error of x" about the true value can be computed without great difficulty.

Let us suppose that the probability that the respondent who has made h visits to the physician became aware of the illness on visit g but not before, is

$$(1-p)^{g-1} \; p, \; g=1,2,\ldots$$

Then the probability that the person who makes h visits to the physician will know his condition is

$$* \quad p_h = 1 - (1-p)^h \; , \quad h=1,2,\ldots, H$$

where p is the probability that a visit results in the respondent learning of the condition.

To estimate p will require a smaller sample than to estimate $p_1, p_2, \ldots, p_H$. Furthermore as a means of guiding improvements in questionnaires and interviewing, the fact that a learning model may be approximately correct would be very useful.

To extend the model summarized by * to include the effects of time since the last visit to the physician would be important in discussing response errors in reporting illnesses; we are here more interested in indicating the use of learning and other behavioral science models as a means of understanding and reducing response error.

This paper, already too long, is part of a larger study on response error which is intended to deal with various aspects of response error measurement, understanding and reduction. The basic approach consists in formulating survey models that include response choice models in order to be able rationally to decide how to allocate the resources of a survey. In so doing the suggestion is made that even artificially simple models for the conclusions, decisions or actions to be taking on the basis of a survey may be adequate for guiding the design of the survey if used with judgment - even though these models are not actually used in the analysis of the survey when made.

SOME METHODOLOGICAL ISSUES IN VALIDATION STUDIES*

Harold W. Guthrie, University of Illinois

Sample surveys designed to collect financial data from households have become an established part of the methodology of research on consumer behavior over the last twenty years. Yet the continued failure to obtain sample data which are consistent with data obtained from other sources has been a serious handicap to researchers advocating new approaches to analysis of consumer behavior -- approaches for which sample surveys are uniquely appropriate. This paper attempts to review briefly the present state of the arts in collecting financial data and to point out some methodological problems which we encounter in attempts to validate survey data. Since sample survey methods are shared by social scientists in several disciplines we have a common interest in research on methods of improving our survey techniques.

Analysts of household financial data benefit from a wealth of data for purposes of comparing estimates of aggregates derived from survey data with parameters measured by other means. On some items the discrepancies are not intolerable; on others, including some very important behavioral variables, the discrepancies are staggering. Since there are differences in coverage between the aggregate data and the survey data, comparisons require skillful adjustments. The resulting comparisons are approximations but they are useful signposts pointing toward weaknesses and strengths in the survey data.

On one important variable, income, the survey data have been reasonably good. The 1950 BLS Survey of Consumer Expenditures underestimated the aggregate for total income by only about 6 percent. The Surveys of Consumer Finances conducted by the Survey Research Center, University of Michigan, underestimated income from 3 percent to 13 percent over the period 1947-1955. While the estimates of total income can be regarded as within a tolerable range of error, the estimates of individual components of income are less encouraging. Data from the 1950 BLS study overestimated entrepreneurial income by 21 percent and underestimated interest income by 74 percent.

Another example of what seems to be a compensation effect in an aggregative measure is provided by saving, the flow of money income into financial and other assets held by households. Saving is one of the more important single dependent variables in analysis of household financial behavior, and our comparisons here are based on the pioneer work of the Survey Research Center in this area. Over a four-year period, 1947-50, the sum of the amounts of saving estimated by the surveys exceeds the comparable sum of amounts of personal saving reported by the Securities and Exchange Commission by only 4 percent, but the two major components of saving show gross discrepancies. The survey data estimate saving in the form of liquid assets at $-22.5 billion against the SEC aggregate of $+7.9 billion. This shocking underestimate is compensated in the measure of total saving by a survey estimate of non-liquid saving in the amount of $71.3 billion against the SEC aggregate of $39.4 billion.

In the early days of sample surveys, attempts to measure the flow of saving were based on the recall method. That is, respondents were asked, for example, the balance in their checking accounts as of a given date, and also the balance as of the same date one year prior to the given date. The recall method was largely abandoned in later surveys in favor of a reinterview procedure. Under this procedure respondents were interviewed in two successive years and they were asked for their balances for given dates, one year apart, in each interview. The flow of saving was therefore measured as the difference in the two reports of holdings in the given asset. This procedure focused attention on the wealth holdings of families, or taking into account their debts as well, on their net worth position.

Survey data on holdings of assets and debts, when compared with similar external data, suggest some of the methodological problems now under attack by many people and institutions concerned with surveys of household financial data. The survey data estimate the number of automobiles owned very well. Proceeding down the scale with other examples the survey data underestimate the number of checking accounts by 24 percent, the number of savings and loan accounts by 64 percent, the value of stock holdings by 75 percent.[1]

The problem of improving our sample data on household finances is now under attack in two major areas. The first area concerns sample design. We know that the size distributions of many of our important variables, such as holdings of common stock, are highly skewed to the right. The sample designs used by the Survey Research Center in the early years of household financial surveys stratified households by a measure of income and over-sampled high income families in an attempt to be able to describe this segment of the population. Since the results did not measure up to the expectations of many analysts, stronger medicine seemed to be called for. In 1963 the Board of Governors of the Federal Reserve System conducted a Survey of Financial Characteristics in which, with the cooperation of the Bureau of the Census, a much more drastic scheme of stratification was used. The top income stratum of this

*In preparing this paper I have used unpublished material prepared by my colleagues, Robert Ferber and E. Scott Maynes. I am grateful for their permission to do so.

[1]Federal Reserve Bulletin, September, 1958, pp. 1041-51.

study includes households in which income is $150,000 or more. The data from this study should tell us much about the potential gains in reliability to be expected from a high degree of stratification of highly skewed populations.

The second major effort to improve the reliability of household financial data is a growing number of validation surveys. The remainder of this paper summarizes some of the recent developments in this area and points out some methodological difficulties in conducting validation surveys.

The prime mover in several important studies of errors of response and nonresponse in household financial data has been the Inter-University Committee for Research on Consumer Behavior, a committee of scholars who have an active interest in using survey data.[2] This Committee, through its Consumer Savings Project, has sponsored three major studies. The first of these studies was conducted by the Survey Research Center under the direction of John B. Lansing. The report of this study states the urgency of the need for validation research: "When the problem is one of obtaining the most accurate possible estimate of a parameter, and the interviewing technique being used tends to result in an underestimate of that statistic by 25 to 50 percent, it is not an efficient use of resources to take a large sample. Money spent on a large number of interviews could much better be spent on a smaller number of interviews plus a study of response error."[3] The studies of response error should be aimed at measuring the errors and developing analytical techniques by which biases might be taken into account in analysis of the data.

A second major phase in the studies sponsored by the Inter-University Committee was a set of panel studies directed by Robert Ferber of the University of Illinois. In the panel studies in three midwestern cities and a farm area several split-sample experiments were designed to discover survey methods with which errors of response and nonresponse were correlated. The samples for these separate surveys were drawn from lists of families who were known to hold savings accounts, checking accounts, consumer debt, life insurance, or farm debt. A third major study under the Consumer Savings Project is directed by Ferber and conducted with the assistance of the Response Research Branch, Bureau of the Census. This study focuses attention on two of the items of household finances which have large biases: savings accounts and holdings of common stock. No built-in experiments on methods were included in this study. Rather, an attempt was made to approximate as closely as possible the field operations and data processing methods used in the Federal Reserve Board's 1963 Survey of Financial Characteristics.

The findings of these studies have been, or will be, reported elsewhere. I shall draw upon these studies to illustrate some methodological problems encountered in validation studies. I shall also attempt to point out some analogues of these problems in validation studies conducted in substantive areas other than household finances.

The first issue that we must cope with in financial validation studies is the greater complexity of our variables as compared with the variables of primary concern in studies of medical conditions or voting behavior. In the latter two kinds of studies, the variable of primary concern is essentially an attribute. Does the element have an illness? Is the element a registered voter? Did the element vote in a given election? Our primary variables are of a similar nature but we also need to measure errors in the amounts. First, does the element have a savings account? Second (and like which kind of illness) is it in a commercial bank, a savings and loan association, a building and loan association, or some other kind of institution? Finally, how much was the balance as of a given date? The bias observed in the comparisons of aggregates can be attributed, in part, to errors at any one of these three stages of specification of the variable. Validation studies aimed at explaining the bias must therefore break down total nonsampling error into at least three components: errors of nonresponse, reporting error, or error in reporting the attribute, and response error, or error in stating the amount of the variable.

Two findings from recent validation studies are related to the nature of the variable and the reduction of response errors in the observation of values of the variable. First, many respondents are sufficiently motivated to give correct values, but they simply do not remember the information required. We have found that it is feasible to ask respondents to check their records so that they can give accurate data. A second useful finding is that a second interview frequently can correct errors resulting from misunderstandings existing at the time of the first interview. A similar finding was noted for the study of data on hospitalization in the National Health Survey.[4]

Since repeated interviews have become a standard procedure in studies of financial data we have also extended our validation research into panel studies involving as many as 5 waves of interviews at intervals of 3-6 months. Although panel studies suffer from declining response rates, Ferber's work has shown that there is a substantial reduction of reporting error in

[2]Members of the Inter-University Committee are: Guy Orcutt, Chairman, Lincoln Clark, Robert Ferber, George Katona, Theodore Newcomb, Howard Raiffa, and James Tobin. Raymond Goldsmith was a member of the Committee until his departure to join OECD in Paris in the summer of 1963.

[3]John B. Lansing, Gerald P. Ginsberg, Kaisa Braaten, An Investigation of Response Error. Studies in Consumer Savings, No. 2 (Urbana, Ill.: 1961), p. 201.

[4]U.S. Public Health Service, Health Statistics from the U.S. National Health Survey, Series D, No. 4. (Washington, D.C.: 1961), pp. 39-52.

successive waves.[5] Apparently those who remain in the panel become increasingly cooperative and more willing to report the existence of assets and debts. Whether the gain in reliability is worth the price of panel mortality is still to be resolved.

A second methodological issue concerns the unit of analysis and the interviewing unit. The obviously relevant unit of analysis in a study of election behavior is the individual person. Although the analyst might be interested in interactions between persons to find out how behavior is determined he needs to be concerned only with the individual for validation purposes. Similarly in studies of health conditions the primary variable is unique to the individual person. Given reasonably good identification data the matching of an interview report with a validation record is not too difficult.

The relevant unit of analysis in studies of household finances, both for substantive analysis and validation studies, is a group of persons, either the spending unit or the family. The necessity of observing the larger unit arises, in part, from the complexity of ownership of assets. Several persons in the same family may own a given type of asset, and joint ownership of savings accounts and corporate stock, not only by husband and wife, but by many different combinations of family members, is very common. While the legal status of the ownership of the asset is clearly defined by the records of the financial institution, the respondent's perception of ownership is frequently much less clear. Wives, aged parents, and children are often not aware of assets which they own singly or jointly and the ownership patterns frequently extend outside the immediate family.

We find that nonsampling errors can be measured meaningfully only if we consider the pattern of ownership of multiple units of the given asset among all members of the reporting unit. The several reported units of the asset must frequently be matched with an unequal number of units reported by the financial institution. The discrepancies between the two reports with respect to ownership require a complex matching of validation reports with respondents' reports.

Given that any of several members of a family may be holders of assets, which one should be designated as the respondent? The increased response error resulting from using a single respondent as a proxy for other family members was noted for the National Health Survey.[6] We face the same hazard in financial surveys. The general practice in financial surveys has been to choose the husband as the respondent in families containing married persons, and the wife or other family members are allowed to be present during the interview. The presence of other persons in the family during the interview improves the accuracy of response in some cases, and reduces accuracy in others.

Two recent developments in this area are worth noting. In the Federal Reserve Board's Survey of Financial Characteristics respondents with extensive and complex wealth holdings were encouraged to ask their financial advisers to fill out a self-enumeration form. For many wealthy families this seems to be the only feasible way to obtain accurate data. The second development is an attempt in a validation study to correlate response errors with the degree of participation of various family members in the interview. The results of this attempt are not yet available.

A third methodological issue is the link between the individual interviewing unit and the aggregate which is the parameter to be estimated. Again, a contrast between validation studies of household finances and studies of voting behavior and health studies illustrates the difficulty. Registration for voting provides ideal conditions for a validation study. Here the validation records are open to the public and local. The aggregate number of registrations to be estimated for a county can easily be related directly to a validation study of individual persons and their interview reports. Validation of voting behavior in terms of how votes were cast is, of course, made difficult by the lack of a validation record. Health studies can also be locally oriented because the market for health care is generally a local market.

One very important consequence of the local and open character of validation records is that both primary and secondary validation can be performed in a single study. That is, the study can determine both the errors of underreporting (respondent reports that he is not registered when, in fact, he is) and overreporting (respondent reports he is registered when, in fact, he is not).

Only rarely do we find very favorable conditions for validation of household financial data. First, the records necessary for validation are never public records, and, indeed, the sources of the financial records are generally business firms who take seriously the ethic of guarding confidential personal data. Moreover, the firms are not easily convinced that validation research is a legitimate justification for their spending the necessary effort to cooperate in a validation study. Second, for many kinds of financial variables, the market is national rather than local. An ideal validation study of life insurance holdings, for example, would require access by the researcher to the records of hundreds of insurance companies. Perhaps at the other extreme are certain kinds of consumer credit for which the risk element requires a local market rather than a national market.

Up to the present time our validation studies have not attempted to cover completely a single local area. We have conducted primary validation

[5] Robert Ferber, "Does a Panel Operation Increase the Reliability of Survey Data: The Case of Consumer Savings," 1964 Social Statistics Proceedings of the American Statistical Association, pp. 210-216.

[6] U.S. Public Health Service, op. cit., p. 8.

studies from known holders of given assets in a given institution. One important cost of using this approach is that we can say very little about the incidence of overreporting. We know that overreporting is typical for some behavioral characteristics as shown by the Survey Research Center election studies which measure voting turnout.[7] A related error is also indicated in financial validation studies which show that some respondents overstate the amount of their asset holdings.

Although overreporting of an attribute may be an important phenomenon in other areas of inquiry, we feel that it is not a serious source of error in financial data. We are sure that our net biases are negative because of the results of comparisons of aggregates. Also, in two surveys under the Consumer Savings Project in which secondary validation was attempted, no clear cases of overreporting occurred. Additional evidence that overreporting of amounts, if not attributes, can be reduced by reinterview is provided by Lansing's finding that gross overreporting was almost eliminated by a second interview.[8]

A fourth methodological issue arises from the fact that one important objective of validation studies is to develop interviewing procedures which will reduce nonsampling errors. Most validation studies of household finances have collected data on many different characteristics of the interview situation. Some attempts to correlate nonsampling errors with these characteristics have suggested possible minor improvements. Many of us who conduct surveys, however, look to the interviewer-respondent interaction as the most plausible explanation of errors. We tend to think that if we could just get perfect interviewers and field operations our error problem would be much smaller. So far in our validation studies of household finances we have not made much progress in this area of research. Our validation studies have been conducted within the context of normal procedures for interviewer selection and field control. We have included in our survey design neither controlled experiments nor random assignments of sample addresses to interviewers. The development of special purpose studies which combine validation, experimental controls, and random assignment of interviewers to respondents is one task which we, and perhaps researchers in other fields as well, might undertake in the future.

Although I advocate making an attack on the problem for the same reason that men climb mountains, I do not expect any significant results. There is undoubtedly a large number of possible types of respondent-interviewer interactions, and advance data about the respondent to guide optimal assignments of interviewers to respondents is usually scanty. Thus we would have to find a very strong and strategically useful relationship between nonsampling errors and type of respondent-interviewer interaction before these research efforts would pay off in substantial gains.

The last issue to be considered is partly methodological and partly theoretical. We noted earlier that comparisons of survey based aggregates with external aggregates produced widely variant results. Examples illustrated a range from an underestimate of about 75 percent for stock holdings to an underestimate of about 6 percent for total income. The data suggest a continuum extending into overestimation. Our ultimate goal in studies of nonsampling error might well be an explanation of differences in degree of error along the continuum and among different items of information.

For example, the underreporting of mental and nervous disorders in the National Health Survey[9] and the underreporting of financial data may have a common cause, a desire to suppress secret information. One of Lansing's results throws some light on this issue. He found that respondents who grew up in families in which financial information was concealed from the children tended to conceal information from the interviewer.[10] Perhaps mental and nervous disorders and holdings of common stock are not very far apart on a continuum of sensitivity and, if so, it may be that similar techniques to reduce response errors will be successful for both kinds of information.

To summarize this brief review of methodological issues, we in the sample survey field have made substantial progress in measuring nonsampling errors but we must go much farther. We have learned how to conduct useful but imperfect validation studies by paying careful attention to some methodological pitfalls. We have persuasive results to indicate that motivation of the respondent is important, but that we must also take into account errors due to lack of information, deliberate concealment, and other causes. These concepts form the basic structure of a theory of nonsampling errors. Now we must search out and identify specific reasons for differences in errors among different items of information in order to develop feasible remedies. Thus we must take advantage of every opportunity to incorporate validation research into surveys conducted primarily to obtain substantive data. At the same time, some methodological issues, such as the interaction between respondent and interviewer, seem to require specifically controlled validation studies.

[7] Angus Campbell, Philip E. Converse, Warren E. Miller, Donald E. Stokes, The American Voter. New York: 1960. pp. 93-96.

[8] John B. Lansing, et al., op. cit., p. 186.

[9] U.S. Public Health Service, op. cit., p. 54.

[10] J. B. Lansing, et al., op. cit., p. 184.

DISCUSSION

Philip J. McCarthy, Cornell University

Producers and consumers of survey data (whether derived from a sample or census) have long been concerned with errors of measurement, particularly those that may produce bias in the survey estimates. It has also been well recognized that independent and random response errors would contribute to the variance of estimates and would be mirrored in the ordinary variance estimation schemes, except for finite population effects. The papers presented at this session summarize and extend some of the recent developments in these areas.

In reviewing the present papers and some of the material to which they refer, I have been impressed by four main points. These are:

1. Considerable ingenuity and care have been devoted to the development and application of models of "response variance," particularly by the staff of the U. S. Bureau of the Census. Large-scale applications of these models have been made in connection with the Re-interview Program of the Current Population Survey and the 1961 Canadian Census of Population. The models, by themselves, demonstrate that the effects of correlated "response variability" decrease only slowly with sample size, and that ordinary sample variance estimates are understatements of true variability. Furthermore, studies of response variance appear to be able to pinpoint areas in which there are difficulties in the measurement process.

2. Numerous large-scale validation studies have been carried out in several areas. Dr. Guthrie has referred to those in the economic area. The National Center for Health Statistics has also published accounts of validation studies relating to material collected by interview in the National Health Interview Survey -- e.g. on the accuracy of reported hospitalizations. It appears to me that all of these efforts can only serve to improve the available data on a particular subject, and that these concerns are becoming more and more widespread.

3. On the negative side, I have also been impressed by the fact that validation studies usually come up with highly qualified conclusions, indicating ways in which a particular type of data can be "improved", and warning that the results cannot be applied in an unqualified way to other situations and circumstances. In this respect, I can't see that we are any nearer a theory of non-sampling errors, to which Dr. Guthrie refers, than we were 10 or 15 years ago. The individual or organization conducting a single survey can draw somewhat upon this accumulated experience, but can in no way guarantee results "free from measurement bias," any more than he can produce one hundred per cent coverage of sample cases.

4. Dr. Madow has presented us with a penetrating discussion of these problems, and has contributed the details of a model by means of which a survey may, as an integral part of its design, contain its own small validation study. This is, of course, particularly welcome in view of the previously noted fact that generalizations from a specific validation study to a new survey are extremely difficult to make. It is also true, however, that even small validation studies are expensive, and although validation results may be better than ordinary measurements, they will probably never be perfect. Dr. Guthrie's paper illustrates these points very well.

In conclusion, I would simply like to call attention to the fact that, contrary to what some may believe, not even all of the formal statistical problems associated with sample survey design and analysis are as yet solved. We do have a large body of theory with which to produce efficient estimates of population parameters. We have, however, yet to produce completely satisfactory bridges between sampling and estimation techniques (stratification, two-stage sampling, ratio estimation, etc.) and standard techniques of analysis (analysis of variance, regression, nonparametric approaches, etc.)

X

PROBLEMS OF DESIGN IN HEALTH DEPARTMENT SPONSORED ILLNESS SURVEYS

Chairman, PHILIP S. LAWRENCE,
National Center for Health Statistics

BALTIMORE HEALTH SURVEY

Elizabeth B. Kelley, Baltimore City Health Department

Administration of public health programs requires a background of information today quite different from that considered necessary three decades ago. At a time when the dominant problems were communicable diseases for which no effective preventive agents were known, the medical intelligence needs were appropriately met by a vital events registration system and a reportable disease notification system.

The successful development of vaccines and inoculants against diphtheria, pertussis, tetanus, poliomyelitis and measles, and the reduction of the importance of such diseases as pneumonia, streptococcal infection, and tuberculosis have changed considerably the direction of public health interest and effort. This change, towards a concern for eradication of diseases for which a specific preventive agent exists, towards an improvement in general medical care, and towards a heavy involvement with chronic and degenerative diseases, has set forth new requirements for data necessary to guide rational decision making.

By and large the new data requirements have centered about the need for (1) estimate of inoculation levels among children, (2) the need for information on the current practices of the population in respect to use of medical care facilities and (3) a requirement for information on the prevalence of acute respiratory disease. By and large the requirement for a high level of precision in such data has not been urgent, for the number of alternative decisions to be considered are limited.

Soon after the Public Health Service launched its National Health Survey the Research and Planning Section of the Baltimore City Health Department felt that a continuous survey of the population could provide the Health Department with a new intelligence system for certain types of information. It also felt that for such a survey to be a feasible and practical undertaking for a local Health Department it must be designed to operate within the existing framework of the Health Department staff.

Aware that accuracy of response depends upon the respondent's knowledge of and willingness to impart the desired information, it was decided to limit the objectives of the survey to (1) obtaining an estimate of the inoculation levels against diphtheria, tetanus, pertussis, and poliomyelitis among the child population, (2) studying the epidemiology of acute respiratory disease, (3) obtaining basic demographic information including data on the mobility of the population, and (4) obtaining information regarding such questions of timely importance as (1) the day care provided for children of working mothers, (2) safety practices in the home and (3) sources of medical care for low income children.

Originally consideration was given to the use of either the public health nurses or the sanitarians as interviewers. The public health nurses were selected on the basis that their common educational background and familiarity with taking personal histories argued for accuracy of response. The Nursing Department agreed to assume the interviewing task for a continuous survey provided that the load was one which could be absorbed into their ongoing routine.

It was decided that 100 families could be reached each month. A sample of 100 households was expected to yield information on approximately 300 persons a month. This amount of information was considered sufficient to meet the requirements for precision for the types of data sought.

Since the population under study was to be the non-institutionalized population residing within the Baltimore City limits, the city directory (Polk) was available to serve as a sample frame. The sampling procedure developed calls for the systematic selection of 1200 dwelling units per year from this directory. These 1200 Households are then systematically sub-sampled to obtain the 100 interviews to be conducted during each month. Other sampling schemes were considered, namely that described by Serfling and a similar modified cluster sampling plan, but it was found that the anticipated savings in travel time did not materialize. The public health nurses had time limitations and could not interview all units in a cluster at one time.

The survey, inaugurated in January, 1960, has now been in operation for five full years. Each year during this period approximately 83 per cent of the selected households have been successfully interviewed. The refusal rate among households eligible for interview has been between 4 and 5 per cent, which is very close to the rate experienced by the National Health Survey. Approximately 57 per cent of the successful interviews have been obtained at the time of the first visit to the household, but up to three call-backs have been required in order to maintain the

83 per cent completion rate. On the average, approximately 147 visits are made to obtain 83 completed interviews during a one month period. Each month between 15 and 25 interviews must be performed in the evening by a part-time professional interviewer hired for the purpose of contacting those households which could not be reached during the nurse's usual working day.

A check on the validity of the sampling scheme is provided by a comparison of the age and racial distribution of the population in the sample with that derived from the annual estimate of the city's population based on natural increase and migration. For example, during 1964, 39.5 per cent of the persons in the sample were under 20 years of age compared to 38.4 per cent in the annual population estimate for 1964. The proportion of non-white persons has been slightly higher in the sample--42.3 per cent compared to 39.3 per cent in the annual population estimate.

There has been some concern regarding the problem of updating the city directory. From 1960 to 1964 building permits were issued for approximately 16,000 new dwelling units which were not in the directory used for the sample. Arrangements have been made with the Department of Building Inspection to obtain listings of new housing and demolitions so that these changes may be incorporated into the sampling frame.

The sample size of 100 households per month is sufficient to produce estimates of the prevalence of acute respiratory illness with a standard error of estimate of less than 5 per cent for any month. However, it is necessary to collect information over a one year period in order to obtain estimates of inoculation levels among pre-school age children with this degree of accuracy. More important is the fact that, for meaningful program planning, reliable estimates of these inoculation levels are required for subdivisions within the city. Unfortunately, the present sample size does not allow for this.

Although the sample size is sufficient for the determination of the prevalence of acute respiratory illness during a one month period, the accuracy of the prevalence figure presupposes an even sampling over the month. This we have not been able to obtain since the interviewing must be performed as it can be fitted into the nurses working schedules. This lack of scheduling prevents early identification of changes in the prevalence trend.

The use of public health nurses as interviewers has not produced any significant other problems. Initially each nurse is carefully briefed on the purpose of the survey, the procedures to be followed, and the need to avoid interviewer bias. We feel that there are some distinct advantages to having the nurses as interviewers as well as some disadvantages. Among the advantages are (1) the expense of employing professional interviewers is eliminated and (2) their knowledge of vaccines and how they are administered results in more accurate information on inoculation status. Among the disadvantages are (1) the nurses are responsible to a department other than that responsible for the function of the survey and consequently close control of the fieldwork is not possible other than through immediate checks of the completed questionnaires, (2) the assignment of the nurses to definite geographical areas prevents call-backs for repeatability checks, and (3) each nurse receives only 1-3 interviews to complete each month which prevents the development of close familiarity with the survey.

In spite of some of its shortcomings, the survey has produced some distinct advantages. For example, the survey sample developed over the five year period has been used several times for special ad hoc surveys--one carried out in the lower socio-economic areas of the city in an attempt to identify pockets of young children with inadequate protection levels against poliomyelitis and one carried out on a city-wide basis to evaluate the adequacy of existing child day care centers to provide care for the children of working mothers. Recently, the sample was used in a city-wide survey to determine the proportion of young children who are still at risk to measles.

THE SAMPLE DESIGN OF THE NEW YORK CITY POPULATION HEALTH SURVEY

Irving Sivin, New York City Health Department
Paul M. Densen, New York City Health Department

I INTRODUCTION

In the first week of January, 1964, New York City's Department of Health began a probability sample survey of the city's residents. Its chief purpose was to provide hitherto unavailable data, on a continuing basis, about the health status of the population of the city. Because the Health Department is an important provider of medical services to the community, we wanted to obtain a clear picture of how the public and certain groups within it obtain their medical care. Consequently we focused on the area of medical economics. Our questions were designed to determine not only the amount of medical care received by New Yorkers and the diseases for which this care was required, but also to determine the auspices under which it is given, and how families finance their care.

The survey was conceived as a necessary supplement to the traditional vital and service statistical activities of the Department. We hoped that the data obtained in the survey would prove useful in measuring the effectiveness of the Department's current service programs, and could aid in planning new programs. We also felt that the creation of a sampling unit within the Department would provide it with the materials and skills required to conduct AD HOC surveys as the need arose.

The decision to undertake a continuing health survey of relatively large dimensions was also made in the belief that the data collected would prove useful to other city agencies as well. Reliable current data on population, household size, family income, migration, etc. are not available in late inter-Censal years on a local basis, although there is considerable need for them. Because the Population Health Survey would collect such data on a routine basis, we felt the responsibility to make these measurements with as great precision as possible.

It is the purpose of this paper to describe the sample design and data processing controls which were used to achieve the goals set for the survey.

II SAMPLE DESIGN

A. FRAMEWORK FOR THE DESIGN

Sample design should be married to the purposes for which a survey is undertaken, and should utilize the available resources in the most efficient manner. The initial decision to undertake a health survey of the City was made against a background of limited resources to defray the costs of interviewing, data processing and publication of the data. The survey staff would be regular Civil Service employees of the Department of Health. Detailed sample design had to fit into this framework.

It was decided that the most useful way to investigate the health status of the population in the context of a household survey was to ask questions similar to those used in the National Health Survey. The broad outline of the sample design was also to follow the National Health Survey's design. New York's survey would be an annual sample, divided into 52 equal subsamples. One of these subsamples would be interviewed each week throughout the year. From one year to the next, a different, but neighboring, set of households would be interviewed.

The population to be surveyed was the civilian non-institutional population of the City.

B. DETERMINATION OF SAMPLE SIZE

The size of our budget, when set against the expected cost of interviewing, indicated that the upper limit of our sample size would be about 7500 households per year. Our principal problem was to determine the minimum number of households that would satisfy our needs for reliable detailed data on an annual basis.

The economics of interviewing and the resources at our disposal demanded that we employ a cluster sample. Therefore we had to explore the relationships between the size of cluster and its sampling efficiency, and the costs of interviewing together with the cost of the delineation of clusters. We judged, on the basis of response data available to us from the Washington Heights Survey in New York, that the cost per interview would be relatively invariant for clusters of four or more households and would be about $6 per household. We also concluded that the cost of the delineation of clusters would be proportional to the number of clusters in sample, about $5 per cluster. We interpreted variance data from the National Health Survey to mean that for health characteristics, larger clusters would not seriously inflate the sampling errors, when compared to the cost advantages of having fewer of them. Our principal concern about large clusters lay in their inefficiency with regard to socio-economic information.

Before making a decision with regard to the size of the cluster to be employed, we decided to establish a list of the key statistics which were to be derived from the survey. We hoped that this process would indicate, in terms of simple random sampling, how large a sample we should have, and from this vantage we could extrapolate to the size of cluster sample (and of cluster) that we needed.

The list of the survey's chief concerns was based on questions submitted to us by the heads of major units within the Health Department. Our survey sought to answer the following questions:

1. How do different income and ethnic groups within the City finance their medical care? For these groups, what proportion of physicians' services are financed out of pocket?
2. What proportion of families have hospital insurance for all members of the family? How does this proportion vary by family size, ethnic group and income?

3. What proportion of families have one or more family members hospitalized during the course of the year? How is the length of stay affected by family size, income, hospital insurance and type of disease?
4. What proportion of out-patient medical services in the City are provided by governmental agencies?
5. How many physically handicapped persons are there in New York?
6. How many physician visits are made per person per year? How many dental visits are made?

These questions implied that comparisons would be made between the characteristics of different ethnic and socio-economic groups. Our goal therefore was to be reasonably certain that valid comparisons could be made. We felt that this goal would be achieved if an observed difference of 10% in a characteristic for two different socio-economic groups would prove statistically significant.

The smallest ethnic group for which the Department wished detailed information was the Puerto Rican population, which constituted about 8% of the City's population in 1960. The next smallest ethnic group for which detailed data were desired was the Negro population. There are about twice as many Negroes as there are Puerto Ricans in New York. A simple random sample of families would yield about twice as many Negro families in sample as Puerto Rican families, and the reliability conditions outlined above would be satisfied approximately by the equation

$$\frac{PQ}{n} + \frac{PQ}{2n} = (.05)^2$$

Since the highest value of PQ is .25, a simple random sample that yielded 150 Puerto Rican families and 300 Negro families would suffice. The total sample size needed to produce these numbers would be 1875 families. Were cluster sampling to prove only one-third as efficient as simple random sampling in obtaining the economic characteristics of these groups, then a sample of about 5625 families would be required. We speculated that cluster sampling would operate at about this efficiency, if the cluster size were between 6 and 10 units.

A somewhat different orientation was given to the problem of sample size when we realized that data from the survey would be used as an adjunct, if not the principal source, in obtaining current estimates of the total population of the City. We felt that estimates from the survey would prove useful if the total population could be estimated with a relative error of 2% or less, that is, with a two sigma error of about 300,000 persons. We felt this to be adequate because by the end of 1963 the difference between the Census Bureau's projections of the population and those of the City Planning Commission had risen to about 280,000 persons.

To achieve a relative error of 2% on the estimate of the total, as well as for the sake of providing uniform interviewer assignments, it was necessary to make the clusters contain approximately equal numbers of housing units and of persons. Doing this would diminish the variance between cluster totals which is the principal source of sampling error in estimates of the population. The first and chief step in controlling the size of cluster could be made by using Census Blocks as our primary sampling units and selecting these with unequal probabilities, proportional to the number of units reported in them by the Census in 1960.

To achieve a relative error of 2% for the population total, we felt that it would be necessary to achieve an error of about 1% on the estimate of the total number of housing units in the City. If we could obtain a coefficient of variation of the size of cluster of .25 through the use of PPS sampling and careful delineation of clusters, then 625 clusters would suffice to yield an error of 1% for the entire sample. We felt that it would be possible to do this using clusters of about eight units. It seemed to us that were the cluster size smaller, there would be a higher underlying coefficient of variation, which would necessitate the inclusion of more clusters in sample.

We, of course, had to take into account not only the variation of the number of units per cluster, but also the variation of the number of people within the units. We assumed that the relative variance of the estimated population total would be a function of the form

$$V^2_{\substack{\text{Total}\\\text{Population}}} = V^2_{\substack{\text{Total}\\\text{Housing Units}}} + V^2_{\substack{\text{Persons}\\\text{Per}\\\text{Unit}}}\left(1+\delta(\bar{n}-1)\right)$$

where V^2 is the relative variance
δ is the within-cluster correlation coefficient of persons per unit.
$\bar{n}$ is the average number of units per cluster.

We assumed that the relative variance of the number of persons per housing unit was equal to 0.4, a value somewhat higher than the figure for the number of persons per occupied unit, which we derived from the 1960 Census. Under the assumption that cluster sampling would be only 50% as efficient for this characteristic as simple random sampling, $1 + \delta(\bar{n} - 1)$ equals 2.0. Therefore we would expect that 625 clusters of 8 units would yield a relative error of 1.6% on the estimate of the total population. Under the assumption that the efficiency would prove to be only 33% of simple random sampling, that is with $1 + \delta(\bar{n} - 1)$ equal to 3.0, a sample of 625 clusters of eight households would yield a relative error of 1.9%, on the estimated total population.

Ultimately, as a compromise between the indicated values, and for the sake of simplicity in estimation, we agreed on a sample with an overall fraction of one in 500. This would produce about 5700 housing units and 5400 households in sample each year. The sample would have about 700 clusters of eight housing units each.

C. GEOGRAPHIC STRATIFICATION

We arrived at our sample size under a set of assumptions which are relevant to a simple random sample of clusters. Actually, from the very beginning we had envisaged that the Population Health Survey would be a stratified sample, in which the strata were to be the 30 Health Districts into which New York City is divided. Stratification would ensure that the representation of the various socio-economic and demographic groups in the City would be close to their level in the population. Because the Health Districts differ greatly in their demographic and socio-economic composition, they also differ greatly in the type and amount of public health services provided by the department. We desired to produce data, from time to time, on the health and medical care characteristics of groups heavily and lightly serviced districts, and therefore needed stratification.

The introduction of geographic stratification would not greatly affect the overall sample size we required for inter-group comparisons. For city-wide estimates, stratification along geographic lines, would produce at best only modest gains in the overall efficiency of the survey, and we felt no need to reduce the sample size on this score.

D. EFFECT OF ESTIMATION PROCEDURES ON THE SURVEY DESIGN

The processing resources at our disposal precluded the use of all but the simplest estimation procedures for the data from the survey. Both for the sake of the efficiency of the sample and for simplicity in processing, we determined to have a self-weighting sample. All estimated totals derived from the survey would therefore, be the sample totals multiplied by the reciprocal of the overall sampling fraction. All estimated rates, percentages or proportions would have sample totals for both the numerator and the denominator. Since these simple techniques of estimation were to be employed, we could not count on any gains in the efficiency of the survey through their use.

E. NEW CONSTRUCTION STRATA

In a survey where the estimation procedures are simple inflations of survey results or just sample proportions, the actual efficiency achieved by the sample rests heavily on the amount of detailed work entering into the final design, and the degree to which the physical realization of the sample conforms to the blueprint.

To insure that the totals estimated from the sample have the precision we required, we realized that we had to create a separate selection frame, unrelated to the Census Block Statistics from which the sample in our geographic strata would be chosen. We needed a second sampling frame because New York adds about 30,000 new housing units to its inventory each year, a rate better than 1% per annum. Most of these additions are in the form of large apartment buildings or large developments of private homes. Were only one of these large developments to appear unexpectedly within a sample cluster, the precision of our estimate of the total number of housing units in the City would easily be cut in half. For example, let us assume that, aside from the presence of a single cluster containing 60 housing units, the sample would have achieved its goal: a coefficient of variation of the number of units per cluster of .25 on an individual cluster basis. Then, the presence of the single large cluster would have increased the relative variance of the sample of 700 clusters from .000089 to .000174, an increase of 94%. The argument is detailed below.

Let $\bar{x} = 8$. This is the average number of units per sample cluster, provided no large cluster is encountered.

Let $\sigma^2 = 4$. This is the variance of the number of units per cluster under the assumption that we have achieved a coefficient of variation of .25 in the size of the cluster. The relative variance is .0625 on an individual cluster basis.

