I

DATA EVALUATION PROCESSES

Chairman, Leslie Kish, Survey Research Center

MEASUREMENT ERRORS IN CENSUSES AND SURVEYS

By: Morris H. Hansen, William N. Hurwitz, and Max A. Bershad
Bureau of the Census

SUMMARY

Introduction

In a census or a sample survey we may obtain observations through personal inquiry, direct questionnaire, or other methods. The set of measurements or observations recorded in the collection operation ordinarily are examined for internal consistency and acceptability, certain "corrections" may be made, and some of the entries may be coded to identify them in a classification system. The results are then summarized into totals, averages, correlations, or other statistical measures. Taken together the collection and processing operations constitute the measurement process and are the source of any measurement errors.

In considering measurement errors we shall regard a survey as being conceptually repeatable, such that repetitions may relate to the same point or period in time and such that carrying through the operation once does not influence results obtained in repetitions of the operation. A particular survey result or estimate is the result of one trial. This conception provides the basis for defining a variance and bias due to response, processing, or other sources of measurement errors. Such a postulation can reasonably approximate actual conditions for a single survey, even though in practice independent repetitions of the survey may be impracticable or impossible.

Measurement errors may arise from many different types of causes, and depend on the conditions under which the survey is taken. Some of the conditions under which a survey is taken may be beyond the control or specification of the survey designer. Other conditions can be controlled so as to influence the quality of survey results in the sense that various aspects of the conduct of the survey are specified. These are typically imposed or attempted to be imposed in an effort to insure adequate quality. Conditions subject to control in survey design but which might be regarded as varying between the conceived repetitions of the survey are the particular choices of interviewers (if it is an interview survey) and other personnel chosen to do various aspects of the work, the particular assignments each is given, and other similar variable factors. These conditions in the conceived repetitions of the survey determine the averages, variances, correlations, and other expected values of various functions of the individual measurements.

The survey may be either a complete census or a sample, and a particular survey is regarded as one trial. For simplicity we shall consider the case of estimating a proportion from the survey. An observation on the j-th unit in the survey is designated by x_{jt}, which has the value of 1 if the j-th unit is assigned to the particular class under consideration on the t-th trial, and otherwise has the value 0. An estimate of the proportion of the population having a characteristic from a survey or trial is

$$p_t = \frac{1}{n} \sum_j^n x_{jt} \tag{1}$$

where n is the number of units in the sample (with n = N for a complete census).

We have, then, the total mean square error of the estimated proportion is

$$MSE_{p_t} = E(p_t - p_s)^2 + E(p_s - P)^2 + 2E(p_t - p_s)(p_s - P) + B^2 \tag{2}$$

where

$$p_s = \frac{1}{n} \sum_j^n P_j \tag{3}$$

is the mean, for the particular set of units included in the s-th sample, of the P_j, where P_j represents the expected value of the observed results for the j-th unit over all possible results and samples in which the j-th unit appears, and

$$P = \frac{1}{N} \sum_j^N P_j$$

is the average result over all units and all possible observations under the conditions under which the census or survey is taken.

The first term in Eq. (2) is defined as the response variance contribution to the total variance of p_t, i.e., the response variance of p_t is defined as

$$\sigma^2_{\bar{d}_t} = E(p_t - p_s)^2 = E(\frac{1}{n} \sum_j^n d_{jt})^2 = E\, \bar{d}_t^2 \tag{4}$$

and is a function only of the response deviations, $d_{jt} = x_{jt} - P_j$. In the case of a complete census $p_s = P$, the second and third terms are zero, and this response variance term is the total variance of a census average.

The second term in Eq. (2) is defined as the sampling variance of p_t. When the indicated expected value of the second term is taken this becomes the usual sampling variance formula for the appropriate sample design as given else -

where.[1] The P_j's are the unique values associated with the units, j=1, ..., N, assumed in the usual theory for sampling from finite populations. In the case of a complete census, $p_s = P$ and the sampling variance term becomes zero.

The third term in Eq. (2) is twice the covariance between $\bar{d}_t$ and p_s. This term is not necessarily equal to 0. It will be zero for a complete census, or when repetitions of a survey are defined only for a fixed sample of units, and we shall ignore the effect of this covariance term in this paper.

The final term is the square of the bias of the survey estimate.

Correlated and Uncorrelated Response Deviations

The response variance given in Eq. (4) can be restated in the following form which separates the effect of uncorrelated and correlated response deviations:

$$\sigma^2_{\bar{d}_t} = \frac{1}{n}\sigma^2_d\,[1 + \rho(n-1)] \qquad (5)$$

where

$$\sigma^2_d = E\,d^2_{jt} = \frac{1}{N}\sum_j^N P_j(1 - P_j) \qquad (6)$$

is the variance of the individual response deviations over all possible trials, and

$$\rho = \frac{E\,d_{jt}\,d_{kt}}{\sigma^2_d} \text{ (for j not equal to k)} \qquad (7)$$

is the intraclass correlation among the response deviations in a survey or trial.

The response variance contribution from uncorrelated response deviations is less, often in practice much less, than (PQ)/n, and if there are important contributions to response variance, they arise from the factors involving correlated response deviations.

The possible impact of even a very small intraclass correlation is substantial, as can be seen from an examination of Eq. (5). For example, if the intraclass correlation among response deviations is zero, the response variance of p_t is σ^2_d/n. Suppose, on the other hand, that the intraclass correlation is, say, .01 (a correlation so small that it might in other applications be regarded as of no consequence whatever). Suppose, also, that the sample or census involves the enumeration of n = 3,000 cases. In this case, the impact of the correlation is to increase the response variance by a factor of $\rho(n-1)$ = .01(2,999) = 30, or 3000 percent! Thus, even if the response variance with uncorrelated response deviations is relatively small, when multiplied by such a factor it may be quite large.

As an example, an interviewer's misunderstanding of his instructions, carelessness, or a tendency to introduce his own judgments into a survey, may cause his results to differ from those of other interviewers, and thus be a source of correlated response deviations. A supervisor's interpretations of instructions that are passed on to interviewers under his jurisdiction and that differ from those of another supervisor may be another cause of correlated response deviations; the varying interpretations of different coders or other processors may be another cause.

We have carried through experimental studies that provide approximate rough estimates of the various correlated and uncorrelated response variance contributions to the MSE of estimates of various items in the 1950 Census. Our estimates can, in fact, be shown to be such that they will lead to understatements in general, of the various terms of the MSE, but are rough approximations and have proved useful in guiding census plans and further research.

As an illustration, we shall choose the proportion of persons classified as farmers and farm managers, which for April 1950, was estimated to be .039 in the 1950 Census. The corresponding estimated proportion from the Current Population Survey was .042. The difference of .003 is our estimate of the bias in the Census for this item (a relative bias of about 7 percent).

We shall now develop an approximation to the total mean square error for this illustrative item by making certain additional assumptions as to the repeated trials. We shall assume, first, that interviewers are independently selected and assigned in each repeated trial, but that other aspects of the staffing, procedures, etc., remain fixed. Also, we shall ignore any contributions to the response variance of correlated response deviations other than within the work of interviewers, and shall, as a consequence, understate the total response variance.

For this special case the total MSE can be written approximately

$$MSE_{p_t} \doteq \frac{\sigma^2_d}{n}\,[1 + \rho(\bar{\bar{n}} - 1)] + \frac{N - n}{N - 1}\,\frac{\sigma^2_S}{n} + B^2 \qquad (8)$$

where N is the total persons of the area, n is the number of persons in the sample (n = N for a complete census), and $\bar{\bar{n}}$ is the number of persons covered by each interviewer. Also, ρ is the intraclass correlation among response deviations within the work of an interviewer, σ^2_d is the response variance per unit given by Eq. (6),

$$\sigma^2_S = \frac{1}{N}\sum_j^N (P_j - P)^2$$

is the sampling variance per unit, and B is the response bias.

N and n depend on the size of the population of the area under consideration, and approximate values for other parameters relating to the illustrative item we selected above, i.e., proportion of persons classified as farmers and farm managers, are given below:

$\bar{\bar{n}}$ = 1000 for a complete census and = 250 for a 25 percent sample

$P = .04$ and $PQ = .0384$

$\sigma_d^2 = .13PQ = .005$

$\rho = .03$

$\sigma_S^2 = PQ - \sigma_d^2 = .033$

$B = .003$

The values for ρ, σ_d^2, and B are based on material presented in certain other sources. [2], [3], [4], and [5]

We shall assume that the general conditions of the survey are substantially the same whether the coverage is based on a 100 percent or on a 25 percent sample -- i.e., that the above approximate values will hold in either case.

The accompanying table compares results of a 100 percent and of a 25 percent sample survey, for populations of different sizes. The above values are assumed to hold in each case, and the numbers in the table were obtained by substituting the values assumed above in Eq. (8).

The following inferences are drawn from such results for items as that illustrated (this item was selected for illustration because it was roughly typical of a good many types of measurements in the census):

1. The combined sampling variance and the response variance contribute significantly to the MSE for very small tabulation cells.
2. The $\sqrt{MSE}$ with a 25 percent sample is not substantially greater than $\sqrt{MSE}$ for a complete census, even for the smaller cells.
3. The response bias is the important contributor to the errors of census statistics, especially for large tabulation cells.

Inferences such as these were important factors in the introduction of sampling and in the development of other modifications in census methods that have been introduced into the 1960 Population and Housing Censuses. However, much more research is needed, and extensive work is planned, to evaluate the effectiveness of the present and alternative methods.

We expect to report more fully on the appropriate theory and empirical results in a forthcoming paper. Extensive experimental work is being planned in the 1960 Censuses that will provide much fuller information than now available on response variances and biases, and the experiments to produce these results are described briefly in another paper in these Proceedings.

Size of population N	Complete Census							25 Percent Sample						
	n	$\sigma^2_{\bar{d}}$ ($\bar{\bar{n}}$=1,000)	σ^2_p	B^2	MSE	$\sqrt{MSE}$	$\frac{\sqrt{MSE}}{P}$ (%)	n	$\sigma^2_{\bar{d}}$ ($\bar{\bar{n}}$ = 250)	σ^2_p	B^2	MSE	$\sqrt{MSE}$	$\frac{\sqrt{MSE}}{P}$ (%)
1,000...........	1,000	.000 155	--	.000 009	.000 164	.0128	32	250	.000 169	.000 099	.000 009	.000 277	.0166	42
2,500...........	2,500	.000 062	--	.000 009	.000 071	.0084	21	625	.000 068	.000 040	.000 009	.000 117	.0108	27
5,000...........	5,000	.000 031	--	.000 009	.000 040	.0064	16	1,250	.000 034	.000 020	.000 009	.000 063	.0079	20
10,000..........	10,000	.000 016	--	.000 009	.000 025	.0050	13	2,500	.000 017	.000 010	.000 009	.000 036	.0060	15
25,000..........	25,000	.000 006	--	.000 009	.000 015	.0039	10	6,250	.000 007	.000 004	.000 009	.000 020	.0045	11
100,000.........	100,000	.000 002	--	.000 009	.000 011	.0033	8	25,000	.000 002	.000 001	.000 009	.000 012	.0035	9

REFERENCES

[1] M. H. Hansen, W. N. Hurwitz and W. G. Madow, Sample Survey Methods and Theory, John Wiley & Sons, Inc., New York, 1953.

[2] U. S. Bureau of the Census, The Post-enumeration Survey: 1950, (in press).

[3] A. R. Eckler and W. N. Hurwitz, "Response Variance and Biases in Censuses and Surveys," Bull. de L'Institut International de Statistique, Tome 36-2e Livraison, Stockholm, 1958, pp. 12-35. Paper presented at the 30th Session of the I. S. I.

[4] U. S. Bureau of the Census, The Accuracy of Certain Census Statistics with and without Sampling, Bureau of the Census Technical Papers -- No. 2 (in press).

[5] W. N. Hurwitz and M. A. Bershad, Self-enumeration with Follow-ups, Bureau of the Census (unpublished memorandum), April 1958.

HEALTH STATISTICS FROM RECORD SOURCES AND HOUSEHOLD INTERVIEWS COMPARED

By: O. K. Sagen, Ruth E. Dunham, and Walt R. Simmons
U. S. National Health Survey

1. INTRODUCTION

Analysis of the differences between ideal knowledge of a quantity and the product of a process for measuring that quantity can take many forms. Indeed the possible hazards of a statistical measurement are almost without bound. One early listing is offered in J. Edward Demings text on sampling.[1] Numerous other lists exist.

Statisticians may find it convenient to break discrepancies into four major components:

(a) Conceptual differences between the universes of the idealized quantity and of the defined measurement;

(b) bias of the measurement, i.e., the difference between expected value of the designed measurement process and the "pure" or "true" defined measurement;

(c) the average squared deviation of a designed measurement from the expected value of such measurements, i.e., the variance of the process; and

(d) the mistakes of measurement, i.e., the non-random errors of execution.

Within each of these major components, it is possible to identify sub-components, and also possible to identify areas of hazard which overlap more than one of the foregoing major components.

It is sufficient for present purposes to note that origins of error are found not just in the commonly recognized unrepresentative character of many "samples" and in inaccuracies of counting, but in every step of the statistical process, from initial birth of the notion that a measurement needs to be made, to the ultimate use which is made of the statistic produced.

General Techniques for Evaluating Data

Again a long list of approaches might be identified. Four general methods are recalled here.

(a) Global Comparisons of Two Sets of Statistics.—Certainly one of the oldest and most commonly utilized techniques is the comparison of end results from two separate procedures which in some sense are thought to be measures of the same phenomenon. The two procedures may both be sample surveys, as for instance is the case when trend of unemployment is measured both by the Monthly Report on the Labor Force, and by a 1-percent sample of claimants for unemployment insurance. Or the comparison might be between two presumptively complete enumerations: e.g., the total value of real estate in an area as indicated on the one hand by tax assessments and on the other by insurance evaluations. Or the comparison may be between results from a sample procedure and those from a census. One continuing, formalized version of this type is engaged in by the Bureau of Labor Statistics which compares annually its current estimates of employment from a sample of business establishments with totals from social security records on which taxes are paid. Adjustments of the former are made to the latter.[2] Each reader can provide other illustrations.

Except as it may be supported by subsequent further exploration, this type of evaluation is indeed but a comparison of two sources; it does not determine in itself which if either is "correct." Nor does it usually identify components of the discrepancy. It is a device for quantifying the global difference between two figures.

(b) Replication.—The traditional procedure of the physical scientist for validation of findings is repetition of the experiment. Thus a discovery by one physicist is checked by a second or third or more researchers who attempt to duplicate as nearly as possible the original procedure to test whether the original findings are reproduced. The same general process is of course common in laboratory and clinical trials in other fields.

There are many variations of this central notion. Much of the purpose of the entire technique of Design of Experiments, popularized by R. A. Fisher and employed especially throughout the biological sciences is to discover whether an observed number is truly meaningful; i.e., whether it is a valid finding. In one manner or another, the process seeks through statistical controls over selected factors to discover whether repeated observations on another factor do produce an acceptably stable characteristic.

In the social sciences, four further variations may serve as reminders of other aspects of repetitive checks. One technique, used by a number of experimenters, including especially those at the Survey Research Center of the University of Michigan and at the U. S. Census Bureau, is the procedure of reinterview of respondents in a household survey. While these reinterviews differ in purpose from one survey to another, the prime purpose usually is to attempt a repetition of the original interview by use of a more skillful interviewer and thereby to discover if the original interviewer was successful in carrying out his original assignment.

Another technique also employed widely in household sampling, is comparison of results from two or more sets of questions, each set being given to a different random sample of the same population by a fixed panel of interviewers under a fixed operational procedure. Here one seeks to discover whether the formulation of the question produces a stable result.

A third variant seeks to discover interviewer-variability through random assignment of interviewers to random subsamples of the population, and comparison of the results among interviewers. A description of some aspects of this matter is given in these Proceedings in the paper by Hansen, Hurwitz, and Bershad.

The fourth variation is one in which differences between respondents are noted by utilizing common interviewers, questions and procedures, but addressing the inquiry to different types of respondents.[3]

(c) Record-Check Studies.—This title, though coming to be widely used, is a misnomer for the class of validations which are treated here under this label. The reason for the title is made clear with two illustrations: (1) In the study described in the body of the present paper a comparison is made for each respondent between the answer given by the respondent in an interview survey and the written record of the fact about which the respondent was questioned. (2) Age as reported in the Population Census was checked against the record of birth registrations for a sample of cases in the 1950 Census.[4] The idea behind this class of validation is that there exists some different superior type of measure to the one being utilized in the main survey, and that this superior measure, through a case-by-case comparison, will provide an estimate of the bias of the main survey measure, and in addition may indicate the cause of error in the latter. "Record check" was termed a misnomer because in some instances of applying the same general process, the main survey measure is itself a written record, while the presumed superior standard may be a different sort of measure.[5] Normally, it would be supposed that the "superior" measure was unavailable on a comprehensive basis—for cost or other reasons—or it would be used as a substitute for the main survey measurement.

(d) Internal Consistency.—In some types of investigations, data from the survey itself afford an opportunity for an internal consistency check. For example, in the BLS Consumer Expenditure Studies it is necessary that revenue of all types approximately balance with outlays, if the statistics are to be valid. In housing statistics, the census of dwelling units at one point in time augmented by new construction, and adjusted for conversions and demolitions, should equal approximately the census at a later date, definitional changes being taken into consideration.

Commentary

Thus, a considerable and growing number of methods are available for assessing gross error and the many components of statistical error. Progress on a very difficult front is being made. But the authors state their belief that much needs to be done, both in undertaking and reporting specific evaluation projects, and in synthesizing error analysis.

Most schemes known to us suffer from either one or both the difficulties of unsatisfactory theoretical formulation and operational execution.

The theoretical difficulties are of many types, with perhaps two being especially prominent: (1) the task of establishing conceptual identity between a measurement and the phenomenon measured, and (2) the building of models which disentangle confounding components of errors in a tractable manner.

Operational difficulties also can be severe. They vary with the type of validation. As examples, note three illustrations:

(1) In a large scale survey the administrative burden of randomizing interviewers over geography is heavy, and the process is costly.

(2) In a record-check study, it is not always easy to match individual cases; further, it may not be clear that the "superior standard" is in fact the better measure.

(3) In replication procedures which involve human respondents and frequently human interviewers, can one be assured that conditioning of respondent or other agent has not taken place?

2. MEASUREMENT PRINCIPLES IN A HEALTH SURVEY

Consider the matter of how many persons in the United States are sick. There is first the idealized concept of this quantity, which in some sense is the number of persons who are unwell or unhealthy. Then there is a defined measurement which may be applied to each person. The measurement process may consist of obtaining answers on interviews to a specified set of questions, including perhaps such inquiries as : (1) were you sick at any time last week or the week before? (2) Last week or the week before did you have any accidents or injuries, or suffer from the effects of a previous injury? (3) Last week or the week before did you take any medicine or treatment for any condition, or see a doctor concerning any condition, or cut down on your usual activities because of some health condition? A positive answer to any of the questions may be used to classify the person as ill.

There clearly are possible differences between this defined measurement and the idealized concept of illness. For example, the person might have glaucoma and be completely unaware of it.

Focus attention next on a survey designed to produce the intended measurement. This survey, let us say, is on an area-type, household multi-stage stratified probability sample design, in which the answers to questions are obtained through personal interview. This survey process and its associated estimation procedure has an expected value which may differ from both the idealized concept and the defined measurement, for a variety of reasons including elements from sampling, training, respondent conditioning, the classification of replies, non-response, the formal estimation process, and many others. These same sources of potential error are also responsible for the variance of the measurement process from one application to another, as different samples are drawn, different interviewers selected, different temporal factors introduced. Finally, a gross error in execution can contribute to differences among the possible measurements: such phenomena as an interviewer who cheats; a storm which prevents interviewing in an entire sector of the nation; a faulty circuit in the electronic computer which edits the data.

To carry the illustration further, turn again to the area of uncertainty which arises from the interviewer. One needs to be careful in labeling interviewer deviations. Quite different interpretations are possible depending on what is accepted as the universe of interviewer-recordable replies. There is, for example, a within-interviewer variance which reflects variation in treatment of the same question with different respondents or at different times by a single interviewer. That same interviewer may have a bias which is defined as the difference between his expected product and that of some group of interviewers of which he is a member. But this same difference becomes a part of the between-interviewer variance when viewed at another level of analysis. In turn, the expected value of the product of all interviewers in the groups combined may be a bias or a component of variance depending on the universe designated as frame of reference.

In dealing with these problems of measurement, we face not only a great variety of types of discrepancy but also many possible dimensions of analysis for evaluating them.

The U. S. National Health Survey is making a number of attempts to better understand and evaluate the data obtained from its household interview survey on health conditions. One of these is the study identified as the H.I.P. record check, which seeks to explore an area which, in essence, is the intersection of two fundamental theses. The first thesis is that the concept of illness inherent in a written medical record created in the course of rendering medical services differs from the concept of illness contained in a household respondent's replies to a specific line of questioning by a survey interviewer. Certainly the interview questionnaire on health conditions and the medical record would be expected to produce some different statements as to health or illness status of the individual concerned. In contrast, the second thesis is that the household interview report and the medical record have some elements in common and that the information from one source is correlated with that from the other source. At least, for certain classes of individuals and conditions, the two information sources ought to furnish measures of the prevalence of illness which vary together from class to class. In summary, interview and medical record information ought to be different but still similar to each other in important respects.

3. THE 1958 H.I.P. RECORD CHECK ON THE NHS HOUSEHOLD INTERVIEW

A stratified sample of families enrolled in the Health Insurance Plan of Greater New York (H.I.P.)[6] was interviewed during May and June, 1958 to compare the illness conditions and hospitalization experience reported on interview against the comparable information on the records of H.I.P. Since H.I.P. enrollees do not necessarily obtain all their medical care from H.I.P. even though they are entitled to full care, this study is largely confined to a one-way comparison, namely, how much of what appears on the H.I.P. record gets reported on interview. However, an upper limit on the extent of overreporting can be inferred from the data on interview-reported conditions which were not on record with H.I.P.

There was a conscious and strong attempt to duplicate the conditions present in a regular National Health Survey household interview so as to make the study representative of the ordinary interview process, an impossible task, to be sure. Interviewing was spread out over a two-month period, interviewers had the same instructions and the same questions for the items under study, and interviewing was conducted by the Bureau of the Census under the supervision of the regular Census Bureau supervisor for the New York City area. Furthermore, this supervisor conducted reinterviews on a sample of the interviewed households just as in the regular survey.

Despite all this, there are a number of specific differences between the interviewing for the record-check study and the interviewing in the regular survey. First, in our study the interviewers were given a surname and a specific address whereas in the regular survey the interviewers approach six dwelling units in a cluster of addresses without any surnames. This fact, together with others, makes it impossible to conceal from the interviewer that the study interviews are something special instead of routine. Second, certain questions were deleted from the regular questionnaire and others added for the study. For example, the study needed to obtain signed consent for examination of the family's medical records and it was also necessary to identify the physician who rendered service for a medically attended condition; third, 12 interviewers were used in the study whereas only three interviewers normally worked on the regular survey in the New York City area. This circumstance made it necessary to recruit and train a small corps of new interviewers for the study. Also, the study sample was stratified to yield more reports of chronic illness than appear in the regular survey.

Even if the ordinary conditions of interviewing could be completely and faithfully reproduced for this type of study, there are several basic problems which again can at best be handled imperfectly. One of the difficult problems rests on how to set up a reliable correspondence between the terminology used by physicians for medical records and the descriptions used by the whole gamut of respondents to a household interview. For example, when the medical record shows "duodenal ulcer and hemorrhoids" while the respondent mentions only "stomach trouble" is there any, some, or no correspondence between the two?

When the medical diagnosis is "club foot" and the respondent reports "flat feet" how should we handle the report? This indicates a need for very broad categories of diagnosis for classifying the diagnostic statements on the medical record coupled with a scheme for grading the degree of correspondence. The man with the club foot at least reported the affected part of the body correctly and no one quarrels with the view that a person with an ulcer has stomach trouble. Accordingly, for this study three degrees of match were used to classify household interview reported conditions which had a possible counterpart in the medical record.

The problem of classifying responses is also complicated by the ground rules for interviewing.[7] The interviewer asks about (a) illness, accidents, injuries, other conditions and medical service in the past two weeks; (b) ailments and conditions that have been bothering one for a long time; (c) a check list of 26 conditions possibly present sometimes during the previous 12 months; and (d) a check list of 9 impairments. A chronic condition is one which is elicited by this procedure and which either appears on one of the check lists or had its onset more than 3 months before interview.

It is easy to assume that the medical record is the perfect source of reliable information on the medically treated conditions which were encountered by the population under study. However, as in any record system we must face the possibility of errors in the criterion document. There are the errors of omission arising from the failure to record pertinent information on the medical record. More serious for this type of study are the errors committed by mis-recording a diagnosis, recording a tentative diagnosis, and mistakenly assigning one patient's ills to another.

The criterion record used in this study is the H.I.P. "Med-10" report, which is not a medical record in the usual sense. The Med-10 is a line entry report for each patient seen by the doctor during a single day; it is designed for statistical and administrative purposes only. For each visit, the doctor identifies the patient, the place at which he was seen (office, home, or hospital), the nature of the service rendered (preventive, obstetric, surgical, and other) and an abbreviated diagnosis.[6] Various studies in H.I.P. have shown that the Med-10 is a reliable record and, for statistical purposes, is both more reliable and convenient than the complete clinical record.

In order to more fully investigate factors which might be connected with interview reporting of medical information and to check up on the validity of the Med-10 record as a criterion record, a physician interview phase of the study was designed. Herein, the H.I.P. physicians who rendered care to the subjects of the household interview were questioned about the case of each such patient. Because from 7 to 8 months necessarily elapsed between the household interview with the enrollee and the interview with his physician it was not feasible to again go back to the household respondent and question him further about discrepancies. Also, H.I.P. policy considerations prevented a check back to the respondent after having compared his interview response with the Med-10 record.

Full analysis of the study is far from completed. Furthermore, it should be borne in mind that a single study of this type cannot be definitive, and can at best be indicative. Nevertheless, the broad outlines of some preliminary findings can be described at this time.

4. THE INITIAL HOUSEHOLD INTERVIEWS AND RECORD CHECK

The persons for study were drawn from the subscribers and their dependents enrolled in H.I.P. on June 30, 1957, a total of 513,000 enrollees. Interviews were conducted on about 5,000 of these enrollees in 1,413 households. Of these households 959 were in the "chronic" stratum (i.e., at least one member had service for a chronic disease after June 30, 1957) and 454 households were in the "non-chronic" stratum. In the population represented by this sample, an estimated 44 percent had at least one chronic condition on a Med-10 record after June 30, 1957.

Table 1 summarizes the principal components of the interview which were checked against the Med-10 records. It is noteworthy, although not surprising, that the experiences which are least difficult to describe, viz., hospitalizations and recent acute illness or injury, are far more fully reported than are the chronic conditions.

The percent of conditions reported varies considerably with the kind of illness. Among the acute conditions, respiratory ailments are 73 percent reported while acute eye and ear conditions are reported only 40 percent of the time. Among

Table 1. Correspondence between experience on record and household interview reports by type of respondent for 6,609 persons (weighted frequencies)

Class of experience	All respondents			Self-respondents[a]			Proxy respondents[a]		
	Med-10[b] record	HHI[c] report	Percent	Med-10[b] record	HHI[c] report	Percent	Med-10[b] record	HHI[c] report	Percent
Non-chronic conditions	201	127	63	58	35	60	143	92	64
All chronic conditions	4648	1481	32	2222	791	36	2426	690	28
Class 1--Check list, unqualified	1872	826	44	878	419	48	994	407	41
Class 2--Check list, qualified	1231	340	28	605	216	36	626	124	20
Class 3--Non-check list	1545	315	20	739	156	21	806	159	20
Hospitalization episodes	350	306	87	205	180	88	145	126	87
Days hospitalized									
All episodes	2634	2397	91	1533	1438	94	1101	959	87
Unreported episodes	258	0	0	133	0	0	125	0	0
Reported episodes	2376	2397	101	1400	1438	103	976	959	98

[a]When the respondent and the subject of the interview are the same person, the case is called self-respondent; if the respondent is a responsible adult member of the subject's family, the case is called proxy-respondent.

[b]Refers to data from Med-10 except for "days hospitalized." Hospital stay was obtained directly from the hospital record because the Med-10 only records the days on which the doctor saw his patient in the hospital. Only stays of overnight or longer were counted.

[c]HHI = Reported in Household Interview.

the chronic conditions asthma and hay fever are 76 percent reported while just one-third of the cancer is reported and only 26 percent of mental illness is reported. In general, however, the more medical service is given for a chronic condition the better it is reported. All chronic conditions receiving only one service were 20 percent reported while those with ten or more services were 80 percent reported. None of the cross classifications lead to any surprising inferences but at the same time there is no simple classification scheme which lends itself to a convenient definition of items that are best reported on interview.

The reinterview phase of the study did not contribute enough to better reporting to be taken into consideration here. While some additional conditions were elicited on reinterview, there was not enough to appreciably affect the completeness of reporting when that information was added to the information from original interviews.

5. THE PHYSICIAN INTERVIEW

The individuals about whose illnesses the physicians were interviewed were a sample of 341 adults, drawn randomly from the two strata of people in the record-check study and who met the following criteria: (1) were self-respondents in the household interview; (2) gave written permission during the household interview for review of medical records; and (3) received service from H.I.P., according to Med-10 data, during the study year for one or more "queriable" conditions.

These conditions, termed queriable, were chosen for purposes of clarity of results and ease of interviewing. They were limited to certain conditions defined in the National Health Survey as chronic conditions or as impairments. Excluded were acute conditions (such as pneumonia), certain minor chronic conditions (such as fungus infection of the feet) for which an interview would be pointless or awkward for the participants and produce little useful information, and diagnoses indicated as tentative on the Med-10.

Personal interview of the attending physician was chosen as the means of obtaining the information sought. This choice was made on the basis that personal interview (1) would result in greater response than would other methods, e.g., mail queries; and (2) would be likely to result in better information of the kind needed. This latter point was related closely to the choice of interviewer personnel.

Physicians were decided on as interviewers because it was believed that they could be expected to gain entree to the attending physicians readily, could establish a good relationship with them quickly, and could obtain the needed information in the shortest time. Six Public Health Service medical officers were specifically trained for this purpose and conducted the physician interviews.

The interview used a 4-page questionnaire, part of which was quite rigidly structured to provide categorical answers and part of which allowed free responses. Each questionnaire dealt with one patient and provided space for comments by the physician on three conditions. In some instances, a physician was interviewed regarding several patients during one appointment, and for a number of patients more than one physician had to be interviewed for services rendered the same person. The physician interviews took place during 3½ weeks in January and February 1959. Two hundred and eighty-four H.I.P. physicians in 30 of the 32 H.I.P. medical groups were seen.

6. RESULTS FROM THE PHYSICIAN INTERVIEW

There were 341 persons whose chronic illnesses and impairments were investigated in the physician interview. About nine percent of the chronic conditions recorded on the Med-10 represented only tentative diagnoses and another four percent of the diagnoses had been ruled out and the patient so informed before interview. Although the tentative and ruled out diagnoses were not considered appropriate as items to check against interview reports, nonetheless 30 percent of these diagnoses did get reported on interview.

In 97 percent of the cases with a definite diagnosis, the patient had been informed of the diagnosis in some fashion, accounting for 452 conditions. If the medical terminology had been given the patient, 63 percent of the conditions were reported in the household interview, whereas if the

Table 2. Influence of complaints, medication, interference, and distress, as determined in physician interview on reporting of chronic conditions by class of condition (weighted frequencies)

	Total conditions		Class 1 Checklist conditions unqualified[a]		Class 2 Checklist conditions with qualifications[a]		Class 3 conditions not on checklist	
	Number of conditions	Percent reported	Number of conditions	Percent reported	Number of conditions	Percent reported	Number of conditions	Percent reported
Total conditions	452	53	275	62	63	56	114	32
Conditions with:								
Complaints or symptoms	359	58	213	66	56	55	90	40
Medication or regimen	356	59	219	68	53	57	84	37
Interference in usual way of life	186	69	108	77	34	68	44	52
Physical or emotional distress	296	62	172	70	49	78	75	47
Conditions without:								
Complaints or symptoms	91	35	60	47	7	57	24	0
Medication or regimen	96	31	56	36	10	50	30	17
Interference in usual way of life	264	42	166	51	28	43	70	19
Physical or emotional distress	151	37	100	48	14	50	37	3

[a]A check-list condition is termed unqualified, if it is listed simply and without limiting restrictions; e.g. "Heart Trouble", a condition listed with any limiting description is called qualified; e.g. CHRONIC bronchitis.

Table 3. Completeness of reporting chronic conditions among cases investigated in the physician interview according to how many times medical service was received in the 12 months preceding interview (weighted frequencies)

	All cases		Cases with some botheration[a]	
	Number	Percent reported	Number	Percent reported
Total conditions	452	53	156	72
Medical service given:				
Once	137	36	31	65
2-4 times	164	46	44	55
5-9 times	73	64	37	76
10 or more times	78	87	44	93

[a]Complaints or symptoms, medication or special regimen, interference with usual way of life, and physical or emotional distress.

condition had been described in some other way only 46 percent were reported.

An important objective of the physician interview aspect of the study was to see if certain characteristics of a chronic illness might have a bearing on how well the condition was reported. The physicians accordingly were asked if the condition:

(1) caused the patient any complaints or symptoms, or
(2) required any medication or special regimen, or
(3) might have interfered with the patient's usual way of life, or
(4) might have caused the patient physical or emotional distress.

The findings are summarized in Table 2. Each of these four characteristics which might accompany a chronic condition produced about the same pattern of response for the three classes of conditions, a situation which does not obtain in the absence of these characteristics.

A little different axis of trouble associated with a chronic condition is the volume of medical services required by the condition. Table 3 demonstrates that the most troublesome conditions get reported best.

The physician interview study could not help greatly in interpreting factors in the reporting of specific diagnoses because of the small frequencies For example, it would not be correct to say that the study demonstrated more nearly complete reporting for diabetes than for hemorrhoids, although the data do suggest this and it is not an unreasonable finding. However, when we confine attention to the cases confirmed by the physician as definite chronic disease, some items of interest were very well reported. For example, asthma and hay fever were 88 percent reported, sinusitis 73 percent, diabetes 93 percent, ulcers 80 percent, heart disease 70 percent, and arthritis and rheumatism 61 percent.

7. DISCUSSION

The overriding conclusion to be drawn from the preliminary results of the H.I.P. record-check study is that the prevalence of chronic illness obtained from household interviews must essentially rest on an operational definition determined by the interview procedure itself.

That is to say, the prevalence of heart disease as estimated from the survey is not its prevalence in the clinical sense of one disease but the frequency of positive answers to a specific line of questioning, carried out in the survey context.

Comparison of the household interview data with H.I.P. records shows definite and substantial underreporting of chronic disease for which medical service was rendered. However, the evidence is just as clear that substantial correlation exists between the two sources and that the correlation is greatest for the conditions which combine a large volume of medical service with considerable travail and inconvenience to the patient. The complexity and multiplicity of factors affecting correspondence between the two sources makes it a formidable task to describe their precise relationship. Nevertheless, preliminary results justify the hope of satisfactorily describing classes of conditions for which correspondence is high, others for which it is lower, and so on. A table, which arrays conditions along an axis graded by degree of importance or severity, showing percent of correspondence, definitely establishes the basis for this hope.

Percent correspondence between household interview and medical record for definitely diagnosed chronic conditions of which the patient had been informed by his physician

Characteristics	Percent correspondence
1. No symptoms, distress, medication or interference with usual activity	15 percent
2. Symptoms only	19 percent
3. Symptoms and distress only	41 percent
4. Symptoms, distress, and medication only	52 percent
5. Symptoms, distress, medication, and interference	72 percent
6. All four characteristics together with 10 or more medical services for the condition	93 percent

While this paper does not take up the question of overreporting, viz., the reporting of conditions in the household interview which are not reflected in medical records, the full study does show that such possible overreporting fails by far to compensate for the underreporting just discussed. It is fairly evident that any diagostic information which an individual is able to report about his illness is dependent on his having had medical care for the ailment and to a large extent on his being given the medical terminology for the condition by the physician. Otherwise, all the individual can do is to report complaints which may or may not be significantly symptomatic of a diagnosable disease.

Several investigators have discounted the household interview survey technique for its inability to measure the prevalence of chronic disease and they advocate surveys using clinical examinations as yielding more information and a truer picture.[8-12] At the same time this has tended to obscure the fact that the data obtainable from clinical examination also rest on an operational definition of its own which is inherent in the type of examination conducted, the length of time over which the individual is observed, and the point in time at which the examinations are concluded.

The situation with respect to acute illness and hospitalization is somewhat different. Both types of experience are much more clear cut. Limitation of the time reference to the two weeks preceding interview for acute conditions not only holds down memory loss but makes the items inquired about much more specific and understandable to the respondent. A similar time restriction for chronic disease would tend to ignore the main feature of much chronic illness, i.e., that episodes of acute distress come and go and that once the individual has adjusted to his chronic difficulty he tends to consider the situation normal for him.

In the current H.I.P. study it was found that known hospital episodes were 13 percent underreported in the household interviews while the aggregate days of overnight or longer hospital stays reported in household interviews were 8 percent less than on record at the hospital. In San Jose, California, Belloc[13] observed, respectively, 13 percent and 7 percent underreporting of the two items. In addition, preliminary results from another current NHS study* yield almost identical figures on underreporting. All three of these studies used a one year recall period for the hospitalization questions and confined themselves to hospital stays of overnight or longer. Belloc found that, among San Jose residents, hospital episodes were 11 percent overreported, thereby very nearly compensating for the underreporting. However, neither the H.I.P. nor SRC studies find overreporting to anywhere near this extent.

8. SUMMARY

A discussion of concepts and principles involved in using record sources as a standard of comparison for reports from survey interviews has been presented to evoke interest in further research and study of this aspect of social survey techniques. The problems of concept and design for a study of health interviews compared against medical records have been described and some preliminary results from the study have been presented. Even these rather fragmentary findings leave little room for doubt that in an interview survey of individuals with respect to their health, strictly operational definitions apply to the data on the prevalence of chronic conditions by diagnostic categories. Correspondence between data from household interviews and medical service records varies with the degree of travail and inconvenience to the patient and the volume of medical service entailed by the condition.

Comparisons with alternative sources of data in the same general area can illuminate the operational definitions but cannot replace them with more nearly absolute concepts.

Data on acute illness and on hospitalization present less severe problems and show closer correspondence between reports from household interview and medical record of service than is the case for chronic disease.

*A contract study by the Survey Research Center, University of Michigan utilizing a sample of discharges from 21 hospitals scattered over the United States. The households of the discharges were interviewed in 1959 by the Bureau of the Census according to the regular NHS specifications.

REFERENCES

1. Deming, William Edwards. Some Theory of Sampling, New York, John Wiley & Sons, Inc., 1950.

2. Simmons, Walt R. "Measurement of Industrial and Commercial Employment in the United States," The American Statistician, Vol. 10, No. 4, October 1956.

3. Nisselson, H. and Woolsey, T.D. "Some Problems of the Interview Design for the National Health Survey," Journal of the American Statistical Association, Vol. 54 (1959), pp. 69-86.

4. Marks, Eli S., Mauldin, W. Parker, and Nisselson, Harold. The Post-Enumeration Survey of the 1950 Census; a Care History in Survey Design, Journal of the American Statistical Association, Vol. 48, June 1953.

5. Simmons, Walt. R., "Concepts and Response in Statistics on Business Establishments." Study described at the annual meeting of the A.S.A. in Detroit, September 1956.

6. Division of Research and Statistics. HIP Statistical Report - 1958. Health Insurance Plan of Greater New York, 625 Madison Ave., New York, N.Y., November 1959.

7. U.S. National Health Survey. Concepts and Definitions in the Health Household-Interview Survey. Health Statistics. Series A-3. Public Health Service Publication No. 584-A3. Public Health Service, Department of Health, Education, and Welfare, Washington, D.C., September 1958.

8. Elinson, Jack, and Trussell, Ray E., "Some Factors Relating to Degree of Correspondence for Diagnostic Information as Obtained by Household Interview and Clinical Examinations," American Journal of Public Health, Vol. 47 (1957), No. 3, pp. 311-321.

9. Krueger, Dean E., "Measurement of Prevalence of Chronic Disease by Household Interviews and Clinical Evaluations," American Journal of Public Health, Vol. 47, No. 8 (1957), pp. 953-956.

10. Commission on Chronic Illness, Chronic Illness in the United States Vol. III: Chronic Illness in a Rural Area, The Hunterdon Study. Harvard University Press, Cambridge, 1959.

11. Commission on Chronic Illness, Chronic Illness in the United States Vol. IV: Chronic Illness in a Large City, The Baltimore Study. Harvard University Press, Cambridge, 1957.

12. Eichorn, R.L., and Morris, W.H.M., "Respondent Errors in Reporting Cardiac Conditions on Questionnaires," Proceedings of the Purdue Farm Cardiac Seminar, Purdue University Agricultural Experiment Station, Lafayette, 1959, pp. 46-48.

13. Belloc, Nedra B., "Validation of Morbidity Survey Data by Comparison with Hospital Records," Journal American Statistical Association. Vol. 49 (1954), pp. 832-846.

II

LABOR FORCE MEASUREMENT AND INTERPRETATION

Chairman, Gladys L. Palmer, University of Pennsylvania

RECENT CHANGES IN PROGRAMS FOR LABOR FORCE MEASUREMENT

By: Margaret E. Martin, Office of Statistical Standards, Bureau of the Budget [1]

In July of this year, the Department of Labor assumed a large share of the responsibility for the Federal government's current program in labor force measurement. At that time, full responsibility except for data collection was transferred from the Bureau of the Census to the Bureau of Labor Statistics. The Monthly Report on the Labor Force is now planned, guided, budgeted for, analyzed and published by the BLS, with Census acting as agent in the collection and tabulation of the results, as part of the monthly Current Population Survey.

A word on past history may help give perspective. The Monthly Report on the Labor Force was originated late in the 1930's by the WPA. The survey was just well started when it became evident that the WPA would shortly be terminated. At that point, the Bureau of the Budget undertook a review of the Survey to determine: firstly, if it should be maintained, and secondly, which department or agency should assume the function. The Bureau found general agreement that the survey should be continued as a part of the government's statistical program, and also found that more than one agency was interested in undertaking the activity. After consideration of possible alternatives, the function was assigned to the Census Bureau on the grounds that a household survey of the population and the development of the required sampling techniques were both activities which would benefit from close connection with that Bureau's major interests. The Census Bureau itself would benefit by having at hand a population planning and analytic staff, by making use of the sample survey in planning the decennial censuses and by maintaining a field force experienced in household enumeration as a core for decennial census operations.

It was recognized from the start that other government agencies had a vital interest in the survey. When the Census Bureau was given the function, it agreed to continuous review of its plans by two interagency committees, one dealing with occasional policy issues and chaired by the Assistant Director for Statistical Standards of the Bureau of the Budget, the other on technical problems chaired by Miss Gladys Palmer, consultant to the Bureau. The Committee on Labor Supply, Employment and Unemployment Statistics, familiarly and affectionately known within the Government as "The Palmer Committee", is composed of technicians drawn from Federal agencies interested in labor force measurement and is still actively functioning.

The Department of Labor was not content with the decision to assign the survey to the Bureau of the Census. Measurement of employment and unemployment, the core of the survey, were considered to be subjects of primary concern to that Department, because of its responsibilities for labor matters--for employment and unemployment policy, for the program of unemployment insurance, for the operation of public employment offices and other related policies. On a number of occasions, the Secretary of Labor requested the Bureau of the Budget to reconsider its determination. Meanwhile, the Department of Labor continued to participate in and, in fact, to take a leading part in the interagency planning and review of the survey at both the technical and the policy level.

During the years of the Census Bureau's stewardship, the program for collection of monthly information on the labor force became an integral part of the Government's statistical program. Early problems of sample design were solved and the sample expanded. Many problems of field collection of the data were resolved. The household survey developed into a vehicle for asking a variety of supplemental inquiries in areas only distantly related, if at all, to labor force analysis - for example, the taking of polio shots, or anticipations to buy certain items of household equipment. And there was a continual expansion in the amount and detail of labor force information provided simultaneous with growing public interest and reliance on the results of the survey.

With the acceptance of the survey as providing particularly useful current economic indicators, increased emphasis has been placed on comparison of changes in the labor force figures with other monthly series, especially the monthly employment estimates based on establishment reports which are prepared by the Bureau of Labor Statistics. To eliminate confusing public announcements on monthly employment and unemployment changes, a joint monthly release of the labor force estimates, the employment series, and the insured unemployment reports was arranged. From the spring of 1954 until last summer, staff of the Bureau of the Census, the Bureau of Employment Security, and the Bureau of Labor Statistics met once a month under the chairmanship of the Bureau of the Budget to draft a combined statement for release by the Secretaries of Commerce and Labor.

Despite these cooperative ventures, it became apparent in recent years that additional use of labor force measurements could be made within the Department of Labor for development of policy, for program-planning purposes and for current labor market analysis. Flexibility in meeting these needs would be facilitated if the planning and analysis of the figures were primarily the function of the Labor Department. At the same time, comparisons between the household survey results and the establishment payroll report results would be encouraged and more light thrown on reasons for differences if the two series were the responsibility of a single department. Finally, the Department of Labor has a publication program designed to reach those interested in the functioning of the labor market which would bring the labor force statistics and related analytic studies to a wider audience.

At the same time that these considerations were becoming more important, a survey in an entirely different field provided a reassuring example of a continuous survey for which the Bureau of the Census is responsible for the field collection and tabulation of the data but which is planned, analyzed and published by another agency. This is the National Health Survey, for which the Bureau of the Census acts as collection agent for the Public Health Service.

It was therefore determined, by agreement between the Department of Commerce and the Department of Labor, with the concurrence of the Bureau of the Budget, that responsibility for labor force, employment and unemployment statistics should be consolidated in the Department of Labor. The Census Bureau is continuing the collection of the information as part of the Current Population Survey, as agent for the Labor Department. In addition, the Labor Department is given explicit responsibility for the annual supplements on work experience during the preceding year and on multiple jobholding. Supplements on income, migration and school attendance remain a responsibility of the Bureau of the Census as does the planning, analysis and publication of demographic data obtained through the Current Population Survey mechanism. Both the technical and policy committees under the aegis of the Bureau of the Budget continue to function as a means of reflecting government-wide interests in labor force measurement and in the Current Population Survey.

The shift in function did not require any formal reorganization action, since the Labor Department already had sufficient statutory authorization. The transfer of function was accepted by the Congress in passing on the appropriations for the fiscal year 1960 and put into effect in July, 1959. The mechanics of the transfer were worked out smoothly by the two Departments and as far as users of the data are concerned, the only evidence so far of the change in organizational responsibility for the series is in the publication program. The MRLF data are now published monthly in BLS publications. 2/

Although from one point of view, this transfer in functions is perhaps the most notable event in recent years affecting labor force measurement programs, from the viewpoint of the nongovernmental user of the statistics it is of relatively minor import. For those who are interested in what data are available, what improvements in adequacy, reliability, or detail have been made, what additional items of information are presented, a number of other steps have been taken in recent years, of which I should like to mention two briefly, and describe the third at some length.

First, you will recall that the number of households interviewed in the Current Population Survey sample was expanded by two-thirds in 1956 (following an earlier change in the sample design in 1954). At the same time that the number of households was increased, the sample was spread more widely through the country, increasing the number of primary sampling units from 230 to 330 areas. The result has been to reduce the sampling error of the major labor force categories by about 20 per cent. The sample expansion has also permitted showing of more detail in cross-classifications or finer breakdowns, such as the provision of a considerably expanded list of occupations, estimates of labor force rates by regions, and considerably more information on the characteristics of the unemployed.

The second improvement I referred to is the development of seasonal adjustment factors for the major labor force categories. This program is still being actively worked on, in the effort to find improved methods, particularly for the seasonal adjustment factors for unemployment. Meanwhile, users have been provided factors which are of considerable assistance in analyzing changes in labor force data.

Finally, during 1954 and 1955, a general review was undertaken of the programs for collecting labor force, employment and unemployment statistics. Concern over labor force measurement always rises as business activity declines and the recession of 1954 was no exception. The interest in this case was further accentuated by the problems which had been encountered in changing the sample design early in that year. A subcommittee of "The Palmer Committee," called the Review of Concepts Subcommittee, was established by the Bureau of the Budget to review the Federal government's work in the measurement of employment and unemployment and to recommend improvements. I should like to describe the work of this Subcommittee briefly and then indicate where we stand at this date with regard to the recommendations which the Subcommittee made.

The Review of Concepts Subcommittee was composed of personnel drawn from various Federal agencies, those that used as well as those that produced the statistics under review--that is, the Current Population Survey, the current monthly estimates of employment in nonagricultural establishments, the weekly insured unemployment reports and the monthly estimates of farm labor. Charles Stewart, then an Assistant Commissioner of the Bureau of Labor Statistics, was chairman of the Review of Concepts Subcommittee and, in addition to the "producing" agencies, staff of the Council of Economic Advisers and of the Federal Reserve System participated. I served for the Bureau of the Budget, a fact to be kept in mind in considering my comments on the Subcommittee's activities and conclusions.

Our assignment was to review the concepts of the various series, and make such proposals for changes or other improvements which we thought desirable. At the time the Subcommittee was convened, we had been discussing these statistics for a number of years. An earlier subcommittee, also under Mr. Stewart's chairmanship, had reviewed the labor force concepts in 1948, so there seemed little point to spending a great deal of time merely talking among ourselves again. We, therefore, made a determined effort to obtain the views of users of the statistics from outside the Federal government. We invited comments by letter from a large number of business and labor analysts, research organizations, Federal Reserve Banks, State employments security agencies and other organizations and individuals presumed to be using one or more of the series. In addition, open invitations to submit comments were carried in a number of professional journals. After assimilating the comments received, and outlining some tentative proposals, we held an open "hearing" where those interested in presenting their views in person could discuss relevant issues with the Subcommittee. Finally, in October 1955 we issued an "Interim Report". 3/

Opinions may vary as to whether the Subcommittee went far enough in recommending improvements in the labor force and employment series. Nevertheless, there was general agreement with the desirability of most of the recommendations. It, therefore, seems appropriate to summarize these recommendations and note what progress has been made in implementing them.

From this point of view, the Subcommittee's "Interim Report" leaves something to be desired. The various recommendations were not ordered in accordance with any concept of priority - whether of importance or of chronology. Recommendations which could be adopted by immediate administrative action of a single agency are cheek by jowl with recommendations for long-term cooperative research involving two or more agencies, considerable advance planning, and the provision of additional funds. For this reason, the keeping of a numerical scoreboard on the number of recommendations put into effect would be misleading. As a supplement to my paper, I have quoted each recommendation in its summary form, and indicated briefly its current status. 4/ That brief summary cannot do justice to the reasoning on which the recommendations were based, which is explained in the "Interim Report" itself. In the remaining paragraphs of this paper, I shall touch on what seem to me to be the most important points in connection with the labor force survey.

Firstly, the Subcommittee recommended the continuation of each of the series it reviewed using the same basic concepts as had been developed in the past. The Subcommittee thus recognized the unique contributions of the survey of the population in obtaining employment and unemployment estimates based on concepts of labor force activity in a current week; the surveys of establishments in obtaining industrial detail on employment, hours and earnings; and the unemployment insurance records in providing additional information on unemployment weekly with geographic detail. 5/

The Subcommittee then proceeded to outline a proposal for a change in question wording to measure the activity of "looking for a job", a change which it thought should be tested on an experimental basis, as a possible method for improving the measurement of unemployment (leaving the basic concepts unchanged). The experimental work recommended has never been undertaken. It was recognized at the time that it could not be started immediately, since the Census Bureau was in the process of planning its sample expansion and all concerned agreed that experimental tests to change definitions should not be undertaken concurrently. The sample expansion was completed by the summer of 1956. At that time, planning for the 1960 Census of Population which involved a number of experimental field surveys was getting underway, and little more than a year later the CPS series indicated the beginning of the recent recession. Although small-scale experimental surveys might possibly have been undertaken at that time, a tryout of new questions on a national scale was not considered advisable, lest the changed definitions influence the regular series which was being so closely followed to watch the course of the recession. For this mixture of economic and administrative reasons, therefore, this recommendation of the Review of Concepts Subcommittee has not been implemented.

Meanwhile, there has been put into effect the recommendation which met with almost universal support of the users contacted by the Subcommittee, the proposal to classify persons on temporary layoff or persons waiting to start a new job as unemployed rather than as employed in the "with a job but not at work" category. The questionnaire was redesigned so that separate information could continue to be reported for these groups included among the unemployed, and the overall employment and unemployment estimates were readjusted to account for this change in concepts back to 1947.

Other recommendations with regard to the Current Population Survey concerned primarily additional information which the Subcommittee thought it desirable to have. As a result, whether any of the time off was being paid for, of persons reported as "with a job but not at work", is now asked monthly and the information shown by reason for time off. Thus, for nonfarm wage and salary workers, we know that 75-80 percent of vacation time off is on paid leave, while about 35-40 percent of persons on sick leave receive pay. This information is of interest in itself; it is also useful in making seasonal comparisons in movements between the labor force estimates of employed persons and the reports of employees on the payrolls of nonagricultural establishments, since the former includes persons on leave, whether or not they are paid for the time off, and the latter includes only those who are paid for the time off.

The Subcommittee also recommended that more detailed and more regular information be obtained on multiple jobholding. A survey has been undertaken each year since. In this case also the information is desired both for the light it throws on labor force activity and because it assists in understanding of the differences between the population-type and the establishment-type of employment estimates under varying business conditions. Thus a number of recommendations for the regular collection of additional facts about the labor force have been accepted.

Finally, the Subcommittee proposed that a number of occasional supplementary surveys or special research studies should be undertaken in connection with the labor force enumeration, mentioning in particular surveys of "potential workers who would be in the labor force and looking for work under specified conditions, with special attention to persons who have dropped out of the labor force because of discouragement, illness, etc." and also suggesting a number of additional characteristics and facts concerning the unemployed which could be studied in more detail in occasional surveys than is possible in a repetitive monthly collection of data.

To date, little has been accomplished in developing these special research projects as a result of the Subcommittee's report. In my view, this is the most important area for future development of the labor force measurement program. Such work should, in my opinion, take precedence over the proposal to re-examine the unemployment question wording. My reasons are simple. It is unlikely that any feasible change in the monthly unemployment questioning would identify adequately persons who leave the labor force for long periods because they become discouraged and discontinue a search for work. Under normal conditions this may not be a national problem of any great magnitude. But on a local basis, in depressed areas, where employment opportunities may be severely limited even in good times, the conventional measurement of the labor force may be misleading without supplementary information on the work and job-hunting experience and possibly other characteristics of persons not currently in the labor force.

Experimental work to develop survey plans, followed by the accumulation of data for a number of depressed areas in accordance with these plans, should, in my view, be the next major goal of the labor force survey system. Such information will be of particular importance if any national policy is to be established or action is to be undertaken with respect to individual depressed areas. Such information should also add to our store of knowledge on the labor force activity of various groups in the population, possibly provide insights on the labor force motivation of marginal workers and thus help interpret changes in labor force participation rates.

FOOTNOTES:

1/ The views expressed are those of the author, not the official position of the Bureau of the Budget.

2/ See "Monthly Report on the Labor Force" and "Employment and Earnings".