The relative variance of a sample of 700 clusters under these assumptions is .000089

Now in a sample of 700 clusters which had achieved the above characteristics, the value of the sum of squares is 47,600, because

$$\sum^{700} X_i^2/700 - 8^2 = 4$$

where X_i is the number of units in a cluster

If one cluster with 60 units had been encountered instead of a cluster with 8 units, then the sum of squares in the sample would have increased by 3536 to 51,136.

The sample mean number of units would now be 8.08, not 8.00

The variance between the sample cluster values would increase and now be 7.93, not 4.00

The relative variance on an individual cluster basis would now be .1217, not .0625

The relative variance of the sample total would now be .0001738, not .000089.

The inflation of sampling error caused by a large cluster occurs not only in one or two characteristics, but is quite general. This is a result of the fact that the uniform rental or price structure of a development attracts a relatively homogenous population to it.

Since field work for the survey was to start almost four years after the Census date, we had to expect that we would encounter not one, but several clusters of new construction in the course of our work.

It was possible for us to create a separate selection frame for units built subsequent to the 1960 Census through the use of certificates of occupancy issued by the N.Y.C. Department of Buildings. These certificates indicate the address and the number of housing units contained in each new residential structure. They also contain a tax block number, which enabled us to determine how many new units were built in each tax block. Within each borough tax blocks were then selected for inclusion in the new construction sample with a probability proportionate to the number of new units in the block.

Since the sample derived from the 1960 Census is an area sample, it is necessary to prevent any new units which lie within the area clusters from being given a double chance of selection, should they be also represented in the new construction strata. Therefore each Census Block in the area sample was also identified by tax block number. These numbers are screened against the list of tax numbers in the new construction strata. New structures which appear in both the sample frames are then excluded from the area sample.

Every six months the new construction sample frame is updated. Since the Population Health Survey is a continuing survey, some of the newer units may have already been interviewed as part of an area cluster. If that has happened, the particular structure is not included in the updated new construction sample frame. This process assures us that the two sampling frames remain unduplicated. Clusters are selected from the updated frame with the requisite conditional probabilities, and are interviewed in the second half of the year.

F. ASSIGNMENT OF MEASURES TO CENSUS BLOCKS

Since Census Blocks in New York City vary from zero to nine thousand units each, we could neither achieve an efficient sample nor maintain uniform interviewer workloads by giving each block an equal chance of coming into the sample. Therefore, we assigned to each block a number of measures, proportional to one-eighth of the Census count of housing units. Additional measures were assigned to those blocks which contained non-institutional group quarters, in order to keep the number of persons per measure relatively constant from block to block. Census Blocks with only a few reported units were amalgamated on the sample frame with adjacent blocks to prevent clusters with very few units from coming into sample. We also took pains to be sure that blocks which had not been recorded in Census tabulations (because they had no population in 1960) were given non-zero probabilities of selection. This was done by joining them to neighboring blocks.

The assignment of measures to individual blocks and the inclusion of all blocks in the sample frame were completely verified. The total number of measures assigned to blocks within a health district was also checked against a control number of measures, based on health district tabulations derived from the Census.

G. SAMPLE SELECTION

In a stratified sample with a uniform sampling fraction, sample selection is usually made independently from stratum to stratum, and a simple random sample is taken within each stratum. We wanted to plan for the year to year change in our sample, for expansion of the sample in particular strata, and for the selection of special samples in connection with other Health Department studies. We also wanted to avoid, to as great a degree as possible, any tail-end variances which independent selection in our thirty odd strata might induce in the estimation of borough-wide and city-wide totals. A systematic sample of measures carried over from stratum to stratum was the best answer we could devise. Systematic sampling, however, has two inherent disadvantages. Periodicity in the population might be a multiple of the sampling interval, and therefore, a single systematic sample might be imprecise. Secondly, with systematic sampling, there is no unbiased way to estimate the sampling error. To overcome these difficulties we accumulated the block measures throughout the City in an order determined by three stages of randomization. In the first stage, a random permutation determined the order in which the five boroughs would have their measures cumulated. Within each borough another random permutation designated the order in which the health districts would appear in the cumulation. Finally, a third set of random permutations determined the order in which 357 small geographic areas would appear within the 30 health districts. Given this ordering, we employed just one random number and a systematic interval of 500 to select our sample of clusters for the entire City. Therefore the systematic effects, if any, could occur only within our smallest unit of randomization. We would still be left with a great number of degrees of freedom to estimate the sampling error because of the randomizing procedure. The whole process of sample selection in which we engaged may therefore be viewed as a single stage sampling. The cluster included in sample within a sample block corresponds to a translation of the cumulative random number which had selected the block.

Once the entire sample was selected for the initial survey year, weekly subsamples were established. We employed constraints in the subsampling process in order to insure that the geographic distance between the clusters interviewed in any two successive weeks would be as great as possible. As stated earlier, we scheduled a systematic half of the housing units in each cluster for interview during the first half of the year, and the other half for interview twenty-six weeks later.

III REALIZATION OF THE SAMPLE

A. CREATION OF CLUSTERS

Following the selection of a block, the address, inclusive of apartment number, of every residential unit in it was listed by our field staff. The count of the number of residential units was compared to the Census report, and the list of units was accepted when the count was within five percent of the Census value. Listings in disagreement with the Census report were reconciled by reference to Sanborn maps and to other information. Blocks with faulty listings or with irreconcilable ones were independently relisted, then reconciled.

Clusters were created out of the block listings so that the maximum size of any cluster within a block exceeded the minimum size by no more than three units. Wherever possible, we delineated compact clusters. In apartment houses to avoid ambiguity with regard to cluster boundaries and to keep the number of units per cluster constant, the ultimate sampling unit was often a systematic portion of a larger cluster.

The tightness of our control over the size of clusters was the final step which we could take to obtain the smallest sampling error from our survey. In the actual conduct of the survey all our efforts were bent to achieve the smallest possible biases in the data. We shall now turn to some of the efforts which we have put forward to control the mean square error of the Population Health Survey.

B. COVERAGE CHECKS

One of the important contributors to survey error is under-coverage of the target population. There are two components to under-coverage, missed households and missed persons within households. We have not designed a procedure to cope with the latter problem, but we do check on a sample basis for missed units within the clusters. To date, however, the best means of detecting the existence of missed units has been the second set of interviews taken in the sample clusters 26 weeks after the first set by a different interviewer. Published housing unit totals for New York City are in close agreement to our inflated sample values. We do not seem, therefore, to be suffering greatly from the under-coverage of units.

C. RESPONSE RATE

Perhaps the greatest contribution to the mean square error of the Population Health Survey has been made by non-response of sample households. In the first year of survey operations the response rate has proved to be 88% of the eligible households. In Manhattan, response has been 83%. While we are not chagrined by these results, they are well below the standard set by Census Bureau operations in New York. We are making every effort to improve our record in the second year of the survey operation.

D. INTERVIEWER EFFECT ON THE MEAN SQUARE ERROR

In terms of its weekly sample the Population Health Survey is a small operation, requiring only a few interviewers. To increase the number of people who collect the data, and thereby decrease the effect of a single interviewer on the mean square error on the statistics published, we have deliberately kept the weekly workload of each interviewer as small as possible. Because it has not proved feasible to assign interviewers to work outside their home boroughs, estimates of borough values represent the product of very few hands, and are subject to high variability on that account. Within the boroughs, however, we randomly assign clusters to the interviewers, thereby reducing the effect of interviewer-area interactions.

Our budget does not permit us to engage in a large re-interview program. It is therefore impossible for us to assess accurately the impact of so few interviewers on the data produced. Nevertheless, we find it useful to produce tabulations of health data by interviewer each quarter to discern gross differences in performance, particularly within boroughs where we have randomized assignments.

E. DATA PROCESSING CONTROLS

After the initial input of data onto the questionnaires, an inflation of the mean square error of the survey is bound to occur during the course of data processing. By setting up stringent quality control checks we have attempted to minimize this inflation.

The completed questionnaires are given routine check-in edits, and the fact that the interview has indeed taken place is verified, primarily by phone. The segment lists are checked to determine that all units scheduled for interview have been properly included in the sample.

Since the questionnaire employed in the survey is not an instrument which can readily be key punched, the information contained in it is transcribed, after coding, onto forms suitable for further processing. The professional staff of the survey reviews the medical and occupational coding on a sample basis. The accuracy of the transcription is also reviewed on a sample basis.

Once the data has been key punched onto cards, a computer program is employed to detect inconsistencies in the data. Inconsistencies within individual cards, inconsistencies between the cards of a single person, and inconsistencies between the cards for different persons within the same family are detected by this program. Before any tabulations are produced, all the errors detected by this program are corrected, and the program is re-run, to be sure that the indicated changes have been made. To date only four percent of all the cards processed have had detectable errors.

IV. ESTIMATION AND SAMPLING ERROR

A. ESTIMATION

As stated earlier (Section II D) we use only simple inflation estimates for totals. Sample values of proportions, rates or percentages are used directly for population values of the same types.

Two different inflation factors must be employed. The factors correspond to the two different sampling fractions used in collecting the data. In order to estimate statistics such as the number of persons in New York or the number who have been discharged from a hospital the sample data is simply multiplied by 500. To estimate population values of other types of data which are collected with a two week reference period, such as the number of physician visits or the number of dental visits made in the City in a year, the sample total must be multiplied by 13,000. This factor is the product of the basic inflation factor, 500 times 26, since the year is conceived as 26 two week periods.

The inflation factors are applied to the sample data only after adjustment has been made for complete non-response of eligible households. This adjustment is made by duplicating the information collected for a responding household in the same cluster as that which had the non-response. When no information is available about the characteristics of the non-responding household we have duplicated at random one of the interviewed households in the cluster. We have frequently obtained information about the number of persons in the non-responding household, and when this information is available we have selected for duplication a household within the same cluster that contains the same number of persons.

We have examined the results of this duplication process by tabulating health data with the duplicated cards in the deck, and by tabulating the same data without using the duplicated cards. Differences in such statistics as the percent of persons currently medically attended, the percent hospitalized in the past year, and the percent with a current limitation of activity have been trivial. We expected to find, as we did, that there were significant differences between the two sets of tabulations in economic data, since high income clusters in Manhattan's East Side have proved difficult to interview.

The failure of respondents to answer certain questions during the course of the interview, either because they do not know the answers or because, as in the case of income questions, they refuse to state an answer, increases the means square error of the survey and causes additional problems in estimation. When non-response is large - with income it has run to 8% of our respondents - we exhibit characteristics for the class of persons with income and the characteristic is used frequently in cross tabulations we set up a separate category for income unknown. However, when we seek to establish median family income figures, non-respondents are allocated to age - race - occupation groupings before estimation begins.

For items such as hospital insurance coverage the non-response rate has been under 1%, and we have simply considered the percentage of persons covered to be that number giving positive responses to our questions divided by the total number of persons in sample. For hospitalization rates, however, we have excluded persons with hospital status unknown before computing the percentage of persons hospitalized.

B. SAMPLING VARIANCES

Each statistic produced by the Population Health Survey has a specific sampling error. It is quite beyond our means to produce exact estimates of the error of each survey statistic. We have therefore limited ourselves to estimating the variances of seventy important totals and 100 key rates or percentages. From these variance tabulations we attempt to generalize our findings so that they serve as guides to the sampling errors of other items published by the survey.

The relative variance of an estimated total, $\hat{X}$ derived from the survey is computed as

$$V_{\hat{X}}^2 = .998 \sum^{34} \frac{k_h \sum^{k_h} x_{hi}^2 - (\sum x_{hi})^2}{k_h - 1} \Big/ \Big(\sum^{34} \sum^{k_h} x_{hi}\Big)^2$$

Where x_{hi} is the enumerated total value for all elements in cluster i of stratum h.

K_h is the number of clusters in stratum h.

34 is the total number of strata, of which 30 are geographic and 4 are new construction strata.

.998 is the finite sampling correction factor.

For a ratio of the form $\sum\sum x_{hi} / \sum\sum y_{hi}$, the estimated relative variance is computed as

$$V_r^2 = V_{\hat{X}}^2 + V_{\hat{Y}}^2 - 2 V_{\hat{X}\hat{Y}}$$

This form of estimation probably overstates the true sampling error within the health districts (strata), because it treats the sample within the districts as a simple random sample of clusters. It is a measure of the variance between as well as within the smaller geographic areas used in the third stage of the randomizing process. We do not feel, however, that the overstatement would be large or important with regard to health characteristics.

The items for which we computed relative variances were chosen to represent families of items which we believed would have different sampling efficiencies. These families comprised data on demographic, economic, two-week condition, hospitalization and medical attendance data.

For each family of data we fit a curve of the form

$$V_X^2 = a + b/X$$

X is an estimated total; a and b are values determined by minimizing the squared relative residuals of the function. Two or three iterations of the process are often necessary to produce a good fit. In using a curve of this form as in many other instances, we have followed the lead of the National Health Survey.

V. EVALUATION OF THE SURVEY DESIGN FROM SURVEY RESULTS

At this date, tabulations are available from the first six months of data collection in 1964. From the results it would seem that the survey is functioning at a level of efficiency somewhat higher than we had expected. The relative variances of some principal demographic statistics are shown below:

Item	Relative Variance	Relative Error
Occupied Housing Units	.000149	1.2%
Family Heads	.000302	1.7%
All Persons	.000298	1.7%
Unrelated Individuals	.002370	4.9%
Non-White Family Heads	.005083	7.1%

From these figures it seems clear that we have achieved our goal of estimating the population of the City with a relative error of under 2%, while employing only half the number of interviews we had regarded as necessary. The chief source of our over-estimate of our requirements in this regard lay in our assumption about the homogeneity of the size of households within clusters.

On the other hand with regard to the health and medical care characteristics of small groups in the population, the sample size is none too large. For example in the first published report of the Population Health Survey the hospital insurance coverage of non-white persons in New York was estimated to be 50.7%, an estimate which was not significantly different from the 42.2% coverage reported for persons of Puerto Rican birth or parentage. These values are from data collected in the first six months of the Survey's operation, and differences as large as 8.5% should prove significant at the conclusion of a complete cycle of enumeration. Smaller differences near the 50% level between these two groups will not be significant.

VI POSSIBLE REDESIGN OF THE SURVEY

In order to have more detailed information about certain groups in the City, it is possible to redesign the Survey somewhat, when the first two years of survey operation have produced information about the characteristics of the City as a whole. Under consideration is a plan to expand the sample in the six or seven health districts with the most severe public health problems. This expansion could be accomplished without changing our budget greatly, by reducing the sample size in the rest of the City to two-thirds of its present level. Such a design would still permit us to produce reliable city-wide estimates while sharpening our knowledge of the health status and medical care economics of the population in these districts.

SUMMARY STATEMENT

The Population Health Survey has produced detailed health and demographic data about the City of New York which is not available from any other source. While accomplishing this, it has been able to design and realize samples for special studies both within the Health Department and for other governmental agencies of the City of magnitudes ranging up to 25,000 households. The creation of such an instrument for a local health agency in a community in which no other central data collection center exists is well worth its relatively small cost.

THE PUERTO RICO MASTER SAMPLE SURVEY OF HEALTH AND WELFARE

Raúl A. Muñoz,
Puerto Rico Department of Health
and
Jack Elinson,
Columbia University

Two decades ago the acute health problems of Puerto Rico were so overwhelming that almost any health service provided anywhere was bound to have an obvious effect. Half of all deaths were attributed to diarrhea and enteritis, tuberculosis, pneumonia and malaria. Today, malaria has been completely eradicated and great progress has been made in the control of other infectious diseases so that today less than one-fifth of all deaths are attributed to these causes.

Currently, chronic diseases are the dominant disablers and killers in Puerto Rico. Prevention and control of chronic diseases and efforts toward the organization of care for the chronically ill have increased in priority in Puerto Rico. A substantial proportion of the Puerto Rico Department of Health and Welfare's budget is geared toward the prevention and control of chronic diseases and to the care and maintenance of the chronically ill. Still, funds are scarce in relation to the magnitude of the emerging problems and difficult decisions need to be made as to priorities of how wisely to appropriate scarce funds for massive problems.

Traditional Data

The information traditionally available to the Secretary of Health and Welfare in which to ground such decisions include: conventional statistics on mortality and population growth and change; registries of reportable diseases such as cancer and tuberculosis; and operational statistics from the various bureaus and agencies particularly charged with selective aspects of the overall program. These data are important but fail to provide the basic intelligence badly needed to keep abreast of the total chronic illness picture. What is needed, then, is systematic information which can:

1) characterize the Island-wide population, not a small portion of it,

2) reveal the dimensions and magnitudes of the chronic illness and other health problems of the island as a whole,

3) indicate the net effect of the various programs created to deal with these problems in the way of prevention, control and care, and

4) characterize the effects of chronic illness on family units as well as on individuals.

To develop information of this sort a "Master Sample Survey for Puerto Rico" was initiated as a Community Health Service Project. The subject matter areas with which the Master Sample Survey is primarily concerned are the chronically ill and aged in the context of family health and welfare.

Survey Objectives

The specific objectives of the "Master Sample Health Survey" are four-fold.

1) As an aid in overall planning, evaluation and assesment of priorities by providing quantitative Island-wide information. The data collected and analyzed will point out, and place in perspective, needs and gaps in health and welfare services for the chronically ill and aged.

2) As an opportunity and method for directors of health and welfare department programs concerned with various aspects of chronic illness and aging to insert key questions into the Master Sample Survey.

3) As a way for the Secretary of Health and his staff members to identify areas of health and welfare needs not adequately met by current programs.

4) As a procedure for the planning and conduct of certain kinds of epidemiologic and social research investigations on chronic conditions and related health and welfare programs.

Sampling [1/]

A multistage stratified area probability sample was designed in such a way as to permit a series of representative samples of the population to be drawn. Demolition of slum areas and new construction were taken into account. Advice on optimal sampling designs was sought from Dr. Ira H.

[1/] Marks, Eli, Ira H. Cisin, and Jack Elinson. Recommendations on Sample Design for the Master Sample Survey of Health and Welfare.- Mimeographed.

Cisin and Dr. Eli S. Marks, Statistical Consultants for the survey. The basic source of the sample is the labor force master sample of the Puerto Rico Department of Labor recently revised under the technical Direction of the U.S. Bureau of the Census. The population under study is the non-institutionalized people residing in the island. A sample of approximately 3,000 families was interviewed in 1958 in conection with the "Study of Medical and and Hospital Care in Puerto Rico"* conducted in collaboration with the staff of the School of Public Health and Administrative Medicine of Columbia University. It was decided to devise a sampling design which would incorporate roughly half of those 1958 households into an integrated current sample with 1964 households. The rationale backing this decision lies in the definite advantage of both measuring internal longitudinal and cross-sectional changes. The final design allows for annual reinterviewing of half of the households after the completion in 1964 of the original 1958 households. The design calls for quarterly subsamples of around 800 households which can be treated as individual samples or added for joint analysis.

Field Operations

A carefully selected field staff of college graduates was thoroughly trained in interviewing technics for health survey and given a field trial in a pilot survey conducted in the municipality of Guaynabo, of the North-eastern Region for Health and Welfare of Puerto Rico. The purpose of the pilot survey was threefold: 1) pretest a preliminary questionnaire designed for obtaining basic data on illness experience of the survey population; 2) provide actual field experience for training, evaluation and final selection of the interviewing staff; 3) obtain needed data based on a household approach to compare with the operational statistics and records of the Guaynabo Health Center.

Survey Instruments

The questionnaire used in the pilot survey was basically similar to the one used in the 1958 household survey conducted as part of the "Study of Medical and Hospital Care of Puerto Rico". Additionally several questions were included intended to measure the extent of use of the individual clinical and hospital services of the Guaynabo Health Center and some general information of the disabling effects of both acute and chronic illness.

The final questionnaire used during the first year (1964) included, besides the basic illness questionnaire just mentioned, three additional sections: 1) one dedicated to a thorough exploration of chronic conditions, the medical care sought and received, the disabling consequences of the illness, and it's economic effects; 2) another section concerned with a thorough exploration of some specific problem area of health: dengue, mental illness, cancer, prenatal care, including awareness of symptoms, knowledge of facilities and resources and actual utilization and experiences, and 3) a section on areas of social concern such as migration, public assistance, health insurance, etc.

Presented below in chronological outline are the general subject matter covered and the nature of the population queried in each of the quarterly samples of the first year of field operations.

First Year (1964)

First Quarter	Questionnaires	Sample Unit
November	Core Health	Entire household
December	Disability	Head of family
January	Mental Health	Random adult
	Migration	Entire household
Second Quarter	**Questionnaires**	**Sample Unit**
February	Core Health	Entire household
March	Disability	Head of family
April	Dengue	Any adult
	Pub. Assistance	Entire household
	Migration	Entire household
Third Quarter	**Questionnaires**	**Sample Unit**
May	Core Health	Entire household
June	Disability	Random adult
July	Cancer	Random adult
	Health Insurance	Entire household
	Pub. Assistance	Entire household
	Migration	Entire household
Fourth Quarter	**Questionnaires**	**Sample Unit**
August	Core Health	Entire household
September	Disability	Random adult
October	Migration	Entire household
	Pub. Assistance	Random adult
	Mental Health	Random adult

* Medical and Hospital Care in Puerto Rico,- A Report by the School of Public Health and Administrative Medicine of Columbia University and the Department of Health of Puerto Rico, 1962.

Rationale for the questionnaires

The core health questionnaire which is intended as the central portion of essentially every field interviewing operation is familiar in content to health surveyors. Resemblances will be seen to the current U.S.National Health Interview Survey, the current Health Survey in New York City, and other local surveys both past and ongoing,viz: California, Hunterdon, Baltimore, Washington Heights, Hawaii, etc. The core health questionnaire is in some ways similar also to the Island-wide Puerto Rican household interview survey conducted during 1958, in order to see what changes have taken place in the past five years. The opportunity to do quarterly surveys in 1963-64 as against a one-shot survey covering a year's period as was done in 1958, has made it possible to reduce the memory burden for the respondent and presumably to increase the accuracy of the data obtained.

It has been the philosophy of the Project from the beginning to involve program planners and directors in the Department of Health and Welfare in the determination of subject matter content for the supplementary quarterly surveys. A meeting was held of the Advisory Committee on January 23, 1964 at which priorities for subject matter to be included in the surveys were discussed. It was essentially on the basis of this and subsequent meetings that the subjects of the ensueing quarterly surveys, cancer, mental health, immunization, were chosen. Subsequently, discussions were held by the Project Consultant, and the Project Director, with the Head of the Division of Cancer Control, and the Chief Cancer Epidemiologist,where suitable specific objectives and subject matter for the cancer questionnaire were developed,and a questionnaire constructed adapting some past efforts in this regard (in particular, the Makover, Kutner and Crocetti study of Delay in Cancer in New York City).

The supplementary questionnaire dealing with mental illness and mental health services was worked out in consultation with the medical director of the Psychiatric Hospital and the Medical Consultant to the Project. The Project Consultant had been engaged in a parallel investigation on this subject at Columbia University in cooperation with the New York City Community Mental Health Board and supported by the Health Research Council of the City of New York. Interesting comparisons of conceptions of mental health and perceptions of community mental health services among Puerto Rican living in New York and in San Juan are becoming available, probably for the first time, as a consequence of these two surveys.

The second quarterly supplementary questionnaire came about as a reaction to the epidemic of dengue which hit the island in the summer-fall of 1963. The Secretary of Health and Welfare, Dr. Guillermo Arbona, requested that knowledge of dengue prevention be included in the next wave of interviewing to complement the epidemiologic investigations being carried out by a special team from the Public Health Service's Communicable Disease Center in Atlanta. The use of the Master Sample Survey in this way illustrates its quick adaptability in meeting needs for immediate health intelligence on topical questions as well as for coverage of longer term objectives. An intensive educational and operational campaign is currently underway to eliminate the dengue mosquito, also financed by Federal funds.

Inquiries have been received from a variety of sources as to the possibilities of inclusion of health-welfare relevant questions in the Master Sample Survey. Among them, for example, is a request from the Population Council (letter dated June 12, 1964) about the possibility of learning something about the "Enko" program through the Master Sample Survey. A request was received from the Director of the Division of Economic and Social Analysis of the Puerto Rican Planning Board as to the possibility of adding some questions to the next quarter sample on the subject of problems of older people, such as housing, etc. Discussions of these and other requests are underway.

Preliminary data from the Master Sample Survey on cigarette smoking were requested by and furnished to the Cardiovascular Research Group at the School of Medicine.

Sample Outcome - First Quarter

The table below shows the sample outcome for the first quarter. It is clear that respondent cooperation to health interview surveys appears to be less a problem in Puerto Rico than in some cities and States and compares favorably in this regard with experience of the U.S. National Health Interview Survey. The 97% completion rate is actually conservative. An additional one percent could not be located and therefore a determination could not be made as to eligibility; they have nevertheless been included here as eligible, but not obtained. Only four out of 863 families refused to be interviewed and in four families no mentally competent adult could be found to answer the questionnaire.

It was sometimes necessary to interview personally as many as three different individuals in a household in the first quarter: a household respondent, the head of the house if under 65 and a randomly selected adult between the ages of 20-64. Sometimes, by chance, the three designated respondents were the same person.

Evaluation

The first comprehensive evaluation of the Master Sample Survey was completed in February 1965 by Louis Moss, Director of the Social Survey Division of the Central Office of Information in

England and Theodore Woolsey, Deputy Director, National Center for Health Statistics, U.S. Public Health Service. The evaluation team recommended the incorporation of the Survey into the regular activities of the Department of Health; the strengthening of the resources for data analysis and reporting and the reinforcement of the continuing contacts with the users of the Survey results.

These main recommendations are at present in the process of instrumentations with particular emphasis on the one referring to data analysis and reporting. The Survey is converting to computer processing and two additional persons have been recruited for report writing. An effort is now under way to alter the basic concepts and analytic methods slightly to make them more consistent with those in the U. S. Health Interview Survey. There will be great added value if the statistics can be made comparable with those in the U. S. Survey.

The evaluators had some reservations about the feature of the plan that calls for devoting a half of the sample in any one quarter to re-interviewing at addresses where there was an interview the year before and following up certain of these families that have moved. The question of whether this procedure is worth the extra cost and complexity depends upon how the Survey results will chiefly be used. The staff is currently planing to re-examine this feature of the design to decide whether it will pay for itself in additional useful data.

The Secretary of Health has officially endorsed the Master Sample as a regular function of the Department and has under way the financial arrangements for its support. To date, two official reports have been published on illness data and three special ones are about to be released: dengue, cancer and mental health. Forthcoming reports will deal with conditions, public and private medical care, health insurance costs, and immunization.

TABLES

Sample Outcome -- First Quarter

A. Units eligible for interview among household units drawn:

TOTAL UNITS	NUMBER	PER CENT
Drawn in sample	988	100.0
Eligible for interview	863	87.4
Ineligible units	125	12.6

B. Reasons for ineligibility of household units

	NUMBER	PER CENT
Total Ineligible Units	125	100.0
Moved to U.S.	20	16.0
Moved to other area	29	23.2
Unoccupied	28	22.4
Building destroyed	14	11.2
Head deceased	13	10.4
Occupied by ineligible head	18	6.4
Vacant lot	4	3.2
House under construction	2	1.6
House moved	2	1.6
Two houses converted into one	2	1.6
Converted to business	2	1.6
No longer head of family	1	0.8

C. Household interviews completed

	NUMBER	PER CENT
Total Eligible Household Units	863	100.0
Household interviews completed	840	97.3
Household interviews not completed	23	2.7
Could not locate	9	1.1
Could not contact family	6	0.7
Refused	4	0.5
Mentally incompetent	4	0.5

D. Household heads interviewed

		NUMBER	PER CENT
Household Interviews Completed	840		
Ineligible heads (i.e. 65 and over or under 20)	157*		
Eligible heads		683	100.0
Interviewed		662	96.6
Not interviewed		21	3.4

* 155 family heads were 65 years or over; and 2 were under 20 years.

E. Reasons for non-interview of heads of household

Not found	12
Refusal	2
Absent	2
Died	1
Seriously ill	1
Spouse of head interviewed	1
Not determined	2
	21

F. Random adults interviewed with supplementary questionnaire

		NUMBER	PER CENT
Household interviews completed	840		
Total adults (19 and over)	1,997		
Total eligible adults (20-64)	1,657		
Eligible adults in sample (one per household)		778*	100.0
Interviewed		741	95.2
Not interviewed		37	4.8

* 62 household had no adults 20-64 years.

G. Reasons for non-interview of random adults

Not found		16
Mentally incompetent		11
Deaf and/or dumb		2
Institutionalized		3
a) prison	2	
b) hospitalized	1	
Other		5
a) refusal	1	
b) continental	1	
c) wrong selection of adult	2	
d) absent from Puerto Rico	1	
		37

While certainly not typical, the degree of cooperation is illustrated by the granting of an interview by a respondent whom an interviewer had difficulty locating since he was known only by the cognomen of "El Gran Brujo" (The Great Sorcerer). He was finally located on an obscure mountain cliff by the field supervisor (Félix Cotto González) who was asked by the respondent as to whether he carried any guns. The respondent at first refused to talk since he feared Sr. Cotto represented the police. But when Cotto explained he was from the School of Medicine and was interested in conducting a health interview, the man assented. His occupations included gambling, selling liquor, and running a prostitute service. He paid his girls $35.00 weekly. It was necessary for Sr. Cotto to return to complete the interview since he did not carry a sufficient number of hospitalization forms with him.

DESIGN AND PROCEDURES IN THE HAWAII HEALTH SURVEILLANCE PROGRAM

Charles G. Bennett and Paul T. Bruyere
Hawaii State Department of Health

The Hawaii Health Surveillance Program, begun in May 1964, is a continuing monthly health survey of the resident non-institutional population of the Island of Oahu, on which are located the City and County of Honolulu. The Bureau of the Census has designated the Island a metropolitan statistical area. It contains about 80 per cent of the population of the State.

During a three year period the program is being supported in part by the Bureau of State Services (Community Health) of the Public Health Service as a demonstration project. Its objectives are "to institute, develop and demonstrate the feasibility and utility of continuing health surveillance by means of interviews conducted in small random samples of households, independently selected each month; to provide sensitive, up-to-date measures of morbidity, population characteristics, health attitudes and the degree of health information in the community, and other knowledge useful in health planning, evaluation and research." After the demonstration period, the program may continue on a permanent basis using State funds entirely. Whether it so continues, obviously, will depend to a large degree upon results obtained during the present demonstration period.

The program is largely an outgrowth of the federal Health Interview Survey which was extended on an amplified scale to Oahu for one year during 1958 and 1959. Our questionnaire, instructions to interviewers, concepts and procedures, to a large extent, are based on those of the National Health Survey and we have received a great deal of constructive advice from consultants of that agency.

Sampling

The main sampling frame is a list of electric light meters. One power company serves all of the Island and the number of families without electricity is negligible. The few we have identified can be sampled separately. The company maintains a computer tape of all addresses to which bills are sent. We use this on a 1401 computer to draw semi-annual samples from which monthly subsamples are drawn. The semi-annual sample is in the form of a listing of every fiftieth meter billing address after a random start. From this we select a similar systematic subsample monthly of about 230 addresses, which constitutes a ratio of 1 address to about 575 households in the population. On an annual basis this ratio becomes about 1 to 48.

Some apartment buildings, housing developments and the like do not have individual household meters and all of these "demand accounts" are shown separately on the listing from the tapes. In such cases we ascertain the number of dwelling units in each building or account and, in effect, add this number to the sampling frame. If one or more sample addresses fall in one of these buildings, interviewers are given instructions indicating which households to interview. For example, if the sample includes the 10th living unit, the interviewer ascertains which one that is from a list of numbered units or by actual counting of the units inside the building or housing area.

An item on the face sheet of our questionnaire asks, in effect, whether more than one household receives electricity through a single meter. This question is asked because occasionally what the electric company designates as a household may include more than one, according to survey definition. In case the interviewer finds more than one household using a meter, where this was not previously known, instructions have been prepared on how to select the one to be interviewed. If there are no more than 2 or 3, coin tossing suffices.

Interviewers

Three full-time employees of the project do some of the interviewing together with office work such as coding. The rest of the interviewing is accomplished by 50 public health nurses stationed in 5 districts on Oahu. By agreement the nurses give an average of at least two hours each per month to the work. During the first year they averaged 1.6 completed questionnaires per month as compared with 3 estimated in the initial planning. This outcome has been due chiefly to the large number of households where no one has been found at home during working hours. (More than 40% of the females aged 17 and over on Oahu work outside the home.) When no acceptable respondent is at home on a second or occasionally a third visit, the nurse returns the uncompleted questionnaire to the statistics office for an evening call. As a result, about 47 percent of the blank questionnaires sent to the nurses for use at designated sample addresses are returned uncompleted, and the nurses submitted only about 40% of all questionnaires completed during the first year of interviewing.

Partly in an attempt to increase the proportion of interviews by nurses, we have enclosed a stamped return postal card with the advance letter from the Director of Health sent to each sample address. This card asks when the interviewer will be likely to find an adult member of the household at home. Only about one third of these cards are returned; we suspect that many of them reach the waste paper basket as advertising matter that is not read. Nevertheless we consider it worthwhile to continue sending them, for those which are returned specifying a time to call save the interviewers time, and when an evening hour is indicated the corresponding sample address need not be sent to the nurses.

We are still seeking some means whereby the public health nurses will do all or most of the interviewing, including at least some evening and weekend calls. Perhaps the only way to accomplish this will be to prevail upon the legislature to augment the nursing staff.

We have made some tabulations which we found of interest, comparing the data secured by the nurses with those of the other interviewers. The results indicated, as might have been expected, that the nurses are interviewing, for the most part, larger and younger households. As indicated above, they obtain data only from those households with a suitable respondent at home during the regular working hours. These are the households most likely to include children with the mother at home to take care of them. Predictably also, the nurses found higher acute condition rates, while the non-nurses encountered higher chronic condition rates.

This situation was reversed, however, for one acute condition labeled "the virus, not otherwise specified." Here it was noted that the rate obtained by the non-nurses was 20 times higher than the nurses' rate. Apparently the latter were less willing to accept this ill-defined term as a condition and made more searching inquiry as to what was wrong. Upon being apprised of the situation, the other interviewers, for better or worse, practically stopped making any entries for "the virus."

Questionnaire Content

Our basic questionnaire is in 4 parts. Part I relates to demographic characteristics; Part II consists of health probe questions to find out the kinds of morbidity and hospitalizations occurring; Part III goes into detail about the morbidity conditions reported in Part II; and Part IV secures data on the hospitalizations reported in Part II. Almost every item in this basic questionnaire has been tested in the National Health Survey.

Supplementary questionnaires are to be added from time to time for periods of a year or less. These will concern topics which various units of the Health Department may wish to have investigated. At present, data on the use of insecticides in the home are being obtained. Other supplements on health attitudes and home accidents are in the process of development. At one time we had hoped to formulate a satisfactory short supplement concerning mental health problems, but our attempt in this direction has not turned out well.

Recall Period

The interviewer inquires about acute conditions occurring during the full calendar month previous to the month of interview. Our basis for this choice was that a complete calendar month provides the respondent with an excellent frame of reference for recall; the sample is increased in terms of the number of conditions or events reported; and administrative control is somewhat simpler on a calendar month basis. In order to accomplish this, however, we feel it is important to complete the interviewing promptly at each month's end. To date we have been able to complete only about 65 percent of the interviews in the first week of the following month; we are making every effort we can think of to increase this figure.

In an effort to evaluate the calendar month against the two week period used in the National Health Interview Survey we have included the following question concerning each acute illness mentioned: "Were you sick from (this condition) before the 15th of the month or later in the month?" For injuries the wording is slightly different: "Did (this injury) happen before the 15th of the month or later in the month?" Tabulation of these questions over a 7 month period showed that 40% of the conditions existed before the 15th of the month, 43% later, and 18% both before the 15th and after. On the hypothesis of equal recall over the whole month, the before and after percentages should be about the same, and in fact the difference is well within sampling error. Nevertheless we intend further study using more cases and breaking them down by type of condition and other factors.

Chronic conditions are reported on the basis of 12 months recall. For the most part they are elicited by means of a checklist. The respondent is asked to check "yes" or "no" after each item in a prepared list. In the federal survey conducted on Oahu in 1958 - 1959, the interviewers simply read off the list of conditions to the respondent. Dr. Philip Lawrence, chief of the Division of Health Interview Statistics, National Center for Health Statistics, advised us that his office had subsequently found the checklist approach more effective. Perhaps as a result, we find the overall chronic condition rate in the current survey about 8 percent higher than in the previous federal survey.

Supplementary Mortality Survey

A major segment of the annual statistical reports from this project will be devoted to hospitalization in short-stay hospitals. This is expected to be of considerable value in current local efforts toward planning for future hospital and related health facilities. In order to make such statistics more complete, a subsidiary survey is being carried on each month concerning hospitalization of deceased persons during their last year of life. This is essential because interviews cover only the experience of persons living at the time. Obviously, a considerable number of individuals, especially at the older ages, receive extensive hospital services during the last months of life.