3/ This report was reprinted in full in Employment and Unemployment Statistics, Hearings before the Subcommittee on Economic Statistics, Joint Committee on the Economic Report, November 1955, pp. 6-24.

4/ Copies of this supplement will be supplied upon request to the author as long as the supply lasts.

5/ For a discussion of the reasons why more than one series is needed, and of the effect of differences in concepts, see the testimony of Mr. Raymond T. Bowman, Assistant Director of the Bureau of the Budget, before the Subcommittee on Economic Statistics, op. cit., pp. 24-36.

SOME ASPECTS OF UNEMPLOYMENT CHANGE

By: Seymour L. Wolfbein, U. S. Department of Labor

The past several years have witnessed a substantial growth of interest in and analysis of the complex of factors which generate a given level of unemployment. As a result, there has been a considerable amount of work in disaggregating the total unemployment estimate for any period of time and, in turn, focusing on such matters as the composition of the unemployed, the duration of unemployment, spells of unemployment, types of unemployment (frictional, seasonal, structural), etc. 1/

While by no means a new phenomenon, this emphasis on the disaggregative approach reflects

(a) the greater availability of data on the component parts of the unemployed,

(b) the growing recognition of the importance of tracking current changes in the demographic and economic characteristics of the unemployed and not just the global total of unemployment; that is, the understanding that for any given time period, meaningful perception of what is happening on the current scene as well as in the longer run is derived from an examination of the anatomy in addition to the overall total of unemployment,

(c) the realization that in terms of viable programs for action in this area, answers to such questions as "Who are the unemployed?" and "Where are the unemployed?" and "How persistent is unemployment?" are perhaps the critical ones.

Among the various facets of the disaggregative approach, one which again is by no means new but which is receiving increased attention is the matter of the gross changes which are involved in a given net change in labor force, employment or in unemployment. For example, in April 1958 when unemployment was still on the increase, the Monthly Report on the Labor Force reported as follows:

		(Millions)
Unemployed in March 1958		5.2
Found employment in April	-1.5	
Left labor force in April	- .6	
Reductions in unemployment March-April 1958		-2.1
Remaining unemployed March-April 1958		3.1
Lost employment in April	+1.2	
Entered labor force and unemployed in April	+ .8	
Additions to unemployment March-April 1958		+2.0
Unemployed in April 1958		5.1

One year later (April 1959) when unemployment was on the decline, the MRLF reported as follows:

		(Millions)
Unemployed in March 1959		4.4
Found employment in April	-1.6	
Left labor force in April	- .7	
Reductions in unemployment March-April 1959		-2.3
Remaining unemployed March-April 1959		2.2
Lost employment in April	+ .8	
Entered labor force and unemployed in April	+ .7	
Additions to unemployment March-April 1959		+1.5
Unemployed in April 1959		3.6

The two examples are illustrative of the fact that any given month's unemployment change is apparently a compound of a very substantial amount of gross flows into and out of the labor force and employment; that even during a month of increasing unemployment a very substantial amount and proportion of the unemployed did find jobs; that, on the other hand, during a month of declining unemployment, a very substantial amount and proportion of the unemployed that month were made up of persons who had just lost jobs, etc.

What can perhaps be best described as an exploratory foray into the field of gross change in relation to unemployment is the subject of this paper, which examines the record of one cycle (1957-58-59) and asks the following questions, to which some answers, or at best hypotheses, will be attempted:

1. What is the actual amount--in absolute as well as relative terms--of gross change which finally sifts down into a given month's unemployment level?

2. What is the relative importance, in the midst of all this turmoil, of the people who remain unemployed over the month?

3. Where do those who exit from unemployment go? To jobs? Out of the labor force?

4. From where do those who newly enter unemployment come? From jobs? From out of the labor force?

Cutting across these four questions, we will also attempt to discern--on the basis of an avowedly short span of time and data--any relevant seasonal and cyclical differences.

1. The total number of separate individuals involved in any one month's unemployment figure falls into three groups: a) the stayers--those remaining unemployed from last month; b) the exits--those leaving unemployment by virtue of finding jobs or leaving the labor force; and c) the entries--those newly unemployed by virtue of losing jobs or coming into the ranks of the unemployed directly from outside the labor force.

Adding all three and dividing by the current month's unemployment yields a figure representing the rate of gross change, i.e., the relationship between the total number of persons involved and the resulting net unemployment figure, thus:

$$\frac{\text{Stayers+entries+exits}}{\text{Current month's unempl.}} = \text{Rate of gross change in unemployment}$$

The monthly rates for the period 1957 through 1959 are shown in Table 1 and Chart 1.

The following observations seem to be warranted by the data:

A. The rate of gross change in unemployment is very high--approximately close to 150% during the period under observation. To put it another way: Every 100 persons reported as unemployed during a given month represented the net effect of about 150 persons entering, leaving and staying unemployed, with the entries and exits further compounded by movements into and out of employment, into and out of the labor force.

B. There is an apparent cyclical variation in the movement of the rates of gross change in unemployment--for the total group as well as for men and women separately. As can be seen from Chart 1, during the winter and spring of 1958 when unemployment was moving up, these rates declined perceptibly; during the winter and spring of 1959 when unemployment was turning down, these rates moved up rather sharply; throughout the recession year of 1958 the rates of gross change were below either 1957 or 1959, representing a trough between these two adjacent years.

Our hypothesis in this case would state: Rates of gross change in unemployment vary inversely with the business cycle. To put it another way: The dynamism and change which go to make up a given month's unemployment level tend to slow down somewhat in the business downturn, to accelerate during recovery; the rate of moving around in and out of employment, in and out of the labor force ebbs during a business downturn, rises in recovery.

C. The brief period covered, including as it did rather sharp cyclical change as well as the effects of the steel strike in latter 1959, may have masked a number of seasonal patterns in the rate of gross change in unemployment. One--and the expected one--does stand out: the fairly regular increase in the rate during spring to the peak figure in June each year as students enter the labor force.

D. The differences between the two sexes were also very much in accord with *a priori* expectations, with gross change consistently higher among the women. The difference was substantially widened during the trough of the recession, mostly because of the much sharper drop in the rate of gross change in unemployment among the men than among the women, for reasons which may become apparent from the data presented below.

E. So far, our discussion has been confined to the rates of gross change in unemployment. The absolutes are of some interest and significance, too. Their sheer volume is impressive. In calendar 1957 there were about 18 million separate entries into unemployment status and a somewhat smaller number of separate exits from unemployment. (These, of course, are not necessarily different persons.) In the recession year of 1958, both entries and exits moved up substantially (although relatively less than total unemployment, so that the rates declined as we have already pointed out) and the 1959 figures are also running well ahead of 1957.

Measuring the impact of employment and unemployment change is done in a number of different ways: we, for example, calculate the total volume of unemployment, the proportion it represents of the labor force, the duration of unemployment, spells of unemployment, etc. The data described so far represent still another related dimension. Again, in terms of sheer impact, the fact that there were about 24 million new unemployment experiences during 1958--or even 18 million during the relatively much better year of 1957--is a matter to be reckoned with. By the same token, so is the number of exits out of unemployment--not only in relation to their economic consequences, but to some of the institutional settings which are supposed to deal with this kind of turnover, e.g. the local employment service office or the personnel offices of business and industry, as well as to some of the social consequences, e.g. in attitudes toward unemployment under conditions of substantial turnover.

2. The data cited so far lend considerable weight to the concept of the unemployed as a substantially volatile group. Nevertheless, unemployment does persist among various groups and various areas, and we come now briefly to the question of how much unemployment carries over from one month to the next. What, then, is the proportion of a current month's unemployment which is left over from a previous month (the "stayers" in our terminology)? Table 2 and Chart 2 speak to this point.

A. On the overall, somewhere between about 50 and 60 percent of a month's unemployed were carry-overs from the previous month during the three years under observation. As might be expected from the previous data, women, who have a higher rate of gross change in unemployment, tend to have a smaller proportion of "stayers" in unemployment status in any two adjacent months.

B. Again, it is difficult to assess the presence of seasonal patterns within such a small span of time; but, as expected, the proportion of stayers moves down very perceptibly in June when the students move into the labor force.

C. The change over the cycle is rather clear. With the duration of unemployment moving up significantly during 1958 and early 1959, the proportion still unemployed from the previous month also moved up. This was particularly true among the men. And just as the sex differential in rate of gross change widened during the business downturn, so did the sex differential in the proportion of persons carrying over in unemployment status from one month to the next. This period starts out rather sharply in the middle tier of Chart 2.

Thus, both the level and amplitude of change in the variable under observation in this section was larger for the men than the women. For the men, at least, the change in direction of movement was rather marked at both turns of the cycle.

3. If, on the overall, somewhere between 50 and 60% of a month's unemployed are left over from the previous month, then somewhere between 50 and 40% are newly arrived in unemployment status that month. What proportion of these, in turn, represent persons who have lost their jobs? What proportion have entered unemployment directly from outside the labor force? Table 3 and Chart 3 present the evidence on this score.

A. The difference between the two sexes in this regard is so marked as to warrant quite a different set of observations for each. Among the men, the great majority (about three-fourths) of the new entries into unemployment during a given month are employed the month before.

B. There is a marked seasonal pattern here: The proportion of men moving into unemployment from employment is at a peak (about four-fifths) in January as men exit from seasonal December jobs; the proportion then falls rather regularly to its expected low in June when students come into the labor force and then rises again into the fall and winter when men leave employment from such outdoor activities as agriculture and construction.

C. Among the women, on the other hand, the great majority (about three-fifths) of the new entries into unemployment during a given month were not in the labor force at all during the previous month.

D. As is true among the men, the proportion of the newly unemployed women who came from employment status the month previous reaches a peak in January (the only time the figure is over 50%) after exits from seasonal December jobs, especially in trade; the low point here is also in June when students enter the picture. There are apparently also, however, upturns in the proportion of a given month's jobless among the women who come from the employed rather than from outside the labor force in the spring (April-May) and fall (October). The spring phenomenon may be due to temporary Easter employment, the October one to perhaps the end of the season in such nondurable manufacturing activities as food processing.

E. There is no discernible major cyclical change among either the men or women, although the peak in the winter of 1957-58 is worthy of mention. The unemployed doubled in numbers between October 1957 and February 1958; just between December 1957 and January 1958, unemployment rose from 3.4 million to 4.5 million, an increase of one-third; for that month (January 1958) the proportion of new unemployed among men coming from employment the month previous was not far from 90 percent.

The proportions of the newly-unemployed arriving from a job that was lost (or given up) as against those coming from outside the labor force may be of substantial significance. The disemployed represent persons with work experience, perhaps a different kind of income status, certainly a better chance that there will be a claim for unemployment insurance. This, of course, is not to denigrate the importance of those coming directly from outside the labor force, e.g. the young graduate in search of his or her first full-time job. At any rate, seasonal movements aside, a marked change in these proportions may be worthy of the analyst's attention, both in terms of an assessment of current developments as well as a prognosis for the future.

4. As indicated previously, the large monthly number of entries into the status of unemployment are accompanied by an equally large number of exits from unemployment. Thus, during 1957 there were an average of about $1\frac{1}{2}$ million entries and $1\frac{1}{2}$ million exits from unemployment each month and about 2 million entries and 2 million exits from unemployment each month in 1958. Table 4 and Chart 4 show the information on what proportion of those who left unemployment found jobs and what proportion left the labor force altogether.

A. Here again, the differences between men and women are very marked. The preponderant majority of the men (anywhere from 70 to 80%) move from unemployment into a job; the corresponding ratio among the women is about 50%. Thus, women again represent the major stream of movement into and out of the labor force, into and out of unemployment.

B. The seasonal patterns of movement seem to be fairly regularly defined. Among the men there is a drop in the proportion going from unemployment to employment (and thus an increase in those going from unemployment to outside the labor force) during the winter months when many outdoor employment activities ebb. The same is true during the summer months, representing the well known phenomenon of students leaving the labor force after a short period of unemployment between school terms. The reverse movement (an increase in the proportion among men going from unemployment to employment) is seen regularly in the spring with the increase in outdoor employment opportunities.

C. The seasonal pattern among women is somewhat less regular and less marked in amplitude, but not too far substantively from the one among the men. Generally, the proportions moving from unemployment to employment are lowest in the winter months, highest in the fall, with the summer showing the usual effects of the pattern of student entry and exit.

D. As was true in the case of movements into unemployment, there is no major discernible change over the cycle in the pattern of exits from unemployment into jobs or into nonworker status. Unlike the case of entries into unemployment, however, whatever change over the cycle took place is apparently more notable among the women rather than the men. As can be seen from Chart 4, there was some dampening of the proportion among women moving from unemployment into employment (with a somewhat higher proportion going outside the labor force altogether) during the recession year of 1958.

On the basis of this brief, exploratory analysis of gross change in unemployment which took place during the years 1957-58-59, the following conclusions seem warranted.

A. The phenomenon of gross change (in employment and labor force as well as in unemployment) first became observable in an objective, quantified manner through the operation of the Monthly Report on the Labor Force and the continuity inherent in its sampling procedures as well as in its conceptual structure. Even the cursory examination of the phenomenon in this paper shows it to be an important factor in the analysis of current developments, and perhaps even strategic at certain stages of the employment and unemployment situation.

B. Although perhaps more complicated than some of the other variables reported and analyzed each month, the importance of these data warrants a recommendation that they be presented on a regular, sustained basis and used in the analysis of current, as well as trend information on the employment situation. This presupposes the resolution of the technical problem involved in estimating gross changes from only that part of the sample which can be matched for two consecutive months.

C. The analyst always presses for more information and for more analytical time, and this paper is no exception. Even within the confines of this presentation it would have been very valuable to know, for example, the age differentials in gross change (is it the young who are the most volatile group? Is it the men in the central age groups, however, who become more affected during a business downturn? Is it the older men who move in greater proportions from unemployment outside the labor force altogether (labor force disappearance)?) Similarly, another important dimension of gross change must be in the occupational and industrial (at least farm vs. nonfarm) differentials. Where, in the occupational ladder, does most of the gross change take place? Is it the unskilled? Or perhaps the professional personnel who are so much more mobile? Do these patterns change with the cycle?

One final word: The forces of gross change, as illustrated in this paper, are part of the overall picture of the great mobility of the American labor force. It is a pleasure to present this paper under the chairmanship of Dr. Gladys Palmer, who for such a long time has urged us on to research efforts in this field and who has, herself, conducted some of the most perceptive work in this important aspect of the dynamics of the labor force.

This acknowledges the assistance of Robert L. Stein of the Bureau of Labor Statistics in organizing the basic data and thinking through some of the concepts of gross change.

1/ cf. e.g., "Who Are the Unemployed?" U.S. Dept. of Labor June 1958; "The Unemployed Spring 1959" U.S. Dept. of Labor May 1959; "The Extent and Nature of Frictional Unemployment" by Bureau of Labor Statistics (Study Paper No. 6 of Joint Economic Committee, Congress of United States) 1959; "Unemployment in America," address by Under Secretary of Labor James T. O'Connell, Newark, N. J., May 1, 1959.

TABLE 1
Rate of Gross Change in Unemployment
By month 1957-1959 By sex

Month	Total			Male			Female		
	1957	1958	1959	1957	1958	1959	1957	1958	1959
J	145%	149%	143%	143%	145%	138%	147%	159%	153%
F	145	142	138	141	140	135	152	152	145
M	145	138	137	141	135	132	155	146	147
A	146	137	143	142	132	138	153	149	151
M	152	141	143	147	137	141	160	149	147
J	159	146	155	155	141	153	164	155	158
J	149	140	153	147	136	150	153	147	151
A	154	141	149	151	137	147	159	149	154
S	156	139	149	157	134	145	155	148	155
O	151	143	154	149	139	152	153	151	155
N	156	143		153	139		160	150	
D	148	142		145	139		156	148	

Rate of gross change: Total number of persons entering, leaving and remaining unemployed divided by number of unemployed in each month.
Source: Monthly Report on the Labor Force

TABLE 2
Persons Still Unemployed from Previous Month
As a Percent of Current Month's Unemployment
By month 1957-1959 By sex

Month	Total			Male			Female		
	1957	1958	1959	1957	1958	1959	1957	1958	1959
J	55.5	50.6	57.5	56.8	54.6	61.9	52.7	41.2	47.4
F	55.3	56.5	61.9	58.7	60.4	64.9	48.3	47.6	54.7
M	54.8	62.2	63.4	59.4	65.5	68.0	45.1	53.9	53.4
A	54.3	62.7	57.4	58.1	67.9	61.9	46.7	51.4	49.4
M	47.6	58.9	56.6	52.6	63.0	59.1	39.6	50.6	52.5
J	41.2	53.8	44.8	44.8	58.8	46.9	35.5	44.6	41.6
J	50.5	60.1	47.4	42.7	63.6	49.8	47.3	53.1	43.3
A	45.8	58.9	50.7	48.7	63.3	53.3	41.3	50.6	46.4
S	43.9	60.8	51.0	43.3	66.1	54.8	44.8	51.6	44.7
O	49.4	56.6	46.5	50.6	61.1	47.6	47.5	48.6	44.6
N	44.5	57.2		46.8	60.9		40.3	50.2	
D	52.0	58.3		55.2	61.0		44.3	52.0	

TABLE 3
Additions to Unemployment in the U. S.:
% Entering Unemployment from Employment
By month 1957-1959 By sex

Month	Total			Male			Female		
	1957	1958	1959	1957	1958	1959	1957	1958	1959
J	70.0	74.2	69.8	79.5	86.7	81.7	53.0	51.9	50.2
F	63.4	64.2	65.9	74.3	76.3	78.8	45.7	42.8	41.9
M	58.4	63.8	56.3	69.1	74.7	72.6	42.0	42.9	32.3
A	63.2	61.0	53.3	73.6	71.9	66.7	46.4	45.5	35.4
M	56.6	57.6	56.6	66.4	68.0	68.2	44.5	41.9	40.5
J	47.3	48.1	44.2	52.9	57.2	52.2	39.5	35.5	33.2
J	61.5	59.3	58.3	74.6	71.8	70.2	44.0	40.1	41.2
A	63.9	61.3	65.1	79.2	75.7	80.6	42.9	40.8	42.8
S	63.0	60.0	59.9	76.8	78.3	78.2	40.4	37.6	34.8
O	65.7	61.9	62.0	76.8	73.4	78.4	47.5	46.0	37.5
N	64.4	62.2		78.4	79.7		42.0	36.4	
D	71.0	68.9		82.0	80.0		49.5	47.3	

The difference between the percentage shown in this table and 100% represents proportion entering unemployment from outside the labor force.

TABLE 4
Exits from Unemployment in the U. S.:
% Leaving Unemployment for Employment
By month 1957-1959 By sex

Month	Total			Male			Female		
	1957	1958	1959	1957	1958	1959	1957	1958	1959
J	51.6	56.6	62.4	62.6	68.2	71.1	35.3	38.0	48.1
F	68.3	64.4	63.0	78.2	73.3	75.3	54.5	51.0	45.0
M	67.2	67.2	68.3	79.8	80.0	78.5	46.8	47.1	47.9
A	66.8	71.4	71.0	76.0	80.9	83.0	52.3	51.6	46.6
M	70.8	71.5	64.2	82.1	84.1	78.1	51.6	52.0	43.4
J	69.3	67.1	69.7	80.0	78.4	81.8	55.4	49.6	53.0
J	66.5	64.1	65.8	73.7	75.8	72.1	55.8	48.4	57.7
A	60.5	65.7	60.9	69.8	77.2	70.1	49.5	48.7	48.3
S	68.5	66.1	66.2	75.7	75.0	73.6	58.4	54.0	56.8
O	70.5	71.6	69.3	82.1	84.6	80.9	55.9	54.8	52.0
N	62.0	61.6		74.6	73.4		44.9	46.8	
D	59.8	62.8		70.1	75.7		48.9	48.0	

The difference between the percentage shown in this table and 100% represents the proportion leaving unemployment to go outside the labor force.

CHART 1

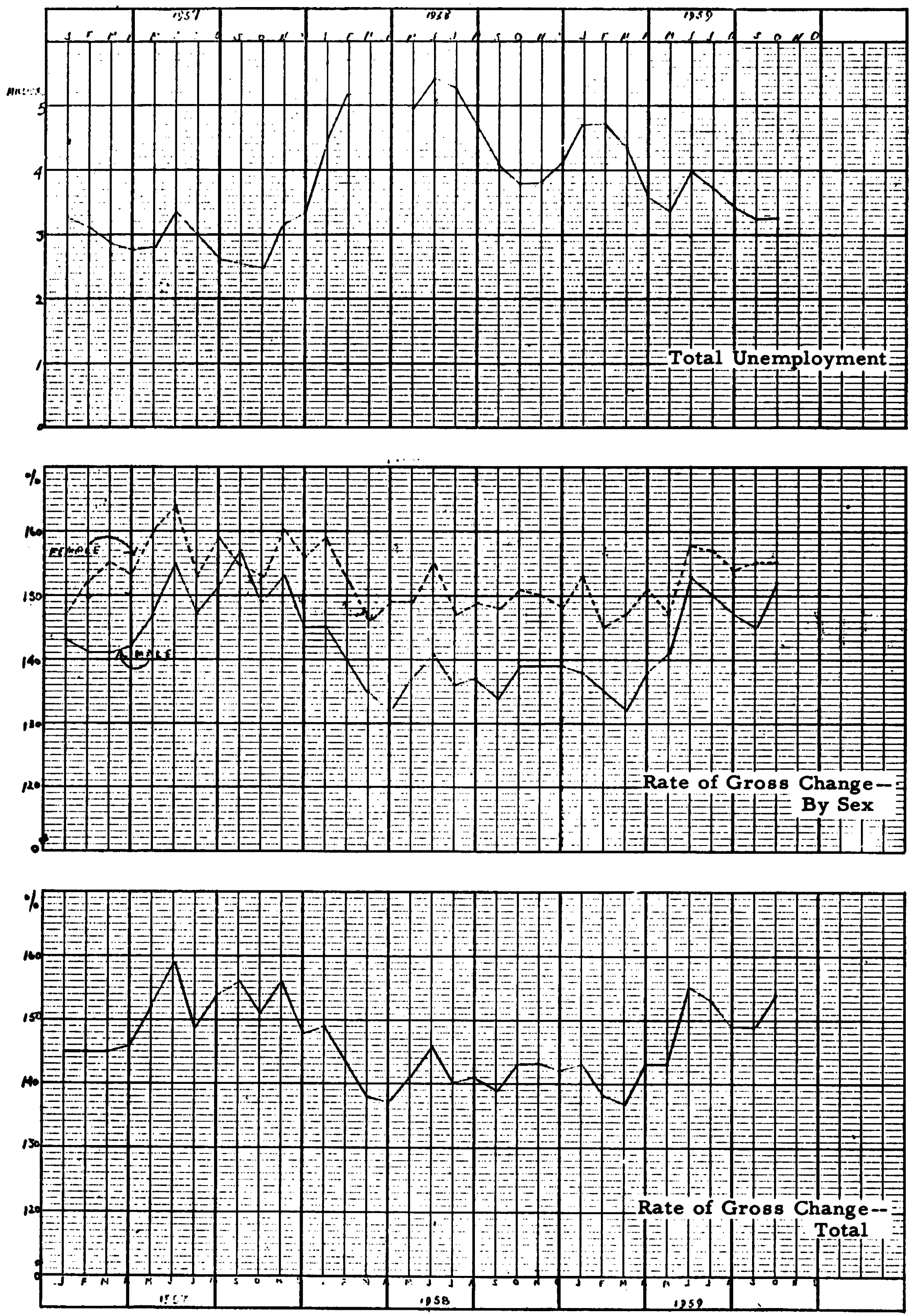

1957
1958
1959
Total Unemployment
Rate of Gross Change--
By Sex
Rate of Gross Change--
Total

CHART 2

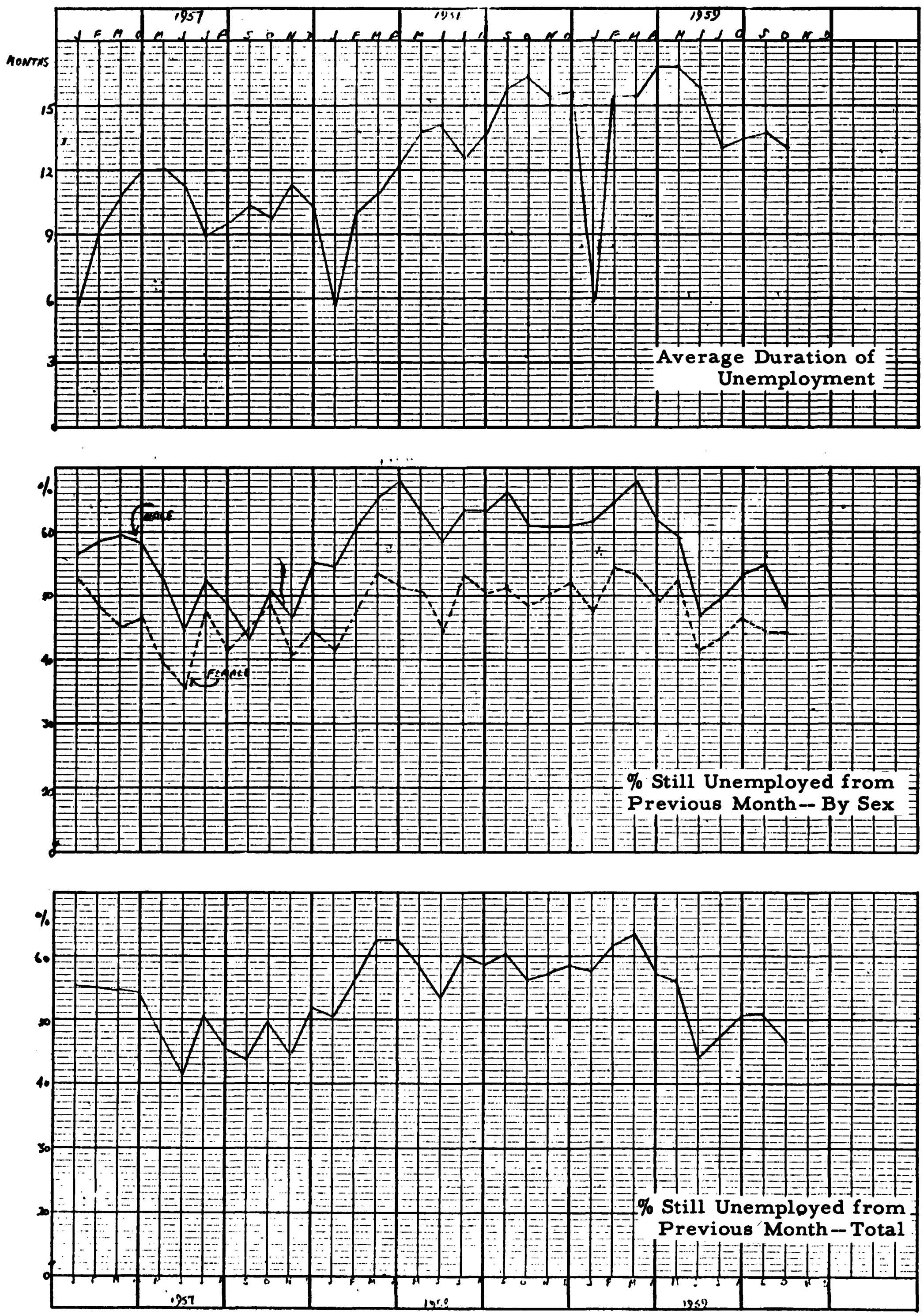

1957
1958
1959
MONTHS
15
12
9
6
3
Average Duration of Unemployment
60
50
40
30
20
% Still Unemployed from Previous Month-- By Sex
60
50
40
30
20
% Still Unemployed from Previous Month—Total