Information is being secured from hospitals and families of the deceased on a ten percent sample of deaths occurring each month. A little time is saved by using the same ten percent sample which is sent to the National Division of Vital Statistics for its current mortality index. As in other aspects of our project, procedures used for securing this supplementary mortality information are similar to those used in connection with the National Health Interview survey.

All hospitals are cooperating and about 90% of the families queried send in the desired information.

Data Processing

Mimeographed code sheets have been made up to record data for person cards, condition cards and hospitalization cards. Items concerning the household as a unit, such as its size, are placed on the code sheet for the person card of the head of the household.

Code sheets for each acute or chronic condition and for each hospitalization that a person has had are stapled behind the person sheet. The card punch operator completes a person card and then automatically duplicates 40 columns of it on each condition and hospitalization card. Many individuals, of course, require only a person card, whereas others may have a dozen conditions and several hospitalizations.

We are using the Medical Coding Manual and short index developed by the National Health Survey. It is an adaptation of the International Classification of Diseases, Injuries and Causes of Death to the kind of data secured in interview type health surveys. Although we have experienced coders, thoroughly familiar with the International Classification, they have had to study and gain experience with this adaptation of it to use it easily and accurately.

Thus far we have run all tabulations on a rather slow speed counting sorter available in the health department. This has been especially convenient for short run monthly reports and for experimenting with various kinds of tabulations. For annual reports involving about 18,000 cards, the sorter alone will require too much time. Since a 1401 is available and we have funds to pay for its use we are developing suitable programs for the annual data.

Projected Uses of the Data

In a survey of this type which produces a numerically rather small sample, the most informative detailed reports can be prepared only with data from at least a year of interviewing. We have not had time to issue any reports as yet on a full year's work. Tabulations have been completed and a report is now being prepared on the demographic and health characteristics of persons in military households. Although this group constitutes better than 15 percent of Oahu's population, little is known about its composition and health. The forthcoming report on these subjects should be of interest to both civilian and military authorities and possibly to other areas having a large population of military families.

Some topics of other reports we are working on in addition to annual tables of incidence and prevalence of morbid conditions by age and sex, are these: respiratory conditions including asthma and hay fever; time lost from work due to illness by various population categories such as government and non-government employees; hospital statistics; income and morbidity; accidents; activity limitation due to chronic conditions; health characteristics of persons 65 and over; bed disability among various classes; and differences among ethnic groups in the incidence and prevalence of morbidity.

Because of the variety of ethnic groups living in Hawaii, the last mentioned topic is expected to provide stimulus toward further research studies. For example if one or two of the ethnic groups such as Japanese, Chinese, Filipino, Portuguese, Puerto Rican or Polynesian, appear more susceptible than others to some condition, follow-up studies can be made in an attempt to find reasons for the difference.

An important objective of the surveillance program is to produce up-to-date demographic data between decennial census years. Already, at their request, we have supplied some items to the State Department of Planning and Economic Development, including an age and sex breakdown of Oahu's population, the age and sex of persons in military households, distribution of family incomes and ethnic composition.

In 1967 we hope to obtain an appropriation from the Legislature with which to make this project a permanent operation. At that time we shall have a large body of data and its analyses in the form of various reports, to show that the program is producing; in addition, we expect to have convincing evidence that the data are actually being used in public health, medical and demographic work. To secure such evidence, we plan to request information from our mailing list on just how any of our data have been used. We are also, of course, keeping a tally of special requests for information, of which we have already received several dozen.

We realize that the health interview survey has many shortcomings which have been discussed at considerable length in the literature. Nevertheless, we believe that the data it produces may reflect with a considerable degree of accuracy that aspect of morbidity in a community which has had an economic or social impact upon the individuals and families concerned. In this day of extensive planning in public health and other areas, this type of survey may continue to provide essential data not usually available from any other source.

XI

THREE YEARS AFTER THE GORDON REPORT: PROGRESS IN LABOR FORCE MEASUREMENT

Chairman, WALTER F. RYAN,
U. S. Office of Statistical Standards

RESEARCH IN LABOR FORCE CONCEPTS

Robert L. Stein, U.S. Bureau of Labor Statistics
Daniel B. Levine, U.S. Bureau of the Census

Following some sharp criticism of the Government's unemployment figures during the 1961 recession, a Committee to Appraise Employment and Unemployment Statistics (the Gordon Committee) was established by President Kennedy. The Committee submitted its report in September 1962, strongly urging that the Government undertake a program of experimentation to sharpen the measurement of unemployment. 1/ Although the Committee approved of the underlying concept of unemployment being used, it pointed out that some of the procedures used to measure this concept were inadequate--in particular, they relied in too many instances on volunteered information, and they depended on questions which were not sufficiently detailed.

The Committee acknowledged that no single measure of unemployment that would satisfy all users of the statistics could ever be devised. However, the Committee did see a need for some more reasonable working rules to set the boundaries between the unemployed and those not in the labor force and for detailed classification within each of the two groups so that different users of the statistics could combine the data to fit their particular requirements.

The Committee set forth 5 general criteria to be used in defining the concept of unemployment:

1. The concept should correspond to objectively measurable phenomena and should depend as little as possible on personal opinion or subjective attitudes.

2. The concept should be operationally feasible.

3. The definition used should be readily understood and broadly consistent with the common understanding of these concepts.

4. The definition should not be so inclusive that it yields figures which are difficult to interpret.

5. The concept should reflect the usual market criteria used in measuring the national output--an unemployed person would be one seeking work yielding a monetary reward.

The Committee recommended the establishment of a separate sample for experimenting with a sharpened definition of unemployment, and for testing questions which would yield greater accuracy and more information about all components of the employed, the unemployed, and persons outside the labor force.

The research sample was placed in operation by the Bureau of Labor Statistics in cooperation with the Census Bureau in April 1964. This sample, which is called the Monthly Labor Survey (MLS), was selected in the same manner as the Current Population Survey (CPS); that is, it is an area probability sample of the entire United States. 2/ The initial sample size comprised 8,750 households per month--one-fourth the size of the CPS--located in 105 sample areas, as compared with 357 for the CPS. The sample size was increased in the summer of 1965 to 17,500 households per month distributed among 197 areas. A completely independent staff of interviewers is used in the operation of this sample survey, to avoid any possibility of affecting the continuing CPS results, and to avoid confusion in the enumeration and in the interpretation of the findings. In each month, interviews have been conducted in the same enumeration week as is used for the CPS--the week containing the 19th day of the month--testing various forms of questions relating to employment status during the preceding calendar week--the week containing the 12th. As recommended, the experimental survey has also attempted to test questions designed to increase accuracy in other items and to provide information not previously available.

It should be re-emphasized that the Committee felt that most of the definitions and procedures used to collect the current labor force information through the household survey were well formulated and reasonable. Thus, in large measure, these same concepts and techniques were carried over for use in the MLS. The Committee endorsed the CPS definition of employment, which is based on work activity or job attachment during a specified calendar week. It considered but rejected a number of proposals to change the definitions, such as excluding from the labor force 14 and 15 year olds or persons working very few hours, although it placed great emphasis on the importance of providing separate estimates for these groups so that users could subtract them from the totals. Moreover, even with respect to unemployment, where the definitional problem is concentrated, the Committee gave its approval to most of the concepts underlying the present measurement, for example:

(1) Basing the unemployment definition on current labor market activity or status rather than on need or financial hardship;

(2) Including secondary workers (i.e., persons not permanently attached to the labor force or not the primary earners in their families) as unemployed when they look for work although, again,

1/ President's Committee to Appraise Employment and Unemployment Statistics, Measuring Employment and Unemployment, U.S. Government Printing Office, September 1962.

2/ A detailed description of the CPS sample is presented in Census Technical Paper No. 7, The Current Population Survey: A Report on Methodology, U.S. Bureau of the Census, 1963.

separate identification of such groups was to be made insofar as possible. (The Committee's recommendations that unemployed heads of households and unemployed persons seeking part-time work should be separately identified in the statistics have actually been in effect since January 1963.)

(3) Counting as employed those persons on part time for economic reasons. The Committee commended the BLS for its publication of the labor force time lost index which reflects the combined effect of unemployment and involuntary part-time work for economic reasons.

(4) Including among the unemployed persons on layoff or waiting to start new jobs within 30 days (except those in school).

(5) Including among the unemployed persons who were not working and were looking for work even though:

a. they had rejected previous job offers or were selective in the jobs they were willing to accept;

b. they were discharged for cause or quit their last job to seek another;

c. they might be considered unemployable by certain criteria or under certain labor market conditions.

(6) Keeping the issue of under-employment separate from the definition of unemployment. Extension of research in the area of under-employment was recommended.

In short, according to the Committee's recommendations, the basic concept of unemployment would continue to be persons without jobs who were looking for work.

Research Findings

The report that follows should be regarded as an interim progress report on the research undertaken during the past 18 months. None of the changes in procedure which are being tested has yet been adopted for use in the CPS survey (which provides the only official statistics), nor have any final decisions been made as to which features of the experimental program will be recommended for adoption. Nevertheless, it was thought to be useful at this stage of experimentation to report on what has been learned thus far.

Employment. In the area of employment, only one small definitional change was introduced in the MLS. Included as employed were persons absent from their jobs the entire survey week because of illness, vacation, bad weather, labor dispute, or personal reasons even if they looked for other jobs. In the CPS, persons absent from their jobs who are reported as looking for work are counted as unemployed. This change brings the classification of this small group into line with the treatment of persons who were at work but looked for other jobs--they are still employed.

The MLS concept--that is, all persons with jobs are employed--probably corresponds more closely to the public impression as to what is being measured in CPS. In order to evaluate the effect of this change, persons absent from their jobs were asked whether they also were looking for work. Results indicate that this change would increase the employed by less than 100,000. 3/

In addition to this single definitional change, questions were tested which yield additional information about the employed or to increase the accuracy of the statistics on the composition of the employed. For example, a question was added to collect information on whether persons with a job but not at work usually work full time or part time at their present jobs. This would permit more complete estimates of the full-time and part-time labor force, by combining this information with the data for those at work, and with the data on whether the unemployed are seeking full-time or part-time work.

Previous research into the problem of obtaining accurate reporting of hours worked 4/ has shown that many persons tend to report usual or scheduled hours rather than hours actually worked during the survey week. In the MLS, a series of probing questions was added to remind the respondent of time taken off during the survey week because of holidays, illness, or personal reasons; of overtime worked; or of hours spent on a second job. Mainly as a result of these probes, the

3/ Data presented in this paper, unless otherwise stated, are averages based on results for the first 6 months of 1965. The use of 6-month averages increases the reliability of the comparisons by reducing the sampling variability. Although most of the present MLS procedures have been followed since August 1964, the 1965 data are more representative of the results that could be expected from MLS because enumerators had gained training and experience and because there have been no further changes in questions or question wording since the beginning of 1965.

4/ At this point it is appropriate to mention that the Gordon Committee also stimulated the creation in the Census Bureau of a continuing field experiment in measurement techniques and related survey problems, including the reporting of hours worked and feasibility of collecting additional data from persons not in the labor force. Following their development in this experimental program, called "The Methods Test," these questions were incorporated into the MLS. This program is described in an article by Robert B. Pearl and Joseph Waksberg, New Methodological Research in Labor Force Measurements, prepared for the 1965 meetings of the American Statistical Association.

number of part-time nonagricultural workers reported in the MLS (those working under 35 hours) has been running some 1.8 million above the CPS level; most of the additional part-time workers (1.6 million) were working short hours for noneconomic reasons. The number with overtime hours has been 1.1 million higher in the MLS (see table 1). The number reporting between 35 and 40 hours, on the other hand, was 2.6 million lower in the MLS as compared with the CPS results. Average hours were 39.6 in MLS, 40.0 in CPS.

In the CPS, estimates of the self-employed have been too high because they included some persons who were the operators of small incorporated family enterprises, and regarded themselves as proprietors, rather than as wage or salary workers. The misclassification of these wage and salary workers as self-employed has been one of the major reasons for the discrepancy of some 2 million between household and establishment statistics on wage and salaried workers. In the MLS, an additional question was asked for all persons reported as self-employed in a nonfarm business as to whether the business was incorporated. The effect of this question has been to place the MLS estimate of nonfarm self-employed approximately 1 million below the CPS level, and to yield a correspondingly higher estimate of nonfarm wage and salary workers. The MLS procedure reduces the gap between the household and establishment survey estimates by about 50 percent, on the average.

As noted earlier, the experimental program retained the same basic definition of employment, with the exception of one minor change. It is not surprising, therefore, that comparisons of the MLS and CPS estimates of both total and nonagricultural employment have been well within the expected sampling error.

Unemployment. There is, of course, no question but that the genesis of the Gordon Committee was the criticism of the measurement of unemployment. Much of the criticism at that time, and subsequently, reflects a fundamental misunderstanding of the nature and purpose of the statistics. The assumption underlying most of the attacks on the statistics has been that unemployment must necessarily be equated with need or hardship, whereas the actual basis for the official statistics is that unemployment must be an accurate measure of currently available, unutilized manpower resources. Only by examining the regularly tabulated data on the characteristics of the unemployed is it possible to differentiate unemployed persons with very different kinds of employment and financial problems.

The concepts and methods used in the government's employment and unemployment statistics have been subject to periodic review by technical committees. In recent years, three outstanding groups of experts--the Stephan Committee, 5/ the Review of Concepts Committee, and the Gordon Committee--have thoroughly investigated the concepts and methods and have arrived at the general conclusion that the system currently in use was of a very high quality. All 3 groups, and particularly the Gordon Committee, have suggested the need for a number of significant changes designed to refine the statistics. Many of these improvements have been incorporated into the CPS during the past 10 years. The research carried out over the past year and a half has had as its goal a still more accurate system of measurement.

The definition of unemployment currently in use in the CPS includes all persons 14 years of age and over who did not work during the survey week but were looking for work (or waiting the results of a job application made within the last 60 days). Also counted as unemployed are the following:

(1) Persons on layoff waiting to return to work.

(2) Persons waiting to start a new job within 30 days (except those in school).

(3) Persons who would have been looking for work except that they were temporarily ill, or they believed no work was available in their line of work or in their community. These groups are the so-called "inactive" unemployed.

The information is elicited by asking for persons not reported as working last week, "Was ...looking for work?" Persons on layoff or waiting to start a new job are identified by a question on the reason they did not work at their job last week. This question is directed to those who did not work or look for work but were reported as having a job from which they were absent.

The Committee noted critically that the time period for seeking work is not explicitly spelled out, that no evidence is given that steps were actually taken to look for work, and that for the "inactive" unemployed there were no questions that would elicit the relevant facts. Only if the respondent volunteers the information or raises questions can these groups be identified under present procedures.

The Committee proposed an alternative definition for testing--an unemployed person would be one who did not work during the survey week, but who had looked for work within a specified period of time--30, 45, or 60 days--and who was still available for work. Persons on layoff and those waiting to start new jobs within 30 days would

---

5/ The Measurement of Employment and Unemployment by the Bureau of the Census in its Current Population Survey, Report of the Special Advisory Committee on Employment Statistics, August 1954.

also be counted as unemployed, as they are in CPS. All other persons who had taken no definite steps to find work within the specified time period would be excluded from the unemployed.

One specific approach recommended by the Committee for high priority in the testing as a replacement for the present single question involved asking people who did not work during the survey week whether they wanted to work at the present time, whether they had looked for work within a specified recent period, and what they had done to look for work. Those who wanted to work and who had taken steps to find work would be called unemployed. Persons who were reported as wanting to work but not having looked in the past 4 weeks were asked why they had not looked, in order to identify the "inactive unemployed" who are included in the CPS definition but would not be in the definition to be tested.

The results obtained by this approach appeared to be particularly unreliable. Professional staff who observed their use in actual interviews reported that the question on wanting to work drew affirmative answers that appeared unrealistic, and these were sometimes supported by unlikely claims to work-seeking activities. Moreover, interviewers found the question awkward because in many households the wanting to work seemed to be just a vague hope.

This procedure was thus rejected and a new one adopted for testing that was believed to adhere to the spirit and purpose of the Gordon Committee's recommendations and to yield more objective results. For a person not employed or on layoff during the survey week, information was obtained as to whether he looked for work within the past 4 weeks, what he did to look for work, and whether there was any reason he could not take a job during the survey week. According to this procedure, those who took definite steps to find work within the past 4 weeks are counted as unemployed unless they were not available for work during the survey week. 6/

6/ There would be one minor exception to this rule--persons who looked for work within the past 4 weeks but were not available for work during the survey week because of temporary illness would be included as unemployed. This group is very small, amounting to only 0.1 percent of the labor force.

In MLS, the question on reason for absence from a job was changed to "Did he have a job from which he was temporarily absent or on layoff last week?" This is more explicit, in terms of identifying persons on layoff waiting to be called back than is the CPS question "Even though ...did not work last week, does he have a job or business?" In addition, there is a place on the MLS schedule to record the fact that a person was on indefinite or more-than-30-day layoff. As in the CPS, there is also a specific place to record the fact that a person was on temporary layoff with definite instructions to return to work within 30 days. In both surveys, both types of layoffs are included as unemployed but in the MLS the questioning is more precise.

In contrast to the CPS definition, the MLS definition of unemployment excludes persons who would have been looking for work except for belief that no work was available in the community or in their line of work. The basis for this exclusion is the difficulty of measuring this group on a monthly basis with a reasonable degree of objectivity. A somewhat related but more broadly defined group is identified by a series of questions to be described in detail below, and is broken out as a separate component of the total outside the labor force.

Finally, in order to improve the reporting on duration of unemployment, the MLS includes a question as to the date unemployed persons last worked at a full-time job. This is in addition to the regular CPS question on the number of weeks they have been looking for work. If the time since the last job is shorter than the duration of unemployment as reported, the interviewer asks further questions to obtain the correct answers.

To recapitulate, the definition of the unemployed that is currently being tested in the MLS is: Persons without jobs who took specified steps to look for work in the past 4 weeks and were still available for work in the survey week, plus those waiting to be called back from a layoff, or waiting to start a new job in 30 days (unless in school) and available for work in the survey week.

The following table summarizes the various components of the present definitions of unemployment, and those being tested:

Present definition of unemployed	Definition of unemployed being tested
1. Persons not at work last week but looking for work (time period not specified).	1. Persons not employed (i.e., at work or absent from a job) last week who looked for work during the past 4 weeks and were available for work last week. Some definite work-seeking activity must be reported.
2. Persons waiting to start a new job within 30 days (unless in school).	2. Same - If available for work last week.
3. Persons waiting to be called back from layoff.	3. Same - If available for work last week. Question wording more explicit.
4. Persons who would have been looking for work except for temporary illness.*	4. Unemployed if actually looked for work within the past 4 weeks.
5. Persons waiting to hear the results of a job application made within 60 days.*	5. Unemployed if actually looked for work within the past 4 weeks and were available for work last week.
6. Persons who would have been looking for work except they believed no work was available in their community or line of work.*	6. Not in labor force.

*Classified as unemployed if information is volunteered; no specific question asked or identification made.

The discussion below amplifies the significance of each of the changes in procedure being tested in MLS.

(1) Persons on temporary layoff with definite instructions to return to work within 30 days averaged about 100,000 in both surveys despite the change in question wording.

Persons on indefinite or more-than-30-day layoff 7/ averaged 300,000 in the first half of 1965 in MLS. The size of the indefinite layoff group cannot be estimated from CPS.

The MLS data suggest that seasonal cutbacks are responsible for most of the indefinite layoffs under current economic conditions. Persons waiting to be called back from an indefinite layoff averaged only about 100,000 during the summer months of 1964 (June-September), started climbing in October to a peak of nearly 500,000 in February, turned down again in April, and returned to 100,000 by May and June 1965. Only 20 percent of the indefinite layoffs reported unemployment lasting 15 weeks or more. In fact, two-thirds reported less than 2 months of joblessness.

(2) Unemployed persons seeking work at some time during the past 4 weeks and still available for work in the survey week averaged 2.2 million in MLS. The size of this group cannot be estimated from CPS. However, it is a plausible inference that the use of a fixed time period of 4 weeks increases the count of marginal workers among the unemployed. In the June 1965 MLS, it was determined that 650,000 persons who had looked for work in the past 4 weeks, and were still available, did not do anything to find work during the survey week itself. Only 100,000 of these were men in the prime working age groups.

Of course, the CPS (which does not specify the time period for job-seeking activity) also includes some unemployed persons whose work-seeking activities predated the survey week. However, it seems likely that most respondents would assume that the CPS question "Was ...looking for work?" relates to the survey week since it follows 2 questions which specifically mention last week. In any case, since most of the new procedures being tested in MLS tend to reduce unemployment, it must be inferred that the 4-week reference period works in the opposite direction.

(3) The question on current availability in the MLS eliminated an average of 500,000 persons reported as seeking work during the past 4 weeks. Such a question has never been asked in the CPS, but it is certain that some of the persons now reported in the regular survey as looking for work (particularly those in school in the Spring months) would not be available to take a job in the survey week. In MLS, the number of persons eliminated from the group reported looking for work within the past 4 weeks because they could not take a job during the survey week ranged from about 200,000 in January 1965 to about 1 million in June, rising steadily with the approach of school vacations. The insertion of a question on availability changes the seasonal pattern for teenagers, reducing their number sharply in May and June from CPS levels. Such a question would have little effect in the summer, however, whereas the 4-week approach significantly raises teenage unemployment at that time of year.

7/ This combined group is referred to hereafter as "persons on indefinite layoff."

Students constituted 85 percent of the persons eliminated from the unemployed in MLS by the availability question. The remainder were unavailable for such reasons as pregnancy, child care, other family responsibilities, personal business, and vacations.

(4) Persons who reported that they had looked for work were asked what they had been doing in the last 4 weeks to find work. (The methods were listed on the questionnaire but were not read to the respondent.) All persons who said they were looking for work reported some specific activity. A substantial proportion (40 percent) reported that they had done more than one thing to find a job.

The most common method used was to check directly with an employer. After that, checking with a public employment agency was most frequently reported (see table 2).

The fact that all work-seekers reported how they looked for work gives some additional assurance that the figures are not inflated. However, the question on methods does not provide any evidence as to how vigorously work was sought.

(5) The experimental definition used in the MLS does not include inactive work-seekers (theoretically counted in the CPS, but without explicit questions) who would have been looking for work except for belief that no work was available. Under the MLS definition such persons are not in the current labor force if they took no steps to find work in the past 4 weeks.

Originally, the inclusion in the definition of unemployment of persons who would have been looking for work except <u>they believed none was available in their line of work or in their community</u> was meant to refer to discouraged workers in stranded areas or occupations. It has always proved a difficult group to measure because of the subjective nature of the concept. Attempts are being made in MLS to identify a somewhat broader group of presumably "discouraged" workers. The definition of the "believe no work available" group has been expanded from the original CPS definition to allow for the inclusion of workers idled by a seasonal lull and those who believe they can't get jobs because of racial discrimination, lack of education, inadequate training, or lack of skills or experience.

The precise quantitative effect of each specific change cannot be measured because the present CPS cannot be broken down in terms of each of the MLS components included in the unemployment definition. On an overall basis, the <u>net</u> effect of all changes in procedure appears to be relatively small. For the first 6 months of 1965, both surveys yielded a jobless level close to 4 million and an unemployment rate of about 5 percent, not seasonally adjusted. The differences in average level and rate between the two surveys were within sampling error. For adult men 20 years and over, the jobless rates were identical (3.8 percent).

For teenagers, the MLS and CPS rates of unemployment were also within the range of normal sampling variability; both rates were close to 16 percent. The first half year comparison for teenagers, however, was affected by the elimination of a large number from the unemployed in MLS because they were not available for work in the survey week. On an <u>annual</u> average basis, the level of teenage unemployment as estimated by MLS procedures would probably be slightly higher than the CPS level.

For adult women 20 years and over, the MLS rate was consistently higher, averaging 5.6 percent, as compared with 5.0 percent.

Altogether, the new approach brings in more persons seeking part-time jobs: 21 percent of the total in first half 1965 MLS, 16 percent in CPS.

The long-term unemployed were 65,000 fewer in MLS, probably because of corrections made in the reported duration of unemployment as a result of the additional question on date last worked.

<u>Persons not in the labor force</u>. The Gordon Committee was strongly in favor of obtaining more information relating to persons not in the labor force. Their past work experience, reasons for leaving their last job, and their intentions to look for work again all were suggested as useful facts in helping the analyst understand the dynamics of the labor force.

Since the situation for the vast majority of people not in the current labor force remains unchanged for long periods of time (e.g., the disabled, the retired, mothers of very young children), these questions are not appropriately asked of the same individuals month after month.8/ Accordingly, additional questions to be asked of this group were designed for use in households entering the sample for the first time or returning to the sample for their second four-month period of interviewing. Thus, on a monthly basis, this information would be available for one-fourth of the sample, which could be adjusted to represent the universe or, preferably, could be accumulated over several months.

The test questions developed in the experimental program include determining when each person not in the labor force last worked at a regular full- or part-time job. For those who worked within the past 5 years, the reason for leaving and the occupation and industry of that job are recorded. Everyone is asked whether he intends to look for work in the next 12 months,

8/ Households selected for the sample are interviewed for 4 consecutive months, drop out for 8 months, and then return for an additional 4 months of enumeration.

and if the answer is "yes", "probably", or "maybe", why he is not looking for work now.

Beginning in July 1965, all those whose last job was terminated because of economic reasons (for example, completion of a seasonal job or temporary nonseasonal job or because of slack work or business conditions) are asked why they are not looking now, even if they indicate no intention to look for work.

Based on an average of the results for the first six months of 1965, the test questions show that almost 15 percent of the persons not in the labor force (excluding those reported as unable to work)--some 8.5 million persons in all--say that they will or may look for work in the next 12 months. About one-third of this group never held a regular job and one-fifth had not held a regular job for more than a year.

Slightly under half of those planning to look for work--about 3.8 million--had in fact worked within the past 12-month period. On the basis of the answers given as to why they left their last job within the year, it appears that for about 1 million or one-fourth of the group, the last job was terminated for economic reasons--for example, it was a seasonal or temporary job, work was slack, or the company merged or went out of business. Relatively few of this group (about 200,000) were adult men; nearly a third were teenagers, many of whom it may be assumed were reporting on summer employment.

To date, the results of these experimental questions do not suggest that a very large number of persons on the margin of the labor force have been discouraged from looking for work because they believe no jobs are available. 9/ Of the total who intend to look in the next 12 months, only 100,000 were not looking at the time of the interview because of this reason; almost 5 million were in school and most of the remainder mentioned such factors as illness and family or household responsibilities (see table 3).

From the questions on when persons last worked, it was learned that about one-third of the persons not in the labor force, some 18.5 million, had worked at a regular job in 1960 or later. Most of those who had worked since 1960 had left their last jobs for noneconomic reasons--ill health, retirement or voluntary reasons--but about one-sixth (3.2 million) gave reasons that suggested that the job was terminated for seasonal or "economic" reasons. The age distribution of this group is shown in table 4. Of particular significance is the fact that only 300,000 of these persons were adult men 20 to 64 years of age.

Conclusions

To date, then, the experimental program through the use of the MLS has developed techniques which will lead to improvements in measures of hours worked, the self-employed, and the duration of unemployment. Questions have been developed which expand the amount of information available about persons not in the current labor force. The MLS classification of persons with jobs who looked for others during the survey week seems to be more logical and more in line with the underlying concepts--all persons with jobs are employed and the unemployed are the jobless seeking work.

With respect to unemployment, the experimental program has incorporated two fairly noncontroversial Gordon Committee suggestions--the inclusion of specific question wording to identify persons waiting to be called back from a layoff and a specific question to determine the steps unemployed persons took to find work.

The most far-reaching features of the definition of the unemployed used in the MLS testing program are:

(1) Spelling out the time period and fixing it at 4 weeks.

(2) Injecting a test of current availability.

(3) Shifting persons who believe no work is available out of the labor force.

All 3 of these proposals move in the direction of implementing the Gordon Committee's recommendations that the defintions be made more precise and objective. These aspects of the experimental definition sharpen the distinction between the unemployed and persons not in the labor force by establishing specific rules for classification, at the same time minimizing the need for probing and sophisticated judgment by the enumerator.

It was believed desirable in the testing program to experiment with a time period for jobseeking beyond the survey week itself since, by its very nature, jobhunting does not necessarily involve specific identifiable activity every day or every week. The more typical pattern of behavior probably involves periods of activity (i.e., checking with employers) followed by periods of waiting. Some forms of looking are continuous, i.e., registration with public employment agencies, but others are not.

The 4-week cutoff is at the lower limit of the various alternatives suggested by the Gordon Committee. This was done for 3 reasons: (1) to minimize the inclusion of persons with very loose attachments to the labor force, (2) to keep the time reference for jobseeking from getting too far out of line with that of jobholding, (3) to minimize the memory problem.

Results obtained so far indicate that spelling out the time period at 4 weeks and the other MLS procedures do not have any significant net effect on the count of unemployed adult men.

9/ The modified procedure adopted in July will, of course, yield a somewhat higher estimate.

Since they are for the most part family breadwinners or at least responsible for their own support, they are likely to be seeking work intensively when not employed and can be readily identified by any reasonable set of questions. This pattern is reinforced by the social pressure on adult males to seek work, so that even those who do not look very actively are probably reported as looking for work.

Specifying the time period at 4 weeks appears to operate in the direction of increasing the count of unemployed women and of teenagers seeking part-time jobs. Because this kind of change in the composition of the unemployed is a matter of concern, further research into the time period of jobseeking activity is being planned. Beginning in November, unemployed persons will be asked when they last took steps to find work during the last 4 weeks. It will then be possible to study this information in relation to whether they were seeking full- or part-time jobs; their age, sex, and marital status; and other characteristics.

Although further research is necessary, it is clear that the definition of unemployment developed thus far tends to be sharper and more objective than that used in CPS. Moreover, substantial gains in the accuracy and scope of the information on the employed and persons not in the labor force appear to be feasible through the extension of the questions.

It should be pointed out that some of the changes being tested, if adopted for the regular statistics on employment and unemployment, would involve breaks in the historical series. So much of our economic policy depends on the analysis of trends in the employment status of the population that even minor discontinuities could be serious. It is essential, therefore, to continue the research on the MLS for a sufficient period to permit a fuller evaluation of the effects of the changes (with appropriate modifications if necessary) before proposing their adoption. Once the final decisions have been taken with respect to specific definitional changes, the next step would be to merge the MLS and CPS samples, with an ultimate size of some 50,000 households per month, distributed among 445 sample areas.

Table 1.--Persons employed in nonagricultural industries, by hours worked: first half average, 1965 (Millions)

	CPS	MLS
Total	66.8	67.0
With a job but not at work	2.6	2.6
At work	64.2	64.5
1-34 hours	12.8	14.7
Economic reasons	1.9	2.1
Other reasons	10.9	12.5
35-40 hours	30.4	27.8
41 hours or more	20.9	22.0
Average hours	40.0	39.6

Note: Detail does not necessarily add to totals because of rounding.

Table 2.--Methods used by the unemployed to look for work: first half average, 1965

	Number of methods (In thousands)	Percent
Total	5,275	100.0
Public employment agency	1,211	23.0
Private employment agency	316	6.0
Checked with employer	2,176	41.2
Placed or answered ads	691	13.1
Checked with friends or relatives	575	10.9
Other	307	5.8

Table 3.--Persons intending to look for work in the next 12 months, by reason for not looking now: first half average, 1965
(In millions)

Total	8.5
Believe no work available	0.1
Temporary illness	0.6
Family or household reasons	0.9
School	4.9
Other and NA	2.0

Table 4.--Persons whose last job was terminated for economic reasons
(In millions)

Total	3.2
14-19 years	0.6
20-64 years	2.1
Men	0.3
Women	1.8
65 years and over	0.5
Men	0.3
Women	0.2

NEW METHODOLOGICAL RESEARCH ON LABOR FORCE MEASUREMENTS

Joseph Waksberg and Robert B. Pearl, Bureau of the Census

I. INTRODUCTION

Among the accomplishments of the Gordon Committee has been the stimulation of a sharply expanded research program at the Bureau of the Census devoted to labor force measurement techniques and related survey design problems. Some research has, of course, always been built into the Current Population Survey (CPS) program but its character and intensity have been limited by budgetary considerations and, perhaps even more, by the absence of a separate vehicle for the conduct of field experiments.

Budgetary stringencies were alleviated to some extent by the creation in the Bureau of a special Research Center with funds to support basic and applied research relevant to survey methodology in general. The allocation from this source was supplemented by savings arising from operating efficiencies and from curtailment of some less essential aspects of the CPS program.

These decisions led to the creation of a continuing field experiment--the CPS Methods Test--in three geographic areas in which a variety of explorations have been and are being undertaken. Also of consequence are various investigations of the impact of nonresponse on labor force measurements. In addition, a good deal of research is continuing on sample design, estimation procedures, and related mathematical statistical problems, but these activities are outside of the scope of the present article.

II. CPS METHODS TEST

The CPS Methods Test is a vehicle for studying the accuracy of alternate enumeration procedures or questionnaire formats. It has been in operation continuously for almost two and one-half years in three areas--Boston, Massachusetts; Charlotte, North Carolina; and Marion County, Ohio. The areas were purposively selected to provide a range of urban and rural areas without overburdening any one Census field office.

The design, in brief, calls for 6 interviewers in each of the three areas, each conducting about 24 household interviews a week for 3 weeks of each month--a total of about 1,300 interviews a month for the entire experiment. Each interviewer uses a different procedure in each of the three survey weeks of the month, repeating the same pattern in subsequent months. Assignments have been randomized by interviewer, week of the month, and location within the area to control for as many extraneous elements of difference as possible. Households in the experiment were initially interviewed for four consecutive months. Later, the rotation cycle was reduced to three months to increase the percentage of new households each month and thus provide additional experience on differences in reporting between such households and those that have been previously enumerated.

First Series of Experiments.--The first series of experiments, conducted during the period April 1963-December 1964, related to certain modifications in questionnaire design and content and interviewing procedure. The interviewers in each area were divided into two groups with each group testing two alternative procedures against the standard one used in the Current Population Survey. (It was felt inadvisable to train each interviewer on all the procedures to be tested.) The procedures were as follows:

Group I Interviewers

Procedure No. 1 - Current CPS procedure with the standard questionnaire and independent interviews each month.

Procedure No. 2 - More detailed questionnaire with probing questions on hours worked, duration of unemployment, detection of marginal labor force participation, background information for persons outside the labor force to assist in proper classification of such persons, etc.

Procedure No. 4 - In this approach, the interviewer also used the more detailed questionnaire (Procedure No. 2) as the first step in the interview. However, she also had a summary of the previous month's information which was to be consulted after completion of this first step. Additional questioning was specified to determine whether reported changes in employment status or job attachment from a month earlier had actually occurred or were the result of response variation.

Group II Interviewers

Procedure No. 1 - Same as for Group I Interviewers.

Procedure No. 3 1/ - Advance form covering key items mailed to family with request that it be filled and held for interviewer to pick up. Later changed to request mail return of completed form by certain date (if not forthcoming, telephone or personal follow-up was conducted).

Procedure No. 5 - Essentially the same as Procedure No. 4 above except that the standard CPS questionnaire was used for the first stage of interview rather than the more detailed one.

1/ In July 1964 this procedure was suspended because of rather unfavorable results and Procedure No. 2 was substituted for it to increase the sample size available for that approach.

Results have been tabulated separately for each of the procedures, some for 12-months and others for 18-month periods, and procedures have sometimes been combined in tabulation to increase statistical reliability. For example, the Procedure No. 1 results for both groups of interviewers have been combined for some purposes. Procedures 2 and 4 results have sometimes been merged to study the effects of the more detailed interview, and the same has been done for Procedures 4 and 5 to appraise the effect of access to the previous month's information.

Over-all Findings of the First Series of Experiments.

Perhaps the most interesting finding of this experiment is that, in spite of the sharp differences in approach, there do not appear to be any major differences among the procedures in the distributions of the sample by employment status. The labor force rates and unemployment rates for the experimental procedures are not significantly different from those for the standard procedure (Table 1). 2/ Of course, the sampling errors could conceal differences which, although small, may be important from an analytical standpoint.

Another measure which has not varied significantly among the procedures is the rate of month-to-month gross changes (Table 2). This is defined as the percentage of persons interviewed in each pair of consecutive months who reported a basic change in employment status from one month to the next. 3/ Some theories have held that the volume of "gross changes" is considerably exaggerated as a result of response variability and that a lower level might be indicative of an improved, or at least more consistent procedure, provided that the statistics were otherwise unaffected. However, regardless of the validity of the hypothesis, no important differences emerged among the approaches tested.

2/ For an explanation of employment status concepts used in the CPS, see joint publication BLS Report No. 279 - Current Population Reports, Series P-23, No. 13, "Concepts and Methods Used in Household Statistics on Employment and Unemployment from the Current Population Survey," Bureau of Labor Statistics, U.S. Department of Labor - Bureau of the Census, U.S. Department of Commerce, June 1964.

3/ These represent the summation of shifts, in either direction, between an employed and unemployed status, between employed and not in labor force, and between unemployed and not in labor force.

Still another evaluation device was a systematic reinterview program, whereby a subsample of the cases assigned to each procedure was interviewed a second time, a week later, by a supervisory person. Comparison of the original and reinterview results, while subject to quite large sampling variation, showed no evidence that any of the experimental procedures produced smaller differences than the standard approach (Table 3).