CHART 3

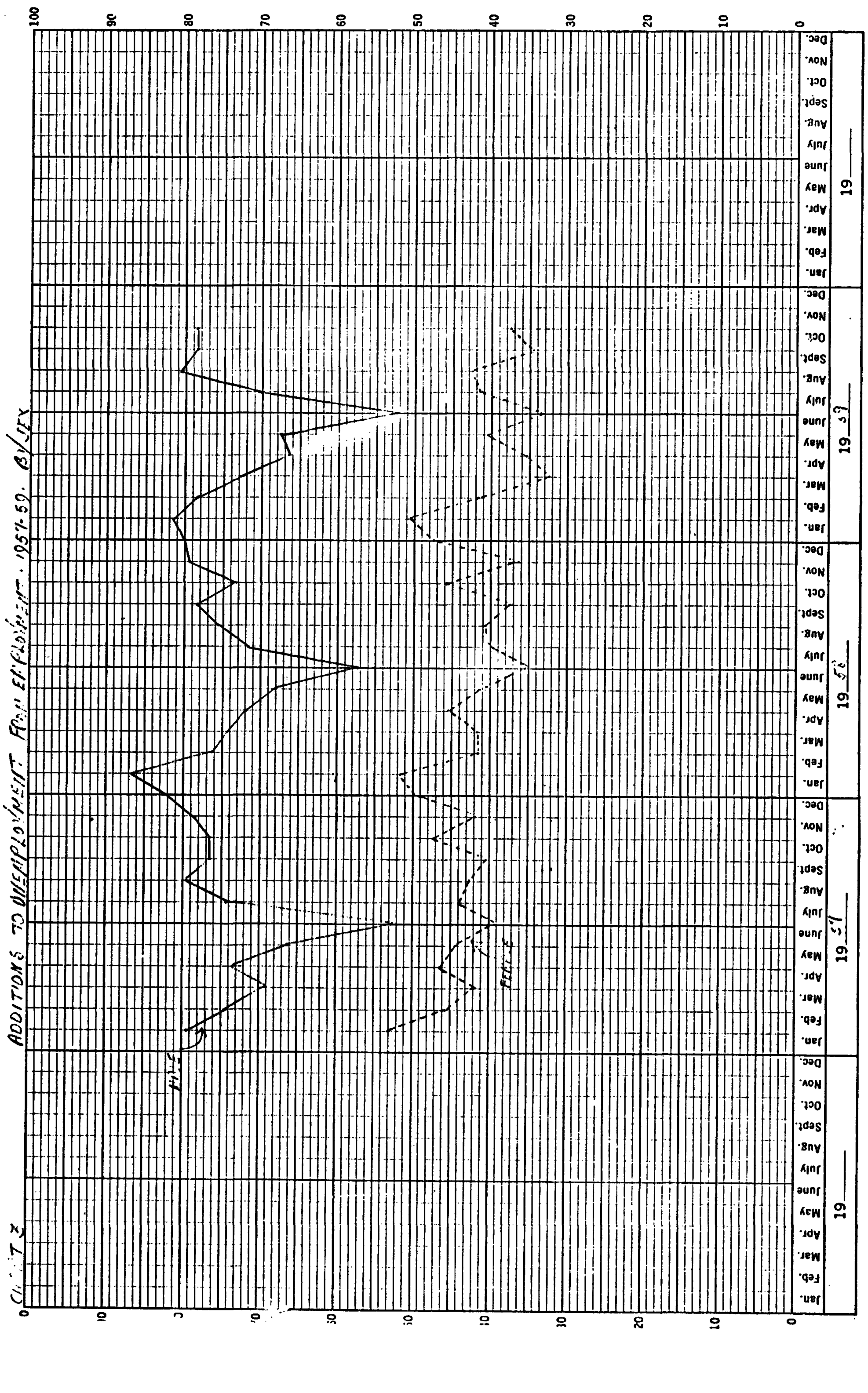


CHART 4

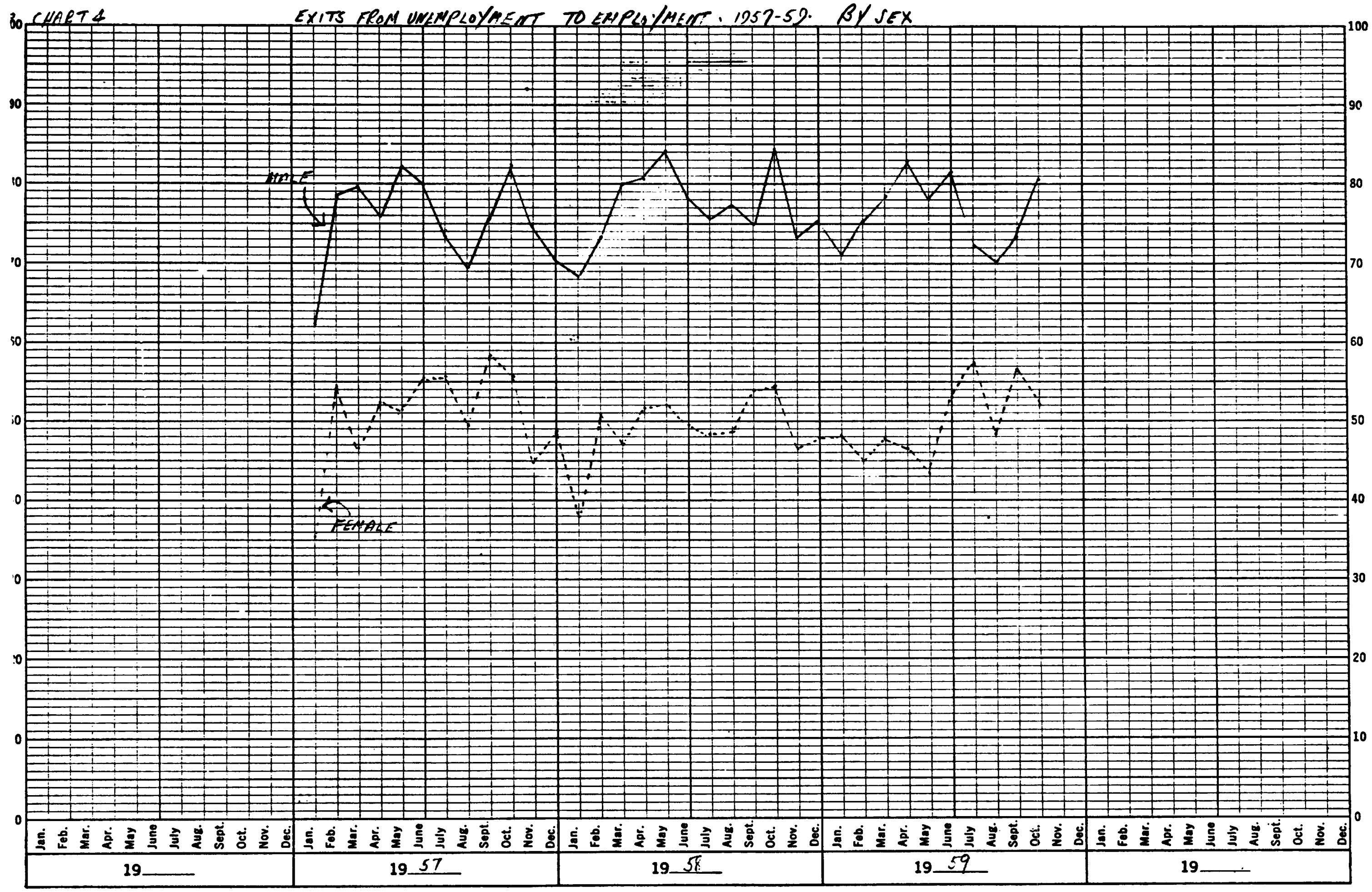


CHART 4
EXITS FROM UNEMPLOYMENT TO EMPLOYMENT · 1957-59. BY SEX
MALE
FEMALE
100
90
80
70
60
50
40
30
20
10
0
Jan. Feb. Mar. Apr. May June July Aug. Sept. Oct. Nov. Dec.
19__
19 57
19 58
19 59
19__

FACTORS IN LABOR FORCE GROWTH

By: Gertrude Bancroft, U.S. Department of Labor

Twenty years ago, in December 1939, the field staff of the Works Progress Administration's Division of Social Research conducted the first run-through of the first monthly survey of a sample of the population preparatory to the official enumeration in April 1940. The purpose of this operation was to compile a direct and objective measure of unemployment. Our Chairman, Gladys Palmer, was a member of the advisory group that watched over the survey in its embryo and infancy stages, and is the only person who has remained in that advisory role during all of the years since. She has seen the survey grow from a 41-county sample to a sample of 330 areas, extending into 638 counties. In the beginning, about 15,000 households were interviewed; now the number is 35,000. Over the years, the growth in the output of the survey in terms of information on labor force and population behaviour and characteristics has been even more spectacular. During these two decades, Miss Palmer has provided the leadership for the discussions which produced developments in concept and measurement techniques, bringing to each problem a fresh eye and a rich knowledge of real life outside the Washinton statistical offices. At the same time, she has firmly pressed for preserving the degree of continuity needed to measure important changes. A great deal of what we know about the labor force today can be credited to her stimulating questions and her unflagging interest, in her role as chairman of the interagency Committee on Labor Supply, Employment and Unemployment.

My purpose in this paper is to note briefly some of the things we have learned over these years from the survey and from two decennial censuses about the factors affecting labor force growth and to raise a question about the meaning of its rate of growth. Of course, the primary factor in the growth of the labor force is the changing number in the population of working age and the changing age-sex composition of the population. Within that frame, what determines the proportions of various types of people who will be in the labor force?

Looking again into the past, I reminded myself that some 13 years ago, a paper by Wolfbein and Jaffe was read at the annual meeting of the American Sociological Society, entitled "Demographic Factors in Labor Force Growth". Comparing gainful worker rates in 1890 and 1930, and using a simple standardization procedure, the authors concluded that changing labor force rates could not be fully explained by changes in the composition of the population—i.e., changes in age, sex, color, nativity, and for women, marital status. A similar comparison for 1940 and 1946 led to the same conclusion. There are, in fact, changing propensities to work whose true causes, the authors speculated, must be sought in "nondemographic" or "socio-economic" factors—social acceptance of women in employment outside the home, the growing importance of white-collar jobs, changes in technology opening factory jobs to women, the changing role of women, the transfer of many household tasks into the commercial field, and so forth. For men, the decreases in worker rates, the authors believed, were the result of longer schooling and earlier retirement.

This pioneering work was further extended by John Durand in his 1948 study. Using a slightly more refined technique, and also taking account of the shifts of the population from farm to nonfarm residences, he too measured the effects of basic demographic changes, and concluded that nondemographic factors were dominant in explaining changing labor force rates between 1920 and 1940. I reached the same conclusion for the period 1940 to 1950 for white women, whose labor force rates, of course, rose, and for nonwhite men, whose labor force rates declined. For white men, labor force rates were almost unchanged during the decade. Nonwhite women would have decreased their labor force participation had they not moved from farm to nonfarm areas in such numbers. Clarence Long, in his very detailed study published last year, made many similar calculations for the United States and other countries.

Some of the other demographic factors affecting the labor force growth that have been measured or discovered will be mentioned briefly.

Whether or not a woman is married and has children is not nearly so important in determining her labor force status as is the age of her children. Women under 45 who are, or have been, married and who have no children under age 6 are two or three times as likely to be in the labor force, age for age, as are women with preschool children. This is true even though the majority of married women under 45 with no preschool children have children of school age. Accordingly, it is not enough to examine the changing proportions of married women or of married women with children in order to predict or explain labor force growth in the age groups 20 to 35 or 40.

For this reason, age at marriage is another factor that helps to determine the size of the female labor force. Where age at marriage is comparatively late, the young female labor force will be large, and will change in relation to the size of the population in the age-group between school-leaving and marriage. Where age at marriage is early, the female labor force will be relatively small in the late teen years and early twenties. However, young brides are also likely to be young mothers and to complete their families before they are out of their twenties. This means that those who want to return to the labor force before they are 35 may do so with comparative ease. They are still below an age where age itself is a criterion for hiring. They are young enough to be thought worth training and not too far away from their school days.

A bit of evidence on this point is the comparative labor force rates in 1950 of urban married women over 35 in the Western States and in the Northeastern States. Age at marriage is earlier for women in the West than in the Northeast. In the age group 35 and over, 25 percent of the Western married women were in the labor force in 1950, but only 20 percent in the Northeast. Later data are not available by age but 1959 data show that urban married women of all ages are somewhat more likely to be in the labor force in the West than the Northeast.

Potentially, women in their late twenties and thirties have a long working life ahead. Of course, those who married at 19 or 20 probably did so before they completed or even entered college, and so may be unsuited for many professional or technical jobs, where so much of the expansion in the next ten years is expected. But in other occupations their prospects are brighter. On balance, under present employment patterns, early marriage by itself probably fosters growth in the number of women in the labor force because it frees for employment outside the home, the large number past 30 or 35 years. At the present time, the labor force participation rates of married women are considerably higher in the twenty-year age span, 35 to 54, than in any of the younger groups. For 1958, the average labor force rate for married women 20 to 24 years was 32 percent, for 35 to 44 years, 37 percent, and 45 to 54 years, 40 percent.

If, however, early marriage also means larger completed families, which just now appears to be true, then the probability of a young married woman's return to the labor force on a more or less continuous basis may be reduced. It is not necessary to point out that the demands on a mother to stay at home and function as nurse, chauffeur, teacher, den mother, and so forth, increase with the number of children she has.

The movement of the population from farms to cities has had its effect on the size of the labor force, as many have pointed out. This type of migration tends to reduce the proportion of men in the labor force but to increase the proportion of women. Other things being equal, in the future off-farm migration will continue to have its effect on labor force growth, particularly for the nonwhite population, although with the forthcoming change in the definition of the farm population, in connection with the 1960 Census, it will be difficult for aw ile to measure its effect. What about the other form of migration that each of us is conscious of every day—the movement from central cities to suburbia or exurbia? Current estimates of the Census Bureau show that the greatest relative increases in population since 1950 have been in Standard Metropolitan Areas outside central cities and in portions classified as rural in 1950, now probably urban.

We have no current data on the labor force participation rates of the population in central cities as compared with the outlying parts of Standard Metropolitan Areas. But we know that in 1950 the rates for women living in the fringes of urbanized areas—that is, the contiguous built-up areas surrounding cities of 50,000 or more and sometimes called suburban—were 5 or 6 percentage points below the rates for those living within the city limits. Suburban married women with young children are not very likely to be wage earners, but their labor force rates in 1950 were not different from those of mothers of young children in the central cities—despite differences in income, ethnic stock or race, education, etc. Suburban married women without young children had labor force rates well below those of women living in central cities except in the age group 45 to 64 where they were about the same. We shall know much more about this after the 1960 Census is completed but it looks at this point as if the movement of the population away from large cities to their suburbs would have a depressing effect on the labor force activity of younger married women. Job opportunities, however, are moving out too and we may find greater similarity in the future between urban and suburban women.

For men, there is little difference between central cities and fringe areas, except for the younger group. Those under 25 living in central cities had labor force rates about two percentage points higher than those in the suburbs. Again, the movement to the suburbs may have a slightly dampening effect on the growth of the male labor force.

Another factor that has been measured in recent years is the impact of increasing high school and college enrollment on labor force growth. Young people of high school and college age are two to three times as likely to be in the labor force if they have left school. Yet the proportion continuing in school has grown prodigiously since 1950—from 40 to 50 percent of the total—and is expected to continue to rise. Except for the boom year 1956, the last decade has seen a decline in labor force rates of teenagers and almost no change for young people 20 to 24 years. In the latter age group, there has been an increase in student employment but the rising rate of school and college attendance has kept down the overall rate. Relatively few students, of course, work full time. Thus, in the last 10 years or so, the number of workers supplied by the age classes under 25 has declined, relatively speaking, and their output in terms of hours worked has also diminished.

The advancing educational level of the population may also be a factor favoring labor force growth, particularly among women. Between 1940 and 1950 the greatest increases in labor force activity were for high school graduates and women with some college training, after allowance is made for increasing marriage rates and childbearing in this group. This tendency for the better educated women to enter or return to the labor force in relatively greater numbers was also seen in the big expansion of the 1950's, particularly for middle aged or older women. There is some evidence from the latest data, however, that the labor force rates may have leveled off, at least temporarily, at the upper end of the educational ladder. Even if there are no further increases in the labor force rates for women with high school education or better, the growing number of these in the population would point to some expansion, if there are no offsetting developments.

As far as men are concerned, there have been no marked changes among the better-educated, but those with minimum amounts of schooling are less likely to be labor force members than in former years. This may reflect the continuing decline in in the labor force participation of nonwhite men that has been observed for more than half a century. It may also reflect the reductions in the labor force participation of elderly men whose average educational attainment is below that of the younger population.

The great flexibility of our labor force is a characteristic that was not revealed until we had a repetitive current measure. Not only is it capable of rapid expansion, as we learned during World War II, and again in the Korean War and the postwar boom, but a shortage in one segment seems to call forth workers from reserves in other segments. For example, during the 5-year period 1950 to 1955, the working age population under 25 did not grow at all but remained just over 24 million. A small increase (750,000) in the 14 to 19 age group was more than offset by a shrinkage in the 20 to 24 year age group. Partly as a result of the lack of population increase and partly because of some reduction in participation rates of teenagers, the labor force in these entrance ages declined by one-half million in these five years. A further shortage factor was the high marriage and birth rate which restricted the labor force availability of young women 25 to 34 years old. Yet the demand for workers and for military manpower were very strong during most of these years. These shortages were met by an unprecedented increase of 2.2 million in the number of women workers over 35, the extent of whose availability for jobs and acceptability to employers could not have been forecast. True, their labor force participation rates had shown a slow upward trend which World War II sharply stimulated, but no one thought the war time increases would continue or even be sustained.

Some analysts think that the availability for employment of a large number of relatively well educated women not only made up for the shortage of young workers, but may also have hastened the retirement of older men. During the same 5-year period, the labor force rates of men 65 and over dropped from 44 to 39 percent—also an exaggeration of long-term trends. What was cause and what was effect here, no one can know from any information we now have. In any case, the past decade has seen an astonishing amount of adaptation of labor supply and demand to meet shortages originally demographic in origin. What will happen during the next decade when young people will start flooding the labor market should be of great interest. Is it possible that the growth of the female labor force over age 35 will slow down or cease altogether, apart from population changes? Present projections assume some continuation in the rise in labor force rates for this population group, and it seems unlikely that the long-term upward trend would be reversed.

The shortage of young workers may have reduced the barriers to employment set up against middle-aged and married women, but what increased the willingness of these women to take jobs? When we turn to nondemographic factors affecting propensity to be in the labor force we are on much less solid ground statistically. We know that, at any given time, the labor force participation of married women varies inversely with husband's income but that this relationship does not hold for changes over time. Rising income has not dampened growth; on the contrary, for a time in recent years, the greatest increases in labor force activity of women have been at the upper ends of the income scale. But increases have occurred all along the line. So we have come to the belief that changing social attitudes and customs which permit or encourage a wife to take a job to help raise family levels of living or to enliven her daily life may be the important factor affecting the propensity of married women to work.

Certainly the change in attitude has been widespread. In every large city in the United States except one, the labor force participation rates of married women increased between 1940 and 1950, with the average rate rising from 18 to 26 percent. The increase tended to be greatest where the average earnings of men were highest, where perhaps demand for women workers was at a maximum to fill the lower paid jobs which men refused to take. It will be interesting to see what the 1960 Census shows about the impact of the great expansion of this decade in different types of areas. Is there an upper limit to the number of women who can be in the labor force in an American community with our present standards and institutions? In general, past growth has been related to the size of the reserves and has been relatively greater in those areas or population groups where there had been a comparatively low rate of labor force activity initially. In the past 5 years, however, (measuring from third quarter 1954 to third quarter 1959) the amount of increase in the labor force rates for women between 30 and 50 years of age has been almost the same for each 5-year cohort even though at the beginning of the period, the rates ranged from 33 percent for age 30 to 34 to 43 percent for ages 40 to 44 and 45 to 49.

The almost universal change in attitude toward the employment of married women must itself have had a variety of causes. Two major and familiar causes for which we have some statistical evidence are the vast expansion in agreeable types of jobs like office and sales jobs, and the greater prevalence of part-time arrangements for work. Between 1950 and 1958 when total employment of women increased by 3 million, professional and technical workers increased by 700,000, clerical and sales workers by 1.7 million. There was also an increase of one million in service occupations—baby sitters, practical nurses, beauty operators, waitresses, and so forth. The occupations that lost workers were the farm occupations and operatives, mostly factory workers. The rising importance of service-producing activities has provided the kinds of jobs opportunities that both well-educated and untrained women could fill. At the same time, the proportion of part-time workers has risen—from 15 percent of the total at work in 1950 to 20 percent in 1958. Part-time jobs are a much more significant factor in the employment of teen-age than of adult workers but at least one in four women in each age class 25 or over was working part-time in 1958—the majority as regular part-time workers. Except for farm work, part-time arrangements are found most frequently in sales and service jobs and to the extent that these types of jobs expand in the future, we may expect further impetus to labor force growth. After all, many but not all, part-time arrangements, particularly in retail trade, require the employment of two workers where under other labor market conditions, one worker—with or without overtime—could have done the job. Thus, a shortage of persons available for full-time work has been a real factor in the increasing numbers in the labor force, without a correspondingly large increase in man-hours worked. Between 1950 and 1957, the average number of persons at work increased by 8.2 percent, but the man-hours worked by those persons increased only 6.4 percent.

Availability of part-time work is not always associated with labor force expansion. The increasing employment of high school and college age students has been largely of a part-time nature, but on balance, the activity of the age group has declined. Elderly men, 65 and over, also frequently work part-time, although the great majority who remain in the labor force are full-time workers—almost 80 percent of all employed in nonfarm jobs. Nevertheless, the labor force rate of this age group has been dropping ever since the end of World War II—from 51 percent in 1945 to 35 percent in 1958. The major liberalizations of the Social Security Act in 1950 and 1954 seem to have been a factor in this reduction. Migration away from farms may also have had some effect on the labor force rates of this age group, but there is some reason to think that the elderly farm resident who moves into town to live does so only after he has given up his farming activities.

In short, for certain types of persons the availability of other income seem to encourage withdrawal from the labor force. Men on the edge of retirement whose health and vigor are diminishing are in this group. Nonwhite women who are widowed, divorced, or separated also appear to be leaving the labor force as pensions or other types of Social Security payments for needy and dependent persons become available. Between 1940 and 1950, the labor force rates for this type of women living in nonfarm homes fell sharply, with the greatest decline for the younger age groups where there were probably children to care for. Evidence since then suggests that there have been further decreases, except in the recent recession period.

Between now and 1975 the projected growth of the labor force is about 23 million, if conditions of high employment prevail and if there is no major war or other large-scale catastrophe. This projection, made by Sophia Cooper and Stuart Garfinkle of the Bureau of Labor Statistics, takes account of several of the demographic factors that I have mentioned, in addition to age and sex—marital status and presence of preschool age children for women, school enrollment for young people. It also projected changes in labor force participation rates assuming certain changes in the propensity of various groups to be in the labor force. For example, some rise in labor force participation of married women and a further decline for men 65 and over are expected. In order to test the relative importance of purely demographic changes in a period when an enormous expansion can be expected because of the growth in the population of working age, I have estimated what the 1975 labor force would be if there were no changes in propensity to work. If marital status and presence of young children, as well as school enrollment, are allowed to vary in the expected fashion but specific rates are held constant at 1955 levels, and 1975 labor force would be 90,375,000 instead of 94,775,000 as projected. In other words about 4.5 million of the projected 23 million growth reflects changing propensity to work. This 4.5 million is made up of a 5 million increase for women and a half million drop for men.

Although the projections imply a fairly steady, smooth growth in the labor force determined largely by demographic factors, actual year-to-year changes are very irregular. Periods of rapid growth occur when the Armed Forces are being mobilized, or in the initial stages of a boom. They are followed by periods of slow growth which include recessions, but are not limited to them. For example, the year-to-year increases in the labor force that averaged over a million each quarter between 1950 and 1951 dropped off markedly between 1951 and 1952, more than a year before the 1953-54 recession started. Recovery from that recession was well under way before the annual growth in the labor force reached the million mark again. The over-the-year expansion of the labor force in the period between the third quarter of 1954 and the third quarter of 1956 was unprecedented in peacetime years, so far as we know. Between 1956 and 1957 the over-the-year growth measured quarter by quarter had dropped to 400,000 or less, even though the recession did not start until the middle of the year 1957. In the first two quarters of 1958, when unemployment rates were at their maximum, labor force growth picked up again, but subsided with recovery until the middle of this year.

At one time, it was thought that only family adversity could bring additional workers into the labor force but by now there is fairly general agreement that attractive job opportunities are more important. From this conclusion, it has been argued that failure of the labor force to grow in the short-run in accordance with a high employment trend line can be taken as evidence of economic maladjustment and that the difference between actual and projected growth is in fact a measure of hidden unemployment. The thesis is that those additions to the labor force that did not materialize comprised persons who were discouraged from seeking jobs because of the lack of opportunities, and the withdrawals which did take place were largely involuntary.

There is nothing wrong with comparing actual with projected labor force levels but I would question the interpretation of deficits as a form of unemployment. First, the estimates of the labor force in any month—and therefore, the over-the-year change for a month, or even for a quarterly average—may be affected by accidental factors such as particularly good or bad weather, by the occurrence of holidays, and by extremes of sampling fluctuations. Second, having lived through the 1930's, I would hesitate to estimate any critical indicator by subtracting a measured quantity from an estimated trend. Unemployment so estimated, including negative unemployment, was bad enough. What should we make of negative labor force deficits? Or labor force surpluses, so easily translated to mean "abnormal" workers as was done in the immediate post-World War II period? Whose trend line should we use?

But finally, and most important, in spite of all the survey results and the spate of words on factors in labor force growth, we cannot yet explain why various types of persons enter or fail to enter the labor force, or why they do or do not remain in. Some elderly men retire, some do not; some childless married women in a given income—education class work, some do not. The research in this whole field has been quite elementary, so far as I know, and the measurement of the motives, whatever they may be, is little better now than 20 years ago. Perhaps in the next golden age of research in labor force measurement, this area of ignorance should be the target for a concerted attack.

NEW ASPECTS OF PUERTO RICAN MIGRATION

By: Robert O. Carleton, Puerto Rico Planning Board

Three years ago the Bureau of Labor Statistics of the Commonwealth of Puerto Rico initiated a new and vastly improved sample survey of migration between Puerto Rico and the United States. Data for the first two years from this survey bring to light several very important and for the most part surprising characteristics of Puerto Rican migration.

First of all, it has been discovered that net migration between Puerto Rico and the United States — to a much greater extent than anyone supposed — is the resultant of a substantial two-way migration movement both into as well as out of Puerto Rico. The reasons why no considerable in-migration was expected to exist are well-known. Net out-migration itself has been so large (averaging 45,000 for the decade and at times reaching flood proportions) that it has been difficult to imagine it to be only the net figure and that the actual number of migrants might be considerably larger. Furthermore, the economic advantages of migration from Puerto Rico (where unemployment continues high, unskilled labor abundant and wages relatively low) to New York (with its scarcity of unskilled and semi-skilled labor) have been so obvious that the existence of large-scale return migration of migrants from earlier years seemed hardly conceivable.

Yet data from our survey reveal an in-migration (not counting the return of seasonal agricultural migrants and of others who had been away from Puerto Rico less than a year) in the neighborhood of 60,000 persons in both 1957 and 1958. Some of this in-migration was undoubtedly temporary and related to the conditions of recession and unemployment prevailing on the Mainland at that time. The major portion of it, however, cannot be understood without taking into consideration the rapid progress of economic development in Puerto Rico and the accompanying growth of economic opportunities — especially for those with above average education and occupational skill. Also, there is the easily understandable fact that most Puerto Ricans would prefer to live in Puerto Rico with only a fairly decent income (if they can get it), than to earn maybe 50% more and have to put up with all the disagreeable conditions which confront Puerto Ricans in New York.