Improvements evident in detailed questionnaire - While findings have been inconclusive with regard to the basic employment status classifications, some clearcut improvements have been evident in the detailed questionnaire (Procedure 2) in secondary, although still important, items of information. Most striking of these is the case of the information on hours worked. Although the standard CPS questionnaire calls for hours actually worked during the reference week, there have been numerous indications that many persons report their usual workweek and overlook deviations such as time taken off, overtime, or hours on second jobs. The detailed procedure contained systematic questions on these possible deviations from the norm and for correcting the initial responses accordingly. The results (Table 4) show a marked decrease in the 35-40 hours category (the "usual" workweek for most persons) in the sample using the detailed questions, accompanied by a sharp increase in the part-time group and some evidence of a rise among those on overtime. It was also revealed that the deficiency in the measurement of the part-time work force in the standard procedure was primarily among persons who had taken time off for personal or family reasons and not among those on short time because of slack work or other economic factors.

An effort was also made in the detailed procedure to seek improvements in the information on duration of unemployment, which is known to be subject to appreciable response error. The approach used was to record date worked on last previous job as well as the number of consecutive weeks seeking work (the latter being the standard approach) and to have the interviewer reconcile and, if necessary, correct apparent inconsistencies between these items. Because of the miniscule number of cases that might be affected, the experiment gave little evidence of the utility of this revised approach. However, some more recent evidence, on a larger scale, suggest that it might have some merit in improving reporting.

Some probing was also introduced in the item on "class-of-worker", that is, whether the person is an employee, a self-employed worker, or an unpaid worker in a family enterprise. By definition, all workers in an incorporated business--regardless of size and dispersal of stock--should be reported as employees. It has been suspected, however, that many small entrepreneurs who have incorporated their businesses for various legal or other advantages may continue to report themselves in the CPS as self-employed, as this may still best exemplify their status in their minds. 4/ The detailed form contained a question addressed to those reported as self-employed as to whether their business was incorporated. The results revealed close to 10 percent of the group were operators of incorporated businesses (a fact which was substantiated in large measure by checking against independent lists of corporations). If inflated to national totals this would involve some 750,000 persons, a number which would materially close the gap between employee totals from the CPS and those based on establishment payroll reports.

Besides these various validity checks, the detailed questionnaire contained a number of additional items of information, some of them key Gordon Committee recommendations. First was an inquiry on the specific steps unemployed persons had taken to seek jobs, with virtually all able to report one or more types of activities. Perhaps more significant were the additional questions addressed to persons currently outside the labor force, such as date last worked, description of most recent job and reasons for leaving it, and job seeking intentions in the succeeding year. The experiment established the feasibility of making these inquiries and there seems little doubt that these would be useful analytical aids.

These several improvements evident in the detailed questionnaire were incorporated in the national experimental sample, the Monthly Labor Survey, on which a separate article is being presented. It is most likely that they will eventually find their way into the basic CPS program.

Results of dependent interviewing procedure - Some new findings have been forthcoming from Procedures 4 and 5, which involved a comparison of the responses for the current month with those provided a month earlier and a reconciliation of reported changes in status. One of the objectives of these procedures was to explore the theory that many apparent changes in status were the result of response variation and that this could be reduced by special probing on over-the-month occurrences. Although these types of response variations are generally offsetting, and may have comparatively little effect on the levels of the statistics or on net changes, they could seriously impair the validity of "gross change" data from month to month or over other periods. In fact, publication of gross-change data from the CPS was suspended a number of years ago, and only cautious use has been made of them since for analytical purposes, because of uncertainty concerning their reliability. 5/

Contrary to expectations, the findings from this experiment tended to confirm the validity of the reported changes in status and to ascribe only a small proportion to response variability. Of all reported month-to-month changes in employment status over the period studied, over 85 percent were validated after the substantial probing provided by these procedures. Moreover, of the respondents who after further questioning decided a change had not really occurred, virtually all pointed to the previous month's information as incorrect, which might suggest a deficiency in the reconciliation stage rather than in the original survey data. The results of this single investigation are, of course, insufficient to reach definitive conclusions, but if substantiated in subsequent tests could make available a valuable analytical tool which has largely been kept under wraps.

Self-enumeration procedure - Procedure 3, which involved a version of self-enumeration, showed rather disappointing results for the most part. The initial approach used (starting in the second month in sample) was to mail an advance form to the household, listing the known members, and requesting that the specified labor force information be recorded and held for the interviewer to pick up the following week. The reason for the pick up was partly to permit a review of the entries but mainly to meet the urgent type of time schedule required in the regular monthly survey. Over some 7 months of operation of this procedure, the proportion of eligible respondents who had completed the advance form averaged about 30 percent. The low response apparently was not attributable to objections to the form or the procedure but rather to the pressure of other activities and the usual human tendency to procrastinate.

In an effort to improve response, the procedure was modified to request respondents to mail back the completed forms (in a postage-free envelope, of course) by the end of the reference week. Various other studies had suggested that this approach generally elicits higher response than the earlier version. Where no response was received by the middle of the following week (or where incomplete forms were returned) the interviewer instituted an intensive telephone or personal follow-up to secure the missing information. A favorable finding was that it was possible with this combined approach to complete the workload within the usual timetable for the survey. However, the proportion who returned their forms by mail continued to be low--about 30 percent, on the average--and close to half of these returns were incomplete in one or more important items.

4/ There is, in fact, some thinking that their classification as self-employed would be more meaningful from an analytical standpoint (for example, National Bureau of Economic Research, Forty-Fifth Annual Report, "The Task of Economics," June 1965, page 9.) This problem could be solved in other ways, however, such as by some occasional measures on size of enterprise.

5/ For a discussion of the potential uses and the problems 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.

Partly because of these results, the self-enumeration procedure was discontinued in July 1964 and replaced by one of the other and more promising approaches. Of course, the Bureau conducts numerous other surveys, often quite complicated ones, by mail with generally satisfactory results. The main problem with the current labor force survey is that the timetable is such that there is little or no opportunity to conduct the reminders by mail or other means that contribute so much to the high self-response rates in other programs.

Second Series of Test.--The first series of tests, described above, was terminated in December 1964 pending analysis of the results, and a second series was instituted at that time using the same survey vehicle. In this case, attention was focused on the selection of the respondent for the interview. Under the present system, a single respondent, generally the housewife, reports the labor force information for the entire family. Although this may be desirable from a number of standpoints--certainly in terms of cost, accessibility, and perhaps cooperativeness--questions have been raised about the validity of reporting on certain aspects of the activities of other members, such as hours worked, occupation, and similar items.

For purposes of this second experiment, 6 interviewers again were used in each of the 3 test areas with the same volume of interviews spread over 3 weeks of the month. In this instance, there were only 3 test procedures and each interviewer conducted a different one of the procedures in each of the three survey weeks of the month. The assignment of procedures to different weeks and locations within the areas were randomized, as before, to reduce irrelevant sources of difference. The same questionnaire was specified for all procedures, in this instance the detailed form used in the national experimental Monthly Labor Survey (essentially the detailed Procedure No. 2 form discussed earlier but with revised questions for the unemployed and a somewhat different order of items).

The three procedures (numbered 6,7, and 8 here to avoid confusion with the first series of tests) are the following:

Procedure No. 6 - Essentially the present CPS approach with a single respondent for the household.

Procedure No. 7 - A procedure whereby each person was to be interviewed for himself, insofar as possible. If at home at the time of the interviewer's visit, each such person would be interviewed directly. Those absent at that time were to be contacted later by telephone, if possible, or if not, by return visit. As a last resort, at the conclusion of the enumeration week, the interviewer would accept the information for outstanding cases from the usual household respondent rather than omitting the case entirely.

Procedure 8 - Still another version involving some self-enumeration was attempted in this series of tests. In this case, its objective was to alleviate the anticipated high costs of interviewing each person for himself. In this procedure, an advance form containing only selected items was sent to the household, requesting each person to record the information for himself, and asking someone in the family to hold the completed form for pickup by the interviewer. If the requested items were filled by the time of the interviewer's visit, she was to transcribe the information to the full questionnaire and complete some subsidiary items for all members by interviewing a suitable respondent who was present. If the requested information was not recorded on the advance form, the interviewer was to proceed as in procedure 7, by attempting to interview each adult member for himself.

At this writing, there is insufficient information to report on the results of this second series of tests. In addition to the usual criteria for evaluation, a key factor here will of course be the comparative unit costs for the various procedures.

III. NONINTERVIEW RESEARCH

Research into another type of methodological problem was instituted at about the same time as the Methods Test. This was a study of possible alternative ways of dealing with noninterviews, that is, occupied households that are in the current sample but for whom no information at all is obtained for various reasons, such as inability to find anyone at home during the survey period, refusal to cooperate, and the like. 6/ There are two separate aspects of this subject: (1) can any methods be developed to reduce the number of noninterviews below current levels, and (2) are any improvements possible in the methods now used to adjust for noninterviews.

For the first item above, general plans have been formulated for an experimental study but it has not yet been implemented to any important extent. The present discussion will therefore be restricted to the second item - that is, to methods of adjustment. Time will not permit a detailed analysis of the data, but we would like to sketch out the methods of approach and some of the principle findings. A future paper will describe the project in greater detail.

6/ Cases where a household is interviewed but where some of the information is omitted inadvertently or for other reasons are not classified as noninterviews. This type of omission is relatively minor in the case of the CPS labor force information, rarely exceeding a few tenths of one percent for any given item.

Some background information on the scope and effect of noninterviews in the CPS will be useful in providing a perspective on this subject. Because of the risk of serious biases arising from differences between interviewed and noninterview households, the Bureau has always placed great emphasis on keeping the noninterview rate at a minimum consistent with budget and time considerations. During the past 10 or 15 years, the rate has averaged about 4 to 5 percent, ranging from a seasonal low of 3 to 3½ percent in certain spring and fall months to a high of around 6 percent in the summer when many people are away on vacation. About 1 to 1½ percent, on the average, reflects outright refusals. The remainder are households that the interviewer cannot contact, primarily because the household members are on vacation, temporarily away from home for some other reason, difficult to contact because of peculiar working hours, or rarely home for other reasons.

Because the total level of noninterviews is comparatively low in the CPS, small differences between interviewed and noninterview households could not have any perceptible effect on the statistics; almost any reasonable method of adjustment would be satisfactory under those conditions. It is only if large differences existed that consideration of alternate methods becomes important. An essential part of the study undertaken therefore, was to determine the approximate size of the differences between interviewed and noninterview households with regard to their demographic and labor force characteristics.

The present method of adjusting for noninterviews in the CPS is as follows:

(1) The 357 primary sampling units (PSU's) in the CPS sample are classified into 76 groups as the basis of similarity of population and labor force characteristics.

(2) The noninterview units in each group of PSU's are classified by color of the occupants and by urban, rural farm, and rural nonfarm residence. Each noninterview unit is then given the same characteristics as the average interviewed unit in the same residence-color class within that group of PSU's.

Several alternate methods are being considered as replacements for the current procedure. Among these are:

(1) Use information supplied by the noninterviewed household in the nearest preceding month if that household had ever been interviewed in the CPS program.

(2) Subsample noninterviews as they occur during an assignment period and subject the selected units to more intensive follow-up, still within the assignment period.

(3) Instead of giving the noninterviewed unit the characteristics for an average of <u>all</u> units in the residence-color class, use the average for some <u>sub-group</u> of this population whose characteristics will correlate more closely with the noninterview units.

(4) Compute noninterview adjustments on the basis of the characteristics of the persons in the noninterview households (such as sex-age-color, assuming this information was available from a previous interview) rather than on the basis of household characteristics.

We have attempted to get labor force characteristics of noninterviews in order to study the effect of the different kinds of adjustment procedures. This was done by following up a subsample of noninterviews intensively during the week after they were to be interviewed for CPS. Information was obtained for close to 40 percent of the cases in the subsample. For both the 40 percent and the remaining 60 percent of the sample cases, we determined the interview status during preceding and succeeding months, and obtained labor force data in those months for the cases that were interviewed on one or more such occasions. Tabulations of these two sets of data were then made.

The fact that we were only able to get data for 40 percent of the noninterviews presents a serious qualification on the analysis of the results. We are proposing to eliminate or reduce this qualification by repeating this study, with modifications that may increase the proportion of successful interviews. Mainly, we hope to do this by increasing the length of time for follow-up to several weeks (instead of one week) and by enlarging the staff assigned to the follow-up operations. This study will probably be conducted this fall.

Meanwhile, a number of special tabulations were made of the CPS data to assist in the analysis. The most important one consisted of a separation of the data actually reported by interviewed households from those imputed for the noninterviews. Normally, this process of imputation is performed automatically on computers and only final total labor force figures are available for analysis.

A second series of special tabulations represented a different type of breakdown of the CPS data. Separate figures were made available on the labor force status of persons in households requiring only one visit for a completed interview, those requiring two visits, etc. This was done to test the suggestion that households requiring several visits were more like noninterview households than those requiring only one visit, and that an improved method of imputation might be to use the characteristics of 2 or 3 visit households rather than all households. 7/

Summary of Findings to Date

As a result of the qualifications mentioned earlier, the findings to date must be considered as tentative. We hope that firmer evidence will become available in the next few months.

Evidence on the following points, however, seem to be emerging.

1. Earlier concerns that the labor force characteristics of noninterview households might differ dramatically from those of interviewed households do not appear to be borne out, with one exception. The exception is the group "with a job, but not at work" (employed persons temporarily absent from their jobs during the reference week) for which the noninterview household rate was about three times that in interviewed households. A large difference of this type would be expected during the summer months, when vacations account for a large part of the noninterviews, but the pattern apparently persists at other times of the year as well.

   The unemployment rate in noninterview households appears to be slightly higher than for interview cases, and there appear to be minor differences in a few of the "not in labor force" categories, such as those "unable to work". However, these differences are small enough so that any reasonable imputation procedures would probably produce satisfactory data for these items.

2. The fairly elaborate adjustment scheme used in CPS does not produce data significantly different from what would result from a simple assumption that noninterview households overall have the same characteristics as interviewed ones; at least this appears to hold for the U.S. as a whole. It is possible that the more elaborate method produces improvements in regional data, statistics for nonwhites, or other subclasses of the population, and a further analysis along this line is being made.

3. There is so far, no support for the hypothesis that as the number of visits required to enumerate a household increases, the household takes on more of the characteristics of noninterviews. The characteristics of households interviewed on 1 visit differ in many respects from those requiring 2 or 3 visits, but the differences are not necessarily in the direction or of the magnitude required to support the hypothesis. It should be noted, however, that data are so far available for only one month. Further information on this subject is planned.

4. Substitution of CPS data as reported by the noninterview household in a neighboring month does appear to provide a modest improvement over the current technique. Virtually every employment class is closer to the "true" figure, although none of the improvements are startling.

7/ See "A Method of Allowing for Not-At-Home Bias in Sample Surveys" by D.J. Bartholomeau, in Applied Statistics, Vol. X, No. 1 March 1961.

TABLE 1 - EMPLOYMENT STATUS OF THE POPULATION 14 YEARS AND OVER, BY SEX, AGE, AND PROCEDURE USED: CPS METHODS TEST, JULY 1963-DECEMBER 1964

(Units in total number of sample observations in period covered)

Employment Status	Test Procedure Used 1/							
	Procedure 1			Procedures 2 and 4				
	Total	Group I Interviewers	Group II Interviewers	Total	Proc. 2	Proc. 4	Proc. 3 2/	Proc. 5
Total Population 14 +	16,981	8,387	8,594	16,592	8,335	8,257	5,475	8,305
In Labor Force	9,757	4,814	4,943	9,536	4,743	4,793	3,135	4,819
Percent of total	57.5	57.4	57.5	57.5	56.9	58.0	57.3	58.0
Unemployed	578	287	291	527	272	255	204	282
Percent in Labor Force	5.9	6.0	5.9	5.5	5.7	5.3	6.5	5.9
Total Males	7,834	3,907	3,927	7,600	3,804	3,796	2,588	3,903
In Labor Force	6,120	3,064	3,056	5,990	3,004	2,986	1,997	2,988
Percent of total	78.1	78.4	77.8	78.8	79.0	78.7	77.2	76.6
Total Females	9,147	4,480	4,667	8,992	4,531	4,461	2,887	4,402
In Labor Force	3,637	1,750	1,887	3,546	1,739	1,807	1,138	1,831
Percent of total	39.8	39.1	40.4	39.4	38.4	40.5	39.4	41.6
Total Population 14-19	2,391	1,106	1,285	2,405	1,219	1,186	744	1,244
In Labor Force	869	409	460	907	466	441	260	484
Percent of total	36.3	37.0	35.8	37.7	38.2	37.2	34.9	38.9

NOTE: For an explanation of the procedures, see text, pages 2 & 3. None of the differences between procedures are statistically significant at the 95 percent probability level. Rough orders of magnitudes of the standard errors of the individual procedures are: for labor force rates - 0.8 for the total population, 0.9 for males, 1.1 for females and 2.0 for teenagers; for the unemployment rates, the standard error is about 0.4.

1/ For an explanation of the procedures, see text, pages 2 & 3.

2/ Procedure 3 conducted only during period, April 1963-June 1964.

TABLE 2 - MONTH-TO-MONTH GROSS CHANGES IN EMPLOYMENT STATUS,1/ BY SEX AND PROCEDURE USED: CPS METHODS TEST, APRIL 1963-DECEMBER 1964

(Units in total number of identical sample cases in each pair of consecutive months)

Item	Test Procedure Used 2/									
	Procedure 1			Procedure 2			Procedure 3 4/	Procedures 4 and 5		
	Total	Group I Interviewers	Group II Interviewers	Total	Group I Interviewers	Group II Interviewers 3/		Total	Proc. 4	Proc. 5
Both Sexes-Total Persons	12,632	6,314	6,318	7,209	6,012	1,197	4,455	11,751	5,682	6,069
Gross changes: Number	883	423	460	584	485	99	356	836	407	429
Percent of total	7.0	6.7	7.3	8.1	8.1	8.3	8.0	7.1	7.2	7.1
Male - Total	5,893	2,975	2,918	3,322	2,770	552	2,134	5,581	2,729	2,852
Gross changes: Number	446	209	237	270	217	53	169	410	214	196
Percent of total	7.6	7.0	8.1	8.1	7.8	9.6	7.9	7.3	7.8	6.9
Female - Total	6,739	3,339	3,400	3,887	3,242	645	2,321	6,170	2,953	3,217
Gross Changes: Number	437	214	223	314	268	46	187	426	193	233
Percent of total	6.5	6.4	6.6	8.1	8.3	7.1	8.1	6.9	6.5	7.2

1/ Gross changes represent a summation of persons who changed in either direction between one month and the next, i.e., between an employed and an unemployed status, between employed and not in labor force, and between unemployed and not in labor force.

2/ For an explanation of the procedures, see text, pages 2 & 3.

3/ During the period July-December 1964, Group II interviewers substituted procedure 2 for the then discontinued procedure 3.

4/ Procedure 3 conducted only during period, April 1963-June 1964.

TABLE 3 - GROSS DIFFERENCES 1/ IN EMPLOYMENT STATUS BETWEEN ORIGINAL INTERVIEW AND REINTERVIEW, BY SEX AND PROCEDURE USED: CPS METHODS TEST, APRIL 1963-MAY 1964

(Units in total number of sample observations in period covered)

Item	Test Procedure Used 2/							
	Procedure 1			Procedures 2 and 4			Procedure 3	Procedure 5
	Total	Group I Inter- viewers	Group II Inter- viewers	Total	Proce- dure 2	Proce- dure 4		
Both Sexes-Total Persons	1,856	953	903	1,818	938	880	890	914
Gross differences: Number	67	39	28	96	43	53	43	35
Percent of total	3.6	4.1	3.1	5.3	4.6	6.0	4.8	3.8
Male - Total	878	455	423	851	442	409	434	433
Gross differences: Number	31	17	14	53	24	29	19	20
Percent of total	3.5	3.7	3.3	6.2	5.4	7.1	4.4	4.6
Female - Total	978	498	480	967	496	471	456	481
Gross differences: Number	36	22	14	43	19	24	24	15
Percent of total	3.7	4.4	2.9	4.4	3.8	5.1	5.3	3.1

1/ Gross differences represent all cases in the reinterview subsample for which reported employment status (employed, unemployed, or not in labor force) was different in the reinterview than in the original reinterview. In this test, differences were not reconciled with the respondents so that they would tend to be somewhat exaggerated. The approach used probably tended to favor the standard procedure (No. 1) over the others since the standard questionnaire was used for all reinterviews. This was changed later in the test periods but further results are not yet available.

2/ For an explanation of the procedures, see text, pages 2 & 3.

TABLE 4 - DISTRIBUTION OF PERSONS AT WORK, BY NUMBER OF HOURS WORKED IN REFERENCE WEEK AND BY PROCEDURE USED: CPS METHODS TEST, JULY 1963-DECEMBER 1964

(Units in total number of sample observations in period covered)

Hours Worked	Test Procedure Used 1/							
	Procedure 1			Procedures 2 and 4				
	Total	Group I Interviewers	Group II Interviewers	Total	Procedure 2	Procedure 4	Procedure 3 2/	Procedure 5
Total at Work	8,479	4,148	4,331	8,268	4,161	4,107	2,699	4,209
Part Time (1-34 hours)	1,829	920	909	2,089	1,105	984	544	975
For economic reasons	288	140	148	354	205	149	84	169
For other reasons	1,541	780	761	1,735	900	835	460	800
Full Time (35-40 hours)	3,943	1,888	2,055	3,462	1,692	1,770	1,322	1,893
Overtime (41 hours or more)	2,707	1,340	1,367	2,717	1,364	1,353	833	1,341
Percent Distribution								
Total at Work	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Part Time (1-34 hours)	21.6	22.2	21.0	25.3	26.5	23.9	20.1	23.1
For economic reasons	3.4	3.4	3.4	4.3	4.9	3.6	3.1	4.0
For other reasons	18.2	18.8	17.6	21.0	21.6	20.3	17.0	19.1
Full Time (35-40 hours)	46.5	45.5	47.4	41.9	40.7	43.1	49.0	45.0
Overtime (41 hours or more)	31.9	32.3	31.6	32.8	32.8	33.0	30.9	31.9

1/ For an explanation of the procedures, see text, pages 2 & 3.
2/ Procedure 3 was terminated in June, 1964.

TABLE 5 - PERCENTAGE DISTRIBUTION OF EMPLOYMENT STATUS OF PERSONS 14 YEARS AND OVER, AS REPORTED IN CPS AND FOR VARIOUS CLASSES OF NONINTERVIEW HOUSEHOLDS

(Most data represent monthly averages for Feb.-June 1963. See footnotes for variations from this period.)

Employment Status	Total Pop. as tabulated in CPS	Interviewed households as tabulated in CPS 1/	Data for Noninterview Households: As Computed in CPS 1/	Data for Noninterview Households: Interviewed in following week 2/	Data for Noninterview Households: CPS Data for households in CPS in neighboring month 3/	Data for interview households, by number of visits required to obtain interview 4/: Total	1-visit	2-visits	3 or more visits	Telephone int. and n.a.
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
In labor force	56.3	56.6	56.1	57.5	58.2	57.4	56.2	60.3	60.7	57.8
Employed	52.9	53.1	52.6	53.4	54.4	54.6	53.3	57.3	57.9	55.5
Working full time	40.0	40.2	40.0	37.5	50.8	39.8	38.6	41.9	42.2	41.1
Working part time	10.6	10.6	10.2	9.2		9.1	9.2	9.7	9.0	8.3
With a job, not at work	2.2	2.3	2.3	6.6	3.6	5.6	5.4	5.8	6.7	6.1
Unemployed	3.5	3.4	3.5	4.1	3.8	2.9	3.0	3.1	2.8	2.3
Not in labor force	43.7	43.4	43.9	42.5	41.8	42.6	43.8	39.7	39.3	42.2
Keeping house	26.8	26.8	27.5	25.9	26.5	27.2	28.0	25.2	25.2	26.6
In school	8.3	7.9	7.8	7.2	6.2	0.7	0.7	0.8	0.7	0.7
Unable to work	1.4	1.4	1.4	0.7	1.0	1.3	1.4	1.0	1.0	1.2
Other	7.1	7.3	7.2	8.7	8.1	13.4	13.7	12.7	12.4	13.7

1/ Data cover the period March-June 1963. Based on weighted CPS data but using a simpler estimation method than in CPS (age-sex-color ratio estimates and composite estimates have been omitted.)

2/ Represents unweighted tallies of the 40 percent of the Feb.-June noninterviews that were interviewed in the week following the regular CPS enumeration period.

3/ Represents unweighted tallies of about 60 percent of the Feb.-June noninterviews that were included in CPS in either the month prior to or immediately following the month of noninterview.

4/ Data are for period Aug. 1964. Based on unweighted tallies of the entire CPS for that month. The high percentages shown for "with a job, not at work" reflected persons on vacation that month. Similarly the low figures for "in school" reflect the season, as do the high proportion for "other"-not in labor force, a group which includes school children on summer vacation. The level of these items cannot therefore be compared with the level during other months of the year.

DISCUSSION

Albert Rees, University of Chicago

Before turning to the paper by Mr. Stein and Mr. Levine, I should like to congratulate all of the agencies of the Federal government on the splendid way in which they have followed up the recommendations of the Gordon Committee. Where possible, recommendations were implemented promptly. Others have been followed up through the kind of research reported here this morning, and those that have proved unworkable have been intelligently modified. The only part of the government that should perhaps be excepted from this commendation is the White House, which has on occasion ignored the recommendation that labor force statistics should be released by technicians according to a predetermined schedule, and not be announced prematurely by policy making officials when they show especially favorable developments. It might be supposed that the recommendations of any such committee would be pursued or explored with vigor, but the experience of the Price Statistics Review Committee (the Stigler Committee) suggests that this is not always the case.

The paper by Mr. Stein and Mr. Levine reports on a large-scale test of an experimental definition of unemployment. Let me underscore some of the points made in the paper concerning the shortcomings of the present official definition. Although that definition is generally satisfactory, it is misleading because some of its elements have no counterpart in the measurement procedure. For ex. nple, according to the official definition, a person awaiting the results of a job application made within the last 60 days is considered unemployed. However, there is no question in the Current Population Survey interview designed to elicit this information. If the information is volunteered, it is used, but this will occur in a random fashion. Volunteered information may therefore be a source of "noise" in the present procedure, having an effect similar to that of an increase in sampling variability.

In proposing an experimental definition the Gordon Committee suggested the use of a screening question on whether the person concerned wanted to work in the reference week. The authors report that this question produced unsatisfactory results. With the wisdom of hindsight, I am not surprised at this finding and feel that some of us on the Gordon Committee should have anticipated it. Asking whether a person wants to work is likely to produce defensive reactions in those who don't want to work, but feel that it may be expected of them.

In general, the definitions used in the Monthly Labor Survey seem to me to be excellent. They are clear, unambiguous, and accord well with the general understanding of the terms defined. One feature of the new definition that was not suggested or even discussed by the Gordon Committee seems to me to be especially desirable. This is the transfer from unemployed to employed of the people who were looking for work during the reference week but who had a job from which they were absent for such reasons as illness, vacation, or bad weather.

It is gratifying to learn that for the first six months of 1965 the MLS definitions produce an estimated unemployment rate very close to that of the CPS. This suggests that there would not be large costs of changing over to the new definitions in terms of the historical continuity of the overall unemployment series, though there would be larger differences for subgroups of the unemployed.

More insight into the experimental definitions will be gained as the results from the present larger MLS sample become available. If these results are as encouraging as those reported here, I would hope that the MLS definitions or some variant of them could before too long become the official definitions.

DISCUSSION

James N. Morgan, University of Michigan

It is a pleasure to comment on these two excellent pieces of research. Both of them are well-designed, limited to a few important objectives, and seem to have been well carried out. Given a mandate from the Gordon Committee, they have shown that there is a little, but not much, overstatement of unemployment as defined. There are a few job-lookers who aren't quite available yet (students), or some who have a job but are looking for a better one. More important, they have shown that the field techniques and probably the quality of interviewer training are sufficiently good so that changes in procedures or adding a few questions do not make much difference. I'm not at all convinced, however, that adding questions on whether the individual was "available" last week and whether he "has a job" are desirable, since they are subjective in a way that not working and looking for work are not.

These two refinements in measurement, reducing the measured "unemployment" still further, raise the question whether we shouldn't be doing research on the extent to which the official measure understates the problem of underemployment. Once a year the Census finds out whether people were unemployed during the previous year, which helps. But what about people who couldn't find full time work? What about those who wanted to work more than 40 hours a week?

The Survey Research Center has experimented with asking questions of this type. The 1963 Survey of Consumer Finances asked:

> "Sometimes people don't work as much as they want to because of illness or unemployment or short work weeks or lack of extra jobs. How about you (head) would you say that you worked less than you would have liked last year?"

The percentages saying they wanted more work than they had varies from 44% for people under 35 without a high school diploma, down to 2% for those with some college education.(1) But they are considerably higher than the official estimates of unemployment, and higher than the proportion unemployed at some time during the year.

In the fall of 1963 and again early in 1965 The Survey Research Center asked a national sample the following:

> "Some people would like to work more hours a week if they could be paid for it. Others would prefer to work fewer hours a week even if they earned less. How do you feel about this?"

Fifteen per cent had no opinion. Among those with an opinion, some 41% wanted more work, and 16% wanted less.

Quantifying such questions requires coding verbatim replies to open-ended questions, under strict control in a central office. But the results are reproduceable. What we need is periodic collection of such measures to provide checks on the whole iceberg of demand for work, of which the official unemployment measure is only the visible and potentially variable fraction.

I do have one suggestion for dealing with the not-at-home problem. It is really a suggestion by Leslie Kish of a variant of a procedure reported by him several years ago at these meetings. If we assume that not-at-homeness is associated with something unusual, but only during the period when the man is not at home, then we should revisit on a later wave some addresses where people were not at home previously, and ask the unemployment questions about the week before the time of the previous visit. These answers would then be substituted for the missing answers of some of the current not-at-homes, about the previous week. This would eliminate the bias that would arise if people were not at home because they were out of town looking for a job. The model assumes that not being home indicates a temporary situation, so that one could not substitute the current situation of previous not-at-homes for current not-at-homes.

My major concern, however, is that we move to a different research question, namely, what is the potential labor force. We may even want to start assessing the extent to which people want to work less!

---

(1) George Katona, Charles Lininger and Eva Mueller, 1963 Survey of Consumer Finances, (Ann Arbor, Michigan: Survey Research Center, 1964), p.33.

XII

RECENT DEVELOPMENTS IN SAMPLING

Chairman, ELI S. MARKS, University of Pennsylvania

A MULTIVARIATE APPROACH TO ESTIMATION IN PERIODIC SAMPLE SURVEYS

Margaret Gurney and Joseph F. Daly
U.S. Bureau of the Census

1. Introduction:

The increasing use of periodic probability-based sample surveys to study the way in which selected characteristics of a population vary with time has raised many interesting problems of sample design and estimation. The degree of interest which this topic has generated may be gauged by the (admittedly incomplete) bibliography at the end of this paper, listing articles that have come to our attention.

Although the fundamental ideas in this area are already implicit in the basic papers of Jessen [10] and Patterson [18], they are often obscured in the literature by the myriad of sampling details surrounding the problems to which they are applied. In this paper, which is essentially expository, we shall examine a specific periodic sample survey in an effort to show how the problem of constructing "good" estimators can be fruitfully regarded as a problem in standard multivariate analysis and treated by means of techniques currently in use for handling general stochastic processes.

2. The Problem:

A sample survey which continues over a period of time is capable of producing, for each time period, many estimates of each of the characteristics of the population being studied. Each individual observation can be used to make an estimate, or the individual observations can be combined in some desired manner to make one or more estimates for a particular time period.

An estimate which does not make use of the survey data for any time period except that period to which the estimate refers may be called an "elementary estimate." It should be possible to improve such an estimate by making use of correlated elementary estimates available from other time periods. The purpose of this paper is to discuss minimum variance unbiased linear combinations of elementary estimates, and to outline a method of computation which will determine the weights to be used on the various elementary estimates to obtain the best linear unbiased estimate, whether for an estimate of level, of change over time, of an average over time, or in general, for any linear combination of the elementary estimates.

3. A Special Case:

The Current Population Survey of the Bureau of the Census is a monthly household survey, which has been in operation for many years. Data are collected on labor force items, demographic characteristics, and on other characteristics of the population for which a household survey is an appropriate vehicle. At each month estimates are made, for many characteristics, of the current level, of changes since earlier months, and of averages over several months.

The survey is based on a rotating sample in which one-fourth of the households are replaced by a new selection of households each month. In each month, one-fourth of the households are new, one-fourth have been interviewed in the preceding month, one-fourth have been interviewed for three consecutive months, and one-fourth have been interviewed for four months. Each fourth of the sample is treated separately and an elementary estimate for the month is made from it.

The following additional information is available for the survey, and can be used to improve the estimates which are made from the sample for a particular characteristic: (a) the elementary estimates for past months, for each panel, (b) estimates of the average correlations over time for observations one month, two months, and three months apart, and (c) an estimate of the average variance for a single panel, averaged over time.

Consider an estimate for a particular month, say January. If the panels are labelled "A", "B", "C", etc., with "A" corresponding to the panel which has just entered the survey in January, the rotation pattern may be described by the following diagram. The numbering of the elementary estimates X_i in the following chart is chosen to facilitate the computations. This numbering makes the covariance matrix of the X's a straightforward direct product of simple covariance matrices. It is obvious how to terminate the numbering, for a finite number of months.

Month	Panel									
	A	B	C	D	E	F	G	H	I	etc.
Jan.	X_1	X_2	X_4	X_7						
Dec.		X_3	X_5	X_8	X_{11}					
Nov.			X_6	X_9	X_{12}	X_{15}				
Oct.				X_{10}	X_{13}	X_{16}	X_{19}			
Sept.					X_{14}	X_{17}	X_{20}	X_{23}		
Aug.						X_{18}	X_{21}	X_{24}	X_{27}	
etc.							etc.			

As a first approximation we shall assume the following characteristics for the survey:

A. The expected value of the estimates of the 4 panels for a given month is the same, e.g., for January $EX_1 = EX_2 = EX_4 = EX_7$.

B. The covariances between estimates from the same panel at two different months depend only on the number of months between the two estimates. For example,

$$\mathrm{var}(X_1) = \mathrm{cov}(X_1,X_1) = \mathrm{cov}(X_2,X_2) = \sigma^2$$

$$\mathrm{cov}(X_2,X_3) = \mathrm{cov}(X_4,X_5) = \mathrm{Cov}(X_5,X_6) = \rho_1\sigma^2$$

$$\mathrm{cov}(X_4,X_6) = \mathrm{cov}(X_7,X_9) = \mathrm{cov}(X_{12},X_{14}) = \rho_2\sigma^2$$

$$\mathrm{cov}(X_7,X_{10}) = \mathrm{cov}(X_{11},X_{14}) = \mathrm{cov}(X_{15},X_{18}) = \rho_3\sigma^2$$

The average values of ρ_1, ρ_2 and ρ_3 are estimated from past data from the survey. For convenience we shall put $\sigma^2 = 1$.

C. The panels are selected independently, so that (for example)

$$\mathrm{cov}(X_1,X_2) = \mathrm{cov}(X_2,X_4) = \mathrm{cov}(X_2,X_6) = 0, \text{ etc.}$$

Under these assumptions an estimate for January may be made in several ways. The simplest is to use only the observations for the month of January for panels, A, B, C, and D. Since the panels are independent, the best estimate in this case is

$$_1\hat{X}_{Jan} = (X_1 + X_2 + X_4 + X_7)/4$$

It can easily be shown that this estimate has the minimum variance of any linear combination of X_1, X_2, X_4, and X_7; its variance is .25 times the variance of an individual elementary estimate.

When such a survey has been in operation for two months, starting in December, the data available for use in making the best estimate for January are the four observations for panels A, B, C, and D for January, and the four observations for panels B, C, D, and E for December. It is possible to determine the coefficients C_i so that the estimate

$$_2\hat{X}_{Jan} = C_1X_1 + C_2X_2 + C_4X_4 + C_7X_7 + C_3X_3 + C_5X_5 + C_8X_8 + C_{11}X_{11}$$

shall have the minimum variance, knowing the relations

$$\mathrm{cov}(X_2,X_3) = \mathrm{cov}(X_4,X_5) = \mathrm{cov}(X_7,X_8) = \rho_1$$

$$EX_1 = EX_2 = EX_4 = EX_7 ; EX_3 = EX_5 = EX_8 = EX_{11}$$

For an estimate of the civilian labor force, with $\rho_1 = .8$, the coefficients corresponding to the minimum variance are shown in Table 1.

Table 1.--ESTIMATE OF CIVILIAN LABOR FORCE FOR JANUARY FROM CPS SURVEY: COEFFICIENTS OF MINIMUM VARIANCE UNBIASED LINEAR ESTIMATES FOR SURVEYS STARTING IN JANUARY, DECEMBER, OCTOBER, AUGUST AND APRIL

Correlation Pattern: $\rho_1 = .8$, $\rho_2 = .7$, $\rho_3 = .65$

Duration of survey	Panel												
	A	B	C	D	E	F	G	H	I	J	K	L	M
1. One month (started in January)													
January....................	.250	.250	.250	.250									
2. Two months (started in preceding December)													
January....................	.219	.260	.260	.260									
December....................		-.052	-.052	-.052	.156								
3. Four months (started in preceding October)													
January....................	.195	.258	.269	.278									
December....................		-.079	-.036	-.032	.147								
November....................			-.063	-.025	-.005	.093							
October....................				-.061	-.021	.004	.078						
4. Six months (started in preceding August)													
January....................	.189	.258	.271	.282									
December....................		-.086	-.037	-.032	.155								
November....................			-.074	-.026	-.002	.102							
October....................				-.075	-.005	.007	.073						
September....................					-.045	-.006	.001	.050					
August....................						-.027	-.013	.000	.040				
5. Ten months (started in preceding April)													
January....................	.187	.259	.271	.283									
December....................		-.089	-.037	-.032	.158								
November....................			-.077	-.026	-.002	.105							
October....................				-.080	-.006	.007	.079						
September....................					-.053	-.006	.003	.056					
August....................						-.037	-.005	.002	.040				
July....................							-.026	-.003	.001	.028			
June....................								-.018	-.002	.001	.019		
May....................									-.012	-.001	.000	.013	
April....................										-.007	-.003	.000	.010

Table 1 shows also the coefficients for an estimate of the civilian labor force for January when the survey has been in operation four, six, and ten months, respectively. It may be seen that the coefficients for the earliest months in the sample are smaller than the coefficients for the most recent months; they approach zero rather rapidly as the number of months increases. Moreover, the coefficients for the observations from the most recent month approach constant values as the number of months increases; if the survey had been started much earlier, the coefficients for January would be approximately equal to those which are shown for ten months. In fact, after six months, the coefficients for the most recent months are close to the corresponding coefficients for ten months. Hence data from the most recent six to ten months will provide practically all of the improvement which can be achieved in the estimate. The speed of convergence depends upon the covariances between estimates for the same panel at different times. If the covariances are very high (for example $\rho_1 = .95$), the time required for convergence is longer; on the other hand, if ρ_1 is small, say .50, the convergence will be much faster, and only three or four months may be required to approximate the optimum coefficients.

If the survey has been in operation for two or more months, it is possible to make a revised estimate for a preceding month which will have a smaller variance than the one originally obtained for that month. For example, an estimate for December, made when data for January are available, will usually have a smaller variance than the original estimate for January. Coefficients for an estimate of civilian labor force for December, using the data for January, as well as all earlier data, are shown in Table 2, for a survey which has been in operation for two, four, six, and ten months, respectively.

In this manner one can make a revised estimate for November, using data through January, which will have a smaller variance than one which used data only through November (or December).

Table 2.--ESTIMATE OF CIVILIAN LABOR FORCE FOR DECEMBER FROM CPS SURVEY: COEFFICIENTS OF MINIMUM VARIANCE UNBIASED LINEAR ESTIMATES FOR SURVEY STARTING IN JANUARY, DECEMBER, OCTOBER, AUGUST AND APRIL

Correlation Pattern: $\rho_1 = .8$, $\rho_2 = .7$, $\rho_3 = .65$

Duration of survey	Panel												
	A	B	C	D	E	F	G	H	I	J	K	L	M
1. One month (started in January)													
No estimate for December													
2. Two months (started in preceding December)													
January	.156	-.052	-.052	-.052									
December		.260	.260	.260	.219								
3. Four months (started in preceding October)													
January	.128	-.052	-.041	-.035									
December		.224	.264	.267	.245								
November			-.060	-.033	-.035	.128							
October				-.042	-.047	-.006	.095						
4. Six months (started in preceding August)													
January	.120	-.051	-.039	-.030									
December		.214	.263	.267	.256								
November			-.072	-.034	-.032	.138							
October				-.060	-.023	-.003	.086						
September					-.063	-.005	.005	.063					
August						-.035	-.015	.000	.050				
5. Ten months (started in preceding April)													
January	.118	-.051	-.038	-.029									
December		.211	.263	.267	.259								
November			-.077	-.034	-.031	.142							
October				-.067	-.024	-.002	.093						
September					-.073	-.005	.007	.071					
August						-.047	-.006	.003	.050				
July							-.032	-.005	.002	.035			
June								-.022	-.003	.001	.024		
May									-.015	.002	.001	.016	
April										-.009	-.004	.000	.013

The variances corresponding to the best selection of coefficients for an estimate of the civilian labor force, for January and for earlier months, are shown in Table 3.

Table 3.--VARIANCES OF DELAYED ESTIMATES OF CIVILIAN LABOR FORCE: SURVEY ENDING IN JANUARY

Correlation Pattern: $\rho_1 = .8$, $\rho_2 = .7$, $\rho_3 = .65$

Estimates	Duration of survey (months)				
	1	2	4	6	10
A. Monthly Level:					
Jan..........	.250	.219	.195	.189	.187
Dec..........		.219	.183	.173	.170
Nov..........			.183	.167	.162
Oct..........			.195	.167	.157
Sept.........				.173	.155
Aug..........				.189	.155
July.........					.157
June.........					.162
May..........					.170
April........					.187
B. Month-to-month change:					
Jan. - Dec...		.125	.122	.122	.122
Dec. - Nov...			.121	.120	.120
Nov. - Oct...			.122	.120	.119
Oct. - Aug...				.120	.118
Sept. - Aug..				.122	.118
Aug. - July..					.118
July - June..					.119
June - May...					.120
May - April..					.122

The best coefficients for any linear combination of the estimates for January, December, etc., are obtained by taking the same linear combination of the coefficients of the best estimates of the corresponding months. For example, the coefficients for the best estimate of changes from December to January, $X'_{Jan} - X'_{Dec}$, are obtained by subtracting the coefficients in Table 2 from the corresponding coefficients in Table 1. The variance of such a difference is given by

$$\mathrm{var}(X'_{Jan} - X'_{Dec}) = \mathrm{var}(X'_{Jan}) + \mathrm{var}(X'_{Dec}) -$$

$2\mathrm{cov}(X'_{Jan}, X'_{Dec})$, where the covariance between X'_{Jan} and X'_{Dec} is a number which is obtained as part of the general solution of the system. Table 3 presents also variances for estimates of month-to-month change. when the survey has been in operation two, four six, and ten months. The variance of month-to-month change is quite stable, as can be seen from the table.

4. General Procedure:

To obtain the coefficients of the desired minimum variance linear unbiased estimate, we find it helpful to use a geometric approach suggested by the work of Parzen [16], [17] on the application of Hilbert space methods to stochastic processes.

In our problem we have a finite set of chance variables $X_1, \ldots, X_n$ whose joint probability distribution is assumed to belong to a family of distributions subject only to the conditions:

A. No one of the X's is essentially a linear function of the remaining X's. (If initially we had

$$X_n = \sum_{i=1}^{n-1} C_i X_i + \text{const.},$$

with probability one, we assume that we have eliminated X_n from the set, and so on.)

B. The covariances $K_{ij} = E(X_i - EX_i) \cdot (X_j - EX_j)$ are finite and known.

C. The expected values $\mu_i = EX_i$ are subject only to certain linear homogeneous restrictions such as

$$\mu_2 = \mu_3, \ \mu_4 = \mu_5 = \mu_6, \text{ etc.}$$

or more generally,

$$\sum_{i=1}^{n} a_{hi} \mu_i = 0 \qquad (h = 1, \ldots, p)$$

Our approach, in brief, is to let the variables $X_1, \ldots, X_n$ correspond to some vectors $\underline{k}_1, \ldots, \underline{k}_n$ which form a basis for an n-dimensional Hilbert space $\underline{V}$. (Halmos [6], [7]). The vector $\underline{m} = (\mu_1, \ldots, \mu_n)$ lies in this space, and is by virtue of condition C free to range over some subspace $\underline{M}$ of $\underline{V}$. To find the coefficients λ_i of the best linear estimator $\Sigma \lambda_i X_i$ of an arbitrary homogeneous linear combination of the μ_i, say $\Sigma C_i \mu_i$, we form the vector $\Sigma C_i \underline{k}_i$, find its projection $\underline{v}^*$ on the subspace $\underline{M}$, and then express $\underline{v}^*$ in terms of the $\underline{k}$'s. The coefficients in this expression will be the required λ's.

To set up the correspondence, we begin by regarding the rows of the covariance matrix

$$K = \begin{pmatrix} K_{11} & \cdots & K_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ K_{n1} & \cdots & K_{nn} \end{pmatrix}$$

as vectors $\underline{k}_i = \{K_{i1}, \ldots, K_{in}\}$ in an ordinary n-dimensional vector space $\underline{V}$. Next we show that these vectors $\underline{k}$ form a basis for $\underline{V}$, so that any vector $\underline{a}$ with arbitrary components $\alpha_1, \ldots, \alpha_n$ can be represented uniquely in the form $a = \Sigma\, \theta_i \underline{k}_i$. This follows from the fact that the $\underline{k}_i$ are linearly independent, which in turn follows from our condition A. above and from the fact that

$$\mathrm{var}(\Sigma\, \lambda_i X_i) = \Sigma\, \lambda_i K_{ij} \lambda_j.$$

Since any vector $\underline{v}$ in $\underline{V}$ is a linear combination of the basis vectors $\underline{k}$, we can define the inner product of any two vectors

$$\underline{a} = \Sigma\, \theta_i \underline{k}_i, \qquad \underline{b} = \Sigma\, \Psi_j \underline{k}_j$$

in $\underline{V}$ by

$$(\underline{a},\underline{b}) = (\Sigma\, \theta_i \underline{k}_i, \Sigma\, \Psi_j \underline{k}_j) = \Sigma\, \theta_i K_{ij} \Psi_j.$$

This inner product can be thought of as defining angles and lengths in $\underline{V}$ by interpreting the inner product of $\underline{a}$ and $\underline{b}$ as the product of the "lengths" of $\underline{a}$ and $\underline{b}$ by the cosine of the "angle" between them. In particular, we shall say that $\underline{a}$ and $\underline{b}$ are perpendicular if $(\underline{a},\underline{b}) = 0$, and shall call $(\underline{a},\underline{a})$ the square of the length of $\underline{a}$. It then follows that the variance of any linear estimator $\Sigma\, \lambda_i X_i$ is equal to the length of the corresponding vector $\Sigma\, \lambda_i \underline{k}_i$:

$$(\Sigma\, \lambda_i \underline{k}_i, \Sigma\, \lambda_j \underline{k}_j) = \Sigma\, \lambda_i K_{ij} \lambda_j = \mathrm{var}(\Sigma\, \lambda_i X_i) \qquad (1)$$

We note further that

$$(\underline{a},\underline{k}_j) = (\Sigma\, \theta_i \underline{k}_i, \underline{k}_j) = \Sigma\, \theta_i K_{ij},$$

which is by definition the j-th component of the vector $\underline{a}$. Hence if $\underline{m}$ is the mean-value vector with components $\mu_1, \ldots, \mu_n$ it follows that

$$(\underline{m}, \Sigma\, \lambda_i \underline{k}_i) = \Sigma\, \lambda_i \mu_i = E(\Sigma\, \lambda_i X_i) \qquad (2)$$

Thus if $\underline{a} = \Sigma\, \theta_i \underline{k}_i$ is any vector in $\underline{V}$, the expected value of the estimator $\Sigma\, \theta_i X_i$ is given by $(\underline{m},\underline{a})$ and its variance by $(\underline{a},\underline{a})$.

Let us now examine the conditions C! on the components $\mu_i = EX_i$ of the vector $\underline{m}$. Suppose, for example, we have the situation (see Section 3)

Month \ Panel	A	B	C	D	E
January........	X_1	X_2	X_4	X_6	
December.......		X_3	X_5	X_7	X_8

We would then expect the mean-value vector to be of the form

$$\underline{m} = \{\alpha, \alpha, \beta, \alpha, \beta, \alpha, \beta, \beta\}$$

so that we could write $\underline{m} = \alpha \underline{u}_1 + \beta \underline{u}_2$ where

$$\underline{u}_1 = \{1, 1, 0, 1, 0, 1, 0, 0\}$$

$$\underline{u}_2 = \{0, 0, 1, 0, 1, 0, 1, 1\}.$$

More generally, the conditions C imply that there are some linearly independent vectors $\underline{u}_1, \ldots, \underline{u}_m$ such that $\underline{m}$ satisfies C if and only if $\underline{m}$ is of the form

$$\underline{m} = \Sigma\, \xi_\alpha \underline{u}_\alpha$$

i.e. $\underline{m}$ ranges over the subspace $\underline{M}$ of $\underline{V}$ spanned by the vectors $\underline{u}_1, \ldots, \underline{u}_m$.

Let us now single out a particular basis vector $\underline{k}_h$ corresponding to the estimator X_h. As we have seen,

$$(\underline{m},\underline{k}_h) = \mu_h = EX_h$$

for any mean-value vector $\underline{m} = \{\mu_1, \ldots, \mu_h\}$. Suppose $\underline{v}$ is any other vector in $\underline{V}$ such that

$$(\underline{m},\underline{v}) = \mu_h$$

for every vector $\underline{m}$ in $\underline{M}$. Since each of the $\underline{u}_\alpha$ is then a possible choice of $\underline{m}$, we must have

$$(\underline{u}_\alpha,\underline{v}) = u_{\alpha h} \qquad (\alpha = 1, \ldots, m) \qquad (3)$$

where $u_{\alpha h}$ stands for the h-th component of $\underline{u}_\alpha$.

We now show that among such vectors $\underline{v}$ satisfying (3) we can find one, say $\underline{v}^*$, which lies in $\underline{M}$. This is equivalent to the problem of finding numbers $\gamma_1, \ldots, \gamma_m$ which satisfy the equations

$$(\underline{u}_\alpha, \Sigma\, \gamma_\beta \underline{u}_\beta) = \Sigma(\underline{u}_\alpha,\underline{u}_\beta)\gamma_\beta = u_{\alpha h}.$$

Unless the matrix of elements $L_{\alpha\beta} = (\underline{u}_\alpha, \underline{u}_\beta)$ is singular, these equations have the (unique) solution

$$\gamma_\beta = \Sigma\, u_{\alpha h}\, L^{-1}_{\alpha\beta} \tag{4}$$

Where $L^{-1}_{\alpha\beta}$ stands for the (α,β) element of the inverse of $(L_{\alpha\beta})$. But if $(L_{\alpha\beta})$ were singular, there would be numbers ζ_α not all zero such that $\Sigma\, \zeta_\alpha L_{\alpha\beta} = 0$. This would imply $0 = \Sigma\, \zeta_\alpha L_{\alpha\beta}\zeta_\beta = (\Sigma\, \zeta_\alpha \underline{u}_\alpha, \Sigma\, \zeta_\beta \underline{u}_\beta)$, so so that the vector $\Sigma\, \zeta_\alpha \underline{u}_\alpha$ would have zero length. However, in view of our earlier discussion of the K matrix, the only vector of zero length in $\underline{V}$ is the zero vector. And the linear independence of the vectors $\underline{u}_\alpha$ makes it impossible to have $\Sigma\, \zeta_\alpha \underline{u}_\alpha = 0$ unless every ζ is zero.

We now have a vector $\underline{v}^*$ which lies in $\underline{M}$ and which satisfies the conditions

$$(\underline{u}_\alpha, \underline{v}^*) = (\underline{u}_\alpha, \underline{k}_h) = u_{\alpha h} \qquad \alpha = 1, \ldots, m.$$

As an immediate consequence of these conditions, we have

$$(\underline{u}_\alpha, \underline{k}_h - \underline{v}^*) = 0 \qquad \alpha = 1, \ldots, m.$$

Geometrically speaking, we have resolved the vector $\underline{k}_h$ into two components: a component $\underline{v}^*$ which lies in $\underline{M}$ and a component $\underline{k}_h - v^*$ which is perpendicular to every vector in $\underline{M}$. In this sense we call $\underline{v}^*$ the projection of $\underline{k}_h$ on the subspace $\underline{M}$.

The fact that $\underline{v}^*$ is the shortest vector satisfying (3) now follows readily. Let $\underline{v}$ be any other vector in $\underline{V}$ which satisfies (3). Then

$$(\underline{u}_\alpha, \underline{v} - \underline{v}^*) = 0 \qquad \alpha = 1, \ldots, m$$

so that $\underline{v} - \underline{v}^*$ is perpendicular to every vector in $\underline{M}$, and in particular, $(\underline{v}^*, \underline{v} - \underline{v}^*) = 0$. Consequently

$$(\underline{v},\underline{v}) = (\underline{v}^* + \underline{v} - \underline{v}^*, \underline{v}^* + \underline{v} - \underline{v}^*) = (\underline{v}^*, \underline{v}^*) + (\underline{v} - \underline{v}^*, \underline{v} - \underline{v}^*)$$

since the usual cross-product term $2(\underline{v}^*, \underline{v} - \underline{v}^*)$ vanishes. Hence

$$(\underline{v}, \underline{v}) \geq (\underline{v}^*, \underline{v}^*)$$

Thus the estimator corresponding to the vector $\underline{v}^*$ will have minimum variance among all unbiased linear estimators of μ_h.

To find the coefficients of this best linear estimator, we must express the vector $\underline{v}^*$ in terms of the basis vectors $\underline{k}_i$.

From (4) we have

$$\underline{v}^* = \Sigma\, u_{\alpha h}\, L^{-1}_{\alpha\beta}\underline{u}_\beta.$$

But $\underline{u}_\beta = \Sigma\, \theta_i \underline{k}_i$ where the θ_i are determined from the relations

$$u_{\beta j} = \Sigma\, \theta^{(\beta)}_i K_{ij}, \text{ or } \theta^{(\beta)}_i = \Sigma\, u_{\beta j} K^{-1}_{ji}.$$

Hence

$$v^* = \Sigma\, u_{\alpha h}\, L^{-1}_{\alpha\beta} u_{\beta j}\, K^{-1}_{ji}\underline{k}_i \tag{5}$$

where $L_{\alpha\beta}$ can be expressed in terms of the $u_{\alpha i}$ and the K_{ij} by noting that

$$L_{\alpha\beta} = (\Sigma\, \theta^{(\alpha)}_i \underline{k}_i, \Sigma\, \theta^{(\beta)}_j \underline{k}_j) = \Sigma\, \theta^{(\alpha)}_i K_{ij}\theta^{(\beta)}_j = \Sigma\, u_{\alpha i} K^{-1}_{ij} u_{\beta j}.$$

From Equation (5) we then see that the coefficient of X_i in the best linear unbiased estimator of μ_h is

$$\lambda^{(h)}_i = \Sigma\, u_{\alpha h}\, L^{-1}_{\alpha\beta}\, u_{\beta j} K^{-1}_{ji} \tag{6}$$

The preceding discussion was formally restricted to finding the best linear unbiased estimator of some particular component μ_h of the mean-value vector. But from the easily demonstrated fact that the projection on $\underline{M}$ of a linear combination of vectors $\underline{v}_h$ is the same linear combination of their projections, it follows that the optimum estimator of some linear combination of the μ's, such as $\underline{\mu}_3 - \mu_1$, is the same linear combination of the optimum estimators of the individual components $\underline{\mu}_h$.

Equations (6) in the matrix form

$$(\lambda) = U^T(U\, K^{-1}\, U^T)^{-1}UK^{-1}, \text{ where } U = (u_{\alpha i}) \tag{7}$$

are readily programmed on an electronic computer to determine the λ's for given μ's and K's.[1]

Two final comments seem in order at this point. In the first place from $\underline{v}^* = \Sigma\, u_{\alpha h} L^{-1}_{\alpha\beta}\, u_\beta$ we readily obtain the variance of the optimum estimator of μ_h:

$$(\underline{v}^*, \underline{v}^*) = \Sigma\, u_{\alpha h} L^{-1}_{\alpha\beta} u_{\beta h} \tag{8}$$

which simplifies considerably in the problem studied in this paper because in general for a given h one of the $u_{\alpha h}$ is one, and the rest are zero. And secondly the matrix

$$(U\, K^{-1}\, U^T)^{-1}\, UK^{-1}$$

[1] The computation is illustrated in Appendix A.

has the properties of a "generalized inverse" of U^T, a notion that is being increasingly exploited in recent research on least squares estimation (Greville [5], Goldman and Zelen [4]).

5. Approximation to the Optimum Estimate:

The purpose of this section is to examine some alternative estimators which approximate the optimum estimate, and to compare them with the optimum. A desirable feature of these estimators is that they are somewhat easier to compute. Moreover, they provide estimates which have most of the gains of the optimum estimators.

Several forms of "composite estimators" will be considered. A composite estimate is a weighted average to two (or more) linear unbiased estimates of the same characteristic for a given time period; the weights are selected so as to reduce the variance, as compared with the variances of the original estimates. These composite estimators are defined recursively, and use only a limited number of elementary estimates, combined with composite estimates which have already been computed. The description below of several composite estimators will illustrate the definition.

A. Simple Composite Estimator:

To form a "simple composite estimator" for a given month, say January,

(1) make a simple average of the elementary estimates for January from panels A, B, C, and D:

$$X'_{Jan} = (X_1 + X_2 + X_4 + X_7)/4$$

(2) make a difference estimate for January by adding to the (already computed) composite estimate for December the estimate of the December-January change, based on identical panels. Let the composite estimator be designated by X*, and the change by $\Delta_{J,D}$:

$$\Delta_{J,D} = (X_2 + X_4 + X_7 - X_3 - X_5 - X_8)/3$$

The difference estimate for January is

$$X^*_{Dec} + \Delta_{J,D}$$

(3) make a weighted average of the estimates of (1) and (2) above:

$$X^*_{Jan} = (1-K)X'_{Jan} + K(X^*_{Dec} + \Delta_{J,D}) \qquad (9)$$

B. Composite Estimator with Change from Three Previous Time Periods:

The rotation pattern of Section 3 permits estimates of change for identical panels to be made for two successive months, for times two months apart, and for times three months apart. An estimator which permits the use of this additional information is the following:

$$X^*_{Jan} = (1-K-L-M)X'_{Jan} + K(X^*_{Dec} + \Delta_{J,D}) + L(X^*_{Nov} + \Delta_{J,N}) + M(X^*_{Oct} + \Delta_{J,O}) \qquad (10)$$

Here X'_{Jan} is defined as before, and the Δ's are self-explanatory; X* as used here is of course different from that in Equation (9).

C. A Modification of the Simple Composite Estimator:

Some improvement in the estimate of Equation (9) can be made if the linear combination of observations for January has more weight on panel A, and less on panels B, C, and D. Such a change will bring the coefficients on the observations for January more in line with coefficients of the optimum estimate. Let the term $(1-K)X'_{Jan}$ in Equation (9) be replaced by

$$\{(1-K+A)X_1 + (1-K-A/3)(X_2+X_4+X_7)\}/4 \qquad (11)$$

which is equivalent to

$$(1-K)X'_{Jan} + \{(AX_1 - \tfrac{A}{3}(X_2+X_4+X_7)\}/4$$

The expected value of the term in braces is zero. This form is called an "AK-Composite Estimator"; except for the addition of the terms in braces, the formula is the same as that in Equation (9).

D. Composite Estimator with Year-to-year Change:

The Current Population Survey is actually based on eight panels rather than four, which rotate in such a way that there is a 50 percent overlap of households from year to year, as well as the 75 percent overlap from month to month. When the year-to-year correlation is high (relative to the 12th power of the month-to-month correlation), appreciable gains may be obtained by the use of year-to-year change in the composite estimate. An appropriate formula is

$$X^*_{Jan} = (1-K-Q)X''_{Jan} + K(X^*_{Dec} + \Delta_{J,D}) + Q(X^*_{Jan-12} + \Delta_{Jan,Jan-12}) \qquad (12)$$

where X''_{Jan} is an average based on the eight January panels, and Jan-12 refers to January of the preceding year. X''_{Jan} may be a simple average, or may have unequal weights on the eight panels.

E. More General Forms of Composite Estimators:

More general composite estimators can be developed; however, an estimate which is too general will lose some of the advantages of the simpler estimates: it may require the use of too many of the elementary estimates, and may require the retention of too many earlier composite estimates.

The four forms of the composite estimator given above are versions of

$$X^*_{Jan} = Z_{Jan} + KX^*_{Dec} + LX^*_{Nov} + MX^*_{Oct} + QX^*_{Jan-12} \qquad (13)$$

where Z_{Jan} is composed of contributions from the estimates of level for January, and from several estimates of change.

F. Comparisons with the Minimum Variance Unbiased Linear Estimator:

How well a particular composite estimator will approximate the minimum variance unbiased linear estimator may be measured by a comparison of the variances which are obtained by the two estimators. One may also compare the coefficients to be used on the elementary estimates, for different estimators.

Table 4 shows the coefficients for the most recent four months (Panels A-G) which are appropriate for several composite estimates, for an estimate of the civilian labor force; it compares them with the coefficients for the minimum variance unbiased linear estimate based on data for ten months. The comparisons are made for the months of January, December, November, and October. In each case, the constants (K, L, M, Q) used for the composite estimate are the ones corresponding to the smallest value of the variance for the particular estimate considered.

Table 4 shows also the variances of estimates for January of the civilian labor force, for the various estimators.[2] The numbers are relative to the variance of the simple average for January. It is seen that the AK-composite estimate approximates the minimum variance unbiased linear estimator quite closely: its variance is only slightly larger than that of the best estimator, and the coefficients are close to those for the best estimate.

The estimate using year-to-year change (Estimate 5) is not strictly comparable to the other estimates in this table, as it uses information ($\Delta_{Jan,Jan-12}$) which is not available to the ten-month minimum variance unbiased linear estimate. It is included to point out that a high year-to-year correlation pattern can effect further improvements in the estimate.

6. The AK-Composite Estimate Used with Several Characteristics:

The AK-composite estimate is a good approximation to the optimum linear unbiased estimate for a characteristic such as the civilian labor force when A is .4 and K is .7. For an estimate such as the change in monthly level of the labor force, or for another characteristic such as unemployment (which has much lower correlations over time than the civilian labor force), these values of A and K will not be the best. Table 5 shows the variances expected for several correlation patterns, for estimates of level and of month-to-month change, for a number of values of A and K.

It may be seen from an examination of Table 5 that the values A = .2, and K = .7, while not the best for all characteristics, still provide appreciable gains over the simple average of elementary estimates, and even over the present composite estimate for the CPS (which uses the values A = 0 and K = .5) when the correlations are moderate or high.

When the correlations are low very small gains may be expected, compared with the simple average; there may be losses if the values of A and K differ appreciably from zero.

7. Variations:

The data in Tables 1 through 5 have been based on average values of variances and covariances, and on equal means for all of the observations at all time periods. In practice the variances and covariances will change over time, the means at different time periods may have a seasonal pattern, or a long term trend, and the expected values of the observations relating to the same time period may not be identical.

A. Changing Correlation Pattern:

Table 6 presents coefficients and variances for the minimum variance unbiased linear estimate for ten months when the correlations are not equal over time, but vary by as much as 20 percent from equality. The correlations used are shown at the top of the table. Even this

[2] The method of computation is illustrated in Appendix B.

rather large departure from equal correlations has almost no affect on the variance of the estimate of the January level: by comparison with Table 3 it is increased from .187 to .189. The variance of the estimate of January-December change is unchanged at .122. The fact that the variances and covariances are not known more exactly is of little importance.

Table 4.--ESTIMATE OF CIVILIAN LABOR FORCE FOR JANUARY FROM CPS SURVEY

A. Coefficients for Recent Months and Variances, for Several Linear Unbiased Estimates

Estimate and month	Panel							Variance of estimate for January (relative to the simple average)
	A	B	C	D	E	F	G	
1. Minimum variance unbiased linear estimate -- 10 months started in April								
January................	.187	.259	.271	.283				.749
December...............		-.089	-.037	-.032	.158			
November...............			-.077	-.026	-.002	.105		
October................				-.080	-.006	.007	.079	
2. Simple composite estimate: K = .6								
January................	.100	.300	.300	.300				.817
December...............		-.140	-.020	-.020	.180			
November...............			-.084	-.012	-.012	.108		
October................				-.050	-.007	-.007	.064	
3. Three-month composite estimate: K = .4, L = .1, M = .05								
January................	.112	.246	.296	.346				.796
December...............		-.088	-.035	-.015	.138			
November...............			-.074	-.039	.023	.090		
October................				-.083	-.007	.023	.067	
4. AK-composite estimate: K = .7, A = .4								
January................	.175	.275	.275	.275				.756
December...............		-.111	-.041	-.041	.193			
November...............			-.077	-.029	-.029	.135		
October................				-.054	-.020	-.020	.094	

B. Coefficients and Variance for Composite Estimate Using Year-to-year Change

Estimate and month	Panel														
	A	B	C	D	E	F	G	...	M	N	O	P	Q	R	S
5. Month-month and year-year change: K = .5, Q = .2															
Jan.	-.037	.121	.121	.121				...	.087	.171	.171	.171			
Dec.		-.065	-.023	-.023	.061					-.040	.002	.002	.086		
Nov.			-.033	-.011	-.011	.030					-.020	.001	.001	.043	
Oct.				-.016	-.006	-.006	.015					-.010	.001	.001	.021
								⋮							
Variance of estimate for January relative to the simple average:								Jan.-12	-.042	-.026	-.026	-.026			
								Dec.-12		-.031	-.016	-.016	-.001		
								Nov.-12			-.016	-.011	-.011	.006	
.709								Oct.-12				-.015	-.007	-.007	.006
								⋮							

Table 5.--COMPARISON OF VARIANCES OF AK-COMPOSITE ESTIMATES FOR SEVERAL CORRELATION PATTERNS, AND FOR SEVERAL VALUES OF A AND K

Correlation pattern	K	Variance, relative to simple average (A = 0, K = 0) of --									
		Estimate of monthly level					Estimate of month-to-month change[1]				
		A=0	A=.1	A=.2	A=.3	A=.4	A=0	A=.1	A=.2	A=.3	A=.4
1. High--											
ρ_1=.95, ρ_2=93, ρ_3=.90	.4	.791	.784	.785	.792		.524	.543	.580	.636	
	.5	.725	.717	.716	.722	.735	.442	.455	.485	.532	.596
	.6	.651	.642	.640	.644	.657	.370	.379	.403	.443	.499
	.7	.576	.564	.560	.563	.574	.309	.314	.334	.369	.419
	.8	.518	.500	.492	.491	.499	.262	.265	.282	.312	.357
2. Moderate--											
ρ_1=.80, ρ_2=.70, ρ_3=.65	.4	.857	.845	.840	.843		.728	.735	.755	.787	
	.5	.829	.812	.803	.802	.808	.690	.692	.706	.732	.770
	.6	.817	.792	.777	.770	.771	.661	.660	.669	.689	.721
	.7	.848	.806	.780	.764	.756	.641	.637	.642	.658	.684
	.8	.978	.920	.874	.837	.812	.635	.627	.629	.640	.662
3. Low--											
ρ_1=.50, ρ_2=.40, ρ_3=.30	.4	.971	.951	.938	.933		.928	.922	.925	.935	
	.5	1.005	.975	.953	.939	.934	.931	.922	.920	.926	.939
	.6	1.084	1.038	1.003	.978	.961	.942	.930	.924	.925	.933
	.7	1.262	1.187	1.130	1.084	1.048	.960	.945	.935	.932	.936
	.8	1.676	1.560	1.458	1.370	1.297	.985	.968	.955	.949	.948

1 Based on difference of two estimates of level.

Table 6.--ESTIMATE OF CIVILIAN LABOR FORCE FROM CPS SURVEY: MINIMUM VARIANCE UNBIASED LINEAR ESTIMATE

Coefficients and Variances with a Changing Correlation Pattern

A. Changing Correlation Pattern

Correlation between:	ρ_1	Correlation between:	ρ_2	Correlation between:	ρ_3
Jan.-Dec.	.80	Jan.-Nov.	.70	Jan.-Oct.	.65
Dec.-Nov.	.76	Dec.-Oct.	.68	Dec.-Sept.	.61
Nov.-Oct.	.72	Nov.-Sept.	.64	Nov.-Aug.	.57
Oct.-Sept.	.68	Oct.-Aug.	.60	Oct.-July	.53
Sept.-Aug.	.64	Sept.-July	.60	Sept.-June	.57
Aug.-July	.68	Aug.-June	.64	Aug. May	.61
July-June	.72	July-May	.68	July-April	.65
June-May	.76	June-April	.70		
May-April	.80				

B. Variances

Estimate	Variance
Jan. level	.189
Dec. level	.176
Jan.-Dec. change	.122

C: Coefficients for Panels (Estimate of January Level)

Month	Panel												
	A	B	C	D	E	F	G	H	I	J	K	L	M
Jan.	.189	.254	.272	.285									
Dec.		-.082	-.042	-.026	.150								
Nov.			-.073	-.039	.014	.098							
Oct.				-.077	-.017	.022	.072						
Sept.					-.046	-.014	.012	.048					
Aug.						-.032	-.010	.011	.031				
July							-.020	-.006	.006	.020			
June								-.016	-.003	.005	.014		
May									-.010	-.002	.002	.010	
April										-.007	-.003	.001	.009

B. Unequal Means at a Particular Time Period:

It may happen that the observations at a single time period do not all have the same expected value. For example, in the Current Population Survey, the first time a household is interviewed it appears to respond differently to the interview, with respect to some characteristics (for example, employment status), than at the second or later interviews. Following the diagram in Section 3, one might have for January:

$$EX_1 = (1+a)\mu$$

when

$$EX_2 = EX_4 = EX_7 = \mu$$

There may be the same kind of bias in the reports for December, November, etc.

When there are unknown response biases in the expected values, both the minimum variance unbiased linear estimator and the various composite estimators may produce estimates which are biased.

If the pattern of bias is constant over time the total bias will approach a limit for each of the estimators which have been discussed here. Table 7 shows the biases to be expected in several cases.

In Table 7 a characteristic is considered which is possessed by about 10,000,000 persons in the population and which has a correlation pattern similar to that of the civilian labor force. The sampling error of this estimate from a simple average of elementary estimates is 200,000 (i.e., about 2 percent). Two patterns of bias in the estimates from the four panels at a single month are considered:

(1) The bias occurs only at the first time at which a household is interviewed. The bias is of the same order at each time period. The pattern of expected values at a single time period is

$$(1+a)\mu,\ \mu,\ \mu,\ \mu.$$

(2) The "newest" and "oldest" panels have compensating biases. The pattern of expected values for each month is

$$(1+a)\mu,\ \mu,\ \mu,\ (1-a)\mu.$$

Table 7 is computed for values of "a" equal to 100,000 and 200,000. The resulting root mean square errors are compared with the standard errors of an unbiased estimate, for several estimators.

Table 7 shows that when the bias in the estimate from the "new" panel is one-half (or even equal to) the size of the standard error of the estimate, the root mean square error is hardly any larger than the standard error of the corresponding estimate. The gains which are achieved in using the minimum variance unbiased linear estimate, or the several composite estimates, persist, even with a bias of this size.

Table 7.--EFFECT OF CONSISTENT BIAS ON THE RELIABILITY OF SAMPLE ESTIMATES OF A CHARACTERISTIC HAVING A CORRELATION PATTERN OF $\rho_1 = .8$, $\rho_2 = .7$, $\rho_3 = .65$

(Size of Estimate is 10,000,000)

Estimator	Standard error of estimate (10^3)	a = 100 x 10^3: Bias in new panel only (1+a, 1, 1, 1): Bias (10^3)	a = 100 x 10^3: Bias in new panel only (1+a, 1, 1, 1): RMSE (10^3)	a = 200 x 10^3: Bias in new panel only (1+a, 1, 1, 1): Bias (10^3)	a = 200 x 10^3: Bias in new panel only (1+a, 1, 1, 1): RMSE (10^3)	a = 200 x 10^3: Compensating bias in new and oldest panels (1+a, 1, 1, 1-a): Bias (10^3)	a = 200 x 10^3: Compensating bias in new and oldest panels (1+a, 1, 1, 1-a): RMSE (10^3)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Simple average............	200	-25	202	-50	205	0	200
Minimum variance unbiased linear estimate-10 months.	173	-21	174	-42	178	-201	265
Simple composite:							
K = .6...............	181	-25	182	-50	188	-200	270
K = .5...............	182	- 8	182	-17	183	- 33	185
AK-composite:							
K=.7, A=.4...............	174	-19	175	-39	178	-222	282
K=.6, A=.3...............	175	- 6	176	-12	176	-149	230
K=.5, A=.3...............	179	+ 7	179	+13	180	- 93	202

If biases should occur in two panels, and should be compensating, (see columns 7 and 8 of Table 7) the simple average of the elementary estimates would be unbiased. The estimators which make the most use of past data--the minimum variance estimator and the AK-composite with a high value of K--show the largest increases due to such a bias. The columns for compensating biases are included in the table to illustrate perhaps the worst situation which could occur; the likelihood of compensating biases is very small.

If the deviations in expected values are known, or can be measured quite accurately, one may consider the advisability of adjusting the sample estimates accordingly. However, the assumption that the bias will continue to be the same in the future as in the past may lead to serious errors. A better procedure is to try to eliminate the response bias, if it is significant.

C. A Different Rotation Pattern: The Census Current Business Reports Survey:

For small retail and service establishments, The Current Business Reports Survey of the Bureau of the Census is based on 12 panels, one of which is enumerated each month. At the time of enumeration, information on retail sales is obtained for the preceding month and for the next earlier month. After a year, the panel which was in the sample a year ago is enumerated again.