As we have already noted, a high level of in-migration implies that gross migration from Puerto Rico to the United States must have been much greater than has generally been supposed on the basis of the net figures. Our data disclose about 85,000 out-migrants in 1958 when net migration was only 26,000. The corresponding figure for 1957, when net migration was 37,000, is estimated at almost 100,000. What was happening during the years prior to 1957 is a matter of conjecture, since there is no data on gross migration for these years. On the one hand, gross migration would tend to be closer to net out-migration since there were fewer Puerto Ricans living in the United States who could have migrated back to Puerto Rico. On the other hand, net out-migration was averaging 50,000 during the period 1950 to 1956 as compared with 30,000 in 1957 and 1958 so that a heavy gross out-migration for these years would not imply such a large excess over net migration. On balance, it seems not unreasonable to suppose a level of gross out-migration similarly high as the 90,000 or so a year level averaged in 1957 and 1958.

Intelligent forecasting of net migration during the 1960's and 1970's requires separate forecasting of in-migration and out-migration. In-migration from the United States to Puerto Rico should increase as the economic development program leads to an accelerated demand for skilled and some categories of semi-skilled labor. But the pressure to migrate from Puerto Rico should continue and maybe even increase among the unskilled. Because of the disappearance of marginal jobs, especially in agriculture and home needlework, economic development has so far not been able to raise the level of employment; its achievement rests in the improved quality of employment as represented by increased productivity and higher wages.

In the long run, the future of net migration should depend primarily on population pressure. To get a clear picture of present trends in Puerto Rican fertility we shall have to wait the results of the 1960 Census. There are indications of a considerable decline in fertility. The number of births annually has decreased from 85,000 in 1950 to 75,000 in 1958, and the crude birth rate has fallen from almost 39 per thousand to less than 32. However, the heavy migration during the decade has so fouled up our estimates of the age-sex distribution of the population that it is not certain whether the decline in crude fertility rates is attributable to a corresponding drop in age specific fertility or merely to a shifting of births from Puerto Rico to the United States.

There is another very interesting aspect of this Puerto Rican migration. A combination of circumstances has brought Puerto Rico so much closer to New York than it used to be — even though the distance on the map remains 1,600 miles. The airplane has cut the travel time from four or five days to three or four hours. The mass use of the new travel facilities by a more prosperous Puerto Rican population has brought the fare down — in this post-war era of inflation — to not much more than one-third of what it was just at the end of the war. The earnings of the average Puerto Rican have in the meantime more than doubled. As a consequence, the cost of a trip to New York — thirteen or fourteen years ago the equivalent of about three months' earnings — now represents only about two weeks' salary to the average worker in Puerto Rico.

In addition, New York has become close to Puerto Rico in terms of kinship, so much so that it is now recognized as the largest Puerto Rican city in the world. As a consequence, travel between Puerto Rico and the United States just for the sake of visiting has become as commonplace as migration.

In 1958 the number of trips between Puerto Rico and the United States amounted to almost 700,000 — slightly more than half of which were departures. These 700,000 trips represented approximately 420,000 different passengers who break down into about 140,000 migrants and 280,000 round trip passengers. The extraordinary thing to note here is that as many as 140,000 of the round-trip passengers — about half of the 280,000 — consisted of Puerto Ricans who visited friends or relatives either in the United States or in Puerto Rico. In other words, the 140,000 migrants and the 140,000 visitors together add up to 280,000 Puerto Ricans who in 1958 made a trip between the United States and Puerto Rico either to migrate or to pay a visit. Together they comprise two-thirds of the 420,000 passengers who travelled between Puerto Rico and the United States. The significance of these figures is tremendous and we shall return to them presently.

Meanwhile, we observe that many, although certainly not all, of these 280,000 Puerto Ricans were either of low income or sufficiently youthful as not to own much property. For such persons the decision to migrate (providing the trip can be made at all) can be a relatively casual affair since the matter of disposing of property is no great problem. This reason together with the factors that have drawn New York and Puerto Rico close together in travel time, travel costs and in kinship relations have combined to make this Puerto Rican migration a large-scale, long-distance population movement in which for perhaps the first time in history many of the participants travel first and decide later what the purpose of their trip was — whether to migrate or just to visit friends or relatives. The purpose of the trip can depend on how one makes out after arriving.

At this point a comment and a point of clarification have to be made. The comment is that as a statistician I want to go on record as deploring this state of affairs. It becomes very difficult to identify migrants in a survey where many of the passengers by the time they board the plane have not yet made up their mind why they are travelling. We have been forced to fall back on indirect methods of identification where the margin of error is uncomfortably large. Although most of the figures used here may well be off by 20% or so, the conclusions we have drawn from them would not be affected. However, we have much other interesting data on such characteristics as age, sex, education and occupation for the in-migrants, the out-migrants and the net migrants; but conclusions based on this data are not reliable and therefore cannot be released.

The point of clarification is that most of this visiting is from the United States to Puerto Rico, whereas the migration is preponderantly in the opposite direction. While 60% of the 140,000 migrants were moving from Puerto Rico to the United States, only about 30,000 (21%) of the 140,000 visitors were residents of Puerto Rico making a visit to the United States. In view of the higher wage levels in the United States and the fact that the Puerto Ricans living in the United States were the ones away from home, it confirms our expectations to find more of the visiting done by them. But the figure of 110,000 is so surprisingly high as to jolt the general impression most of us have of the economic potentialities of Puerto Ricans living in New York.

Some general conclusions may be drawn. The 280,000 Puerto Ricans who travelled between Puerto Rico and the United States in 1958 either to migrate or to visit friends or relatives represent almost 10% of all Puerto Ricans living either in the United States or in Puerto Rico. According to newspaper reports, Dr. Handlin in his book The Newcomers characterized Puerto Ricans as differing from previous immigrant waves by the extent to which they have retained close ties with the land from which they came. The evidence we have presented certainly supports Dr. Handlin's observation.

The observation, however, is easily susceptible of misinterpretation. Those who see in it nothing more than an exaggerated sense of clannishness together with a refusal to adapt to new conditions have missed the whole point. The observation really cannot be understood except in the context of the tremendous economic and social transformation that is going on these days back in Puerto Rico.

The truly bewildering rate of social and economic change in contemporary Puerto Rico is sufficient evidence that Puerto Ricans are neither unwilling nor unable to adapt to new conditions. By its accelerated program of modernization, Puerto Rico is trying to pass through in one generation a process of development which took the United States perhaps a century and a half. It is hard to communicate what this implies as an abrupt calling in question of long-accepted values, attitudes and patterns of behavior. Two radically different ways of life — the pre-industrial and the post-industrial — are being suddenly and drastically confronted face to face, and the separate parts of each are being weighed in the balance.

The significance of 280,000 Puerto Ricans either moving or visiting between Puerto Rico and the United States in the short space of one year is that the adaptation to new conditions — whether through preference or through necessity — is being made collectively as well as individually. The going back and forth serves a double purpose: (1) It permits Puerto Ricans to experience modern industrial life in the more advanced forms it has taken in the United States as well as its manifestations in Puerto Rico. The social transformation is thereby accelerated so that it may keep pace with the economic. (2) It enables a collective sharing and exchanging of experiences among the migrants, the visitors and those engaged on the home front.

Along with much agonizing self-interrogation, an intense debate is taking place among the Puerto Rican people. The basic question, however, is not whether Puerto Rico should industrialize. Hardly anybody in Puerto Rico, I dare say, would dispute this. The question is whether the values historically associated with industrialized society have to be accepted wholesale as a package deal, or whether and to what extent it is possible to pick and choose. Is it possible to have a highly dynamic industrial society founded on mass education, individual initiative and enlightened self-interest without accepting at the same time the vulgarity of the comic strips and the shabby morality of the TV quiz programs? And if it is possible, is it desirable? There is much difference of opinion among Puerto Ricans on this score. And evidently among Americans as well.

COMMENTS ON PAPERS PRESENTED AT A SESSION ON "LABOR FORCE MEASUREMENT AND INTERPRETATION"

By: George P. Shultz, University of Chicago

The papers presented here, with their impressive display of knowledge about the labor force, suggest how much we know now about this subject, despite constant reference to the areas of our ignorance. Dr. Bancroft concludes with a suggestion for the next Golden Age of research in labor force measurement. The implication of this phrase is that the last twenty years have themselves been a sort of a Golden Age for this field. That is certainly true and, at the same time, is no small tribute to the Chairman of our session. If I may quote from a forthcoming paper by my colleague Albert Rees:

> "The person who more than any other has been at the forefront of this effort is Dr. Gladys Palmer of the University of Pennsylvania and the Office of Statistical Standards of the Bureau of the Budget. I know of no one who can be at a statistical forefront less obtrusively or more effectively."

We are all deeply indebted to Dr. Palmer.

The papers by Drs. Wolfbein and Bancroft are substantive in nature and the one by Dr. Martin procedural. Let me start with a few comments on procedural matters.

I would like to add to Martin's informative description of the transfer of the Monthly Report on the Labor Force from Census to Labor a footnote for the political scientist, a note of clarification for the user, and a plea for the Department of Labor.

Martin makes the switch sound so reasonable that one is tempted to accept hers as a complete explanation. In that case, the political scientist must assume that the Secretary of Commerce either does not care about the scope of his activities or is a very poor bargainer. I have heard by the grape-vine that he was neither of these and that part of the explanation for the switch, or at least for its timing, must be found in a move in the other direction of another statistical program. I will leave it to the ingenuity of political scientists to find out what program it was.

For the user, I must add a note of caution. If you are one of those who received the old Census publication, are now on the list for the Department of Labor release, and intend to file what you receive for some months and then dive into the pile to unearth your favorite tabulation, beware! The new release does not carry all the old information. For the full story, you must look in Employment and Earnings, published a little later and available at a price.

My plea is that the Department of Labor not overreconcile the results of its various series on unemployment and non-agricultural employment. Two independent readings on the labor market are useful to analysts. Each has its area of strength and weakness. Of course, where true explanation of differences is possible, that explanation is useful. There is always a temptation, however, to somehow smooth over differences "so as not to confuse the public". It seems to me that too much can be made of this public relations point. If different readings give different answers, it is useful to analysts to know about that.

Martin also mentions the work now going forward on improving the seasonal adjustment of unemployment. This has always presented a ticklish public relations problem for the publishing agency and I suppose that accounts for the fact that, officially, we now adjust an unemployment rate rather than unemployed people. As is generally recognized, however, this method leaves much to be desired when unemployment is changing rapidly. There are strong arguments for first adjusting the various figures themselves and then computing a seasonally adjusted rate from these figures. I hope that the Budget Bureau is considering these arguments.

Finally, I should like to endorse Martin's call for emphasis on improving estimates of local unemployment as preferred to further fussing with the questionnaire used by the Monthly Report on the Labor Force. Local unemployment presents some sharp policy issues and our factual information about the localities involved is shaky indeed. It would be desirable to have more information before legislation is passed on depressed areas; but if this does not come to pass, then we must have in any such legislation provision for improvement in our knowledge about local unemployment.

Bancroft and Wolfbein together offer 1,967 neatly organized and perceptive observations about the labor force. My sample check on the accuracy of their reporting shows that there are 99 chances out of 100 that 99.9 percent of these observations summarize accurately the available evidence. I shall say no more about these statements, except that they seem to me to be in the best traditions of reporting by producers of statistics: Careful, apparently descriptive but inclusive of much analysis by virtue of the descriptive categories selected, and close to the data.

I was delighted to hear Wolfbein, as a responsible and key figure in the statistical program about the labor force, urge that gross change data be made available on a regular basis and I trust that his hedging comment was as the instinctive act of a man attuned to the workings of a bureaucracy rather than as a cautious note from a professional statistician. As his paper shows, it is important to present more than the gross rate, as he computed it, or the gross numbers involved. Analysts will want to disaggregate as he has. It is particularly important, I think, to separate flows between employment and unemployment from those between out-of-the-labor-force and unemployment. Those latter figures might usefully be analysed in conjunction with flows between out-of-the-labor-force and employment.

Bancroft's suggestion for the path of future research is related, I think, to this point, for it is those groups with changing and diverse labor force attachments that interest her and whose motives she wishes to measure. How is this to be done? Let me conclude by commenting on this question.

There are two broad approaches, it seems to me. One is illustrated by the M.R.L.F. and Bancroft's paper. It is indirect. It seeks to identify groups within the population that exhibit reasonably stable and common patterns of behavior: mostly in the labor force, mostly out, entering or leaving as a function of some variable such as children under school age. Once we have a group nailed down in this sense, we seem to feel we have solved the motivation problem. Some jump to the conclusion that the descriptive categories carry the explanation of motives. Others are more cautious and are content with the high probabilities they can attach to statements about a group's labor force behavior. At any rate, this approach has taken us a very long way. It commends itself to us, as a group discussing labor force research conducted by the survey method through Government auspices.

Alternatively, there is the direct approach. Here one seeks knowledge about how individuals have approached and decided on their labor force behavior. It is illustrated, I think, by the local labor market studies conducted primarily by scholars at universities--studies greatly stimulated by the example and counsel of Dr. Palmer. This approach requires that considerable time be spent with individuals, though perhaps the psychologists may develop some useful short cuts for us. At any rate, these studies do not give the kind of solid evidence produced by the M.R.L.F. The results tend to be suggestive rather than conclusive. Nevertheless, they can be most helpful in the design of categories for use by the survey method.

Each of these methods has special advantages that the other cannot duplicate and it is a mistake, I think, to try to force out of one method too great a range of results. The land of interplay between methods that we have had in the past has been productive. It is to this kind of interplay that we should look during the next Golden Age of research in labor force measurement.

DISCUSSION

James W. Knowles, Joint Economic Committee, U.S. Congress*

The privilege of participating in this afternoon's program as a discussant of these four excellent papers on labor force measurement and interpretation is one I value. For ten years I have been closely associated with the interest which the Joint Economic Committee has had, both in the use of measures of the labor force--hence in their interpretation--and in the improvement of the scope, detail, and accuracy of the measures themselves, an interest evidenced not merely by repeated use of the data but also by recommendations for their improvement and by hearings on the progress of these programs, with which you are all familiar. During these years I have been the Committee's chief technical adviser on matters arising in this area.

The quality of the papers is such that I do not have to comment on technical issues. Rather, I would like to comment on the significance from a user's point of view on some of the points which have been raised in these papers.

First, I would like to point out a significant aspect of Miss Martin's recitation of the changes and improvements in programs for labor force management. The changes and improvements she recites, and the list she gives of ideas still on the agenda for future action, illustrate graphically the fact that improved knowledge of the functioning of our economy does not come about by itself or by the mere exercise of hard work or ingenuity on the part of the technicians involved.

These improvements and advances in the measurement of the labor force came about as a result of the interest of public officials and private policy-makers in better measurement of the labor force and of the characteristics of the various participants in the labor force. The information was needed in order to understand the problems of the labor market, and public and private policies which either influence the state of the labor market or which are influenced, in turn, by current or prospective developments in it.

I think all those interested in reducing the area of guesswork in economic decision-making should learn from this experience that statistical programs that provide the raw material for economic analysis, either in the realm of pure research or applied economic research underlying policy-making, will be developed in direct correspondence to the effort which users put into supporting the programs and insuring that they have both the guidance as to needs, and the resources required to provide the increasing quantity and quality of information required. The organization and growth of the Federal Statistics Users Conference might be mentioned here as a good omen in this regard.

A second point that all of these papers illustrate is the fact that improvements must not be allowed to destroy continuity and comparability of statistics over time to such a degree that we lose the precious information that is available from continuous measurement of key variables in our economy. The continuous leadership provided in the field of labor force measurement by our chairman, Gladys Palmer, has been oft cited in this respect and its value can never be overestimated.

Third, the papers indicate also that the information we gain from programs of continuous measurement, such as we have in the labor force field, are now not mere additions to scientific research. Indeed they have continuous impact in the realm of practical economics. For instance, the example of analysis that Seymour Wolfbein gives in his paper on unemployment has just had immediate relevant application in the deliberations of the Joint Economic Committee in connection with its study of Employment, Growth, and Price Levels. This is illustrated in two ways. The Committee commissioned the Bureau of Labor Statistics to make two studies for the Committee on unemployment, one of which--"The Extent and Nature of Frictional Unemployment" cited by Mr. Wolfbein--was our Study Paper No.6, and has already been published; another, on structural unemployment, is still in preparation and we expect to publish it in the coming weeks. These papers found immediate application, as you will discover if you read the Staff Report on Employment, Growth, and Price Levels.

One of the matters of concern to the Committee has been, of course, the question of the amount of and character of unemployment which is unavoidable under the conditions of high or full employment, and of what part of the unemployment can be reduced or eliminated by measures of public and private policy gained at increasing the mobility of labor between areas, industries, and occupations, and to improve its quality in terms of education, health, etc.

Fourth, underlying any consideration of private or public policies must be the prospects for growth. One of the most important elements of growth, obviously, is the supply of labor and its distribution between different age-sex groups and occupations and extent of its education. Hence, the significance of trends, such as those discussed by Gertrude Bancroft, in the immediate practical realm, where decisions must be made as to whether we can count on the labor force growing sufficiently to support a growth rate in aggregate output of 3, 4, or 5 percent, or only 2 percent. Her analysis shows that a significant proportion of the annual growth in the labor force is a reflection of past and current economic

* The views expressed are those of the speaker and do not necessarily represent the views of the Joint Economic Committee or individual members of that Committee.

conditions and opportunities.

Fifth, I would comment on a point made by Miss Bancroft in her paper concerning the comparisons of actual labor force with the projected labor force derived from considerations of longer term trends. While one should, as she points out, be very careful in interpreting the resulting excesses or deficits as being surpluses of employment or concealed unemployment, the distinction between actual and the potential labor force is important as an instrument of analysis as its use in this way by the Joint Economic Committee has amply proved.

By using this potential or projected labor force as one variable along with considerations of hours of work, capital resources, etc., it is possible to arrive at a rough estimate of the potential output of the economy at full employment. This potential can then be used as a summary indicator of the amount by which the economy is succeeding or falling short of achieving the objectives of the Employment Act.

I believe the ten years of its significant use in this way is proof of its value and proof that it can be used without falling into the trap of which Miss Bancroft quite rightly warns us. As a practitioner in the use of these statistics in this way, I strongly reenforce her warning.

Finally, I wish to strike a personal note. As I read Miss Martin's recitation of the recent history of our programs for labor force measurement, I found myself recalling some vivid moments of my own relationship to these programs. I played some role from the Legislative side, especially for one or two of these developments, in particular the step following the spring of 1954, of drafting each month a combined employment-unemployment statement by the Departments of Labor and Commerce. The Bureau of Labor Statistics now has assumed responsibility for the labor force, employment, and unemployment statistics, while the Census Bureau continues to act as the basic collector of the data. As a technician with a continuing obligation to explain labor force developments to a nontechnical audience, I look forward to the further refinement of the unified interpreting each month of the several sets of data on employment and unemployment under the new arrangements.

III

INNOVATIONS IN STATISTICAL RESEARCH IN SOCIAL SCIENCE

Chairman. T. N. E. Greville. Department of Army

THE USE OF COMPUTERS IN STUDY OF SOCIAL STRUCTURE: INTERACTION IN A 3-PERSON GROUP.[1]

By: James S. Coleman, The Johns Hopkins University

The work reported in this paper stems from two sources. The first is a set of propositions made by George Homans in The Human Group, some of which were formalized by Herbert Simon in a paper several years ago. The second is an observation about 3-person groups made by Georg Simmel, and followed up by Theodore Mills and others in recent experiments.

The three-person group as isolate and pair:

Georg Simmel, some years ago, focussed attention on the triad, and made several insightful observations which have been followed up by recent investigators.[2] Simmel's observations tended to suggest that a three-person group breaks up into a pair, between whom association is relatively great, and an isolate, who had considerably less association with either of the pairs. This was only one of a number of configurations which Simmel discussed, but is the one to be examined here. This suggests that a situation of equal interaction between A and B, A and C, and B and C is inherently unstable, and that a stable equilibrium would occur at a less balanced state of affairs, where interaction between A and B was strong, that between A and C was weak, and that between B and C is weak.

This is not at all a trivial proposition, for even if one agrees that a completely balanced state is unlikely, the imbalance might occur in other ways, as indicated in Figure 1.

Theodore Mills has examined Simmel's suggestion in some detail, and found in experimental groups that such a situation did occur; that the three-person group did tend to break up into a highly-interacting pair and a relative isolate.[3]

In quite different investigations, using ad hoc problem-solving groups, Bales has found that in 3-person groups, the rates of interaction between the two highest interacting persons are considerably greater than that of either with the third person (See Table 1.). At the same time, there is much data, both from systematic research and everyday observation, in which 3-person groups do not degenerate in such a fashion.[4] It would be foolish to view the proposition as more than a statement of a tendency, as something which is false if not universally true.

The Homans propositions and Simon model:

The very simplest propositions from Homans were those which Simon undertook to formalize. They concern the mutual reinforcement of social interaction and positive sentiments: as interaction among a set of persons occurs, positive sentiments will develop among them. These positive sentiments, in turn, lead to an increased rate of interaction, over and above that which initially occurred.

Simon formalized these propositions in order to link them together, and thereby construct a system of relations characterizing the group. The formalization consisted of a pair of differential equations which describe the change of positive sentiments (s) with change in interaction rate (r) and vice versa (with parameters a, b, c, g, h, k).[5]

$$\frac{ds}{dt} = a + b\,r - c\,s \qquad (1)$$

$$\frac{dr}{dt} = g + h\,s - k\,r \qquad (2)$$

For appropriate values of the parameters, this will give a system which has a stable equilibrium point, as indicated in Figure 2. The arrows indicate the direction in which interaction and sentiments will move when the group begins at different levels of interaction and sentiments. As Figure 2 indicates, appropriate values of the parameters give a system such that the levels of interaction and sentiment increase, if they begin at a low point, to a stable position (or if they are artifically raised beyond the equilibrium levels, drop to the point where they are mutually sustaining).

It should be emphasized that this is not an empirical analysis, but simply a formalization of propositions about groups, in order to create a system of relations. This system of relations merely shows the joint implications of these two propositions taken together. The empirical adequacy of the propositions is another matter, which will not be discussed here. Though empirical adequacy is important, and certainly the ultimate test of this work, it seems reasonable to withold such tests until after the work has proceeded some distance.

Simon raised the question, as have others, of what might be gained, relative to the above formulation, by carrying out the formalization at the level of individuals and pair relations, rather than the group as a whole. This has some measurement advantages, because sentiments and interaction rates are ordinarily measurements made upon individuals and pairs, not a group. There always arises some aggregation problem if propositions and variables are posited at the group level. Furthermore, by a shift to the individual level, it becomes possible to examine differentiation within the group, which is impossible in the group level formulation. The disadvantages, of course, are the greater complexity of the individual-level formulation, the difficulty of analytic solutions, and the impossibility of graphical representations such as the one above.

This, then, is a second source for the work to be discussed. Together with Simmel's and others' observations on the three-person group, it constitutes the basis for two computer models presented below.

***

Taking these two directions of work together, the following question may be raised: accepting as true the isolate-pair tendency in a triad, might it not be explainable in terms of Homans' propositions about the mutual reinforcement of interaction and (positive) sentiments? To be sure, there are personality differences, and differences in compatibility, in any three-person group, and some imbalance certainly results from these. But it would be particularly striking if the imbalance occurred even when there were assumed no initial differences, but only differences which developed through Homans' relations as the group carried out its activities. The conditions under which this might occur is that in which the total interaction between the three was fixed. Thus although there is mutual reinforcement of interaction rate and positive sentiments, there is also a condition of scarcity: only one of the three pairs can be interacting at any one time. This interaction tends to increase the positive sentiments between these two, but because neither of these is interacting with the third, sentiments do not have a chance to build up between him and either of them. This imbalance might then perpetuate itself, resulting in a situation in which the third person was either completely isolated or partially so. The possibility of complete isolation occurring as an initial imbalance became greater and greater seems intuitively easy to conceive, and easy to construct a model for. However, the empirical results seem to correspond to a situation of incomplete imbalance, such that the isolate is not completely disregarded by the others.

Thus with the aim of using the Homans relations to account for the pair-isolate tendency, two formal models were constructed. The concepts and relations in both of these were necessarily at the level of individuals and pairs, in contrast to Simon's model discussed earlier. One is a rather direct analogue of the Simon model, while the other is a stochastic model more directly mirroring the interaction process.

The 3-person analogue of the Simon model, with pairwise interaction, and fixed total interaction:

Simon's concepts are aggregate or group level sentiment and interaction rates. The analogous concepts at the individual level are:

s_{ij} = the (positive) sentiments of individual i for individual j. (There are 6 s_{ij}'s.)

$r_{ij} (= r_{ji})$ = the tendency toward interaction between individuals i and j. (There are 3 r_{ij}'s.)

Note that r_{ij} is here a tendency toward interaction, rather than the rate itself, as in Simon's equations. This is necessary because the tendency must be an increasing function of sentiments, yet the sum of the interaction rates is fixed. The tendencies are translated into interaction rates by dividing up the total interaction proportionally to the sizes of the tendencies r_{ij}.

The system of equations set up is as follows:

(a) For the change in interaction tendency between i and j, as a function of s_{ij} and s_{ji}:

$$\frac{d\, r_{ij}}{dt} = c_1 (a_{ij} + k_i s_{ij} + k_j s_{ji} - r_{ij})$$

(all parameters are > 0)
(3 such equations)

This equation says that without any sentiments, interaction would develop until it equaled a_{ij}, but that sentiments raise this. The parameters k_i show the amount of effect that i's sentiments have on the interaction tendency, and thus are a measure of the power or control i has over the interaction. The parameter c_1 is a coefficient indicating the overall speed of change.

(b) For the change in sentiments of i toward j, as a function of r_{ij} (relative to other r's):

$$\frac{d\, s_{ij}}{dt} = c_2 \left(b_{ij} + h_{ij} \frac{r_{ij}}{r_{+2} + r_{13} + r_{23}} - s_{ij}\right)$$

(all parameters are > 0)
(6 such equations)

This equation says that some sentiments would develop without interaction, equal to b_{ij}. But the rate of interaction [which is simply $r_{ij} / (r_{12} + r_{13} + r_{23})$] further raises the level of sentiments. The parameter h_{ij} is the amount of effect interaction with j has in increasing i's sentiments toward him - a matter which depends largely on the amount of rewards i gets from the interaction. The parameter c_2 is a coefficient of rate of change.

This system of equations thus consists of three equations for rate of change of r_{12}, r_{13}, and r_{23}, and six equations for the rates of change in sentiments s_{ij} of each toward the other. Given these equations, what can be said about the Simmelian observation about degeneration into an isolate and a dyad? Two conditions can be distinguished:

(a) the condition in which parameters are identical for each person and each pair. In such a circumstance, will there be an equilibrium at balanced interaction (where $r_{12}=r_{13}=r_{23}$)? And if so, will this equilibrium be a stable one? Such a condition provides the most rigorous test, for if there is imbalance when external conditions (represented by the various parameters) are perfectly balanced, then Homan's propositions can under any conditions account for the imbalance which Simmel observed.

(b) the condition in which parameters differ for different persons and pairs. One would naturally expect imbalance in such a case (for example, when the increase of positive sentiments between A and B due to their interaction is especially great). However, two questions are of importance here: does a slight imbalance in the parameters, such as might be found in any group of three, result in a magnified imbalance in the equilibrium interaction? And second, does any type of imbalance in the parameters result in the dyad-isolate configuration, or do some types of imbalance give rise to other patterns, as shown in Figure 1?

These are the questions which are to be investigated with this model. The first condition, of complete balance in the parameters, can be handled by an analytic solution; the second requires numerical analysis, which in this case was carried out on a digital computer. Before examining the results, the second model will be outlined.

<u>A stochastic model of 3-person interaction, with pairwise interaction and total interaction limited:</u>

The model to be described below is not simply the above differential-equation model with a stochastic element introduced. Rather than consisting of a set of simultaneous equations to represent the system, it attempts to mirror more directly the interaction process involved, while capturing the essence of the Homans propositions about interaction and sentiments. The basic processes of the model, as presented in the flow chart of Figure 3, are as follows:

1. Select tentative interaction pair, on the basis of interaction tendencies r_{ij}.

The process is:

a) $p_{12} = \frac{r_{12}}{\sum r_{ij}}$

b) If p_{12} > random number, the 12 pair is selected; if not, p_{13} is computed, added to p_{12} and tested against the random number. One pair will definitely be selected, since $\sum p_{ij} = 1$, and the random number lies between zero and 1.

2. Test for rejection of interaction on the basis of interaction tendency r_{ij} relative to pro- (r_{ij}) and anti- (q_{ij}) interaction tendencies.

The process is:

a) If $\frac{r_{ij}}{q_{ij} + r_{ij}}$ > random number, then there will be interaction. If not, a new interaction pair will be selected according to (1) above.

3. Interaction between individuals i and j produces (stochastically) a reward or punishment for i, depending on fixed parameters. These rewards and punishments cumulate, to constitute the total rewards and punishments received from interaction with j. The rewards and punishments may be considered directly as positive and negative sentiments toward j, as they will be implicitly below, or they may be considered to affect the positive and negative sentiments. Either way is consistent with the Homans propositions, and either way will give the same formal results.

The process is:

a) $p_{ij} = \frac{g_{ij}}{h_{ij} + g_{ij}}$

where g_{ij} = coefficient of reward to i for interaction with j.

where h_{ij} = coefficient of punishment to i for interaction with j.

where p_{ij} = probability of reward.

b) If p_{ij} > random number, reward counter s_{ij} is incremented by 1; if not, punishment counter v_{ij} is incremented.[6]

3.' Same for j as a result of the interaction with i.

4. Modify interaction tendency between i and j on basis of positive sentiments between i and j. This tendency is governed by k_i and k_j, which as in the other model, are i's and j's control over interaction, respectively.

The process is:

$r_{ij} = k_i s_{ij} + k_j s_{ji}$

5. Modify anti-interaction tendency between i and j on basis of negative sentiments between i and j.

The process is:

$q_{ij} = k_i v_{ij} + k_j v_{ji}$

These processes taken together constitute a system of interaction of the following sort: interaction generates either rewards or punishments for the participants. The rewards bias the interaction toward the pair who interacted, in the next selection of interaction; but the punishments act as a "filter" which must be passed before the interaction proceeds.

<u>Results for the two models:</u>

A. The simultaneous-equation model

1. The first question is the equilibrium state when the parameters are identical for all pairs and individuals. It can be easily shown that there is a single equilibrium point, where interaction is exactly balanced. Furthermore, this is a stable equilibrium, as in the original Simon model. If interaction or sentiments are unbalanced to begin with, they will return to the balanced situation, as indicated in (c) of Figure 4.

There is one exception: if there is no externally induced interaction or sentiments (that is, if all a_{ij} and b_{ij} are zero), and the system begins in an unbalanced state, exactly the same imbalance will remain. It will not increase (as

Simmel proposed) nor diminish, but remain exactly as it began.

The first conclusion, then, is that the most difficult test is failed. The Homans propositions, when translated into a deterministic model with no external differences among individuals, do not produce degeneration into a pair and isolate. Quite to the contrary, a completely balanced situation returns when the balance is in some external way upset.

2. When the external parameters are allowed to differ from person to person, the results are quite different. Since the equations cannot be easily solved for the equilibrium state of r's, numerical analysis on the computer was carried out. Two conditions were investigated: a) one in which two persons found greater rewards from their interaction than either did with the third; (h_{13} and h_{31} are greater than other h's) and b) one in which one person had greater control over the interaction than did the other two (k_3 is greater than k_1 or k_2).

The results of these two situations were quite different, as shown in Figure 4. In Figure 4, each point of the triangle represents 100% interaction between two members; the center is completely balanced interaction. In the case where one pair's interaction was mutually more rewarding, the interaction (which began in a balanced state, at the center) moved off toward a new equilibrium, in which r_{13} was high, r_{23} and r_{12} low, shown in (a) of Figure 4. This is, in effect, the isolate-dyad situation. It should be noted, however, that the imbalance appears to be a direct reflection of the imbalance in h's. The values of h_{13} and h_{31} were twice as high as those of the other h's, so that it required a rather large imbalance of the h's to produce even this imbalance in interaction.

In the case where one member has greater control, a very different kind of imbalance occurs, corresponding to (c) in Figure 2. Two pairs (the two in which the strong member takes part) increase their interaction while the third decreases, until a new asymmetric equilibrium is produced, as shown in (b) of Figure 4. This is not at all the kind of asymmetry Simmel discussed, but quite the opposite kind.

On the whole, then, this formalization of Homans' model hardly accounts for the isolate-dyad phenomenon. It does generate various types of disequilibrium, but nothing beyond the imbalances which are introduced as data. Thus the model has little or no explanatory ability with respect to the isolate-dyad phenomenon.

B. The stochastic reward-and-punishment model

The stochastic model gives a somewhat different picture. Only a few runs have been made with the model, so that its general characteristics are not yet known; but these are enough to show some interesting consequences.

The parameters of the model were identical for all persons. The only parameters of the model are these: a) relative power of each person (equal in this case, so that for everyone one unit increase in cumulative reward produces half a unit increment in interaction tendency; and one unit increase in cumulative punishment produces half a unit increment in anti-interaction tendency); and b) the reward-punishment ratio. This ranged from a low of 6:4 to a high of .95:.05 in the runs which have been made.

Making the parameters identical for all is necessary to test the explanatory power of the Homans propositions. If there were imbalance in the parameters, a resulting imbalance in the interaction would be quite expected, as in the deterministic model above.

The initial conditions are important in this model. In all these runs, the assumed existing interaction tendency between each pair was only one unit. Consequently, on the first interaction, the immediate rewards could have as much influence as past experience (one unit to one unit). If the initial "experience" were much greater, a greater tendency toward stability might be expected.

The results of the first run, which used a reward-punishment ratio of 6:4, are shown in detail in Figure 5. There was an immediate and definite drift toward one of the pairs. By chance nine of the first ten interactions were between individuals 1 and 2, and also by chance, both 1 and 2 were rewarded on all nine of these interactions. This initial sequence set into effect a drift which was never overcome, as Figure 5 shows. The drift continues toward the 1,2 pair, though there are interruptions - and cases of rejected interactions. After 100 time periods, the interactions have been far more numerous between 1 and 2 than between any other pair, the rejected interactions have been more numerous, the positive sentiments are far stronger, and greater feelings of hostility have developed. In other words, the 1-3 and the 2-3 relationships simply didn't have a chance to develop. The 1-3 relation was stifled by negative sentiments on 3's part which happened to come about in early interactions. As a consequence, many of the attempted interactions between 1 and 3 simply never came off, thus reducing the likelihood of a relation developing. The 2-3 relation was a different matter - it never got started. There were only 6 attempted interactions altogether, and two of these failed.

The results of this run are not typical, however. In fact, in examining many runs, no typical pattern was observed. Sixty-two runs were made with reward:punishment ratio of 6:4. At the end of 100 trials, some groups were still rather balanced, some were segregated into a dyad and an isolate, and in some, one person dominated the interactions. In other words, the situation sometimes deteriorated into a dyad-isolate combination, sometimes into single-person dominance, and sometimes remained balanced.

A hundred and forty-seven runs were made at a reward:punishment ratio of 8:2, and a hundred at 95:05. There was little change in outcome with these changes, except an apparently slight drift toward imbalance (of both kinds) as the reward:punishment ratio increased.

A quantitative study of the results is not in order until more data are in. These results indicate only that a situation of imbalance frequently does occur, despite the symmetry of the parameters. Surprisingly, the imbalance seems just as frequently to be of the single-person dominance type as of the dyad-isolate type.

Conclusions:

It would be premature to attempt to draw many conclusions from these initial results. Until there are some extensive Monte Carlo runs made, the general characteristics of the model cannot be known. It is clear, however, that in the stochastic model, either a continued balance or a drift toward asymmetry can occur, depending on the initial conditions and upon chance. Yet the asymmetry is not always of the dyad and isolate type, as Simmel and others have written about, but seems equally often to consist of two strong relations and one weak relation, a somewhat surprising result.

It is also clear that under conditions of balanced parameters, the deterministic model does not give asymmetry of any kind. Only with difficulty does it give asymmetry at all, and then an asymmetry which is completely built in by fixing the parameters asymmetrically.

Finally, some points about further development of these models are relevant.

a. This problem has a direct dual, in which the pairs represent individuals and the individuals represent directed relations. (For groups, larger than three in size, the formal duality does not exist, though only a slight variation in the model is necessary for the dual interpretation.) The dual problem is the problem of individual participation in groups. Interaction rate for the pair becomes an individual's participation rate; and the rewards and punishments for the two members in the interaction become rewards and punishments received from the other two members. Though the detailed interpretation will not be discussed here, it is useful to note that the formal model can be appropriate for the dual problem as well as the original one.

b. The model has been set up to investigate specifically a problem of three-person groups; but the same model is appropriate for larger groups where only one pair is interacting at one time. (It is interesting to note that only in a three-person group is this single interaction per unit time dictated by the very numbers themselves. In groups of four or more, simultaneous interactions may occur between two pairs.)

c. The system as it stands is only a bare framework which may be enriched by further processes. For example, Homans and others suggest that the likelihood of an interaction's being rewarding depends upon one's control over it. When power is distributed very unevenly, those with little control over the interaction are more likely to develop hostility. Also, it is known that as sentiments develop in a group, the members extend the relationship into other activities. These are usually activities in which they have common interests, and which consequently have a higher ratio of reward-to-punishment than the original activity.

d. More generally, this model probably will be most important as a piece of other, more inclusive systems. For example, one system now being developed at Johns Hopkins is a worker-management system in which restriction of production by workers is generated by management's reduction of piece rates. But in the model, the normative constraints workers are able to place upon one another depend upon the strength of sentiments which have developed among them. These sentiments are generated over time, according to the configuration of work activities and the stability of personnel. Either of the two models discussed here can be used to generate these sentiments.

FOOTNOTES

1. Work reported in this paper has been facilitated by a grant from the National Science Foundation. Gratitude is also due to the Applied Physics Laboratory of Johns Hopkins, for making its computing facilities (a Univac Scientific, 1103A) available, and Miss Patricia Powers, for aid in running the problems.

2. Georg Simmel, The Sociology of Georg Simmel ed. and trans. by Kurt Wolf (Glencoe: The Free Press, 1950). Work in this area has included that by Fred L. Strodtbeck, "The Family as a Three Person Group" American Sociological Review, 19 (February 1954) pp. 23-29; and Theodore M. Mills, "Power Relations in Three-Person Groups, American Sociological Review, 18 (August, 1953) pp. 351-357.

3. Ibid, and "The Coalition Patterns in Three-Person Groups," American Sociological Review, 19 (December 1954) pp. 657-667.

4. Strodtbeck (loc. cit.) for example, finds that in 3-member family groups this tendency to degenerate into an isolate and a pair is missing.

5. Simon's formulation is more complicated than this, but it reduces to equations like (1) and

(2). Simon also sets up an alternative set of equations, which makes fewer assumptions about quantitative measurement of r and s, and about the form of the relations. Those are not relevant to the present discussion; nor will the present discussion go into questions of measurability of sentiments and interaction, which have been treated in detail by the author elsewhere. (See "Mathematical Models of Small Group Behavior," in Herbert Solomon, ed., Mathematical Thinking in the Measurement of Behavior: Small Groups, Utility, Factor Analysis (Glencoe: The Free Press), to be published Spring, 1960.

6. The random number is rectangular between zero and 1.

TABLE 1.

Participation rates in 3-person groups (data from R. F. Bales - personal communication)

		TO 1	TO 2	TO 3	(to group)
FROM	1		1445	660	2044
	2	1455		504	1091
	3	731	530		874

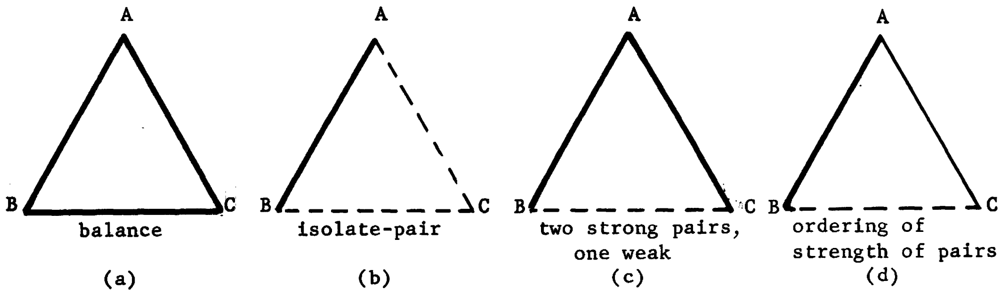


FIGURE 1: Various configurations of interaction in a 3-person group.

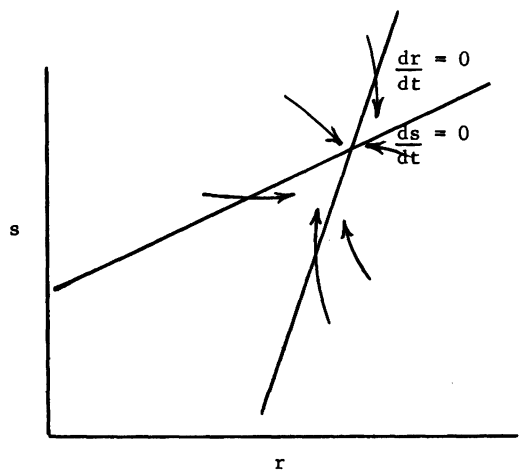


FIGURE 2: The interdependence of sentiments (s) and interaction rate (r) in a group, according to Simon's formalization of Homans' propositions.

SOME STATISTICAL PROBLEMS IN A RESEARCH PROJECT ON ORGANIZATION

By: Harrison C. White, University of Chicago

In order to assess the sampling variability of any measure of sociometric structure, one must define a priori probabilities in the sample space of structures. There is a curious divergence in the literature. In both approaches equi-probable elementary states are defined, but on the one hand only non-isomorphic structures are counted as equally probable,[1] while on the other hand all structures are counted as equi-probable elementary states.[2] It is ultimately an empirical problem to choose between these (and other) alternatives,[3] a problem which apparently has not been recognized in sociometric research. It is suggested that perhaps all structures must be counted if the population to which the measures refer is of groups (of fixed size) studied in concrete field research. On the other hand perhaps only non-isomorphic states should be counted when the target population is an abstract anhistorical one such as is implied by experimental small-group research or research in the evolution of institutional forms of small groups.

This problem can be regarded as one aspect of the general problem of recognizing what states are distinguishable from one another and thus legitimate candidates for a set of equi-probable states.[4] In the writer's research project, in addition to sociometric questions, managers in a firm were asked to estimate the allocation by a named manager of his attention at work among four categories. Here it seemed possible that the psychological process of allocation is such that the counting of only non-isomorphic states is appropriate rather than the more familiar counting of all states. Since this case is much simpler than the sociometric one it will be used as the detailed example to bring out the issues involved in the general problem.

We concentrate on the psychological process of estimation of a typical manager, and in particular on the statistical variations of his estimates around an assumed underlying long-term estimate. Let the null hypothesis be that in some underlying large population of estimates by a manager the mean allocation of attention perceived is the fractions p_1, p_2, p_3, and p_4 for the four categories. We examine one estimate e_1, e_2, e_3, and e_4. What is the probability of obtaining e_1, e_2, e_3, and e_4 if a sample of one is drawn at random from the null population?

We assert that the estimates are discrete rather than continuous variables, so that the problem is one in finite probabilities. The spacings of the possible discrete values of the variables is an empirical question. However, we assume in our crude psychological model that the manager in making an estimate is allocating a number N of "units" of attention among the four categories; it follows that $e_i = \frac{N_i}{N}$ and each e_i has the same equidistant spacing of possible values, $\frac{1}{N}$.

It may seem that the multinomial distribution is now the right form for the answer to our question: i.e.,

$$\binom{N}{N_1,\ N_2,\ N_3,\ N_4} p_1^{N_1}\ p_2^{N_2}\ p_3^{N_3}\ p_4^{N_4}$$

is the probability of obtaining the estimate e_1, e_2, e_3, e_4, where $N_i = Ne_i$. Here N is the number of values taken by each e_i; if the manager answer in integer percentages, N = 100. This form for the answer is equivalent to the following probability model of the process of allocation: the respondent sequentially assigns his N units of estimation to the four categories, each unit having a priori probabilities p_1, p_2, p_3, and p_4 of falling in the various categories, <u>independent</u> of where preceding units have fallen. The four dimensional sample of size one is thus reduced to a one-dimensional sample of size N which is sequential. In order to make clear our objection to this sequential probability model of the psychological process of allocation, and simultaneously to prepare the ground for the model believed more appropriate, we must redescribe the multinomial distribution, in terms of the basic process of counting equi-probable states. This new description is in an abstract form, and a number of different specific empirical interpretations of it can be given.

N balls are to be distributed among Z distinct boxes. The Z boxes are divided into r groups, the j^{th} group containing Z_j boxes. In the concrete problem r would be 4, and Z is the number of micro-categories of attention assumed to exist in the respondent's thinking about attention allocation. Let N_j be the number of balls in the j^{th} group. Then

$$\sum_{j=1}^{r} N_j = N, \qquad \sum_{j=1}^{r} Z_j = Z.$$

If the balls are distinct there are $(Z_j)^{N_j}$ ways of distributing N_j balls among Z_j boxes, with no restrictions on the number of balls allowed in any one box. Thus there are $\prod_{j=1}^{r} (Z_j)^{N_j}$ ways of distributing N_1 particular balls in the first group of Z_1 boxes, and N_2 particular balls in the second group, etc. But there are

$\binom{N}{N_1, N_2, \ldots N_r}$ ways of splitting N different balls into r distinct groups such that there are N_1 balls in the first group, N_2 in the second, etc. Finally, there are

$$\binom{N}{N_1, N_2, \ldots N_r} \prod_{j=1}^{r} (Z_j)^{N_j}$$

ways of assinging N different balls to Z boxes with the restrictions that there be N_1 balls in the first group of Z_1 boxes, etc. On the other hand there are Z^N ways of putting N different balls in Z boxes with no restrictions.

If we make the fundamental assumption that each distribution of N distinct balls among the Z boxes is equally likely, it follows that

$$\frac{1}{Z^N} \binom{N}{N_1, N_2, \ldots N_r} \prod_{j=1}^{r} (Z_j)^{N_j} \qquad \text{(A)}$$

is the probability that there will be N_1 balls in the first group of Z_1 boxes, etc. Two distributions of N distinct balls in Z boxes are different if even one of the balls is in a different box.

Formula (A) is seen to be equivalent to the multinomial formula if we recast it as

$$\binom{N}{N_1, N_2, \ldots N_r} \prod_{j=1}^{r} \left(\frac{Z_j}{Z}\right)^{N_j}$$

and equate $\left(\frac{Z_j}{Z}\right)$ to the a priori probability p_j.

This interpretation of the multinomial model asserts that there are Z elementary categories of attention perceived by the respondent, Z_j of them being included in the j^{th} category of the r categories we defined as observers. The assertion that there is an a priori probability p_j of the respondent assigning one of the N units of attention to the j^{th} category is now equivalent to the assertion that he is equally likely to assign one of the N units to any of the Z elementary categories, if $\frac{Z_j}{Z} = p_j$.

We object to this multinomial probability model of the allocation process because it assumes that the N segments of the attention being allocated are distinguishable, as we can see from the derivation of formula (A), where the N "balls" are assumed distinct. This is not plausible psychologically. It is the categories of activity that are logically distinct in themselves. And the N segments of attention are not distinguishable by sequential order, since in our view the allocation by the respondent of the attention of the named manager among the four categories is not a sequential choice process.

The appropriate probability model for the attention allocation estimate in our opinion is obtained by assuming that the N units of attention are identical. The number of ways of distributing N_j identical balls among Z_j distinct boxes is $\binom{N_j + Z_j - 1}{N_j}$. Our fundamental assumption is that each of the $\binom{N + Z - 1}{N}$ ways of allocating the N indistinguishable segments of attention among the Z elementary categories is equally likely. It follows that the probability of obtaining the percentage allocation estimate e_1, e_2, e_3, e_4 is

$$\prod_{j=1}^{4} \binom{N_j + Z_j - 1}{N_j} \bigg/ \binom{N + Z - 1}{N}, \qquad \text{(B)}$$

where $N_j = Ne_j$. For the same N, Z, and Z_j, formulae B and A give entirely different probabilities for each set of N_j.

Formula (B) is often called the Bose-Einstein distribution, and the multinomial distribution in the form of (A) is often called the Maxwell-Boltzmann distribution (in both cases for zero energy differences among the Z microstates).3 It should be noted that as $Z \to \infty$ for fixed N, formula (B) can easily be shown to approach formula (A) by the use of Sterling 's approximation.

The determination of Z can be treated as an empirical problem similar to the determination of N. Our guess is that Z and N for the concrete problem of managers making allocation estimates would be approximately the same size, about 20, for any of the respondents. With such a value for Z and N it is not possible to approximate formula B by formula A or by a continuous distribution in which Z_j and N_j can be separated.

The major objection to this Bose-Einstein model is that for a given respondent there may not exist elementary categories which are psychologically meaningful and yet equi-probable in the sense given above. The specification of

$\frac{Z_j}{Z}$ is equivalent to stating a null hypothesis

for the given respondent in our new language; i.e. the p_j are replaced by Z and the set of Z_j.

The $\frac{N_j}{N} = e_j$ are of course the observed sample of one estimate.

The nature and total number Z of elementary categories of attention perceived may vary from one manager respondent to another. Whether or not this is true, the number of elementary categories Z_j assigned to the category j imposed by the observer will be expected to vary from one respondent to another: it is this variation which reflects the difference in underlying judgments as to how the named manager allocates his attention (which presumably results from the difference in experiences with the named manager). It should be made clear that the Z elementary categories are of an abstract nature not correlated with the categories imposed by the observer: the simplest example of such abstract elementary categories would be concrete segments of time such as various hours in the day. Selective recall and selective observation are two possible psychological mechanisms which could be used in specific empirical interpretations of the abstract process of allocating N units among Z micro-categories.

This example is in many ways artificial but it is hoped that it clarifies the general problem discussed initially, and that it brings out the assumption of full distinguishability which can be regarded as implicit in the multinomial distribution.

Footnotes

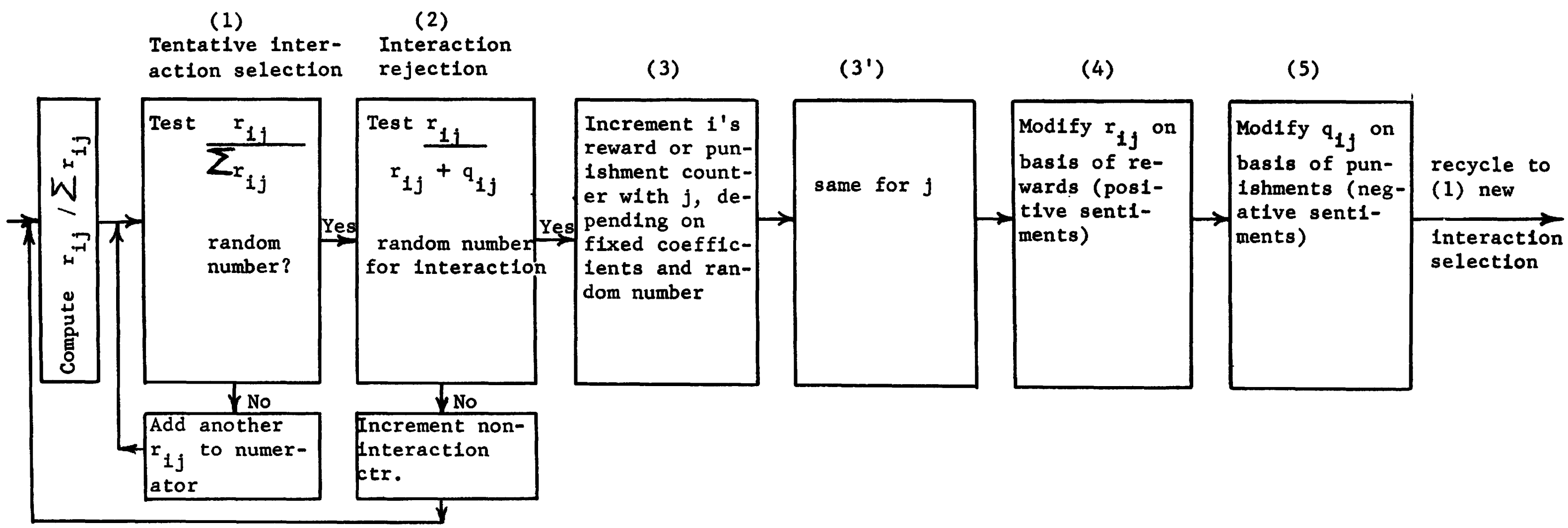
1. e.g., cf. F. Harary, "The Number of Linear, Directed, Rooted, and Connected Graphs," Trans. Amer. Math. Soc. 78 (1956), p. 445.

2. e.g., cf. L. Katz and J. H. Powell, "The Number of Locally Restricted Directed Graphs," Proc. Amer. Math. Soc. 5 (1954), p. 621.

3. W. Feller, An Introduction to Probability Theory and its Applications I (Wiley, New York) 1950, Chapter 3.

4. J. Riordan, An Introduction to Combinatorial Analysis (Wiley, New York) 1958, Chapter 6.

FIGURE 3: Flow chart for stochastic interaction within a triad, depending on rewards and punishments from past interactions.