The rotation pattern is

Month	Panel: Jan.	Dec.	Nov.	Oct.	Sept.	Aug.	July	June	May	April	Mar.	Feb.
January.	X_1											
December	X_2	X_3										
November		X_4	X_5									
October.			X_6	X_7								
September				X_8	X_9							
August..					X_{10}	X_{11}						
July....						X_{12}	X_{13}					
June....							X_{14}	X_{15}				
May.....								X_{16}	X_{17}			
April...									X_{18}	X_{19}		
March...										X_{20}	X_{21}	
February											X_{22}	X_{23}
January.	X_{25}											X_{24}
December	X_{26}	X_{27}										
						etc.						

For this rotation pattern, the minimum variance unbiased linear estimate is very close in form to a composite estimate. If the rotation pattern is altered so that the sample for any month is independent of that for any other month (i.e., the sample for the months of one year are not repeated in subsequent years), then the minimum variance estimate can be expressed exactly in composite form.

For an estimate of January level, we write

$$X^*_{Jan} = X_1 - K_J X_2 + K_J X^*_{Dec}.$$

This estimator will have minimum variance when

$$K_J = \operatorname{cov}(X_1,X_2)/\{\operatorname{var}(X_2) + \operatorname{var}(X^*_{Dec})\}.$$

Here X^*_{Dec} is defined in a similar manner, in terms of X_3, X_4, X^*_{Dec} and a constant K_D, which is defined in terms of $\operatorname{cov}(X_3,X_4)$, $\operatorname{var}(X_4)$, and $\operatorname{var}(X^*_{Oct})$.

These relationships hold, whether the survey has started recently, or whether it has been in operation for a long time. When the covariances are all equal to ρ, and the variances of the individual elementary estimates (X_i) are all equal, say σ^2, the value of K approaches a limit, as time passes:

$$K = \frac{1 - \sqrt{1-\rho^2}}{\rho}$$

and the variance becomes

$$\text{var}(X^*) = \sigma^2 \sqrt{1-\rho^2}.$$

With the altered pattern, the best estimate of month-to-month change (say from December to January) will be obtained by making a revised estimate for December, using the data available from the January survey, and subtracting this from the best January estimated. The revised estimate for December can also be written in composite form

$$X^*_{\text{Dec(rev)}} = (1-K/\rho)X_2 + (K/\rho-K)X_3 + K(X^*_{\text{Nov(rev)}} + X_3 - X_4)$$

where, as before,

$$K = (1 - \sqrt{1-\rho^2})/\rho$$

The variance of the revised estimate is

$$\sigma^2 K/\rho \sqrt{1-\rho^2} ;$$

it is smaller than the variance of the unrevised estimate by a factor of K/ρ.

The best estimate of the December-January difference is then $X^*_{\text{Jan}} - X^*_{\text{Dec(rev)}}$ as was noted by Patterson [18]. In fact, an estimate of this form, which uses all of the available data, has the minimum variance of any linear estimate of month-to-month change, even when the variances and correlations between panels are not constant over time.

APPENDIX A

Computation of Minimum Variance Unbiased Linear Estimate

To illustrate the computation of the minimum variance unbiased linear estimate, consider a set of three elementary estimates, which might be obtained at the beginning of a survey with a rotation pattern like that of the Current Business Reports Survey, as shown in the diagram in paragraph C of Section 7. The observation X_1 is an elementary estimate for the month of January; observations X_2 and X_3 are elementary estimates for the preceding month, December. It is desired to make the minimum variance unbiased linear estimate of level for January, having the following information:

1. $EX_1 = \mu_1;\ EX_2 = EX_3 = \mu_2$

2. The covariance matrix is

$$K = \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

3. The mean value vector is

$$\underline{m} = (\mu_1, \mu_2, \mu_2)$$

which can be written as

$$\underline{m} = \mu_1 \underline{u}_1 + \mu_2 \underline{u}_2$$

where

$$\underline{u}_1 = (1, 0, 0)$$

$$\underline{u}_2 = (0, 1, 1)$$

The matrix of coefficients of the minimum variance unbiased linear estimator for estimating the expected values of the three variables X_1, X_2, and X_3 is (see Equation 7)

$$C = U^T(U\,K^{-1}\,U^T)^{-1}\,UK^{-1}$$

The covariance matrix of the optimum solutions is (see Equation 8)

$$P = U^T(U\,K^{-1}\,U^T)U.$$

The solution is indicated in the following equations

$$K^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho & 0 \\ -\rho & 1 & 0 \\ 0 & 0 & 1-\rho^2 \end{pmatrix}$$

Define

$$L = U\,K^{-1}U^T = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho \\ -\rho & 2-\rho^2 \end{pmatrix}$$

$$L^{-1} = \frac{1}{2}\begin{pmatrix} 2-\rho^2 & +\rho \\ +\rho & 1 \end{pmatrix}$$

Define the matrix

$$P = U^T\,L^{-1}\,U = \frac{1}{2}\begin{pmatrix} 2-\rho^2 & \rho & \rho \\ \rho & 1 & 1 \\ \rho & 1 & 1 \end{pmatrix}$$

and the matrix of coefficients

$$C = PK^{-1} = \begin{pmatrix} \underline{c}_1 \\ \underline{c}_2 \\ \underline{c}_3 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 2 & -\rho & +\rho \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

The minimum variance unbiased linear estimate for January level is obtained when the coefficients $\underline{c}_1$ are used:

$$\hat{X}_1 = X_1 - \rho(X_2 - X_3)/2.$$

The optimum estimates of December level employ the coefficients $\underline{c}_2$ and $\underline{c}_3$, respectively, and lead to identical solutions:

$$\hat{X}_2 = \hat{X}_3 = (X_2 + X_3)/2$$

The covariance matrix P, which gives the variances and covariances between

$$\hat{X}_1,\ \hat{X}_2,\ \text{and}\ \hat{X}_3$$

leads to

$$\text{var}(\hat{X}_1) = (2 - \rho^2)/2$$

$$\text{var}(\hat{X}_2) = \text{var}(\hat{X}_3) = 1/2$$

The variance of the best estimate of change $(\hat{X}_1 - \hat{X}_2)$ is

$$\begin{aligned}\text{var}(\hat{X}_1 - \hat{X}_2) &= \text{var}(\hat{X}_1) + \text{var}(\hat{X}_2) - 2\,\text{cov}(\hat{X}_1, \hat{X}_2)\\ &= (3 - 2\rho - \rho^2)/2.\end{aligned}$$

These results may be verified by conventional methods. For example, one may construct the variance of

$$\hat{X}_1 = \sum_{j=1}^{3} C_{1j} X_j,$$

and determine the values of the coefficients which minimize the variance, subject to conditions 1 and 2.

APPENDIX B

Computation of Variances for Composite Estimates

The general composite estimator, Equation (13), may be written as

$$X^*_1 = Z_1 + KX^*_2 + LX^*_3 + MX^*_4 + QX^*_{13} \qquad (14)$$

where the subscript "1" designates the most recent month, "2" is the preceding month, etc.

To illustrate the computation of the variance of a composite estimate, consider the simple composite estimator:

$$\begin{aligned}X^*_1 &= Z_1 + KX^*_2\\ &= Z_1 + KZ_2 + K^2Z_3 + K^3Z_4 + \cdots \qquad (15)\end{aligned}$$

1. Define, at each time period t, for $t = 1, 2, 3, \ldots$ and $i = 0, 1, 2, \ldots$

$$\begin{aligned}Y_o &= \text{var}(X^*_t)\\ Y_i &= \text{cov}(X^*_t, X^*_{t+i})\\ A_i &= \text{cov}(Z_t, X^*_{t+i})\\ \rho_{iz}\,\text{var}(Z) &= \text{cov}(Z_t, Z_{t+i})\end{aligned}$$

and

$$\rho_{o,z} = 1$$

2. Take covariances between Equation (15) and X^*_1, and X^*_2:

$$\begin{aligned}Y_o &= A_o + KY_1\\ Y_1 &= A_1 + KY_o\end{aligned} \qquad (16)$$

These two simultaneous equations can be solved for Y_o and Y_1 if $|K| < 1$. The solution will give values for $\text{var}(X^*_1)$ and $\text{cov}(X^*_1, X^*_2)$; higher covariances can be obtained successively from

$$Y_i = A_i + KY_{i-1} \qquad (17)$$

for $i = 2, 3, 4, \ldots$

3. The $\{A_i\}$ satisfy a set of covariance equations obtained by taking covariances between Z_1 and $\{X^*_i = Z_i + KX^*_{i+i}\}$ for $i = 1, 2, 3, \ldots$:

$$A_i = \rho_{iz}\,\text{var}(Z) + KA_{i+1} \qquad (18)$$

In particular, for the simple composite estimate (Equation 17)

$$\begin{aligned}A_o &= \rho_{o,z}\,\text{var}(Z) + KA_1\\ A_1 &= \rho_{1z}\,\text{var}(Z) + KA_2\\ A_2 &= \rho_{2z}\,\text{var}(Z) + KA_3\end{aligned} \qquad (19)$$

etc.

4. The form of Z_t is determined by the rotation pattern and the weights assigned to the panels. For the simple composite estimate, with the rotation pattern of the diagram in Section 3, we find

$$Z_1 = (1-K)(X_1+X_2+X_4+X_7)/4 + K(X_2+X_4+X_7-X_3-X_5-X_8)/3$$

or

$$Z_1 = \frac{(1-K)}{4}X_1 + \frac{3+K}{12}(X_2+X_4+X_7) - \frac{K}{3}(X_3+X_5+X_8) \qquad (20)$$

Z_2 is defined similarly, using the elementary estimates with subscripts corresponding to months "2" and "3,", etc.

The variances and covariances

$$\text{cov}(Z_\alpha, Z_{\alpha+i}) = \rho_{iz}\,\text{var}(Z)$$

may be expressed in terms of the variances and covariances of the original (elementary) estimates X_i; that is, in terms of $\text{var}(X)$ and $\{\rho_{ix}\}$.

For the rotation pattern of the diagram, correlations more than three months apart are zero; it turns out that only var(X), ρ_{1x}; ρ_{2x} and ρ_{3x} are non-zero. Starting with $A_4 = 0$, in Equation (18), A_3, A_2, A_1, and A_o may be determined recursively, for particular values of K, and of

var(X) and ρ_{1x}, ρ_{2x}, ρ_{3x}.

5. Having found $\{A_i\}$, Equations (16) may be solved for Y_o and Y_1; in general, $\{Y_i\}$ may be found by applying Equation (17) successively.

6. The form of the $\{Y_i\}$ lends itself readily to the evaluation of a number of estimators, in addition to the estimate of level for a single month. For example:

Variance of monthly level $= var(X_1^*)$

$= Y_o$

Variance of month-to-month change in level $= var(X_1^*-X_2^*)$

$= 2(Y_o-Y_1)$

Variance of year-to-year change in level $= var(X_1^*-X_{13}^*)$

$= 2(Y_o-Y_{12})$

7. By varying the value of K, while keeping $|K| < 1$, it is possible to set up a table of variances for each estimator (of level, month-to-month change, etc.) and to determine the value of K which leads to the minimum value for each. Frequently a compromise value of K can be found which will make gains for several characteristics, and for several estimators, although it may not be the best value for any one.

8. For more complicated composite estimators the system will consist of more than two equations. For example, if year-to-year change is incorporated into the estimate, the system will consist of 13 equations. The form of $\{Z_i\}$ may also be quite complicated, depending on the conditions set on the coefficients of the estimate of level, and on the number of estimates of change which are used.

REFERENCES

[1] Bershad, M.A., and Nisselson, H., "Sampling for Weekly Time Series," Proceedings of the Social Statistics Section, American Statistical Association (1962), pp. 207-213.

[2] Cochran, W.G. Sampling Techniques. New York: John Wiley & Sons, Inc., Second Edition (1963), pp. 341-354.

[3] Eckler, A.R., "Rotation sampling," Annals of Mathematical Statistics,26(1955),pp.664-685.

[4] Goldman, A.J., and Zelen, M., "Weak generalized inverses and minimum variance linear unviased estimation," Journal of Research, National Bureau of Standards, 68B (1964), pp. 151-172.

[5] Greville, T.N.E., "Some applications of the pseudo-inverse of a matrix," Society for Industrial and Applied Mathematics Review, 2 (1960), pp. 15-22.

[6] Halmos, P.R. Introduction to Hilbert Space and the Theory of Spectral Multiplicity. New York: Chelsea Publishing Company, Second Edition 1957).

[7] Halmos, P.R. Finite Dimensional Vector Spaces. Princeton: D. van Nostrand Co., Inc., Second Edition (1958).

[8] Hansen, M.H., Hurwitz, W.N., and Madow, W.G. Sample Survey Methods and Theory. New York: John Wiley & Sons, Inc. (1953), Vol. I, pp. 490-503, Vol. II, pp. 268-279.

[9] Hansen, M.H., Hurwitz, W.N., Nisselson, H., and Steinberg, J., "The redesign of the Census Current Population Survey," Journal of the American Statistical Association, 50 (1955), pp. 701-719. Bershad, M.A. and Gurney, M., Appendix: "Winin PSU variance of the composite estimate."

[10] Jessen, R.J., "Statistical investigation of a sample survey for obtaining farm facts," Iowa Agricultural Experiment Station Research Bulletin, 304 (1942).

[11] Kruskal, W., "The coordinate-free approach to Gauss-Markov estimation, and its application to missing and extra observations," Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. I (1960), pp. 435-451.

[12] Kulldorff, G., "Some problems of optimum allocation for sampling on two occasions," Review of the International Statistical Institute, 31 (1963), pp. 24-57.

[13] Narain, R. D., "On the recurrence formula in sampling on successive occasions, "Journal of the Indian Society of Agricultural Statistics, 5 (1953), pp. 96-99.

[14] Narain, R.D., "The general theory of sampling on successive occasions," Bulletin of the International Statistical Institute, 34 (2) (1954), pp. 87-89, 201-202.

[15] Onate, B.T., "Development of multi-stage designs for statistical surveys in the Philippines," Mimeo-multilith series No. 3: Statistical Laboratory, Iowa State University, Ames, Iowa. (1960).

[16] Parzen, E., "Statistical inference on time series by Hilbert space methods, I," Technical Report No. 23: Department of Statistics, Stanford University (1959).

[17] Parzen, E., "An approach to time series analysis," Annals of Mathematical Statistics, 32 (1961), pp. 951-989.

[18] Patterson, H.D., "Sampling on successive occasions with partial replacement of units," Journal of the Royal Statistical Society, Series B, 12 (1950), pp. 241-255.

[19] Rao, J.N.K., and Graham, J.E., "Rotation designs for sampling on repeated occasions," Journal of the American Statistical Association, 59 (1964), pp. 492-509.

[20] Singh, D., and Singh, B.D., "Double sampling for stratification on successive occasions," Journal of the American Statistical Association, (In print -- 1965).

[21] Tikkiwal, B.D., "Optimum allocation in successive sampling," Journal of the Indian Society of Agricultural Statistics, 5 (1953), pp. 100-102.

[22] Tikkiwal, B.D., "A further contribution to the theory of univariate sampling on successive occasions," Journal of the Indian Society of Agricultural Statistics, 8 (1956), pp. 84-90.

[23] Tikkiwal, B.D., "On the theory of classical regression and double sampling estimation," Journal of the Royal Statistical Society, Series B, 22 (1960), pp. 131-138.

[24] U. S. Bureau of the Census. The Current Population Survey -- A Report on Methodology. Technical Paper No. 7, U.S. Government Printing Office, Washington, D. C. (1963).

[25] Waksberg, J., and Pearl, R.B., "The Current Population Survey: A Case History in Panel Operations," Proceedings of the Social Statistics Section, American Statistical Association, (1964), pp. 217-228.

[26] Woodruff, R.S., "The use of rotating samples in the Census Bureau's monthly surveys," Journal of the American Statistical Association, 58 (1959), pp. 454-467.

[27] Yates, F. Sampling Methods for Censuses and Surveys. London: Charles Griffin and Co., Third Edition (1960), pp. 175-182, 233-235, 260-262.

MULTIPLE PURPOSE OPTIMUM ALLOCATION IN STRATIFIED SAMPLING

H. O. Hartley, Texas A&M University

1. Optimum allocation in stratified sampling*

Consider a finite population of N units subdivided into H strata containing N_h units $(h = 1,2,\ldots,H)$ respectively. Denote by y_{hi} the y of the i^{th} unit in the stratum h and by

$$Y_h = \sum_{i=1}^{N_h} y_{hi} \text{ and } \bar{Y}_h = Y_h/N_h \qquad (1)$$

the strata totals and means and by

$$Y = \Sigma\, Y_h \text{ and } \bar{Y} = Y/N \qquad (2)$$

the
population total and mean. A random sample of n_h units is drawn at random from the h^{th} stratum and denote by

$$y_h,\ \bar{y}_h \qquad (3)$$

the sample strata totals and means corresponding to (2). The customary unbiased estimators of $\bar{Y}_h$ and $\bar{Y}$ are respectively given by

$$\bar{y}_h \text{ and } \hat{y} = \Sigma\, (N_h/N)\, \bar{y}_h \qquad (4)$$

and the variance of $\hat{y}$ by

$$\text{Var }(\hat{y}) = \sum_h (N_h/N)^2\, S_h^2\, (1/n_h - 1/N_h) = \sum_h (N_h/N)^2\, S_h^2/n_h - V_{con} \qquad (5)$$

where the h^{th} stratum variance, S_h^2, is given by

$$S_h^2 = (N_h - 1)^{-1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2 \qquad (6)$$

and V_{con} does not depend on the n_h.

If the cost of drawing the sample is given by the linear cost function

$$C_{TOT} = C_0 + \Sigma\, C_h\, n_h = C_0 + C \qquad (7)$$

then the classical 'optimum allocation' is defined as that set of n_h which minimizes Var $(\hat{y})$ for a given cost C. From classical Lagrangean calculus we obtain

$$n_h^* = C\, (S_h\, N_h\, C_h^{-\frac{1}{2}}) \Big/ \sum_h (S_h\, N_h\, C_h^{+\frac{1}{2}}) \qquad (8)$$

resulting in a minimum variance of

$$V_{min} = C^{-1}\, N^{-2}\, (\sum_h S_h\, N_h\, C_h^{+\frac{1}{2}})^2 - V_{con} \qquad (9)$$

The formal solution (8) will of course, only be of practical use if

$$1 \leq n_h^* \leq N_h \qquad (10)$$

and will, in general, be fractional.

It will be shown in 2. that (8) does indeed yield an absolute minimum of (5) at constant cost.

2. Multiple purpose optimum allocation

Most sample surveys are concerned with obtaining estimates of a fairly large number of population parameters and not just the single linear estimate $\hat{y}$ of $\bar{Y}$. Usually a large number of variables is measured for each sampled unit and not only is it required to estimate the population means for each of these but if the data are used in 'analytic studies' it may be of interest to estimate differences between all or some of the strata means for some or all of the variables but also for other subsections of the population called 'domains of study.' We propose to consider therefore a number of J different estimators $\hat{y}_j$ which are linear functions of some or all of the strata means $\bar{y}_h$ and may involve these means for one or several of the variables. The variance of such linear estimators will be of the form

$$\text{Var }(\hat{y}_j) = \sum_h a_{hj}\, (1/n_h - 1/N_h) = \sum_h a_{hj}\, n_h^{-1} - V_j \qquad (11)$$

where V_j does not depend on n_h. We retain the linear cost function (7) and consider three possible definitions of optimizations

(A)* Minimize a weighted sum of the J variances

$$\sum_{j=1}^{J} W_j \text{ Var }(\hat{y}_j) = \sum_h n_h^{-1} \sum_{j=1}^{J} W_j\, a_{hj} - \sum_j W_j\, V_j = \Sigma\, n_h^{-1}\, A_{hj} - V \qquad (12)$$

at constant cost C.

*See e.g. Cochran, W. C. (1962)

*See Yates, F. (1953) and Cochran, W. G. (1962)

(B) Prescribe values v_j for the variances in the form

$$\sum_{h=1}^{H} n_h^{-1} a_{hj} = v_j : j = 1,2,\ldots,J \qquad (13)$$

and minimize the cost C given by (7) subject to (13).

(C)**Set tolerances for all variances in the form

$$\sum_{h=1} n_h^{-1} a_{hj} \leq v_j \qquad (14)$$

and minimize the cost (7) subject to the inequality restrictions (14). At first sight it may be argued that (B) is not necessary in view of (C) since one would surely not wish to force the variances to attain the upper tolerance v_j if it is possible to achieve smaller variances at the same cost. However, the utility of (B) lies in using its solution for variable v_j, under certain circumstances.

3. The solution to problem (A)

The solution to (A) is, of course, equivalent to a single purpose optimization with the minimum weighted variance (12) attained for

$$n_h^* = C(A_h/C_h)^{\frac{1}{2}} / \sum_h (A_h C_h)^{\frac{1}{2}} \qquad (15)$$

where

$$A_h = \sum_{j=1}^{J} W_j a_{hj}$$

and given by

$$V_{min} = C^{-1} (\sum_h (A_h C_h)^{\frac{1}{2}})^2 - V \qquad (16)$$

In the special case of

$$W_{j'} = 1, W_j = 0 \text{ for } j \neq j' \qquad (17)$$

The problem reduces to the single purpose minimization of Var $(\hat{y}_{j'})$ and (15) becomes

$$n_h (j') = C(a_{hj'}/C_h)^{\frac{1}{2}} / \sum_h (a_{hj'} C_h)^{\frac{1}{2}} \qquad (18)$$

leading to the minimum

$$V_{min}(j') = C^{-1} (\sum_h (a_{hj'} C_h)^{\frac{1}{2}})^2 - V_{j'} \qquad (19)$$

**See Dalenius, T. (1957), Yates, F. (1953) and Cochran, W. G. (1962)

We now show that (19) is an absolute minimum for Var $(\hat{y}_j)$:

Consider an allocation n_h satisfying the given cost condition

$$C = \sum_h C_h n_h \qquad (20)$$

then we have to show that the variance of $\hat{y}_j$ computed from (11) and using the sample sizes n_h will exceed V_{min} (j). Using (11), (19) and (20) we obtain

$$C(\mathrm{Var}(\hat{y}_j) - V_{min} (j)) =$$

$$=(\sum_h C_h n_h) (\sum_h a_{hj} n_h^{-1}) - (\sum_h (a_{hj} C_h)^{\frac{1}{2}})^2$$

$$= \sum_h C_h n_h ((a_{hj}/C_h)^{\frac{1}{2}} n_h^{-1} - Av)^2 \geq 0 \qquad (21)$$

where Av is the weighted average.

$$Av = \sum C_h n_h (a_{hj}/C_h)^{\frac{1}{2}} n_h^{-1} / \sum C_h n_h \qquad (22)$$

Formula (21) shows that the minimum variance is attained if and only if the n_h satisfy (13) and will in general provide the amount by which Var $(\hat{y}_j)$ exceeds V_{min}(j).

4. The solution to problem B

The J linear equations (13) for the H variable n_h can, of course, only be satisfied if H-J of the equations are linearly dependent upon H of them and even then the solutions may not yield positive n_h. We shall therefore confine our discussion of this problem to such specifications of v_i which are 'of interest' that is to situations in which the system (13) has at least a one parametric infinite set of solutions. A necessary condition for a minimum of the cost C under the restrictions (13) is given by (15) where the weights W_j are now to be interpreted as Lagrangean multipliers and must be determined by substituting (15) into (13). It is easy to show by reference to the 2nd order differentials that (15) is also a sufficient condition for the n_h^* to yield an absolute minimum of the cost C provided the W_j are determined to satisfy (13). In practice, however, one would not proceed in this manner but rather start from the Lagrangean weights W_j, and go through the following steps:

(i) Choose weights W_j representing the relative importance of the variances $V(\hat{y}_j)$ and fix a budget C for the survey. Solve problems (A) yielding the optimum allocation n_h^* given by (15).

(ii) By substituting the n_h^* in (15) compute individual variances $v_j - V_j$ for the estimators $\hat{y}_j$. Since the n_h^* are now also the solutions to problem (B) it can be stated that at least these variances $v_j - V_j$ can not be achieved at a smaller cost than the budget C.

(iii) Compare the $v_j - V_j$ with the $V_{min}(j)$ for the same budget given by (19) and increase the weights W_j for such $\hat{y}_j$ where the excess is 'disappointingly large.'

(iv) If the adjustment in (iii) does not lead to a satisfactory set of v_j and if a constant percentage decrease in the v_j is desired the corresponding percentage increase in the budget C will achieve this.

There are obvious limitations to the formulation and solution of the multiple purpose optimization problem in the form (A) and (B), and we may summarize them as follows:

The main reservation about minimizing a weighted variance (as in (A)) puts the onus on the choice of weights W_j which may result in unreasonably high variances of some of the $\mathrm{Var}(\hat{y}_j)$ in the weighted sum. The approach in (B) however, does much to rectify this disadvantage: It not only shows that at least the actually attained $\mathrm{Var}(\hat{y}_j)$ have been met with minimum cost, but it also gives a feed-back for the improvement of the choice of weights. There remain, however, two main disadvantages. First, the procedure described above will in general require that $J<H$, i.e. that the number of estimators entering into the optimization does not exceed the number of strata, and moreover should be moderate or small for convenience in the adjustment of the W_j. Secondly, the solution n_h^* may well exceed N_h and we have so far not discussed what to do in such situations. All these problems can be resolved if we adopt the formulation (C) and solve it by non-linear programming.

5. The solution of problem C by non-linear programming

In finding the minimum cost C under the inequality restrictions (14) we find it convenient to introduce the reciprocals

$$r_h = 1/n_h - 1/N_h \quad h = 1,\dots,H \tag{23}$$

as the elements of our activity vector r which results in a convex activity space with the linear boundaries defined by

$$Ar \leq v - V \tag{24}$$

and the upper 'bounds'

$$0 \leq r_h \leq 1 - 1/N_h \tag{25}$$

where A is the H x J matrix of the a_{hj} and $v - V$ the J-vector with elements $v_j - V_j$ that is the set of tolerances for the variances $\mathrm{Var}(\hat{y}_j)$.

No assumptions need be made concerning the rank of A or the magnitude of H and J except within the framework of available computer codes. The 'objective function,' i.e. the cost now becomes the convex function

$$C = \sum_{h=1}^{H} C_h\left(r_h + \frac{1}{N_h}\right)^{-1} \tag{26}$$

and is of a form described as 'separable' see Charnes and Lemke (1954) and Hartley (1960). This fact would make available the procedure by Hartley (1960) which would involve an approximation to the hyperbolae $\left(r_h + \frac{1}{N_h}\right)^{-1}$ by a moderate number of linear line segments which method has been shown to reduce the problem to linear programming. Alternatively the method recently published by Hartley and Hocking (1963) could be used which does not require polygonal approximations. For the details of the algorithm we must refer to this paper. We confine ourselves here to stating that a new variable is introduced in the form

$$r_{H+1} = -C = -\Sigma\, C_h \left(r_h + \frac{1}{N_h}\right)^{-1} \tag{27}$$

and that r_{H+1} is maximized whilst (27) occurs as a (non-linear) restriction. The problem is solved in the dual form which leads to the following tableau.

Tableau I. Dual tableau for non-linear programming problem

	j=0	j= 1,...J	h= 1,...h=H	h= H+1	h= H+2
h=0	0	$v_j - V_j$	$(1-1/N_h)$	M	C^+
h=1	0		-1	0	
	.		-1	.	
	.		.	.	
	.		.	.	
	.		.	.	
	.	$-a_{hj}$	.	.	$+C_h(r_h+\frac{1}{N_h})^{-2}$
	.		-1	.	
h=H	0		-1	0	
h=H+1	1	0...0	0 . . . 0	1	1

Tableau I will be recognized as the dual tableau in standard form, for maximizing r_{H+1} subject to the restrictions $Ar \leq v-V$ and $r_h \leq (1-1/N_h)$ except that the line h=0 represents the negative of the dual objective function and that the last two columns require some explanation: The column h=H+1 is an 'artificial vector' to supplement the (H+1) x (H+1) identity matrix of slack vectors (not shown in Tableau I) for an initial "basis." Its 'penalty' M will eventually drive it out of the basis. The last column represents the non-linear restriction (27) and is non-standard. Whilst for an explanation of this column we must refer to Hartley and Hocking (1963) we should state here that its first element is given by

$$C^+ = -\Sigma\, c_h(r_h+\frac{1}{N_h})^{-1} - \Sigma\, c_h(r_h+\frac{1}{N_h})^{-2} r_h \qquad (28)$$

and that it is evaluated for varying argument vectors r_h in the course of the simplex process and may contribute several columns for the current basis matrix.

It will be noted that the problem leads to a dual Tableau of size (H+2) x (J+H+3) which is quite a feasible size for a high-speed computer even if H is of the order 50 and J of the order 200. Three small numerical examples are given in Hartley and Hocking (1963).

It must of course be remembered that the algorithm of the non-linear programming technique only yields a numerical optimum allocation n_h^* (in the r_h^* scales) for the particular problem and no general formula for the n_h^*. It may therefore rightly be asked whether there are techniques available which exhibit numerically the effect on the r_h^* of altering the specified variance tolerances $v_j - V_j$. There is indeed such a technique available which is known under the name of 'parametric programming' and which is incorporated in most computer codes. Another question which may be asked concerns the uncertainty in the a_{hj} which depend on the strata variances. Since the strata variances S_h^2 would normally not be known but estimated one may wish to regard the a_{hj} as stochastic variables. Such a model would lead to methods of stochastic programming. Here we are certainly more restricted with regard to the availability of methods and computer codes.

More recently we have obtained some new results on convex parametric programming using a modification of Hartley and Hocking (1963) which will be published shortly. With these methods it will be possible to examine how an alteration of the variance tolerances $v_j - V_j$ effects the optimum allocation n_h^* and the cost C.

References

Charnes, A. and Lemke, C. E. (1954), 'Minimization of Non-linear Separable Convex Functionals.' Naval Research Logistics Quarterly, I (1954), 301-12.

Cochran, W. G. (1962), 'Sampling Techniques' 2nd Edition J. Wiley and Sons, New York.

Dalenius, T. (1957), 'Sampling in Sweden. Almquist and Wicksell, Stockholm.

Folks, J. Leroy and Antle, C. E. (1965), 'Optimum Allocation of Sampling Units to Strata When There are R Responses of Interest,' JASA, 1965, p. 225-233.

Hartley, H. O. (1961), 'Non-Linear Programming by The Simplex Method.' Econometrica, 29 223-237.

Hartley, H. O. and Hocking, R. (1963), 'Convex Programming by Tangential Approximation.' Management Science (in Press)

Kokam, A. R. (1963), 'Optimum Allocation in Multivariate Surveys.' JRSS, Series A, 126 p. 557-565.

Yates, F. (1955), 'Sampling Methods for Censuses and Surveys.' 2nd Edition Hafner, New York.

A MULTIPLE FRAME SURVEY FOR RARE POPULATION ELEMENTS

Joseph Steinberg, Social Security Administration

I. Introduction

This paper discusses one approach to the sample design of a survey for estimating characteristics of a number of rare, but related, target populations. Some of the considerations involved in efficient achievement of study objectives are presented. A number of problems encountered in creating the design are somewhat other than run-of-the-mill and, therefore, may be of interest.

As is described in greater detail both area and list frames are involved in the design. The general plan involved in this sample design is not novel and only in some respects may be considered as having a few new features. In general, when one speaks of using list frames, one envisions certain characteristics--as a minimum, currency of addresses. As a practical matter, many of the lists used for this survey do not have the property of showing current addresses. As a consequence, special procedures are required to insure an unbiased sample. Another characteristic envisioned in the use of list frames is the ability to allocate directly elementary units to first-stage sampling units. As a practical matter, many of the lists (the same ones) can be allocated only to groupings of parts of first-stage units. Again, as a consequence, special procedures are required.

In most plans involving choice between screening for unduplication among frames versus multiple-frame overlap weighting, a decision may be made after consideration of cost and variance. In this sampling plan for the major groupings of list frames and area sample, because of the time relationships of various activities needed for the screening approach, the cost would be extremely large if screening were attempted. Thus, as a practical matter, in these cases only the use of multiple-frame overlap weighting can be used (and at a substantial reduction in cost of securing the necessary information for carrying through the operation). Within one major grouping, screening versus multiple frame overlap weighting exists as a choice; time features would not intrude extensively and as a result after proper consideration of the alternatives screening is being used since it is more efficient.

II. Purpose of Survey

In brief, the purpose of the survey is to study the social and economic consequence of disability among adults. It is estimated that among persons 18 to 64 there are approximately 8% with chronic limitations in their major activities. Among persons in the overall target population, disabilities vary in severity of limitation. The extent of incapacity ranges from limitations in the amount or kind of work which can be performed to total inability for self-care. It is estimated that approximately 1.5% of the adult population is so severely disabled as to be unable to work or keep house. The problems of the disabled population in relation to income insecurity is to be a principal focus of the survey. Thus, the amount of income received by the disabled adult and the sources from which this income is derived (whether major insurance or assistance programs) is of basic interest.

As a consequence of the principal purpose of the survey, six primary target populations were identified. These are characterized by cross-classification of two levels of severity of disability by three categories of source of income. It is speculated that the sizes of these six target populations are approximately as follows:

(In thousands)

Income Source	Unable	Limited
Total	1,500	7,000
Social Security	750	120
Other Public Income Maintenance Programs	600	760
Other	150	6,120

and, therefore, range from about 0.1% to 6% of those 18-64. The primary statistical goals that are to be realized have been formulated as follows:

1. The sample design should permit the determination (at a risk of error of 1 chance in 20) that if 10% of persons in one of the three categories of "unable" have a given characteristic that this is significantly different from an estimated 15% having the same characteristic among another of these three target populations (i.e., that an estimate that 10% of

persons receiving social security disability benefits and classifying themselves as unable to work have a given characteristic is significantly different from a group similarly characterized estimated to be 15% of those who classify themselves as unable to work but who receive no resources from any public income maintenance program).

2. An estimate of any group of approximately 50,000 should be subject to an absolute error of about 10,000 (with a risk of being in error of about 1 in 20).

From this general description, it may be discernible that the usual techniques of area sampling, perhaps supplemented by a single list or two, is not likely to be the most efficient method for dealing with these objectives. A straight area sampling technique would require a very large basic sample (perhaps as much as 250,000 households) to satisfy the study requirements for that target population which is 0.15% of the total population. There are a number of interrelated aspects to the problems of sample design in this situation. Among these are (1) the problems of the availability of certain types of resources such as: Area sample materials and lists. (2) Possible ways of development of special frames. (3) Consideration of double sampling approaches. (4) The possibility of screening among frames for discarding duplicates versus use of multiple overlapping frames with optimal weighting1/ and, (5) Administrative considerations, such as the timing of a variety of activities (i.e., unduplication) as well as the cost factors.

III. The Basic Design

Consideration of the coverage of possible list frames indicated that not all elements of the target populations would be covered by list frames. As a minimum, in order to provide the necessary supplementation for dealing with this gap, it was decided that a multistage, area, probability, double sampling approach would be required. Further, it was decided that the field collection of data in this survey is to be by the Census Bureau. To take maximum advantage of existing sampling materials, the first-stage units are an amalgam of a Census Bureau 197 PSU MLS design and its companion 197 PSU subset of the Census Bureau's CPS--or a 243 PSU first-stage design. Therefore, the 243 PSU's are the areas within which the basic survey is to be done.

Study of the cost factors suggested a ratio of about 20 to 1 between the costs of obtaining and analyzing the detailed characteristics of interest to the field costs of identifying the target populations within the overall population. Pretests have showed the feasibility and moderate costs of a mail plus field follow-up for noninterviews approach as a tool for identifying members of the target populations. Pretests have suggested that among elements of the special list frames 1/2 to 2/3 or more would be members of the target population versus 8% in area sampling. The amount of information to be obtained from respondents, the nature of the other aspects of the survey process (coding and matching to other source data) and the extent of tabulations, suggested a relatively high unit cost for the intensive interview phase of the survey.

The area sampling approach was determined to be sufficiently well accomplished by a first-stage probability design involving 197 strata (243 sample PSU's) with the subsample of units within to be drawn from current survey operations of the Bureau of the Census. (The establishment of the approximate level of the first- and second-stages of the area sampling was part of the joint consideration of components of variability from the area and list frames, as well as costs of all phases such as training, interviewing, matching, etc.) All list samples for field interviewing also have the 243 sample PSU's as the first-stage sampling units. It may be of interest to note that the interaction of availability of area sampling resources and efficiency dictated that in some of the strata (46 out of 197) two probability selections be made. Within the remaining strata 37 consist of a single PSU and in the remainder (114) a single primary sampling unit represents the stratum.

IV. Available Resources and Special Frames

While five of the six target populations are comparatively rare, they vary in difficulty of access.

List frame resources for selection are readily available for those who receive social security benefits (either as disability beneficiaries or as adults whose disability arose prior to their eighteenth birthday). The decisions in regard to sample design between those who classify themselves as unable or limited are simple for this group. Administration of the Social Security Act requires creation of a number of data files. Thus, tape records showing current residence are available for beneficiaries. Therefore, sampling of the population receiving benefits from the Social Security Administration is quite simple.