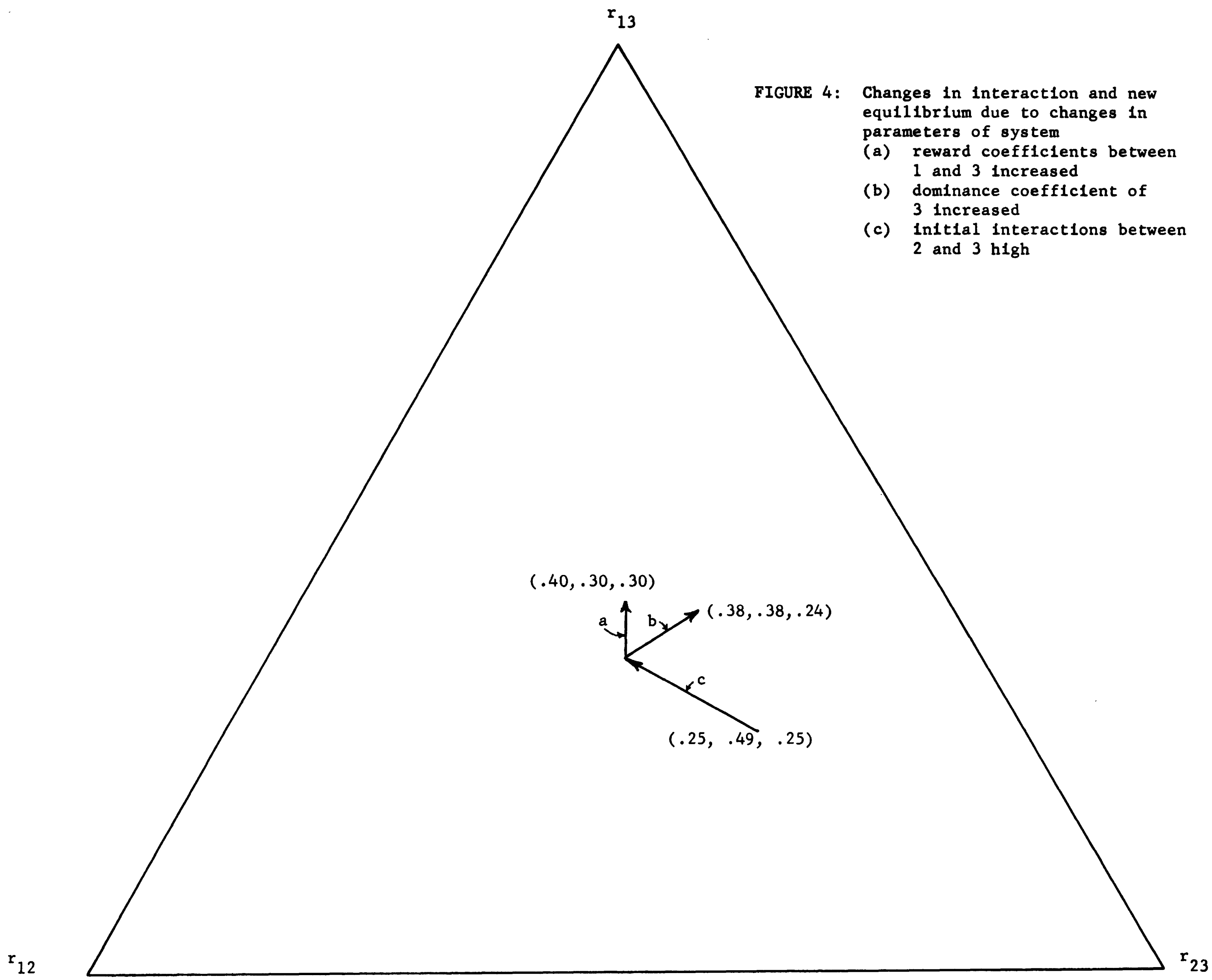


FIGURE 4: Changes in interaction and new equilibrium due to changes in parameters of system
(a) reward coefficients between 1 and 3 increased
(b) dominance coefficient of 3 increased
(c) initial interactions between 2 and 3 high

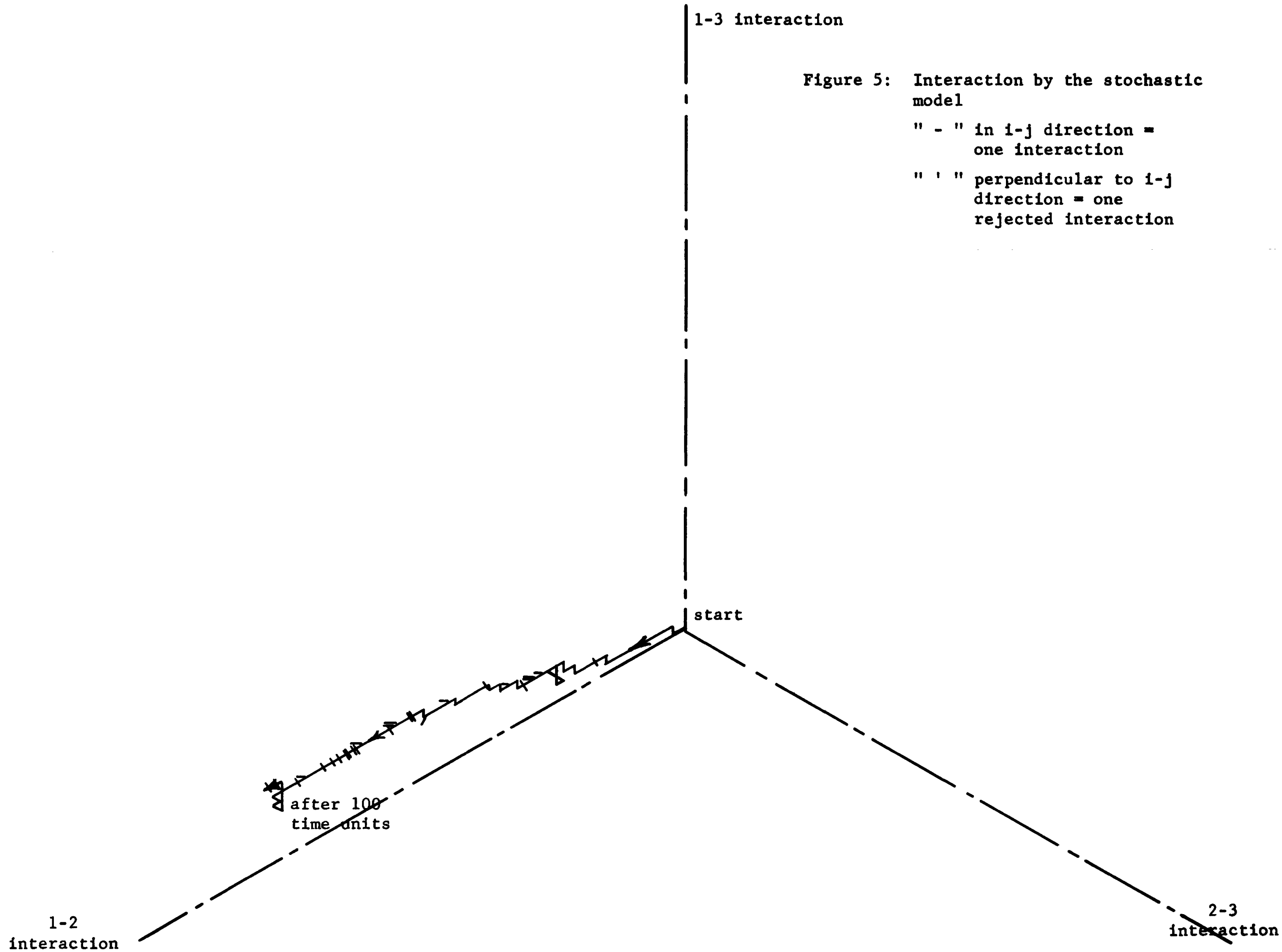


Figure 5: Interaction by the stochastic model

" - " in i-j direction = one interaction

" ' " perpendicular to i-j direction = one rejected interaction

IV

CONTRIBUTED PAPERS

Chairman, Dudley Kirk, Population Council

Sampling With Unequal Probabilities and Without Replacement—H. O. Hartley and J. N. K. Rao, Iowa State University

Empirical Analysis of Segregation Indexes—Karl E. Taeuber, National Cancer Institute

The Construction of "Social Area" Indices: An Evaluation of Procedures—James M. Beshers, Purdue University

Game Theory and Cumulative Voting in Illinois—Jack Sawyer and Duncan MacRae, Jr., University of Chicago

SAMPLING WITH UNEQUAL PROBABILITIES AND WITHOUT REPLACEMENT

H. O. Hartley and J. N. K. Rao
Iowa State University

I. Introduction:

Most survey designs incorporate as a basic sampling procedure the selection of n units at random, with equal probability and without replacement drawn from a population of N units. It is, however, sometimes advantageous to select units with unequal probabilities. For example, such a procedure may be found appropriate when a 'measure of size' x_i is known for all units in the population (i=1, 2, .., N) and it is suspected that these known sizes x_i are correlated with the characteristic y_i for which the population total

$$Y = \sum_{i=1}^{N} y_i \qquad (1)$$

is to be estimated. For example, the sales return for the past year (y_i) of a population of companies may be correlated with the (known) sales returns for the previous year (x_i). Or, again, if the total corn production (y_i) on the i^{th} farm is the characteristic whose population total Y is to be estimated this characteristic is very likely correlated with the total farm acreage (x_i) of the i^{th} farm. One method (though by no means the only method) of utilizing the auxiliary variates (measures of size) x_i is to draw units with probabilities proportional to sizes x_i (pps), a technique frequently used in surveys, particularly for the sampling of primary sampling units in multi-stage designs. Now the theory of sampling with unequal (prescribed) probabilities is equivalent to multinomial sampling provided units are drawn with replacement. On the other hand, it is well known from the theory of equal probability selection that sampling with replacement results in estimators which are less precise than those computed from samples selected without replacement, the proportional variance decrease being given by the sampling fraction $\frac{n}{N}$ (finite population correction'). It has therefore been felt for sometime that similar increases in precision should be reaped by switching to a selection without replacement in unequal probability sampling. However, the theory of sampling with unequal probability and without replacement involves certain mathematical and computational difficulties and has therefore not been fully developed.

The general theory of sampling with varying probabilities and without replacement has been first given by Horvitz and Thompson (1952). Since then, several papers have been published on this topic, but we shall review here only the papers relevant to the particular problem considered in this paper which will be stated later.

Let π_i denote the probability for i^{th} unit to be in a sample of size n. The statistic

$$\hat{Y} = \sum_{1}^{n} \frac{y_i}{\pi_i} \qquad (2)$$

is then an unbiased estimate of the population total Y whilst its variance is given by

$$\text{Var } \hat{Y} = \sum_{i=1}^{N} y_i^2/\pi_i + 2\sum_{i<j}^{N} P_{ij} y_i y_j/\pi_i \pi_j - Y^2 \qquad (3)$$

where P_{ij} denotes the probability for the i^{th} and the j^{th} unit to be both in the sample. Horvitz and Thompson have given an unbiased estimate of the variance, but it was shown by Yates and Grundy (1953) that it can often assume negative values and they proposed an alternative unbiased estimator of variance which is believed to take negative values less often, and is given by

$$\hat{\text{Var}} (\hat{Y}) = \sum_{j>i}^{n} \frac{\pi_i \pi_j - P_{ij}}{P_{ij}} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 . \qquad (4)$$

Most of the published theory is confined to samples of size n=2, owing to considerable computational difficulties involved in the extension to larger sample sizes. Now when the π_i are exactly proportional to the y_i, Var $\hat{Y}$ is zero. This suggests that if we make the probabilities π_i proportional to the 'size measures' x_i i.e. if we put

$$\pi_i = \sum_{j \neq i}^{N} P_{ij} = c\, x_i \qquad (5)$$

that a considerable reduction in Var $\hat{Y}$ will result since 'x_i' is correlated with 'y_i'. Horvitz and Thompson, for the case n=2, propose two methods to satisfy (5) approximately, but their methods have some limitations. Yates and Grundy (1953) also deal with the case n=2, introduce 'modified probabilities p_i'' and set up P_{ij} in the form

$$P_{ij} = p_i' \, p_j' \left\{ \frac{1}{1-p_i'} + \frac{1}{1-p_j'} \right\} \qquad (6)$$

Equation (6) can be realized exactly by making a first draw with probability proportional to the p_i' and a second draw with probability proportional to the p_i' of the remaining (n-1) units. To satisfy (5) they substitute (6) in (5) and solve the resulting system of N nonlinear equations for the p_i' by iteration. This method becomes cumbersome when N becomes large. Des Raj (1956) employs (5) as a set of N equations for the $\frac{1}{2}$ N(N-1) probabilities P_{ij} and determines the latter by minimising (3) subject to (5). This leads to a "linear programming problem" for the $\frac{1}{2}$ N(N-1) positive P_{ij} satisfying (5). The 'objective function' (the variance) involves the population values y_i (which are unknown) and these are replaced by the 'sizes' x_i it being assumed that

$$y_i = \alpha + \beta\, x_i \qquad (7)$$

exactly. Even if these assumptions are accepted the method is clearly unmanageable when n > 2 and /or for large N.

In this paper, we adopt a particular

procedure of drawing the sample for which it is easy to show that equation (5) is satisfied. The probabilities P_{ij} are derived directly from the sampling scheme. Although this scheme which will be described later, is well known to survey practitioners and is for example described by Horvitz and Thompson, page 678, no formulas for the P_{ij} in terms of the π_i are available in the literature, due to mathematical difficulties. These mathematical difficulties were resolved by us and compact expressions for $V(\hat{Y})$ were obtained for moderate size populations N. The sampling procedure is particularly simple for any sample size n and the asymptotic variance formulas derived permit an evaluation of the merits of both the sampling scheme as a design and the estimation (2). It should be pointed out that this method and the results (unlike the procedures previously published) cover the case of general sample size n.

The mathematical derivations of $\hat{P}_{ij}$ and $V(\hat{Y})$ in terms of the π_i will be published elsewhere. Here we shall confine ourselves to describing the sampling procedure with an example, and to stating the formulas for P_{ij}, $Var(\hat{Y})$ and $\hat{Var}(\hat{Y})$. Finally we shall give an example to illustrate the efficiency comparisons.

II. <u>The Sampling Scheme</u>:

It can be shown easily that a necessary condition for a sampling scheme to satisfy (5) is that

$$c = n / \sum_{i=1}^{N} x_i$$

and

$$\pi_i = n\, x_i / \sum_{j=1}^{N} x_j \leqslant 1 \quad . \qquad (8)$$

Henceforth, we shall only consider such 'sizes x_i' and associated probabilities $p_i = x_i / \sum_{j=1}^{N} x_j$ which satisfy the necessary condition (8). The following sampling scheme is now considered:

a. Arrange the units in random order and denote (without loss of generality) by j=1, 2,.., N this random order and by

$$\Pi_j = \sum_{i=1}^{j} \pi_i \, , \; \Pi_o = 0 \qquad (9)$$

the progressive totals of the π_i in that order.

b. Select a 'random start' i.e. select a 'uniform variate' x with $0<x<1$. Then the n selected units are those whose index, j, satisfies

$$\Pi_{j-1} < x + k \leqslant \Pi_j \qquad (10)$$

for some integer k between 0 and (n-1). Since $\pi_i \leqslant 1$ every one of the n integers k=0, 1, .., n-1 will 'select' a different sampling unit j.

<u>Numerical Example</u>:

Consider the population of N = 8 units j=1, 2, .., 8 arranged in random order and with 'sizes x_j' shown in the second column of Table 1. A sample of n=3 units is to be drawn with probabilities proportional to size (pps) and without replacement. The 'total size' is

$$X = \sum_{j=1}^{8} x_j = 300 \quad .$$

Instead of computing the $p_i = x_i/X$ and $\pi_i = 3p_i$ we scale all computations up by a factor of $X/n = 300/3 = 100$. Thus we compute the progressive sums of the x_i and these are shown in column three of Table 1. and correspond to the quantities $(X/n)\Pi_j$ defined by (9). Then we select a random integer (start) between 1 and (X/n) i.e. between 1 and 100

Table 1. An example of the selection of n = 3 units from a population of N = 8 units with probabilities proportional to size and without replacement.

Unit number	Size	Progressive sum	Start=36 Step=X/n=300/3=100
j	x_j	$300\,\Pi_j =$	
1	15	15	
2	81	96	k=o,100x=36
3	26	122	
4	42	164	k=1,100x+100=36
5	20	184	
6	16	200	
7	45	245	k=2,100x+200=236
8	55	300	

and this corresponds to the quantity X x/n. In our example this integer turned out to be 36 and the selection of the three units in accordance with (10) is shown in the 4th column: We must find the lines(j) where the column $300\,\Pi_j$ passes through the levels 100x = 36(for k=o), 100x + 100 = 136 (for k=1) and 100x + 200 = 236 (for k=2). The units j=2, 4, and 7 are thereby selected. This procedure (either with or without the initial randomization a.) has been frequently used but, in the absence of a better theory, is usually treated approximately as a pps sample drawn <u>with</u> replacement.

It can be easily proved that for this sampling scheme, for <u>any</u> ordered arrangement of the N units

$$\Pr\left\{ j^{th} \text{ unit in sample} \right\} = \pi_j \qquad (11)$$

thus satisfying condition (5). It may be remarked that the randomization of the arrangement in step 'a' of the scheme is not required to prove (11). However this is required for obtaining a variance formula for the estimate of Y which does not depend on any particular arrangement of the units.

From equation (3), we see that in order to evaluate Var $\hat{Y}$ in terms of π_i's, we have to find P_{ij} in terms of the π_i . Now we have succeeded in evaluating directly from the sampling scheme the leading term of P_{ij} as well as the term which is of next lower order of magnitude in powers of N^{-1}. This second term represents the gain in precision due to the finite population correction. In particular we have shown that to order N^{-3}, for the case n=2,

$$P_{ij} = \frac{N-2}{N-1}\frac{\pi_i \pi_j}{2-\pi_i-\pi_j}\left(1-\frac{S^2(N-2)}{(2-\pi_i-\pi_j)^2}\right) \qquad (12)$$

where

$$S^2 = \frac{1}{N-3}\left(\sum_{t=1}^{N}\pi_t^2-\pi_i^2-\pi_j^2\right) - \frac{(2-\pi_i-\pi_j)^2}{(N-2)} \qquad (13)$$

By substituting P_{ij} from equation (12) into (3) and simplifying, we obtained

$$Var(\hat{Y}) = \sum_{i=1}^{N}\pi_i\left(1-\frac{\pi_i}{2}\right)\left(\frac{y_i}{\pi_i}-\frac{Y}{2}\right)^2 \qquad (14)$$

to order N^{-1}. This formula is satisfactory for moderately large populations N. Further improvements in $Var(\hat{Y})$ have been made by taking terms of order N^0 also into account, but these details will not be given here. We notice that for sampling with replacement,

$$Var'(\hat{Y}) = \sum_{i=1}^{N}\pi_i\left(\frac{y_i}{\pi_i}-\frac{Y}{2}\right)^2 \qquad (15)$$

so that sampling without replacement and pps is more efficient than sampling with replacement and pps, since the weight factors $1-\frac{\pi_i}{2}$ in (14) are all less than 1.

In order to find an estimate of the variance of $\hat{Y}$, we substitute P_{ij} from equation (12) in equation (4) and after simplification we obtain to order N^{-1},

$$\hat{Var}(\hat{Y}) = \left(1-(\pi_1+\pi_2)+\frac{\sum_{1}^{N}\pi_t^2}{2}\right)\left(\frac{y_1}{\pi_1}-\frac{y_2}{\pi_2}\right)^2. \qquad (16)$$

A check on our formulas is obtained by making all π_i equal i.e. equal to $\frac{2}{N}$ when it will be seen that, to order N^{-1}, equations (14) and (16) reduce to the well known formulas for the equal probability case available for the variance of $\hat{Y}$ and for the estimate of variance of $\hat{Y}$. In the case of general n, we have shown that

$$Var(\hat{Y}) = \sum_{1}^{N}\pi_i\left(1-\frac{(n-1)}{n}\pi_i\right)\left(\frac{y_i}{\pi_i}-\frac{Y}{n}\right)^2 \qquad (17)$$

to order N^{-1} assuming n is relatively small compared to N. The details of the case of general n will be given in a seperate paper.

III. A numerical example for the evaluation of the variance formulas:

Horvitz and Thompson (1952)(pp. 681-3) give an example of a small population of size N=20 for which the data are reproduced in Table 2. below: For N=20 blocks in Ames, Iowa, are given

y_i = Number of households in i^{th} block
x_i = 'Eye-estimate' of number of households in i^{th} block

Table 2. Data for population of size N=20

i=	1	2	3	4	5	6	7	8	9	10
y_i=	19	9	17	14	21	22	27	35	20	15
x_i=	18	9	14	12	24	25	23	24	17	14
i=	11	12	13	14	15	16	17	18	19	20
y_i=	18	37	12	47	27	25	25	13	19	12
x_i=	18	40	12	30	27	26	21	9	19	12

Now the probability π_i for the i^{th} unit to be in the sample is taken proportional to the 'eye-estimated' number of households x_i i.e. we take

$$\pi_i = 2x_i \Big/ \sum_{j=1}^{20} x_j \quad . \qquad (18)$$

In Table 3. below we give the evaluation of the variance of our estimator from formula (14) correct to order N^{1} and shown below this is the value obtained from an improved formula (not shown here) which is correct to order N^0. Above this value we give the variance formula for sampling with probabilities proportional to the π_i but with replacement. And finally (on top) the variance of $N\bar{y}$ i.e. of the estimator when sampling is with equal probabilities and without replacement.

Table 3. Variances of various estimates of the total of the y_i population shown in Table 2.

Sampling Scheme	Form of Estimator	Numerical value of variance of estimator
Equal probability sampling without replacement	$N\bar{y}$	16,219
Probabilities proportional to size x_i, with replacement	$\sum_{i=1}^{2} y_i/\pi_i$	3,241
Probabilities proportional to size, xi; without replacement	$\sum_{i=1}^{2} y_i/\pi_i$	3,025 (14) 3,007

A comparison of the variances in Table 3. shows that sampling proportional to an approximate measure of size is vastly superior to sampling with equal probabilities. It must not be forgotten, however, that there are other devices of decreasing the variance in the latter ease with the help of the known x_i values: Ratio and regression estimation, for example, may be employed. About 7 per cent (235/3241) are gained in precision through sampling without replacement. The two results for our variance viz. 3025 (correct to order N^{1}) and 3007 (correct to order N^0) are in good agreement and illustrate the convergence of our formulas.

References

1. Des Raj (1956), "A Note On the Determination of Optimum Probabilities in Sampling Without Replacement," Sankhya, p. 197.
2. Horvitz, D. G. and Thompson, D. J. (1952), "A Generalization of Sampling Without Replacement From a finite Universe", Journal of the American Statistical Association, p. 663.
3. Yates, F. and Grundy, P. M. (1953), "Selection Without Replacement from Within Strata With Probability Proportional to Size", Journal of the Royal Statistical Society, Series B, p. 253.

Editorial Note: The authors inform us that they are submitting a paper giving full details of the mathematical proofs to the ANNALS OF MATHEMATICAL STATISTICS.

EMPIRICAL ANALYSIS OF SEGREGATION INDEXES (1)

By: Karl E. Taeuber, National Cancer Institute

Residential segregation of whites and nonwhites exists in all U. S. cities. Obviously no one measure can represent, adequately and for all purposes, the patterns of residential segregation in a city. Several "segregation indexes" have been proposed to serve as general summary measures (2). This paper reports on some empirical comparisons of these indexes.

A distinction may be maintained between the behavioral processes involved in residential segregation, and the resulting spatial distribution of white and nonwhite residents. The various segregation indexes share a general approach in the specification of a measurable aspect of the spatial distribution. Several aspects of the spatial patterns might be of interest. These include the extent to which white or nonwhite residents are clustered together, the location within the city or metropolitan area of such clusters, and the general evenness in the distribution of white and nonwhite residents throughout the city. The segregation indexes share a common focus on this latter aspect.

For purposes of measurement, the concept of general evenness in the spatial distribution may be clarified by defining the situations of complete evenness and complete unevenness. Consider a city divided into subareas, such as tracts, wards, or blocks. If each subarea is occupied entirely either by whites or by nonwhites, and no subarea contains residents of both colors, the spatial distribution is said to be completely uneven. A value of 1 may be assigned to this situation, representing maximum residential segregation. If in each subarea nonwhite residents constitute the same proportion of all residents as they do in the entire city, the distribution is said to be completely even. A value of 0 may be assigned to this situation, representing minimum residential segregation.

All of the proposed segregation indexes derive from this general approach. The basic data from which any of the indexes may be computed are the numbers of white and nonwhite residents in each subarea. There are two ways in which indexes computed for a given city may yield different values. The subareas utilized for computation may differ, or the formulas for combining the distributional data into a summary figure may differ. The indexes are all measures of the same general concept, are computed from the same type of data, and share common maximum and minimum values. How important are the differences between index values that may arise from the two sources noted?

It is readily apparent that the choice of subarea will affect the value of a particular index. In general, the smaller the subarea, the greater the unevenness that will be noted, and the higher the index value. For instance, nonwhite residents might occupy approximately ten percent of the dwellings in each census tract, and yet within each tract be clustered together in block occupied only by nonwhites. Such a situation could result in a tract index close to 0 and a block index of 1.

Census tracts and blocks are the two subareas for which data are most readily available. The use of census tracts for computing segregation indexes has been criticized (3), but the general arguments may apply to any areal unit. The boundaries of an areal unit may or may not be useful in delineating residential areas so as to indicate the full extent of residential separation. In delineating census tracts, some tract committees have tried to make tract boundaries correspond with boundaries between white and nonwhite residential areas. Other tract committees have paid relatively little attention to race. In either case, over time the degree of correspondence between tract boundaries and boundaries of white and nonwhite settlement may shift. Clusters of nonwhite occupied dwellings may overlap tract boundaries, or be entirely included within individual tracts. Hence it has been argued that the large size and the arbitrary delineation of tract boundaries mitigate against their use for the computation of segregation indexes. Except that blocks are not generally delineated by committees, the same arguments apply. If nonwhites tend to live in alleys or on side streets, while whites live along the main street fronts, then sole reliance on block data may also be misleading.

Hypothetical argument cannot indicate the empirical importance of the areal factor. It is plausible to argue that extensive concern with the relative merits of blocks and tracts may not be necessary. Although the specific magnitudes of tract and block indexes differ, how much effect does the choice of areal unit have on the comparative rankings of a number of cities?

Comparisons were made for each of 6 different segregation indexes. For five of the indexes, comparisons are between values computed from tract data and those computed from block data for 60 cities, using 1940 data. For the sixth index, the comparisons are based on data for 72 cities for 1950 (4). Table 1 gives the product-moment correlation coefficients between the block and tract values for each of the six indexes. Because tracts are fewer in number, tract indexes are more easily computed. Hence it might be desired to use tract indexes for estimating the values of block indexes. Table 1 also includes the parameters of the regressions of block on tract indexes.

For five of the indexes, there is a definite linear relationship between the block and tract values. The sixth index, Co, is the proportion of all whites in the city who reside in subareas that are exclusively white-occupied. If segregation is defined with reference only to this one aspect of the spatial distribution, the index values and interpretations from them will differ markedly according to the type of subarea used in the computations.

The three indexes, Gh, Rep, and Bell, have in common a distinct dependence on the city nonwhite proportion. The proportion of nonwhites in a city remains the same, whether blocks or tracts are used for computing the indexes. This dependence on a common factor is a partial explanation of the high correlations between block and tract values of these indexes.

The indexes, Gi and D, are, conceptually, the most adequate measures of overall evenness, without specific reference to the proportion of nonwhites. The correlations of block and tract values for these indexes, .74 and .69, indicate only a moderate relationship.

The deviations of particular cities from the regressions of block on tract indexes vary with the index. The differences between block and tract values are functions of the index as well as of the spatial distribution. For cities with large numbers of tracts, the block indexes are only slightly more highly correlated with tract indexes than is the case for cities with few tracts.

For those indexes with a strong dependence on the proportion of nonwhites in the total population (Gh, Rep, Bell), the difference between block and tract values are relatively slight, and intercity comparisons based on either subarea will yield similar results. For indexes that are pure measures of unevenness (Gi, D, Co), the values computed from tract data are an unreliable guide to those computed from block data.

The conclusions derived from studying the rank correlations of tract and block indexes are essentially the same. For gross intercity comparisons of the level of segregation on a given index, block or tract values may be used without greatly affecting the comparisons (except for Co). For more precise analysis, the conclusions will depend on whether tracts or blocks are chosen as subareas for index computation.

The segregation indexes differ in the items of information they utilize and in the formulas by which values between 0 and 1 are assigned to various residential distributions. The Duncans have reported on some of the index properties and on certain algebraic interrelations among the indexes. Thus, some of the indexes utilize the proportion of nonwhites in the city population, while others are pure measures of unevenness in that they utilize only the data on the spatial distribution. One index, the original Cowgill index (Cwg), is distinctive in the definition of maximum and minimum segregation. The extreme distributions are rarely encountered, however, and such a conceptual difference might not result in an empirical difference between the Cowgill index and other indexes. The general question may be raised whether the various differences between the concepts and definitions of the indexes are of significance in empirical investigation. Will a city be found to have approximately the same degree of residential segregation, as compared to other cities, regardless of which segregation index is used?

Table 2 gives the intercorrelations of 8 segregation indexes computed from 1950 block data for 188 cities. The numerous low coefficients indicate that the indexes are not simply minor variants of each other. Rather, the algebraic differences noted by the Duncans tend to be evident in the pattern of intercorrelations.

The first column of the table shows the correlations between each index and the proportion of nonwhites in the city population. The coefficients are positive and moderately high for the three indexes (Gh, Rep, Bell) directly utilizing the proportion of nonwhites in their formulas. Two pure measures of spatial unevenness (Gi, D) have very little relationship to the proportion of nonwhites. This difference between the indexes is obviously of major importance in any substantive analysis of residential segregation.