For those who are unable and receive some public funds (other than from the SSA) the predominant component is those receiving funds through the Aid to Permanently and Totally Disabled or Aid to the Blind Programs. Effective cooperation by the Welfare Administration and state agencies in sampling the latter frames insures meeting the survey objectives. Localization of effort to lists of current recipients of APTD and AB benefits was deemed the only efficient approach. Other comparatively small groups supplement the APTD-AB to comprise this target population. After some investigation, it soon became apparent that the use of other supplementary special list frames for this population would not be efficient. The sole supplementation for this target population (and for its parallel involving persons with limitations) is through the area sampling approach.

The size of the universe of persons who are limited and without public income maintenance support is substantially large so that the area sample portion of the design, alone, permits meeting the specifications.

Those who classify themselves as unable but are not in receipt of funds from any public income maintenance program are a small and most difficult target population. It is for this latter group that the most extensive sampling effort has been required. Thus, the primary need in the development of special frames arose in dealing with this population. Consideration of likely information sources suggested that some knowledge about this group existed at the SSA. However, it turned out that the information was not readily accessible in a convenient single tape for sampling. In order to deal in some way with this group, three different sources (7 lists) are being used. The information available required (a) sampling of elements within those administrative units containing in part one or more PSU's or parts of PSU's; (b) identification of sampled units to sampled PSU's; and (c) a number of other activities to deal with associated problems. These sources arise in a number of ways. Some persons who become disabled approach the SSA either informally or formally to determine whether they are entitled to benefits under the Social Security Disability Program.

A. Some people who are disabled do not qualify for benefits because of lack of insured status under the Social Security Act (3 lists--for 3 separate years for this group are being used).

B. Others who have insured status do not meet the disability requirements (3 lists--for 3 separate years for this group are being used).

C. Some people make inquiry but do not follow through (1 list is being used for this group).

After some consideration and based in part on pretest information, it seemed possible that perhaps as much as two-thirds of the group of persons unable and without public income maintenance benefits might be covered through these multiple list frames.

In addition to the special sampling efforts mentioned briefly above, use of some of these frames involves a number of special procedures to permit setting up an unbiased design. Special procedures are required since these frames are not current with respect to extent or addresses: They contain members other than those of the target population and do not list current addresses as well as not being directly accessible in terms of basic first-stage units. Nevertheless, the needs of the survey seemed unlikely to be met except through utilization of these frames however difficult the problems might be. (It may be worth noting that the last of these frames could be established only by special intervention in an administrative process so that certain control cards normally discarded were maintained specifically for this survey use. The other frames were accessible through computers but only through the use of data tapes containing skeletonized records.)

The basic identification on each record of these seven special frames is a district office code (and in some cases even this is missing). A district office service area may be part of a county, a county, or several counties. As a result, the first-stage of sampling for these 7 frames entailed selection of a sample in those district offices which contained any part of a sample PSU. The sampling rate used is the within PSU rate which satisfies the largest requirement for the rate of within PSU sampling.2/ For those elements with no identification, an overall sample was drawn for subsequent subsampling.

Once the sample was selected, addresses were abstracted from case folders and coded to counties. Some units were established as being outside the sample PSU's. These were discarded from the basic design. For the elements coded to sample first-stage units, subsamples were selected to provide essentially a self-weighting sample.

V. Screening Versus Multiple Overlapping Frames

The multiplicity of frames raised the interesting questions as to the desirability and feasibility of screening among frames to permit discarding versus the retention of elements in overlapping domains with optimal weighting. A number of papers have discussed some of the statistical considerations involved.3/ In a sense, this problem is certainly not new. Any use of frames which overlap raises these questions. The multiplicity of the frames involved in this survey and the nature of the information available as well as the cost and time factors involved has led to an approach which combines and uses screening and discarding in some cases, retention of sample elements in overlapping domains in other cases for differential weighting. Among those frames immediately and directly available to the SSA, the screening for overlap and discarding took place by establishing a hierarchal relationship among the nine SSA lists used for this survey. The basic identification of overlap took place by recognizing the <u>sample</u> elements of a lower order frame in the higher order frames through comparisons of social security numbers. The matching process had access not only to the information available on the skeletonized records on computer tape but, where necessary, to the voluminous paper records which support the tape information and are basic to the administration of the Social Security Act.

The overlap among the major groupings of frames (considering all nine social security frames for this purpose as a single frame) can only be established after the fact. The survey process utilizes double sampling for the area sample to identify members of the target population. Immediately after the first phase of the double sampling process, it will be necessary to move to the second, intensive interview phase for all selected target population elements at which time, social security numbers will be part of the identification information secured. Thus, the area sample supplementation of the samples from other frames as well as the interaction between the other frames can only be done after the fact. Social Security numbers will be used for determining overlapping elements between the area sample and virtually all list frames. (Even where there are problems, administrative techniques within SSA for dealing with incorrect or missing social security numbers assure a high degree of success in identification of overlap elements.) For the APTD-AB list frame overlap, data secured in the intensive interview will be the primary basis for identifying the cases in the overlap domains.

VI. Some Other Problems

As is well known, needs for information for sampling are not always identical with information available in a given list frame. Utilization of incomplete or non-current frames involves problems. Frames which cover substantially more than the target population can be dealt with through double sampling. Incompleteness of frames as described above, is dealt with through supplementation through use of area sampling.

Further, in this survey design some complications arose because the skeletonized tape or other records did not show PSU designations but only the overall service area covered by district offices of the SSA. The method for dealing with this problem was through a multi-stage double sampling approach described in Section IV, above.

The major remaining problem encountered in the use of these non-current frames was in the potential bias of failure to deal with the movers. A number of studies, including the Census Bureau's "Reverse Record Checks"4/ have suggested that it is reasonable to expect high levels of success in finding current addresses when starting with older addresses if sufficient field work is done. Based on this premise, we have tackled the lack of current addresses in these available frames by methods similar to those used in these studies. This problem, the lack of currency of addresses, was approached in several collateral ways. On the one hand, the need was to identify the sample units whose current address is still within a specified sample PSU. Movers, in part, could be sampled by determining the sample elements which had moved to any other sample PSU. For each of these two situations there is then a known probability of selection. The remaining problem lies in dealing with movers who lived in non-sample PSU's at the time of the creation of the record. More specifically, the problem lies in identifying those who lived in non-sample PSU's but who have moved to sample PSU's. To deal with these movers, the approach being followed is to supplement the sample of first-stage non-certainty PSU's with a supplemental sample of PSU's and to select a sample from the same frames in those supplemental PSU's. Those sample elements who are determined to have moved to a specified sample PSU are then part of the basic sample. Thus, the current address determined will, in all cases, be within the original sample of 243 PSU's. To the extent that one can establish current address for those whose addresses were known as of several years ago, we have an unbiased final sample.

VII. Summary

In general, none of the elements of the specified sample design taken by themselves is essentially new. However, some of the indirect approaches for achieving an unbiased design have been deemed to be of more than passing interest. This series of techniques suggests that multiple-frame resource oriented techniques can be found for dealing with rare target populations.

Acknowledgement

The author wishes to thank Harold Grossman for his assistance in the development of the sample design.

FOOTNOTES

1/ Hartley, H.O., "Multiple Frame Surveys" in American Statistical Association, Proceedings of the Social Statistics Section, 1962. Cochran, Robert S., "Multiple Frame Sample Surveys" in American Statistical Association, Proceedings of the Social Statistics Section, 1964.

2/ However, in some cases a lower rate was used to reduce the volume of administrative work involved in the selection of case folders and geographic coding. Where a small PSU with a high within rate was associated in a district office area with a large PSU with a much smaller within rate, a reduced level of sampling was deemed desirable.

3/ Hartley, op. cit.; Cochran, op. cit.

4/ U.S. Bureau of the Census, Evaluation and Research Program of the U.S. Census of Population and Housing, 1960, "Record Check Studies of Population Coverage." Series ER 60, No. 2., Washington, D.C., 1964.

XIII

CONTRIBUTED PAPERS II

Chairman, PHILIP C. SAGI, University of Pennsylvania

A MODEL FOR STUDY OF NURSING ACTIVITY PATTERNS

A. M. Feyerherm, Kansas State University

Measurement of time required to perform various tasks on a nursing unit presents some interesting problems. These include:

1. Simple measures of daily work load, such as number of units produced or processed, are unavailable. However, counts of different types of patients may provide basic data needed to estimate work load,
2. Total work load varies from day to day since the patient mix varies even when the census may show little variation. Furthermore, many nursing tasks are such that they can neither be postponed nor performed ahead of schedule,
3. For a given work load, total time spent in a given activity, by a set of workers of a given skill level, may be influenced by the number of workers available to share the work load,
4. Different skill levels are involved and there may be some overlap of functions and tasks.

Since problems of this type are not unique to nursing units, analytical methods, suggested in this paper, are not restricted to such units.

Our goal in this research was to relate amount of time spent in different nursing activities to a selected set of variables based on staff compositions and patient mixes. Through such relationships we expected to gain insight into: (1) reaction of workers to changing work load conditions, and (2) staff composition needed to meet varying work loads.

Dependent and Independent Variables

The dependent variables in this investigation involved the aggregate times spent by nurses, on a particular nursing unit, while engaged in each of four different major activities. Specifically,

$T_1(N)$ = total time spent by nurses, on a given unit, in physical activities between 7 AM and 3 PM,
$T_2(N)$ = total time spent by nurses, on a given unit, in clerical activities between 7 AM and 3 PM,
$T_3(N)$ = total time spent by nurses, on a given unit, in oral communication between 7 AM and 3 PM,
$T_4(N)$ = total time spent by nurses, on a given unit, in standby activities between 7 AM and 3 PM.

Since these activities were considered to be mutually exclusive and all inclusive, it follows that $\sum_{i=1}^{4} T_i(N) = 480 \cdot N$, minutes per day,

where N = number of nurses.

Similar definitions were used for nurses' aides with

$T_i(A)$ = total time spent by aides, on a given unit, in the i(th) activity between 7 AM and 3 PM; i=1, 2, 3, 4,

and $\sum_{i=1}^{4} T_i(A) = 480 \cdot A$ minutes per day, where A = number of aides.

Physical activities included all activities inside a patient's room, preparations for treatments and procedures, cleaning, restocking supplies, transporting patients and walking. Standby time included lunch period, coffee breaks, and other periods of inactivity. The remaining activities are self-explanatory.

The independent variables measured daily changes in staff composition (N and A, as previously defined) and patient mix. Description of patient mix in terms of care-level requirements led to use of:

P(M) = number of minimal-care patients,
P(I) = number of intermediate-care patients,
P(H) = number of high-care patients,

and in terms of hospital day to use of:

P(1st) = number of 1st-day patients,
P(2-4) = number of (2-4)th-day patients,
P(5+) = number of (5+)-day patients.

Since the subdivisions of patients were considered to be mutually exclusive and all inclusive, it follows that:

$$\text{Daily census} = P(T) = P(M) + P(I) + P(H) = P(1st) + P(2\text{-}4) + P(5+).$$

Daily classification of patients by care level was based on their physical independence as exhibited by their ability to help themselves in bathing, eating, walking and getting up. Classification by hospital day needs no explanation except that (5+) refers to patients present for five or more days.

Mathematical Model

In building a mathematical model to relate time spent in various activities to staff composition and patient mix, one must consider that both nonlinear responses and interactions may occur. The proposed model allowed for nonlinearity by including squared terms and for interactions by including some cross-product terms. For nurses, the model consisted of four equations of the form:

$$T_i(N) = \beta_{0i} + \beta_{1i}N + \beta_{2i}N^2 + \beta_{3i}A + \beta_{4i}A^2 + \beta_{5i}N\cdot A + \beta_{6i}P(T) + \beta_{7i}P^2(T) + \beta_{8i}P(M) + \beta_{9i}P(H)$$

$$+ \beta_{10,i}P(1st) + \beta_{11,i}P(5+) + \beta_{12,i}N\cdot P(T) + \beta_{13,i}A\cdot P(T) + \varepsilon_i; \quad i = 1, 2, 3, 4,$$

where the ε_i's are random errors and β_{0i}, β_{1i}, . . . , $\beta_{13,i}$ (i = 1, 2, 3, 4) are the parameters of the model, considered constant over days, for a particular nursing unit. Similar functions were hypothesized to relate the $T_i(A)$'s to the independent variables.

The nature of the model forces a relation among some of the parameters. Since $\sum_{i=1}^{4} T_i(N) = 480 \cdot N$ minutes per day, it follows that $\sum_{i=1}^{4} \beta_{hi} = 0$ for $h \neq 1$ and $\sum_{i=1}^{4} \beta_{1i} = 480$ minutes. Similar relations hold for parameters associated with activities of aides.

It should be noted that terms P(I) and P(2-4) have been excluded from the model to avoid linear dependence among the independent variables. With the model in its present form β_{8i} measures the difference in contribution to time spent in the i(th) activity between a minimal-care and an intermediate-care patient and β_{9i} is the difference in time between a high-care and an intermediate-care patient. Similarly $\beta_{10,i}$ and $\beta_{11,i}$ measure differences in time needed for 1st-day patients relative to (2-4)th-day patients and (5+)-day patients relative to (2-4)th-day patients, respectively.

Estimation of Parameters

Data were collected from a number of nursing units in two voluntary general hospitals. For each nursing unit, 70 days of data yielded 70 sets of values, on independent and dependent variables, for estimating parameters. Work sampling techniques were used to estimate values for the dependent variables so that the left hand sides in our model contain a sampling error component.

The parameters were estimated by applying the method of least squares to each equation in the model. By assuming, for a given i, the ε_i's over days were independently and normally distributed with zero mean and common variance, multiple regression techniques were applied and tests of significance performed. A term was retained in the set of equations, making up the model, if one or more of its coefficients was significant at the p = 10% probability level. If this condition was not met, the coefficients were set equal to zero and new coefficients were determined for the remaining terms for all equations.

Since the method of least squares was used to produce estimates of parameters it follows that these estimates, like the parameters they estimate, have the property that: $\sum_{i=1}^{4} \hat{\beta}_{hi} = 0$ for $h \neq 1$ and $\sum_{i=1}^{4} \hat{\beta}_{1i} = 480$ minutes.

Furthermore, if we choose to combine two activities, such as "oral communication" and "standby", then the coefficients in the equation for combined activities are simply the sum of the coefficients of corresponding terms in the separate equations.

Discussion of Results

Results shown in equations (1-4) in Table 1 apply to nurses on a nursing unit with all medical patients. Some interesting observations can be made from a study of these equations.

Equations (1-4) describe mathematically the interchange of time among the four major activities as the independent variables change values from day-to-day. For example, suppose the census on this medical unit increased by one intermediate or minimal-care patient and number of nurses remained the same. Predicted changes (see coefficients of P(T)) would be an increase of two minutes in physical activities and five minutes in clerical work. The increases would be offset by a five minute decrease in oral communication and a two minute decrease in standby time. Had the additional patient been a high-care case, both the coefficients of P(T) and P(H) would be involved and changes in activity times would be 15, 6, -13, and -8 minutes, respectively.

Equations (1-4) reflect the simple fact that time cannot be increased for one activity without an equivalent decrease in one or more of the other activities if the number of nurses remains constant. If the coefficient for P(T) in equation (1) had turned out to be zero, it would mean that time needed for an increase of minimal or intermediate-care patients could be absorbed within activity 1 and no time need be borrowed from other activities. Absorption within an activity could be accomplished in several ways. One way would be to interchange time among minor activities within a major activity. Another would be to decrease time spent with some or all patients by either deleting certain optional care measures or working at a faster pace. Both absorption and borrowing may occur and positive coefficients of terms like P(T) and P(H) in equations (1) and (2) represents amounts of time that must be borrowed from other categories because they can't be absorbed.

Table 1

EQUATIONS FOR ESTIMATING ACTIVITY TIMES

	R^2	$s_{y \cdot x}$ (minutes)	Equation No.
MEDICAL UNIT			
$\hat{T}_1(N) = -102 + 158\,N + 2\,P(T) + 13\,P(H)$	.36	66	(1)
$\hat{T}_2(N) = 6 + 72\,N + 5\,P(T) + 1\,P(H)$	.18	57	(2)
$\hat{T}_3(N) = 109 + 154\,N - 5\,P(T) - 8\,P(H)$	.29	84	(3)
$\hat{T}_4(N) = -13 + 96\,N - 2\,P(T) - 6\,P(H)$	.30	51	(4)
$\hat{T}_1(A) = -590 + 482\,A - 34\,A^2 + 24\,P(T) + 28\,P(H) + 7\,P(1st) - 12\,P(5+)$	.67	125	(5)
$\hat{T}_2(A) = 6 + 19\,A - 2\,A^2 - 1\,P(T) + 2\,P(H) - 1\,P(1st) - 1\,P(5+)$	.04	27	(6)
$\hat{T}_3(A) = 888 - 423\,A + 73\,A^2 - 14\,P(T) - 9\,P(H) + 20\,P(1st) + 16\,P(5+)$	.58	112	(7)
$\hat{T}_4(A) = -304 + 402\,A - 37\,A^2 - 9\,P(T) - 21\,P(H) - 26\,P(1st) - 3\,P(5+)$	.41	103	(8)
MEDICAL & SURGICAL UNIT			
$\hat{T}_1(N) = -417 + 244\,N - 24\,A + 20\,P(T) - 0.17\,P^2(T)$	.64	94	(9)
$\hat{T}_2(N) = 1206 + 71\,N - 4\,A - 72\,P(T) + 1.23\,P^2(T)$	.27	66	(10)
$\hat{T}_3(N) = -869 + 102\,N + 19\,A + 66\,P(T) - 1.28\,P^2(T)$	.32	100	(11)
$\hat{T}_4(N) = 80 + 63\,N + 9\,A - 14\,P(T) + 0.22\,P^2(T)$	.39	39	(12)
$\hat{T}_1(A) = -429 + 337\,N - 81\,N^2 + 347\,A + 14\,P(T) - 19\,P(H)$	.92	122	(13)
$\hat{T}_2(A) = 36 - 40\,N + 9\,N^2 + 12\,A - 1\,P(T) + 0\,P(H)$	.16	32	(14)
$\hat{T}_3(A) = 472 - 314\,N + 65\,N^2 + 69\,A - 11\,P(T) + 20\,P(H)$	.36	107	(15)
$\hat{T}_4(A) = -79 + 17\,N + 7\,N^2 + 52\,A - 2\,P(T) - 1\,P(H)$	.38	78	(16)

With number of patients held constant, equations (1-4) also indicate how an additional nurse tended to distribute her time among the four activities. The coefficients of N indicate that 158 minutes go into physical activities, 72 minutes into clerical work, 154 minutes into oral communication, and 96 minutes for standby.

Equations (5-8) include an A^2 term which would suggest decreasing returns from additional aides. A change from three to four aides increased physical activities by 244 minutes while a change from four to five aides increased time in this activity only 176 minutes. Also of interest was the magnitude of demands for time placed on aides by an increase in high-care patients.

Equations (9-12) include the term $P^2(T)$. In this case the fitted model revealed that time needed for physical activities, for increased census, was secured from clerical time when the census was low but for larger censuses it was secured from oral communication.

Two points are brought out in equations (13-16) for aides. One is the effect of additional nurses on time aides spent in physical and oral communication activities. An additional nurse, on this unit, increased time nurses spent in physical activities by an average 244 minutes (equation 9) while aides' time in the same activity decreased by either 68 minutes or 230 minutes depending on whether the change was from two to three or three to four nurses. The other point involves the coefficients of P(H). An increase in number of high-care patients meant a decrease in physical activity time for aides. The inference is that nurses were largely responsible for bedside care of high-care patients.

It should be noted in analyses of the type used, that values for coefficients will, in part, be dependent on overall work load for nurses and aides during data collection. Other factors affecting these coefficients would be the extent to which nurses do aide's work, aides do what nurses might do with fewer aides present and the extent to which nurses and aides consciously keep activities such as standby and oral communication at a minimum when the work load is relatively light.

Minimal and Maximal Staffing

Examination of equations (1-16) indicates that additional time needed for physical and clerical activities resulting from changes in patient counts was usually secured from oral communication and/or standby time. Recognizing that the latter categories include necessary communication, lunch, and other personal time, it is reasonable to set prescribed minimum times on a per nurse (aide) basis for the combined categories of oral communication and standby. Sufficient nurses and aides should be on duty, each day, to keep time spent in communication and standby activities above the prescribed minimums. Minimal staffing occurs when the minimum limits are reached.

Some difficulty may be encountered in determining minimum limits but, for illustration, suppose we set limits at 135 minutes per nurse and 80 minutes per aide for a medical unit. From equations (3-4) and (7-8) one can write down the following inequalities:

For nurses: $$96 + 250\,N - 7\,P(T) - 14\,P(H) \geq 135\,N, \tag{17}$$

For aides: $$584 - 21\,A + 36\,A^2 - 23\,P(T) - 30\,P(H) - 6\,P(1st) + 13\,P(5+) \geq 80\,A\,. \tag{18}$$

Thus, if P(T) = 24, P(H) = 8, P(1st) = 2, P(5+) = 10, N = 2, A = 4, the inequalities are satisfied. If P(H) = 12, instead of 8 neither inequality is satisfied. Rather than adding both a nurse and an aide, the addition of a nurse who spends half her time working as an aide would satisfy both inequalities.

We might also consider the problem of finding the staff size, for a particular nursing unit, which would rarely have to be augmented to care for extra heavy work loads. Such a staff size might be considered maximal. Over time, the number of patients in special categories (high-care, 1st-day, etc.) will usually be between some minimum and maximum percentage of the total number. For example, for the medical unit, the following inequalities are likely to hold:

$$P(H) \leq 0.60\,P(T),\quad P(1st) \leq 0.30\,P(T),$$

$$P(5+) \geq 0.40\,P(T).$$

If these values are substituted into inequalities (17) and (18), the results will be:

For nurses: $$115\,N \geq 15.4\,P(T) - 96 \tag{19}$$

For aides: $$36\,A^2 - 101\,A \geq 37.6\,P(T) - 584 \tag{20}$$

Thus, if census on this unit rarely exceeds P(T) = 25 patients, then for N = 2.6 and A = 4.9, inequalities (19) and (20) and also (17) and (18) will be satisfied, except on rare occasions.

A PROCEDURE FOR AUTOMATIC DATA EDITING

R. J. Freund and H. O. Hartley, Texas A & M University

1. INTRODUCTION

A cursory study of some presently used editing procedures reveals that a great variety of these procedures have been developed. Some of these are based on simple 'one item at a time' consistency checks and imputations using either common sense consistency principles, 'historical data', 'hot-deck' or 'regression estimates'. Others employ complex networks of interlocking checks and associated imputations. This great variety of editing procedures reflects the impact of two often conflicting desiderata:

(a) An effective editing procedure must recognize the particular error patterns as well as inconsistency-and-error correlations arising in a particular survey.

(b) An effective editing procedure must have a comparatively simple logic and must be easy to program as otherwise there is a tendency not to use automatic editing at all.

Desideratum (a) calls for a procedure 'custom made' for the particular survey and thus uses a logic specifically oriented to a particular study; this requires considerable programming effort for each survey. Desideratum (b) calls for a procedure which is easily understood and applied by the programmer, but thereby tends to bypass a detailed scrutiny of the specific error patterns which the specialist in a particular survey area would call for.

In attempting to reconcile (a) and (b) we have tried to make contributions on the following lines:

(a) We have attempted to develop a standard editing procedure which is to be implemented by some simple macro-codes or special forms, and

(b) we have developed a procedure for producing relatively consistent data when several restrictions must be met.

The general procedure consists of:

1. The Gross Check.

2. The Internal Consistency Check.

If the internal consistency check fails, we will correct the data by

3. The Least Squares Correcting Procedure.

1.1 THE GROSS CHECK

Each item of the input record is checked for gross errors. Such errors are characterized by the data being completely unrealistic or out of line. Data items found to violate such gross error checks are imputed, one at a time, by relatively simple and straightforward imputations such as the use of historical data, 'hot decks', the use of ratios known to be usually consistent, corrections for misplacement of decimals or incorrect units of measurement, etc. A record will be kept of all data items for which imputations have been made in this editing phase.

1.2 THE INTERNAL CONSISTENCY CHECK

The data will be checked to see if certain internal consistencies are satisfied within a sufficient tolerance. Thus, for example, acreages in various crops must add up to total acreage in crops, quantity times price must be equal to value, etc. Any time a check fails to be fulfilled, attempts may be made to impute data by some of the same procedures as were used in the gross check. During this phase of the editing procedure, a notation will be made of all data items involved in checks which are not satisfied.

1.3 THE LEAST SQUARES CORRECTING PROCEDURE

A least squares correcting procedure will be used if simple, one-at-a-time correcting procedures do not produce data which will satisfy all internal consistency checks. It is proposed that this procedure will substitute for the complex interlocking networks of checks and imputations which are usually used. It is hoped that the use of this procedure will not occur too frequently since it will take a moderate amount of computer time. However, it will require no complex, custom-made networks and will, therefore, be generally useful for many situations.

This procedure also makes use of the fact, noted above, that some items in the data are more 'suspect' than others due to their either being subjected to corrections in the gross check or being involved in some of the consistency checks which failed to satisfy tolerances. This will be done by the use of weighted least squares, where the weights will indicate the reliability of the data. These weights can, of course, also be used to indicate the 'usual' reliability of individual data points, recognizing that some data are usually more reliable than others.

1.3.1 General Considerations

We start with the previously stated premise that we want to satisfy (to some degree) certain (linear) consistency equations by 'correcting' some of the input variables. Some of the consistency equations are more important than others, i.e., some must be satisfied to a greater degree of accuracy than others. Likewise, some input data are assumed to be more reliable than others and consequently some data should be changed less than others by the correcting procedure, but the corrected data should be as 'close' as possible to the original data. This is to be accomplished by minimizing the weighted sum of squares of the discrepancies of the consistency equations plus the weighted sum of squared differences between original and corrected data; the weights are used to indicate importance of restrictions and/or reliability of data.

The equation for the sums of squares to be minimized is:

$$SSC = \sum_j^n w_j(x_j-y_j)^2 + \sum_i^n u_i(\sum_j^n a_{ij}x_j)^2$$

where

x_j is the j-th corrected datum

y_j is the j-th original datum

w_j is the weight given to the i-th datum, a large weight indicates more reliable data,

$\sum_j^n a_{ij}x_j$ = the j-th consistency equation of the form $\sum_j^n a_{ij}x_j = 0$,

u_j is the weight of the j-th consistency, a large weight indicates an important consistency.

The minimization is with respect to the x_j; thus the corrected data is chosen to minimize this sum of squares.

The minimization is accomplished as follows:

$$\frac{\partial SSC}{\partial x_k} = 2w_k(x_k-y_k) + \sum_i^m u_i\ (2\sum_j^n a_{ij}x_j a_{ik}) = 0,$$

$$k = 1, 2, \ldots, n.$$

Rearranging terms we have:

$$w_k(x_k-y_k) + \sum_i u_i a_{ik} \sum_j a_{ij}x_j = 0,\ k = 1, 2, \ldots, n.$$

In matrix form this can be written:

$$D_w(\underline{x}-\underline{y}) + A'D_uA\underline{x} = 0$$

where

D_w is a (n x n) diagonal matrix of the w weights (for data),

$\underline{y}$ is a (n x 1) vector of original data,

$\underline{x}$ is a (n x 1) vector of corrected data,

A is the matrix of coefficients of the restrictions, hence these can be written $A\underline{x} = \underline{0}$, where $\underline{0}$ is a vector of zeroes,

D_u is a (m x m) diagonal matrix of the u weights (for the restrictions).

Solving for $\underline{x}$ we have

$$\underline{x} = (D_w + A'D_uA)^{-1}D_w\underline{y}$$

This expression is, of course, easily solved by high speed computers. Note that A will be usually predetermined for an entire study; D_u may also be constant for an entire study and hence $A'D_uA$ need be computed only once.

1.3.2 The Determination of Weights

The weights of the data points should exhibit

(a) The basic variability of the datum,
(b) the 'usual' reliability with which the datum is reported, and
(c) the reliability of the datum in a specific record.

(a) Basic Variability

Data which is basically variable is more subject to corrections and should, therefore, receive smaller weights. This type of variability is often associated with size of unit of measurement. Thus small items should receive smaller corrections and large items large corrections; data which should be zero should ideally receive no correction.

(b) 'Usual' Reliability

Some data are naturally recorded more accurately than others. For example, tobacco acreage are very precisely known due to strict acreage controls whereas woodland acreages may not be well known, particularly if woodlands are used partially as pastureland.

(c) Reliability of a Specific Record

It is assumed that the record which is subjected to a least squares correction has failed in an initial, relatively simple, editing-correcting sequence (see above). If a data point has been subject to a gross error correction or has been involved in several unsatisfied consistency checks, it is most likely in error. Thus, weights of items involved in a gross error correction or non-satisfied consistency check will be reduced in proportion to the number of involvements.

The weights for the consistency equations (u_i)

should exhibit the importance of the consistency equations, i.e., the degree in which the equation must be satisfied. Sometimes it is vital for the purposes of the study that a certain consistency check is accurately satisfied while other checks are not as critical. The use of (relatively) large weights for some equations will assure small discrepancies in these equations.

It should be noted that weights are relative and it is the ratio of large to small weights which is of importance. The magnitude of this ratio for practical use is subject to further study; initial experimentation indicates ratios of 5/1 to 20/1 are needed for effective control, i.e., differentiation of magnitude of corrections. It can be further noted here that we are attempting to correct data points to conform to certain consistency checks. Thus it is reasonable that weights for data should be smaller than weights for restrictions. On the other hand, the most reliable individual data points should probably not be corrected. Thus a procedure for assigning weights should assign nearly equal and relatively large weights for the most reliable data points and most important consistency equations; the least reliable equations should have weights possibly 1/5 as large and the least reliable data points 1/20 to 1/100 as large.

The entire procedure outlined above can be summarized in a flow chart as outlined in Figure 1.

2. EXAMPLE

We will use as examples the data from some hypothetical farms, using selected information as recorded in the Bureau of Census, Farm Questionnaire Sample Survey of Agriculture, 1961 (Form No. 60-02-548.4). Fourteen items involving acreages have been selected for use (see Table 1).

The "w weights" indicating the 'basic variability' and 'usual reliability' of the data are given in Table 1. Thus, for example, acres owned and acres of cropland should be reliably recorded since exact knowledge of these is required for taxes and government programs; the rental - sub rental acreages can be considered confusing, and thus, are more likely to be in error.

The restrictions are given in Table 1 as coefficients (reading vertically) of equations that should equal zero. Thus, restriction 5 states that total acres in place is equal to acres owned and not rented-out plus acres rented but not sub-rented. The "u weights" given at the bottom indicate the importance of the restriction; thus it is considered important that acres in place agree both with respect to rental arrangement and land use (restriction 2, u = 85) but breakdown of land use is open to questions of unaccounted land (fences, roads, etc.) and double cropping (restriction 1, u = 10).

2.1 GENERATION OF DATA

We shall attempt to evaluate the above outlined editing procedures by their use on some artificially generated 'incorrect' data. We assume that we have a large number of identical schedules into which we introduce random errors. It is then relatively easy to see how close to the "correct" data the editing procedures actually come.

There are essentially two decisions to be

Table 1

Data For Example

DATA					RESTRICTIONS (coefficients, read down)						
Variables Acres	Item No. from Schedule	Correct Acres	Mnemonic	"w" weight	1 Land Use	2 Tot=Tot	3 Land Use	4 Rent Out	5 All Tenure	6 Rent In	7 Total Acres
"u" weight ⟶					10	85	80	40	35	40	50
Owned	8A	160	OWN	10				+1			
Rented to	8B	10	RENTO	6				-1			
Not rent to	8C	150	RENTONO	1				-1	-1		-1
Rent from	9A	65	RENTFM	3					-1	+1	
Subrent	9B	0	SRENT	6					+1	-1	
Not Subrent	9C	65	SRENTNO	1						-1	-1
Total in place*	10	215	TIP	6		-1			+1		+1
Cropland total	61	180	CROPT	10	+1		-1				
Pasture	61A	60	PAST	3	-1						
Gov't Program	61B	90	GOVTP	10	-1						
Other	61C	20	FFETC	3	-1						
Harvested	61D	10	HARV	5	-1						
Total in place*	66	215	TAIP	6		+1	+1				
Other uses	(62+63+64+65)	35	OWSES	1			-1				
Tolerated Limits					3%	2%	2%	2%	3%	2%	2%

*This is requested twice in schedule.

in the generation of data with errors:

1. Whether a particular item be correct or incorrect.
2. If a particular item is incorrect, what type of error it should exhibit.

The procedure generated errors in two steps. First, a random number was generated to correspond to each item in the schedule and if the random number was less than 1/1.5w, then the item was designated as being incorrect. This procedure does, of course, generate a much larger than usual number of errors (the most reliable items had W = 10, hence over 6% of even these are in error), but we do not wish to waste computer time not correcting many "good" records.

Once an item was designated as being in error a second random number was generated and the type of error was assigned as indicated in Table 2. Thus, for example, if the random number was between 0 and .1 it is assumed

Table 2

TYPES OF ERRORS TO BE GENERATED

Random Number	Type of Error
0 —> .1	Blank
.1 —> .2	+ 10%
.2 —> .3	- 10%
.3 —> .4	+ 5%
.4 —> .5	+ 5%
.5 —> .6	- 5%
.6 —> .7	- 5%
.7 —> .8	* 10
.8 —> .9	* 0.1
.9 -> 1.0	Return

there was a blank in that particular portion of the schedule, a random number between .7 and .8 would indicate that there had been a scaling error of a factor of 10. The random number between .9 and 1.0 was not used for a specific error and hence, if this occurred, a second random number would be generated and the error assigned according to the second random number. This allows one additional type of error to be introduced if it is desired to do so.

2.2 RESULTS

Several different sampling experiments were performed using minor variations of the above outlined editing procedure. In all there were eleven sampling experiments. Since the random number generator always has the same starting value, all samples are based on identical sets of "incorrect" data, a direct comparison of the results is quite meaningful. These comparisons are afforded by the summary in Table 3 which shows, for the originally generated data and the results of various editing procedures, the mean difference between each of the correct and corrected items and the standard deviation of the items. At the bottom of the Table are the sums of absolute mean deviations and the standard deviations, also the sum of squared deviations and sum of variances. It should be noted that these standard deviations are based on the sum of squared differences of individually corrected items and the sample means of the corrected items rather than the sum of squared deviations from the true values of the items. This latter figure can, of course, be generated, but would not be much larger than those given.

Of course, any editing procedure which resulted in a mean difference of 0 and a standard deviation of 0 would be ideal; needless to say this has not been realized and is not likely to be realized. It is difficult to make a value judgment as to whether it is more important that the mean difference be 0 or that the variance be small; since most data of this type is used for summaries, it is most likely more important that the mean differences be small. It should be remembered that the results of the least squares procedure do indeed guarantee that we have consistent results; that is, the various consistency checks will be almost completely satisfied.

In Table 3 the first column provides the correct values and the second set of two columns provides the mean differences and standard deviations of the data as generated by the random error generator. It can be seen that this procedure was quite successful in generating "incorrect" data. There appears to be a definite upward bias in practically all the items; this is due to the fact that we had a ten per cent chance of generating an item ten times too large. The second set of two columns indicates what the data would look like if it were subjected only to the gross check and imputations procedure (item 2 on flow chart). It can be seen that this procedure does definitely improve the quality of the data and it might well be argued that one should stop there. It should be noted, however, that there is no guarantee that the results will be consistent for any given schedule or, indeed, for averages derived from this data.

The next four sets of two columns are the result of the editing procedure as outlined in Figure 1 except that no corrections were made in the consistency check stage (item 5 on flow chart is bypassed). In other words, the consistency equations were checked and if any one equation was not satisfied the least squares procedures was used. All checks were, however, made in order to provide the count of the number of times items are involved in unsatisfied consistency checks.

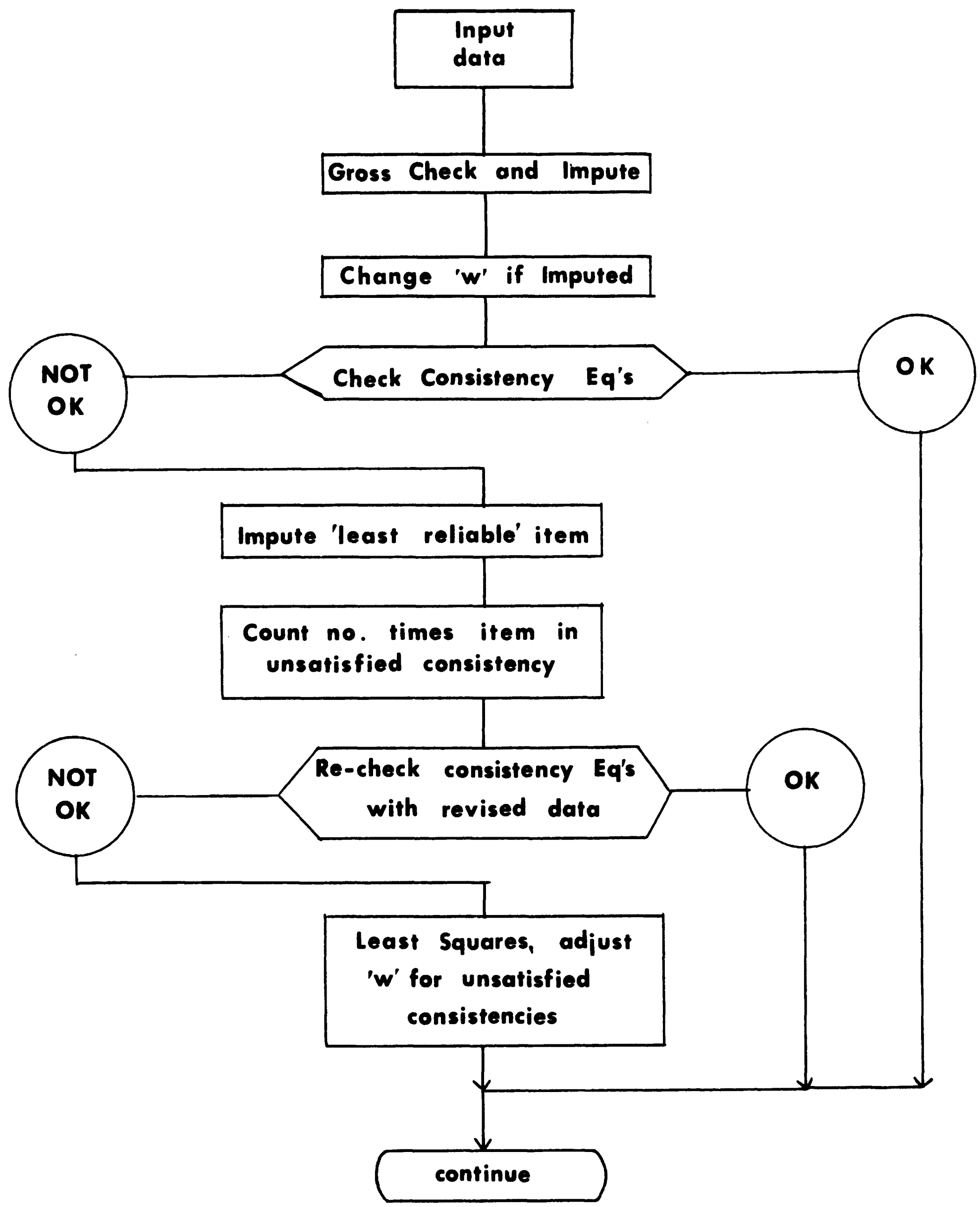
The reasoning behind the elimination of this particular step in the editing procedure is that

Table 3: Results of Sampling Experiment

						Gross Check and Least Squares Correction								Gross Check, Consist. Corr. and L. S.				Gross Check and L. S. all w's start at 10 for L. S.			
		Generated		Gross Check		Original Wts.		Wts. adj. for Gr. Ch.		Wts. fully Adj. WAF#1		Wts. fully Adj. WAF#2		Wts. fully Adj. WAF#1		Wts. fully Adj. WAF#2		Wts. fully Adj. WAF#1		Wts. fully Adj. WAF#2	
ITEM	Correct μ	diff	σ	diff	σ	diff	σ	diff	σ	diff	σ	diff	σ	diff	σ	diff	σ	diff	σ	diff	σ
OWN	160	7.5	126.5	-4.0	23.7	-3.5	20.7	-3.5	20.7	-3.1	19.2	-3.0	19.2	-3.3	20.2	-3.3	20.2	-3.1	19.2	-3.1	19.3
RENTO	10	.2	6.5	.2	6.5	- .5	8.4	- .5	8.2	-1.2	9.0	-1.3	9.4	-1.0	7.8	-1.0	7.9	-1.2	9.0	-1.3	9.4
RENTNO	150	96.5	365.4	-8.8	35.4	-2.8	16.3	-2.9	16.8	-1.8	13.6	-1.7	14.1	-2.3	15.4	-2.2	15.3	-1.8	13.6	-1.7	14.1
RENTFM	65	8.0	77.9	-2.5	12.0	- .7	10.1	- .6	10.1	- .6	9.0	- .5	9.5	- .3	9.5	- .3	9.6	- .6	9.0	- .5	9.5
SRENT	0	0.0	0.0	0.0	0.0	- .9	4.3	- .9	4.3	- .9	4.5	-1.0	4.9	-1.1	4.5	-1.1	4.7	- .9	4.5	-1.0	4.9
SRENTNO	65	37.2	154.3	-5.5	17.4	+ .1	11.7	+ .2	12.0	+ .4	10.7	.5	11.8	+ .8	11.3	+ .8	11.6	.4	10.7	+ .5	11.8
TIP	215	23.1	215.9	-1.1	14.7	-3.6	13.9	-3.6	13.7	-1.4	11.2	-1.1	11.4	-1.5	12.7	-1.4	13.3	-1.4	11.2	-1.2	11.4
CROPT	180	18.5	181.3	-3.0	22.9	-2.4	20.3	-2.4	19.7	-2.2	17.5	-1.7	14.7	2.1	17.8	-1.6	15.2	-2.2	17.5	-1.7	14.7
PAST	60	11.7	85.3	-2.1	10.9	-3.0	12.2	-3.1	12.8	-3.1	13.0	-2.9	12.4	-1.5	13.1	-1.3	12.5	-3.1	13.0	-2.9	12.4
GOVTP	90	- .5	6.0	- .1	8.3	-1.2	8.5	-1.1	8.6	-1.2	8.6	-1.1	8.5	-1.4	8.8	-1.3	8.7	-1.1	8.6	-1.1	8.5
FFETC	20	3.2	25.4	3.2	25.4	+2.3	19.2	+2.4	19.7	+2.4	18.9	+2.6	18.8	+1.6	18.8	1.8	18.7	2.4	18.9	+2.6	18.8
HARV	10	+ .4	6.4	+ .4	6.4	- .2	7.9	- .2	8.2	- .2	8.4	- .1	8.0	- .7	8.1	- .6	7.8	- .2	8.4	- .1	8.0
TAIP	215	12.8	168.5	-1.7	20.0	-3.5	14.0	-3.6	13.8	-1.5	11.2	-1.1	11.4	-1.5	12.7	-1.4	13.3	-1.4	11.2	-1.1	11.4
OUSES	35	21.0	84.4	-2.9	9.6	- .1	17.4	- .2	17.3	+ .7	14.1	+ .5	10.3	+ .7	14.3	.4	10.9	.7	14.1	+ .5	10.3
Sum (abs)		240.6	1503.8	35.5	213.2	24.8	184.9	25.2	185.9	20.7	168.9	19.1	164.4	19.8	174.0	18.5	169.7	20.5	168.9	19.3	164.5
Sum Sq.		12,444.62	302,434.88	166.31	44,410.34	68.40	2,778.33	70.46	2,801.11	42.01	2,277.01	37.03	2,134.86	35.78	2,443.64	32.29	2,299.45	41.49	2,277.01	37.87	2,138.71

Figure 1

Flow Chart of Editing Procedure



the consistency-correcting procedure was quite arbitrary in that if a particular consistency equation failed, the item in that equation with the lowest "w" weight was imputed by subtraction. It is obvious that, in many cases, a correct item may be altered by this procedure, thus creating other inconsistencies which will cause further "corrections" of correct data. Thus this procedure will often create incorrect data which is nevertheless consistent and will thus not be further edited.

In the first two columns of this set (entitled Original Weights) the "w" weights are never adjusted (either in the gross check or in the consistency check) and thus the least squares is entered with the "w" weights as originally indicated in Table 1. In the second set of columns the "w" weights are only adjusted in the gross check phase, where "w" weights are divided by 4 for any item which is imputed at that stage.

The third and fourth sets of columns in this group are results of the procedure (as above) with two different sets of factors used to adjust the "w" weights for the number of involvements in unsatisfied consistency checks. The third set of columns corresponds to the weight adjustment factors as given in Table 4; this Table indicates the divisor for the "w" as a function of the possible number involvements in consistency checks and the actual number of involvements in unsatisfied consistency checks. The fourth set of columns is a result of the use of the weight adjustment factors in Table 5; these adjustments are more "severe" and actually increase "w" weights in case an item is often involved in satisfied consistency checks.

Table 4
Weight Adjustment Factors No. 1

No. of Actual Involvements	No. of Possible Involvements 1	2	3
0	1	1	1
1	5	2	1
2		10	5
3			20

Table 5
Weight Adjustment Factors No. 2

No. of Actual Involvements	No. of Possible Involvements 1	2	3
0	1	1/5	1/10
1	2	1/2	1/5
2		10	2
3			50

In the next two sets of columns the entire editing procedure as given in Figure 1 is used with the two different weight adjustment factors (i.e., Tables 4 and 5, respectively).

The last two sets of columns are intended to show what happens when incorrect "w" weights are used. In this particular sampling experiment the "w" weights as presented in Table 1 are used to generate the data, but for the least squares procedure all weights are initially set equal to ten before being adjusted in the gross check and consistency phases as before with the two different sets of weight adjustment factors. This set of sampling experiments was performed in order to see if it is really very important to initially assign weights indicating prior knowledge of the reliability of items.

The results of Table 3 can be summarized as follows:

a. The gross check does improve the data to a great extent, but the use of the least squares procedure definitely improves the data even further.

b. It appears that the adjustment of weights in the gross check phase is not of much help, but that the adjustment of weights from the consistency equations is useful.

c. From this point on there is not much difference in the results among the procedures and at present it would seem that the elimination of the consistency imputations and the use of weight adjustment factor 2 without prior assignment of differentiated weights is the optimum procedure.

More work of this type is certainly desirable before more definite conclusions can be drawn.

MULTIPLE REGRESSION AND HISTORICAL EXPLANATION

Frank W. Carlborg, Northern Illinois University

The problem which led to this paper was suggested to me by Stanley Parsons of the history department of the University of Missouri at Kansas City. His interest was in describing the Populists in Nebraska in the latter part of the last century. For this purpose, he analysed the elections of the time. Since the 88 counties of Nebraska varied widely in their support of the Populist movement, one might hope to explain the movement by comparing the county-by-county vote with other variables. The method of multiple regression suggests itself in such a situation. This paper is concerned with some of the questions that arise after a satisfactory function has been fitted to the data in the usual least-squares manner. What follows may be divided into three general parts:

A. The regression function for 1890 and the questions raised by it.
B. The general derivation of a procedure for answering those questions.
C. The application of the procedure to the election of 1890.

Part A on the handout gives some of the usual results from a multiple regression analysis. The dependent variable Y is the Populist vote in a county as a percent of the total vote. The six independent variables, the x's, were chosen on historical grounds. They include such things as the percent of farm income in each county paid in interest charges and such as the percent of the population in each county of Protestant cultural background. The quadratic regression function of display (A1) was fitted to the data with the results of display (A2). Before settling on any function in a problem such as this, one must decide when to stop looking for more independent variables; also, one must decide when to stop adding on terms in the equation. These are very important problems which to my knowledge are unsolved. This paper has nothing to say about either of these questions. Rather, I would like to assume that a satisfactory function has been found. Therefore, let me assume that the equation on the handout fits the election of 1890 to a degree that is acceptable to the historian and his critics.

The original means, variances, and standard deviations of all seven variables are given in Table 1. The correlation matrix for the six independent variables is below Table 1. In the regression function of display (A2), the independent variables have been transformed so that each has mean of zero and standard deviation of one; then, the correlation matrix is also the covariance matrix of the transformed variables. Under the assumption that the fit of the function is satisfactory, these figures now summarize the election of 1890. To the historian, however, this summary is probably not satisfactory; he requires further interpretation. One thing he may like is an objective measure of the importance of each independent variable. The remainder of this paper is concerned with developing a procedure for measuring the importance of independent variables. This measure will be a non-negative function of the betas and the marginal distribution. To motivate the desirability of such a measure, consider again the numerical example. If one were to interpret the results, he might look at the coefficients for the linear terms first. β_6 is 8.24, and this suggests that x_6 is important. β_1 is next largest here. Among the coefficients of the quadratic terms, β_{22}, β_{66}, and β_{44} are the large ones. Several of the coefficients of the cross-product terms appear large--particularly, those involving the variable x_5. With one or two exceptions, the pairwise correlations among the x_i's are near zero in this example, but how should these correlations be considered, if at all, in evaluating the importance of a variable? These observations suggest that variable x_6 is important. Other than that, it seems difficult to make a very definite statement. As mentioned before, this paper will suggest a procedure for this situation. Some rather arbitrary steps will be taken. I would like to point them out.

To begin the general derivation, consider the simplest possible situation which is given in display (B1)--that is, the linear regression function with only one independent variable. To the historian, the magnitude of β_0 is irrelevant because he is interested in explaining variations in Y. The absolute value of β_1 suggests more information. The larger this absolute value is; the more important the variable is. Of course, the value of β_1 may be changed at will by performing scale transformations on x_1. The measure of importance should be independent of such transformations. A reasonable way out of this is to multiply the absolute value of β_1 by the standard deviation of x_1. Henceforth, it will be assumed that all independent variables have been transformed to unit standard deviations. Also, the results are more simply stated if it is assumed that the independent variables have a mean of zero. The formulas will be derived for this standardized situation.

They are easily generalized to other situations. This argument has now led to the conclusion that the absolute value of β_1 is a reasonable way to measure the importance of x_1 in this very simple situation. Note that the absolute value of β_1 is also the absolute value of the slope of the regression function with respect to x_1. Thus, the importance of x_1 is measured by the amount of change in the dependent variable which one would expect with a unit change in the independent variable. This way of stating the result seems, to me at least, to be one possible way to evaluate an independent variable in this historical situation. It also provides a base for generalization. The intention here is to proceed from this base. It is certainly an arbitrary decision and is possibly unsatisfactory to some viewpoints. The remainder of this paper is concerned with extending this idea to more general regression functions and with its application to the Populist example.

Consider next the regression function of display (B2). This is a quadratic function with still only one independent variable. The problem is to develop a measure of importance of x_1 for this situation; the result should be consistent with the previous result when β_{11} is zero. Toward this end, consider the slope at any value of x_1 as in display (B3). The prior, linear case suggested the absolute value of this as a measure. However, this absolute value is itself a function of x_1. An overall measure might be taken as the average of the absolute slopes with respect to the marginal distribution of occurrences of x_1. In order to actually perform this calculation, one must return to each actual data-point and compute the slope. Consider instead the following slight variation: average the square of the slope with respect to the marginal distribution of x_1. Display (B4) gives this result. The square root of this is an approximation to the average absolute slope. Note that the calculation of this quantity does not require returning to the actual data-points. Working with the square of the slope is more convenient in several other ways, and this will be done in the future. Investigations have revealed that this is not an expensive convenience.

For the next case, take the regression function of display (A1) and let the problem be that of measuring the importance of any one independent variable, say x_k. At any data-point, the partial derivative of the regression function with respect to x_k is the appropriate slope, and it is given in (B5). This, then, gives the rate at which one expects changes in the dependent variable for small changes in x_k. The average of the square of this with respect to the distribution of data-points appears in (B6). Note that this is a quadratic form in the betas of the regression function. The square root of (B6) is an approximation to the average absolute slope, and it is the desired measure of importance. The quadratic regression function in (A1) is the most general to be considered here. The generalization of (B6) to higher degree polynomials raises no new problems.

The next step in the generalization is to consider the problem of measuring the historical importance of a pair of independent variables. It would be very convenient if the importance of a pair of independent variables were a simple function of their individual measures of importance. If this were not the case, then separate calculations would be required for each pair, each triple, etc. To start the derivation, display (B7) gives a linear regression function with two independent variables, x_1 and x_2. The measure of importance of this pair should be consistent with what has been done above, and the result should reduce to a previous result in degenerate cases. A generalized concept of the slope will be used to generate the needed measure. For a unit change in x_1 and x_2, what is the corresponding change in the expected value of Y? The answer depends upon the direction of the unit change. If Δx_1 and Δx_2 are the changes in x_1 and x_2, a unit change corresponds to the condition (B8). One possibility is to take the unit change in the direction leading to the greatest change in the expected value of the dependent variable. This change is given by (B9). This is the generalized slope previously mentioned. Note that the square of (B9) is the sum of the squares of the separate slopes for linear regression functions, and note that relation (B10) holds for this case. This derivation is really much more general than the linear case. For any regression function with any number of variables, the above steps may be retraced and suitably modified. The chief modification is that β_i of the linear model is replaced by the partial derivative of the regression function with respect to x_i. The details are omitted here. The major result is that these measures of importance combine like orthogonal vectors under addition. That is, relation (B10) is always true. Triples of independent variables are similarly handled.

Item C on the handout gives the numerical results of applying this measure of importance to the Populist example. Table 2 shows the six variables in the first column; the second column gives the individual importances; the third gives the squares of the individual importances. Note that x_2 and x_6 are the two most important variables. Although the measure of x_2 is about twice that for x_1, it should be remarked that since these measures add like vectors, it would take four independent variables like x_1 to equal the importance of x_2. Next look at the importance of the pair x_2 and x_6. This is found by adding 191 and 138 and taking the square root of the sum. This number is 18. Also look at the importance of all six taken together. This is found by taking the square root of the sum of the six numbers in the last column. The ratio of the former for x_2 and x_6 to the latter for all six is .80. This shows that x_2 and x_6 account for 80 percent of what is told by all six variables. This suggests to the historian that he should look at these two variables in further detail. This has turned out to be a good numerical example because so much is suggested about the election by looking at only these two variables. Table 3 shows the result of classifying the original data according to only these two variables. The high quadratic and cross-product betas for x_2 and x_6 suggest displaying the data at three levels of each variable. The nine numbers in parentheses are the actual frequencies. That is, the "7" in parentheses indicates that seven of the 88 counties ranked in the low third according to x_2 and in the high third according to x_6. The other nine numbers are the average vote of the counties in each category. That is, the "55.8" indicates that the seven counties had an average Populist vote of 55.8 percent. There were three major parties in this election. This table shows great variations in Populist vote among the various levels of x_2 and x_6. No historical interpretation of this will be attempted here. This table is probably the major result of the study. It is important to remark that Table 3 does not depend in any way upon the assumptions made in the analysis of Part B above. The analysis has only suggested where to look to find something interesting. The table was constructed directly from the raw data.

To summarize, this paper has presented a procedure to measure the importance of independent variables in a multiple regression. This procedure is intended to be useful when the investigator's purpose is to explain a phenomenon, such as a historical event. The basic idea is to weight an independent variable according to the expected change in the dependent variable resulting from a change in the independent variable. The procedure would not be appropriate in many other regression situations--such as when the investigator's purpose is prediction or control of the dependent variable.

HANDOUT

Part A: Regression function for the election of 1890

(A1) $$E(Y) = f(x_1,\ldots,x_n) = \beta_0 + \Sigma\beta_i x_i + \Sigma\beta_{ij}x_i x_j + \Sigma\beta_{ii}x_i^2$$

(A2)

β_0 = 42.70	β_{11} = - .42	β_{24} = -1.69	β_{44} = 1.68
	β_{12} = 2.20	β_{25} = -4.62	β_{45} = -4.64
β_1 = 4.10	β_{13} = 1.08	β_{26} = -2.69	β_{46} = -1.99
β_2 = 3.32	β_{14} = -4.70		
β_3 = 3.00	β_{15} = 3.83	β_{33} = - .34	β_{55} = -1.08
β_4 = -1.85	β_{16} = 1.27	β_{34} = 2.71	β_{56} = 5.66
β_5 = 2.17		β_{35} = -1.42	
β_6 = 8.24	β_{22} = -7.02	β_{36} = 2.86	β_{66} = 2.73
	β_{23} = -4.45		

Table 1: Original Variables

Variable	Mean	Variance	Std. dev.
x_1	77.6	277.02	16.6
x_2	17.6	40.79	6.4
x_3	5.3	18.92	4.3
x_4	82.8	50.31	7.1
x_5	34.6	135.67	11.6
x_6	91.0	26.10	5.1
Y	37.3	285.94	16.9

The correlation matrix $[\rho_{ij}]$ of the marginal distribution of the x_i's is

$$\begin{bmatrix} 1.00 & .32 & .00 & .30 & -.03 & .22 \\ & 1.00 & .01 & .14 & -.47 & .10 \\ & & 1.00 & -.11 & -.14 & -.03 \\ & & & 1.00 & .04 & .66 \\ & & & & 1.00 & -.07 \\ & & & & & 1.00 \end{bmatrix}$$

The multiple correlation coefficient is .81

Part B: General derivation

Importance of one variable

(B1) Linear function: $E(Y) = f(x_1) = \beta_0 + \beta_1 x_1$

(B2) Quadratic function: $E(Y) = f(x_1) = \beta_0 + \beta_1 x_1 + \beta_{11} x_1^2$.

(B3) $f_1 = \frac{df}{dx_1} = \beta_1 + 2\beta_{11} x_1$

(B4) $\text{Average } (f_1^2) = \beta_1^2 + 4\beta_{11}^2$ with $\text{Average } (x_1) = 0$, $\text{Average } (x_1^2) = 1$

(B5) $f_k = \frac{\partial f}{\partial x_k} = \beta_k + 2\beta_{kk} x_k + \sum_{i \neq k} \beta_{ik} x_i$

(B6) General result for quadratic function:

$\text{Average}(f_k^2) = \beta_k^2 + \text{Average}(2\beta_{kk} x_k + \Sigma\, \beta_{ik} x_i)^2$

where $\text{Average}(x_i) = 0$, $\text{Average}(x_i^2) = 1$, and $\text{Average}(x_i x_j) = \rho_{ij}$

Importance of more than one independent variable

(B7) $E(Y) = f(x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

(B8) $(\Delta x_1)^2 + (\Delta x_2)^2 - 1 = 0$

(B9) Importance of x_1 and x_2 $= \frac{\text{Maximal } \Delta E(Y)}{\text{subject to (B8)}} = (\beta_1^2 + \beta_2^2)^{1/2}$

(B10) [Importance of x_1 and $x_2]^2$ = [Importance of $x_1]^2$ + [Importance of $x_2]^2$

Part C: Application

Table 2: Measures of importance

Variable	Importance	[Importance]^2
x_1	6.5	43
x_2	13.8	191
x_3	6.9	47
x_4	7.3	53
x_5	6.8	47
x_6	11,8	139

Table 3: Vote as a function of x_2 and x_6

	low x_2	middle x_2	high x_2
high x_6	55.8 (7)	51.8 (10)	37.6 (12)
middle x_6	35.3 (9)	38.0 (11)	39.6 (9)
low x_6	25.1 (14)	26.5 (8)	33.3 (8)

SOME INTERESTING DECISION FUNCTIONS

Michael F. Capobianco
Polytechnic Institute of Brooklyn

Basic Considerations of Statistical Decision Theory

The problem is to decide which of q possible states of nature $\theta_1, \theta_2, \ldots, \theta_q$ is the true one by observing the outcome of some experiment which has n possible outcomes $x_1, x_2, \ldots, x_n$. For each θ_j there is a probability distribution vector

$$P_j = \begin{bmatrix} p_{1j} \\ p_{2j} \\ \vdots \\ p_{nj} \end{bmatrix}$$

where $p_{ij} = P(x_i | \theta_j)$. We also form a loss vector for each θ_j

$$W_j = \begin{bmatrix} w_{1j} \\ w_{2j} \\ \vdots \\ w_{qj} \end{bmatrix}$$

where w_{ij} = the loss incurred in making decision d_i when θ_j is the true state of nature; $w_{ij} \geq 0$ and $= 0$ if and only if $i = j$. In order to make decisions we need a mechanism for choosing a θ_j upon observing an x_i. Such a mechanism is called a decision function and can be represented by a matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \cdots & a_{1n} \\ \vdots & & \\ a_{q1} & \cdots\cdots\cdots & a_{qn} \end{bmatrix}$$

where $a_{ij} = P(d_i | x_j)$.

The idea is to find a matrix A that in some sense minimizes the loss. We can compute the expected loss, called the risk, for any A under a given state of nature. This is denoted by $R(A, \theta_j)$ and

$$R(A, \theta_j) = W_j' \; A \; P_j$$

where W_j' is the transpose of W_j. In the absence of any further information one way of choosing a single decision function is by using the minimax criterion, i.e., choose the function with risk equal to $\min_A \max_\theta R(A, \theta)$.

Another possibility is to use the maximum likelihood decision function, i.e., choose the function A such that

$$a_{ij} = \begin{cases} 1 & \text{if } p_{ji} \geq p_{j\ell} \text{ for all } \ell \\ 0 & \text{otherwise} \end{cases}$$

This is an example of an non-randomized decision function; each column of A has a single entry equal to 1 and all other entries equal to 0. In such a case we say that A is non-randomized. Clearly each column of any A must add up to 1.

If an a priori distribution is available, i.e., if one has a vector

$$P = \begin{bmatrix} P(\theta_1) \\ P(\theta_2) \\ \vdots \\ P(\theta_q) \end{bmatrix}$$

where $P(\theta_i)$ = the probability that θ_i is the true state of nature, one can then find the expected risk (or Bayes risk)

$$\sum W_j' \; AP_j \; P$$

where $\sum = \begin{bmatrix} 1 & 1 & \ldots & 1 \end{bmatrix}$, a row vector of q 1's. One now chooses the decision function with the minimum expected risk. This is called the Bayes decision function.

One property that a decision function should have is that of admissibility. To explain this term we introduce first the notion of dominance. If $R(A, \theta_j) \leq R(B, \theta_j)$ for all θ_j and strict inequality holds for at least one θ_j, then A is said to dominate B. A decision function is admissible if it is not dominated by any other one.

Proportional Likelihood Decision Function

We argue as follows: There seems to be a weakness in the maximum likelihood criteria in that it chooses that state of nature θ_j which yields the observed x_i with the highest probability, even though some other state of nature may yield x_i with a probability almost as high. It seems reasonable that it would be better to give all states of nature a chance of being chosen which is proportional to their respective probabilities of yielding x_i. We, therefore, propose to form the matrix A with

$$a_{ij} = \frac{P(x_j | \theta_i)}{\sum_{i=1}^{q} P(x_j | \theta_i)} = \frac{p_{ji}}{\sum_{i=1}^{q} p_{ji}}$$

The following example shows that such a decision function may be admissible.

$$P_1 = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix} \qquad P_2 = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{2} \\ \frac{1}{6} \end{bmatrix}$$

$$W_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad W_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$A_1 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \qquad A_2 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad A_6 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

$$A_3 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \qquad A_7 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

$$A_4 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \qquad A_8 = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

Sample Calculation:

$$R(A_4, \theta_1) = W_1' A_4 P_1 = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix} = \frac{2}{3}$$

Tabulation

	$R(A, \theta_1)$	$R(A, \theta_2)$
A_1	0	1
A_2	$\frac{1}{3}$	$\frac{5}{6}$
A_3	$\frac{1}{3}$	$\frac{1}{2}$
A_4	$\frac{2}{3}$	$\frac{1}{3}$
A_5	$\frac{1}{3}$	$\frac{2}{3}$
A_6	$\frac{2}{3}$	$\frac{1}{2}$
A_7	$\frac{2}{3}$	$\frac{1}{6}$
A_8	1	0

Note that A_2, A_4, A_5 and A_6 are inadmissible. Also A_3 and A_7 are both maximum likelihood functions. Now for the decision function proposed above.

$$A = \begin{bmatrix} \frac{1}{2} & \frac{2}{5} & \frac{2}{3} \\ \frac{1}{2} & \frac{3}{5} & \frac{1}{3} \end{bmatrix}$$

Sample Calculation

$$a_{22} = \frac{\frac{1}{2}}{\frac{1}{3} + \frac{1}{2}} = \frac{\frac{1}{2}}{\frac{5}{6}} = \frac{3}{5}$$

Now

$$R:(A, \theta_1) = W_1' AP_1 = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{2}{5} & \frac{2}{3} \\ \frac{1}{2} & \frac{3}{5} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$$

$$= \frac{1}{6} + \frac{1}{5} + \frac{1}{9} = \frac{43}{90},$$

and

$$R(A, \theta_2) = W_2' AP_2 = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{2}{5} & \frac{2}{3} \\ \frac{1}{2} & \frac{3}{5} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} \frac{1}{3} \\ \frac{1}{2} \\ \frac{1}{6} \end{bmatrix}$$

$$= \frac{1}{6} + \frac{1}{5} + \frac{1}{9} = \frac{43}{90}$$

so that A is admissible. Note that all losses were taken as equal in order to make things most favorable for a maximum likelihood decision function. If we make this assumption in general i.e., suppose W_i is a q x 1 vector with 0 in the i^{th} position and the quantity W in all other positions, then we can see that proportional likelihood decision functions become dominated by maximum likelihood decision functions, and in fact have a risk of W for each θ_i as $q \to \infty$. We say, therefore, that they are asymptotically inadmissible. The proof is as follows: Let A be the proportional likelihood function

$$R(A, \theta_i) = W_i' AP_i = W_i' \left[\sum_{k=1}^{n} a_{jk} p_{ki} \right]$$

$$= W \sum_{j \neq i}^{n} \sum_{k=1}^{n} a_{jk} p_{ki}$$

$$= W \sum_{j \neq i}^{n} \sum_{k=1}^{n} \frac{p_{kj} p_{ki}}{\sum_{\ell=1}^{q} p_{k\ell}} = W \sum_{k=1}^{n} p_{ki} \sum_{j \neq i} \frac{p_{kj}}{\sum_{\ell=1}^{q} p_{k\ell}}$$

$$= W \sum_{k=1}^{n} p_{ki} \frac{\sum_{j=1}^{q} p_{kj} - p_{ki}}{\sum_{\ell=1}^{q} p_{k\ell}} = W \sum_{k=1}^{n} p_{ki} (1 - a_{ik}),$$

which approaches

$$W \sum_{k=1}^{n} p_{ki} = W$$

as $q \to \infty$ because

$$a_{ik} = \frac{p_{ki}}{\sum_{\ell=1}^{q} p_{k\ell}} \leq \frac{1}{\sum_{\ell=1}^{q} p_{k\ell}} \longrightarrow 0$$

as $q \to \infty$.

Now for a maximum likelihood decision function A,

$$R(A, \theta_i) = W_1' AP_i = W \sum_{j \neq i} \sum_{k=1}^{n} a_{jk} p_{ki}$$

where

$$a_{jk} = \begin{cases} 1 & \text{if } p_{kj} \geq p_{k\ell} \quad \text{for all } \ell \\ 0 & \text{otherwise} . \end{cases}$$

Therefore,

$$R(A, \theta_i) = W \sum_{k=1}^{n} p_{ki} \sum_{j \neq i} a_{jk} = W \sum_{k=1}^{n} p_{ki}(1 - a_{ik})$$

$$= W \sum p_{ki} ,$$

where the sum is taken over all k such that $p_{ki} < p_{k\ell}$ for some ℓ. This is less than W for at least one θ_i.

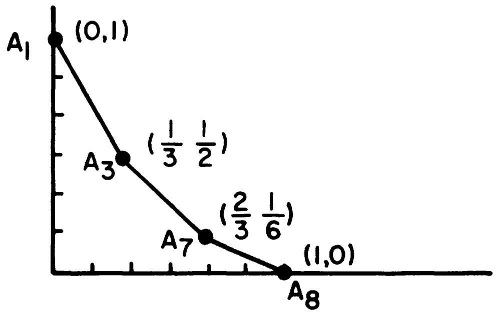
Proportional Bayes Decision Functions

We will assume that in our above example we have on a priori distribution given by

$$P = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \end{bmatrix}$$

and we will show how to find the Bayes decision function.

If we plot $R(A, \theta_1)$ against $R(A, \theta_2)$ for all the admissible A's and join these points by a broken line we have



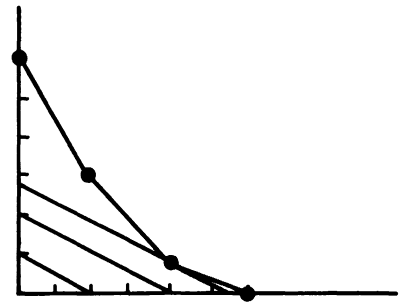
It can easily be shown that any admissible randomized decision function can be obtained from the non-randomized admissi ble ones in the following way: Select two non-randomized admissible functions, say A_3 and A_7, which are jointed by a straight line segment. Choose A_3 with probability α and A_7 with probability $1 - \alpha$. This yields a randomized function A such that

$$R(A, \theta_i) = \alpha R(A_3, \theta_i) + (1 - \alpha) R(A_7, \theta_i), \; i = 1, 2$$

The point $(R(A, \theta_1), R(A, \theta_2))$lies on the line segment jointing A_3 and A_7. Hence, the points of the entire broken line are the risk pairs for all admissible functions. To find which of these is the Bayes function we form the equation

$$\frac{1}{3} x + \frac{2}{3} y = k$$

where, x is the risk under θ_1, and y is the risk under θ_2, and let k vary from zero up until this line first touches our broken line of admissible functions. As we can see from the diagram below, A_7 is the Bayes decision function.



A problem arises in this procedure if the line

$$P(\theta_1)\, x + P(\theta_2)\, y = k \tag{1}$$

has a slope equal to that of one of the line segments of admissible functions. In such a case there will be an infinite number of Bayes functions. In our example this would happen if $P(\theta_1) = P(\theta_2) = \frac{1}{2}$. Then (1) has a slope of -1, and so does the line segment joining A_3 and A_7. We now must choose one of the functions along this segment. To do this we argue as follows:

$$R(A_3, \theta_1) < R(A_7, \theta_1)$$

while

$$R(A_7, \theta_2) < R(A_3, \theta_2) .$$

Therefore if θ_1 were the true state of nature, A_3 would be better, while if θ_2 were true, A_7 would be better. Hence, we propose choosing A_3 with probability $P(\theta_1)$ and A_7 with probability $P(\theta_2)$. It seems that the resulting function is in some sense better than either A_3 or A_7, but it is not clear how this can be expressed mathematically.

The procedures discussed can all be generalized to more than two states of nature. We used the above example in the interests of clarity of exposition.

MINUTES OF ANNUAL MEETING OF THE SOCIAL STATISTICS SECTION

Philadelphia, Pennsylvania, September 10, 1965

Over thirty-five members attended the meeting and heard representatives make their reports. Chairman Eli Marks announced the names of the Section officers for this coming year. These are:

Chairman - Margaret E. Martin (Office of Statistical Standards)
Chairman-Elect - Jacob J. Feldman (Harvard School of Public Health)
Vice-Chairman - John D. Durand (1965-66) (University of Pennsylvania)
Vice-Chairman - Henry S. Shryock (1966-67) (U. S. Bureau of the Census)
Secretary - Philip C. Sagi (University of Pennsylvania)
Representative on Board - Conrad Taeuber (1965-66) (U. S. Bureau of the Census)
Representative on Council - John K. Folger (1966-67) (National Academy of Sciences)

First reported was the action of the Council approving the amendments to the charter of the Social Statistics Section. In summary, these amendments bring the Section Charter into line with the revised Constitution of the Association and provide for adequate liaison between the Council and Board of Directors by making the representatives of the Section to these bodies, members of the Section Committee. The amendments replace a previous provision of the Section Charter which made the Section Chairman and Chairman-elect its representatives on the Council.

A further recommendation was offered to provide that the senior Vice-Chairman serve as program chairman and the junior Vice-Chairman as assistant program chairman. This recommendation was approved by voice vote of members present. The recommended amendment reads as follows:

> Under 'Organization' - Revise third paragraph to read:
>
> The officers of the Section shall consist of a Chairman, Chairman-Elect, two Vice-Chairmen 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. One Vice-Chairman shall be elected each year. Each Vice-Chairman shall serve as Program Chairman of the Section for the second year of his term and as Assistant Program Chairman during the first year of his term. 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, and the Editor of the Proceedings of the Social Statistics Section.

Edwin Goldfield reported that the distribution and sales of the <u>Proceedings of the Social Statistics Section</u> is now at about a thousand copies. In addition, he informed the members of a request made to the Section for support on the inclusion of the question on "race of parents" on the standard birth certificate. The Section recommended that the resolution on Race-Color Designation be published in <u>The American Statistician</u> and sent to other professional organizations which have an interest in this matter.

Suggested program topics for the 1966 meeting of the American Statistical Association include: the poverty program, a continuation of a theme in the 1965 meetings of needed developments in social statistics, the progress of integration, birth control and birth limitation techniques, compliance with the Civil Rights Laws, measurement of the quality of education, the importance of geographic factors in population forecasting and the measurement of the same.

These suggestions are by no means fixed. Additional suggestions are invited by the Program Chairman, Dr. Margaret Martin, Executive Office of the President, Bureau of the Budget, Office of Statistical Standards, Washington, D. C. 20503. To be of assistance in planning the 1966 program, suggestions should be in the hands of the Program Chairman before January 1, 1966, since the preliminary program of the Social Statistics Section, with titles and speakers, is to be submitted to the ASA Program Chairman by January 7, 1966.

Respectfully submitted,

Philip C. Sagi, Secretary
Social Statistics Section
American Statistical Association

<u>1965 Officers of the Social Statistics Section</u>

Chairman:	Eli S. Marks	Secretary:	Philip C. Sagi (1964-65)
Chairman-Elect:	Margaret E. Martin	Proceedings Editor:	Edwin D. Goldfield
Vice-Chairmen:	Jacob J. Feldman (1964-65) John D. Durand (1965-66)		