The index which deviates from the others in the definition of maximum and minimum segregation (Cwg) is not highly correlated with any of the other indexes. The three indexes, Gh, Rep, and Bell, are highly intercorrelated with each other. The nearly perfect linear relationship between Gh and Bell confirms the impression that the Bell index is of the same general type as the other two.

The indexes Co and Oc, although pure measures of spatial unevenness, each depend on one aspect of the distribution. They are empirically distinctive from each other and from each of the other indexes. The other pure measures, Gi and D, behave quite similarly. If only those cities with large Negro populations are considered, the similarities between Gi and D are greater.

Although the indexes are not empirically identical, they are not 8 completely independent measures. Using tract data, the Duncans have reported that observed values of D and the proportion of nonwhites could be used to make quite good estimates of Gi, Gh, Rep, and Bell. It is clear from Table 2 that good predictions could be made of some of the block indexes, given values for the others. The proportion of nonwhites and indexes D and Cwg are relatively independent of each other. Without precise evaluation, it is clear that linear combinations of these three could be used to estimate values of the other six indexes. At least 60 percent of the variance in Oc could be accounted for, and better predictions would result for the others.

Another means of analyzing the interdependence of the nine measures is factor analysis. A brief exploration was carried out, using the Turinsky direct factor program for Univac. The raw data were used, and the resulting factors are not readily identifiable. However, the patterning of the factor loadings on the indexes for the four extracted factors confirms the previous analyses. Gi and D have similar loadings. Cwg stands by itself, except for similar loadings to Co on two factors. Co is close to Gi and D, the other pure measures, on two factors. Cc, however, tends to be similar to the proportion of nonwhites. Bell, Gh, and Rep form a group with similar factor loadings.

The preceding analyses of the empirical behavior of segregation indexes demonstrate very important differences between them. Values of the indexes computed for U. S. cities may differ because the areal units from which they are computed differ, or because of differences in the ways the areal data are utilized. The results of any substantive analysis utilizing segregation indexes will therefore depend both on the index chosen and on the areal unit data from which it is computed.

For some empirical investigations of residential segregation, tract data may have particular advantages over block data. For inter-city analysis of the spatial evenness in the residential distribution, however, block data are generally preferable. Blocks are smaller and less subject to arbitrary delineation. These properties increase the comparability of indexes computed for different cities. Although perfect comparability is not possible with indexes based on areal data, blocks are of sufficiently fine scale that the limitations to comparability are not serious. In addition, block data are more sensitive than tract data to small variations in the evenness of the residential distribution. Block boundaries are more stable over time. Block data are available for both 1940 and 1950 for more cities, although tract data are available for some entire metropolitan areas.

The choice of a particular index must be based on the purposes of the investigation. For most analyses, it is desirable to study separately the spatial evenness and the proportion of nonwhites. The compounding of these two dimensions in a single index is thus an undesirable property. Of the pure measures of evenness, it has already been noted that Gi and D are the most adequate central measures of overall evenness. Although the Gini index is more sensitive to all parts of the spatial distribution, the dissimilarity index (D) is empirically quite similar and is more easily computed. For intercity comparisons of spatial unevenness in the distribution of white and nonwhite residents, either the Gini index or the dissimilarity index, computed from block data, is likely to be the analytically most useful choice (5).

Table 1. Block and Tract Indexes of Residential Segregation between Whites and Nonwhites: Means, Correlations, and Regressions.

	Index and Year					
Measure	Gi 1940	D 1940	Co 1940	Gh 1940	Rep 1940	Bell 1950
Number of Cities	60	60	60	60	60	72
Mean of Block Indexes	.96	.86	.74	.70	.51	.64
Mean of Tract Indexes	.82	.70	.08	.50	.29	.44
Correlation of Block and Tract Indexes	.74	.69	.63	.92	.90	.93
Regression of Block on Tract Indexes						
Slope	.18	.31	.78	.70	.93	.89
Intercept	.81	.64	.68	.35	.24	.25

Table 2. Intercorrelations of Eight Block Indexes of Residential Segregation between Whites and Nonwhites and the Proportion of Nonwhites, 188 Cities, 1950.

	q	Gi	D	Co	Oc	Gh	Rep	Bell
Gi	.19							
D	-.01	.90						
Co	-.50	.58	.75					
Oc	.76	.51	.32	-.13				
Gh	.69	.49	.13	-.32	.75			
Rep	.79	.52	.22	-.31	.85	.95		
Bell	.72	.52	.19	-.30	.81	.99	.98	
Cwg	.20	.55	.05	-.04	.34	.70	.49	.63

For index definitions and references, see Footnote (2). The letter q represents the proportion of nonwhites to total city population.

Footnotes:

(1) This is Paper No. 5 in the series, "Comparative Urban Research", issuing from the Population Research and Training Center, University of Chicago, under a grant from the Ford Foundation. Computation of the indexes was supported by the Laboratory of Social Relations, Harvard University. The research was performed under a National Science Foundation Predoctoral Fellowship.

(2) For definitions of the indexes, analysis of their algebraic interrelations, additional discussion of some of the problems considered here, and references to other sources, see: Otis Dudley Duncan and Beverly Duncan, "A Methodological Analysis of Segregation Indexes," American Sociological Review, 20(1955), 210-217. Cwg, the Cowgill index, is defined in the reference in footnote (3). Oc is a variation of the other indexes, and is equal to the proportion of all nonwhites in the city who reside in subareas occupied exclusively by nonwhites.

(3) Donald O. Cowgill and Mary S. Cowgill, "An Index of Segregation Based on Block Statistics," American Sociological Review, 16(1951), 825-831.

(4) The five indexes based on 1940 tract data were computed by the Duncans. Community Areas were used for Chicago and Health Areas for New York. The Bell indexes for 1950 tract data are given in: Wendell Bell and Ernest M. Willis, "The Segregation of Negroes in American Cities," Social and Economic Studies, 6(1957), 59-75. All block indexes except the Cowgill index were computed by the author; see the reference in (5).

(5) The research reported on here is discussed in greater detail in the author's unpublished doctoral dissertation, "Residential Segregation by Color in United States Cities, 1940 and 1950," submitted to Harvard University, December, 1959.

THE CONSTRUCTION OF "SOCIAL AREA" INDICES: AN EVALUATION OF PROCEDURES[1]

By: James M. Beshers - Purdue University

Shevky, Bell and Tryon have described several procedures for constructing "social areas" from census tract data.[2] These procedures all consist of reducing the demographic characteristics of census tracts to a relatively small number of "dimensions." Various statistical techniques may be employed in this reduction - factor analysis, cluster analysis, Guttman scaling, etc.

Evaluation of these procedures has been undertaken by several authors recently.[3] Most of these investigations consist of replications of "social area" procedures on a number of cities. Evaluation of the procedures is therefore based on the empirical results of using them. A consistency, or reliability, criterion for the procedures is implicit in this work. But one cannot determine from such studies whether lack of consistency reflects deficiencies in the procedures themselves, or variation in social characteristics among U.S. cities.

An alternative approach is to examine the procedures themselves. The mathematical properties of these procedures are all well known - a comparison on this basis is unnecessary. However, one may examine critically the use of the procedures in this particular empirical problem. To do so, one must follow through the details of applying these procedures to at least one empirical case.

The present study is directed to this alternative approach using data from Cleveland, Ohio, in 1950.[4] The results reported here are part of a larger study of statistical procedures used in manipulating census tract data. The choice of Cleveland and the other constraints imposed by the larger study design do not appear to the author to have distorted the results of this part of the research.

STUDY DESIGN

In this study twenty variables were selected from the census tract data. The relationships between these variables were used as a criterion for reducing the twenty variables to a smaller number of constructs - thereby eliminating redundancy. Subsequent calculations are simplified by obtaining a small number of constructs. Further, the theoretical significance of the raw data is summarized in the constructs; thus the subsequent analysis will have direct theoretical relevance. Indices for these constructs will be chosen.

On the basis of exploratory calculations, eighteen tracts were eliminated from the analysis of the 1950 data. Fifteen tracts had less than six hundred inhabitants, while three tracts had large institutional populations.

An intercorrelation matrix was computed for all twenty variables over the 331 remaining tracts. Two techniques were used to analyze the correlation matrix. An effort to reveal the effects of a single factor was made with Guttman s simplex approach.[5] This approach consists of rearranging the correlation matrix so that the largest correlations are all next to the main diagonal; the correlations should decrease in the successive diagonals away from the main diagonal and achieve the smallest value in the extreme corner of the matrix. Other methods of inspecting the correlation matrix were used to supplement the rearrangement of the columns. Table 2 contains the analysed matrix. In order to clarify the relationships a number of the variables have been reversed. For example, X_3, percent non-white, becomes percent white in order to emphasize the positive relationship with the socio-economic status indicators.

The correlation matrix was also analysed by factor analysis. The principal components technique was used. Communalities were estimated by calculating the inverse matrix and using the diagonal elements to obtain the square of the multiple correlation coefficient of a particular variable on the other nineteen variables. Guttman has shown that the communality is an upper bound for the multiple correlation.[6] Therefore the multiple correlation, as a lower bound for the communality, may be used as an estimate of the communality. The rotated factors are reported in Table 3.

MATRIX ANALYSIS

A detailed examination of Tables 2 and 3 will reveal a surprising correspondence between the results of the two methods of analysis. The arrangement of Table 2, following the simplex rule, brings out the dominance of socio-economic status. The variables exhibiting the strongest effect of this dimension are at the left, arranged in approximate order of decreasing effect. If only the first ten variables are examined as a group, an approximate simplex can be seen. With the exception of X_{17}, the large correlations are near the main diagonal, and a general decline in correlations can be seen away from the main diagonal. The next four variables in Table 2 have a general positive relation with the social status group, but no simple pattern is revealed by the magnitude of their coefficients.

The last six variables have little or no consistent relationship with the major socio-economic variables. Four of these, however, have large positive intercorrelations. They seem to represent a life cycle stage in residential areas.

These two main effects in Table 2 have been headed A and B. A further phenomenon of interest can be uncovered by examining the correlations for X_3, the race index. There seems to be a moderate average correlation with the social status group, about .4, but there are high correlations with X_{14} and X_{12} among the social status indicators. That is, non-white is correlated

TABLE 1

List of Variables, Designations when Reversed, and Their Derivation from Raw Data

X_1--Males (females)--Percent males--male/total population

X_2--foreign born white (native born white)--percent foreign born white--foreign born white/total white

X_3--race (white)--percent nonwhite--nonwhite/total population

X_4--large family (small family)--population per household

X_5--education--median school years completed

X_6--same house--percent same house in 1949--same house in 1949 as 1950/persons one year old and over, 1950

X_7--income--median family income

X_8--children--percent males under 15 years--males under 15 years/total males

X_9--old people (not old)--percent males over 65 years--males over 64 years/total males

X_{10}--males married--percent males married--married males/males 14 years old and over

X_{11}--females working (females not working)--percent females in labor force--females in labor force/females 14 years old and over

X_{12}--unemployment (employment--percent males unemployed--unemployed males/males 14 years old and over

X_{13}--professional--percent males professional--male professionals/employed males

X_{14}--laborers (not laborers)--percent males laborers/employed males

X_{15}--owner occupied--percent owner occupied dwelling units--owner occupied dwelling units/all dwelling units

X_{16}--one dwelling unit--percent separate dwelling units--one dwelling unit, detached/all dwelling units

X_{17}--old dwelling units (new dwelling units)--percent dwelling units built 1919 or earlier--dwelling units built 1919 or earlier/number reporting age

X_{18}--crowding (not crowded)--percent 1.01 or more persons per room--number dwelling units 1.01 or more persons per room/number reporting persons per room

X_{19}--median rent--median contract monthly rent with estimates from X_{20}

X_{20}--median value of dwelling units--median value of owner occupied one-dwelling-unit structures with estimates from X_{19}

TABLE 3

First Four Factors: Rotated and Interpreted

	Social Status	Young Family	Females Stability	Race
X_5	.65	.23	-.07	.31
X_{13}	.79	.31	.01	.38
X_{20}	.88	.16	-.01	.29
X_{19}	.77	.22	-.07	.42
X_7	.85	.03	.12	.01
X_{14}	.79	.40	-.04	-.28
X_{12}	.83	.04	.12	-.28
X_{17}	.79	-.29	-.17	.10
X_{16}	.78	-.34	-.25	-.14
X_{15}	.83	-.29	-.04	-.26
X_{18}	.66	.47	.20	-.26
X_{10}	.72	-.39	.21	.05
X_2	.61	.04	-.29	.04
X_3	.58	.38	-.00	-.61
X_1	.32	-.12	.66	.29
X_{11}	.54	-.62	-.05	-.09
X_8	.22	-.85	-.07	-.04
X_9	.13	-.70	-.19	.02
X_4	.01	.87	-.06	-.05
X_6	.09	-.26	.78	-.30

TABLE 2

Intercorrelation Matrix of Twenty Tract Variables with Derived Constructs

	X_5	X_{13}	X_{20}	X_{19}	X_7	X_{14}	X_{12}	X_{17}	X_{16}	X_{15}	X_{18}	X_{10}	X_2	X_3	X_1	X_{11}	X_8	X_9	X_4	X_6
	A: General Social Status										C					B				
	A_1					A_2		A_3												
	Social Status					A ≠ C		A ≠ B			A_2≠C	A ≠B	A	Race	Only X_{10}≠X_6	Young Family				Only X_1≠X_4
	Ed	Prof	DU Val	Rent	Inc	Not Lab	Emp	New DU	One DU	Ownr Occ	Not Crd	M Mar	Nat Wh	Wh	Fem	F Not Wrkg	Chil-dren	Not Old	Large Fam	Same House
X_5		.89	.84	.82	.70	.70	.63	.62	.53	.53	.58	.54	.64	.42	.32	.29	.04	-.02	-.21	-.15
X_{13}			.88	.84	.70	.60	.54	.55	.47	.46	.57	.40	.52	.36	.31	.25	-.08	-.18	-.22	-.09
X_{20}				.86	.79	.63	.65	.71	.63	.65	.59	.53	.49	.37	.29	.37	-.00	-.04	-.08	-.03
X_{19}					.71	.56	.52	.65	.48	.48	.48	.46	.46	.27	.26	.26	-.05	-.05	-.18	-.17
X_7						.66	.67	.65	.64	.72	.61	.59	.42	.49	.28	.46	.12	.03	.02	.19
X_{14}							.77	.48	.45	.57	.80	.47	.52	.81	.18	.18	-.11	-.10	-.39	-.08
X_{12}								.57	.62	.73	.69	.63	.50	.67	.29	.43	.15	.10	-.07	.23
X_{17}									.72	.74	.34	.67	.45	.28	.16	.61	.39	.37	.23	.00
X_{16}										.91	.34	.57	.51	.35	.07	.65	.40	.30	.32	.04
X_{15}											.50	.63	.41	.46	.16	.61	.32	.27	.30	.22
X_{18}												.35	.29	.68	.19	.06	-.30	-.26	-.40	.16
X_{10}													.42	.23	.54	.61	.51	.46	.24	.23
X_2														.39	.12	.31	.23	.14	-.17	-.22
X_3															-.04	.16	-.08	-.22	-.41	.16
X_1																.13	.16	.07	.09	.45
X_{11}																	.70	.40	.55	.23
X_8																		.68	.71	.19
X_9																			.54	-.04
X_4																				.28
X_6																				

The effect of a general status factor can be seen by the magnitude of correlations reading from left to right, a modified Guttman Simplex. However, note the composites, A_2 and C link together X_{14}, X_{12}, X_{18}, and X_3; while A_3 and B link together X_{17}, X_{16}, X_{15}, X_{10}, X_{11}, X_8, X_9, and X_4.

with laborers and unemployment more than might be expected from a general social status effect. This result is in keeping with the findings of sociological research with individuals. Further, the high correlation with X_{18} indicates that non-white is associated with crowding. The designation C has been given to race in the headings of Table 2.

A linkage between A and B may be discovered by scanning the correlations in B. While most of the correlations of B with A are negative, the average correlation of B with X_{17}, X_{16}, and X_{15} is above .3. This linkage of life cycle stage with characteristics of dwelling units is not strong, but it is significant information with regard to the nature of the young family group in the matrix. A similar linkage can be seen between B and X_{10}.

Only two variables, X_1 and X_6, have little to do with A, B, and C. In fact, they have almost as high a correlation with each other as they do with any other variable. Further, there is no clear theoretical reason for the relationship between these two variables, suggesting that they may be excluded from subsequent consideration.

The combined simplex and subsequent analysis suggests that three constructs may represent the relationships in the correlation matrix - socio-economic status, young family life cycle stage, and race. The first and last of these three are familiar constructs in social structure analysis. The life cycle stage is not readily assimilated to social structure, although it may be seen as a status associated with the nuclear family. The life cycle stage may be regarded as a construct for urban social structure analysis, but further research and development of theory are necessary if its relevance to social structure in general is to be discovered. One reason for the neglect of this construct is suggested by the correlation matrix. Except for the relationships discussed above, B is almost independent of both A and C.

The same constructs are evident in the factor analysis. Five factors, representing 97.4 percent of the total estimated communality, were considered for rotation. The first two factors, representing 53.6 percent and 22.1 percent of the total estimated communality, were identifiable without rotation. Reference to Table 3 indicates that these are constructs A and B from the previous analysis. The positive value of X_4 on the second factor is due to the fact that it was reversed in this analysis. The third and fourth factors, representing 9.3 percent and 7.6 percent of the total estimated communality, were clarified by an orthogonal 45 degree rotation with reference to each other. The third factor consists of the relationship between X_1 and X_6. The fourth factor represents race predominantly, with scattered relationships with the other variables. The fifth factor, representing 4.8 percent of the total estimated communality, had low factor loadings on all variables, and showed little possibility of clarification by rotation. It was dropped from the analysis.

COMPARISON WITH OTHER STUDIES

The results of this construct formation process for Cleveland are quite similar to results obtained by Shevky and Bell for Los Angeles, and Tryon for San Francisco.[7] However, there are important differences in detail. The Shevky and Bell constructs of economic status and family status are very similar to the socio-economic group and the young family group isolated in the analysis above. However, the mode of presentation and the method of analysing the relationships between the Shevky and Bell constructs obscures the relationships between the constructs. The linkages reported between A and B in Table 2 are evidently unknown to these investigators. Bell uses factor analysis with a greatly reduced matrix of variables so that these results cannot be seen clearly.[8]

The Shevky and Bell ethnic status construct includes nationality as well as non-white as a component. The analysis for Cleveland shows that foreign-born white and non-white are about equally related to socio-economic status, but the correlation between the two is only .39, suggesting that they should not be combined. Again, the linkages between A and C are not reported by Bell.

Tryon's constructs differ largely from Bell's in that economic status has been replaced by socio-economic independence and socio-economic achievement. Since the correlation between these two constructs is .9, one may question the usefulness of separating them. Otherwise the socio-economic status and young family group found in the present analysis are equivalent to Tryon's constructs of combined socio-economic status indicators and family life. Tryon's assimilation construct includes race along with other related variables, and therefore is roughly comparable to the single item of race.

The linkage between race and laborers in Table 2 may be seen in the components of assimilation. Other linkages from Table 2 may also be found by cluster analysis; correlations may be found between clusters, e.g., family life is correlated .3 with assimilation. But in two respects the method of cluster analysis obscures the relationships between the variables. First, cluster analysis makes no provision for reordering the matrix to demonstrate the effect of one general factor. Because of this lack Tryon's correlation matrix cannot be scanned intelligently; Cluster I, Cluster V, and Cluster VII are all made up of socio-economic items and are very similar, yet they are scattered from one end to the other of the correlation matrix, as their numerical titles suggest.[9] Second, the only method of investigating linkages is the correlation between clusters; this method will overlook cases in which a single variable links two clusters together. Perhaps cluster analysis, by relying on average correlations to rearrange a correlation matrix, has become too rigid a set of rules. Tryon's inability to incorporate other simple techniques in organizing and analysing a correlation matrix is responsible for the neglect of linkage analysis.

INDEX CONSTRUCTION

The choice of indices for the three constructs - socio-economic status (or "social status"), young families, and race - was guided in the present study by two criteria. The statistical relationships in Table 2 and Table 3 were used to determine the importance of variables within the construct groups. Concurrently, the tabulations of the 1940 and 1930 census were used to discover comparable indices in the earlier data. Therefore both a statistical criterion and a comparability criterion were employed.

The approach to index construction used here differs from the usual psychological testing approach. Several variables could be readily weighted and combined to form multi-variate indices for each construct. Instead attention was focussed upon selecting a single variable which could best meet statistical criteria for representing a construct yet could also meet other criteria as well. This procedure has several desirable characteristics. First, comparability with previous census data is more easily achieved. Second, comparison with other studies using 1950 census data is direct; one may test these indices for other cities, select those cities for which the same variables are valid and then compare these cities in terms of the census variables rather than a specially contrived index. Third, in case of complex findings the simpler index should facilitate interpretation. In general, arbitrary operations upon the raw data are likely to obscure subsequent analysis and interpretation. The Shevky and Bell choices of particular cities and censuses to define their Social Area Indices seem to possess this arbitrary quality.[12]

The dwelling unit value with estimates from rent, X_{20}, had the highest intercorrelations among the first five indicators of socio-economic status, and had the largest factor loading on the social status factor. In the 1930 tract data for Cleveland, equivalent monthly rental figures computed by Howard Whipple Green are the only data which correspond to the socio-economic data published in 1950.[10] From these two points of view it seemed that X_{20} should be used as an index of socio-economic status for the 1950 analysis of spatial distribution. A complication arises from this choice. Hoyt's theories and research on urban areas utilize rent as a general socio-economic index.[11] Therefore tests of Hoyt's work with 1950 data should be guided by the rent variable, X_{19}. Since the two variables were constructed as composites, and there is a corelation of .86 between them, little distortion will result from this procedure.

In the young family group children, X_8, has the highest average correlation with the other variables. Age distributions are reported in the 1930 and 1940 censuses, and the classification permits the construction of the same variable as X_8, was selected as the index of the young family group for subsequent analysis.

The index of race, X_3, is itself a single variable, and is tabulated in the 1930 and 1940 censuses. The adoption of this variable as an index for subsequent analysis leads to complications. Percent non-white, X_3, is extremely skewed. There is little to be gained by transforming this variable since fifty-nine percent of the tracts have the value zero. Therefore it may be used in subsequent analysis in its present form. Several techniques of analysis assume the use of normal variables; indeed, the inclusion of this variable in the correlation matrix carried the assumption of normality with it.

SUMMARY

The present critical evaluation may be summarized as follows. The work of Shevky and Bell and of Tryon has led to the identification of three constructs for summarizing the characteristics of census tracts. One of their constructs is novel to the literature and poses a number of interpretation problems. The other two, while familiar, are nevertheless dramatized by their analysis. However, the techniques used in identifying these constructs have overlooked a number of important points of detail, especially some of the more intricate relationships between the constructs. Simplex methodology and systematic inspection of the correlation matrix lead to these conclusions.[13] Finally, the choice of indices for these constructs and some alternative indices are suggested. More empirical cases would be needed to determine the substantive weight of these criticism. Yet even without further research important procedural defects in Social Area analysis are indicated here.

FOOTNOTES

1. This research is drawn from the author's unpublished Ph.D. thesis, "Census Tract Data and Social Structure: A Methodological Analysis", University of North Carolina, 1957. Daniel O. Price, Rupert B. Vance, and James A. Norton, Jr. have helped to formulate these ideas. A grant from the Purdue Research Foundation provided writing time.

2. Eshref Shevky and Wendell Bell, Social Area Analysis, (Stanford: University Press, 1955), Robert C. Tryon, Identification of Social Areas by Analysis, (Berkeley and Los Angeles: University of California Press, 1955).

3. Maurice D. Van Arsdol, Jr., Santo F. Camilleri, and Calvin F. Schmid, "The Generality of Urban Social Area Indexes", American Sociological Review 23 (June 1958).

4. U.S. Bureau of the Census, U.S. Census of Population: 1950, Vol. III, Census Tract Statistics, Chapter 12. Washington: U.S. Government Printing Office, 1952.

5. Louis Guttman, "A New Approach to Factor Analysis: The Radex", Chapter 6 in Paul F. Lazarsfeld (ed.) Mathematical Thinking in the Social Sciences, (Glencoe, Ill.: The Free Press, 1954).

6. Louis Guttman, "Multiple Rectilinear Prediction and Resolution into Components", Psychometrika, 5 (1940), pp. 75-100.

7. Shevky and Bell, op. cit., p. 4, Tryon, op. cit., p. 16.

8. Wendell Bell, "Economic, Family, and Ethnic Status", American Sociological Review, 20 (Feb., 1955).

9. Tryon, op. cit., p. 54 (insert).

10. Howard Whipple Green, Population Characteristics by Census Tracts, Cleveland, Ohio, 1930, (Cleveland: Plain Dealer Publishing Co., 1931).

11. Homer Hoyt, The Structure and Growth of Residential Neighborhoods in American Cities.

12. Shevky and Bell, op. cit., pp. 67-68.

13. A similar methodological statement is contained in Edgar F. Borgatta, "On Analyzing Correlation Matrices: Some New Emphases", Public Opinion Quarterly, XXII no. 4, (Winter, 1958-59).

GAME THEORY AND CUMULATIVE VOTING IN ILLINOIS

Jack Sawyer and Duncan MacRae, Jr.

University of Chicago

Since the theory of games was first made widely available (von Neumann and Morgerstern, 1944), with application to economic behavior, its use has been suggested in many other areas, from the global (Kaplan,1957) to the individual (Simon, 1956).

Its contribution to substantive knowledge in the empirical sciences, however, has been modest, and Luce and Raiffa (1957) judge that its use has been greater in applied mathematics. The area of political behavior—despite the apparent applicability of the notion of conflict of interest—is similarly lacking in studies, although notable exceptions exist in Shapley and Shubik (1954) and Luce and Rogow (1956).

Many of the previous studies have taken the form of defining a situation in terms of game theory and prescribing the proper behavior for a given set of conditions. In the present study, however, it was possible not only to specify a simple game theory model, but to evaluate it in a large number of actual cases. The following sections describe a voting behavior being modeled, the game theory model employed, an empirical test of the model, and implications of the results.

THE VOTING SITUATION

Voting for representatives for the Illinois General Assembly proceeds in a manner which is politically unique (although common in voting for corporate boards of directors). (See Glaser (1959) for an application of game theory to cumulative voting in this latter context.) The system, "cumulative voting," is intended to secure minority representation and does so by providing multiple member constituencies and allowing individuals to "cumulate" their votes on fewer than the total number of candidates to be elected.

In the case of cumulative voting in Illinois, three representatives are elected from each district, and each voter has three votes, which he may distribute 3-0, 2-1, $1\frac{1}{2}$-$1\frac{1}{2}$, or 1-1-1, among the candidates. Each party may nominate for the general election one, two, or three candidates, and the <u>number</u> to be nominated is decided upon and announced prior to the primary. This decision is made more or less independently by separate three-man committees elected by members of each party, with the voters in the primary election then determining <u>who</u> the candidates shall be.

The committee's decision is made under uncertainty as to the percentage of the vote which the party will receive, and often, though not always, the number of candidates which the other party will nominate. It is the behavior of this committee in arriving at a decision in the face of uncertainty which is being examined. In particular, the "rationality" of this decision in terms of its maximizing the number of the party elected is investigated by reference to the theory of games.

Previous studies of cumulative voting (Moore, 1919; Hyneman and Morgan, 1937; and Blair, 1960), while giving major attention to other aspects, have noted that occasionally the majority party fails to nominate two candidates and thereby loses the second seat it could otherwise fill. (As it turn out, however, by far the most frequent source of loss is when a party with a 75% majority nominates only two rather than three candidates.) Hence, in order to systematically examine the behavior of the nominating committee, the present study postulates, for the process of deciding upon the number of candidates, a model, based upon the theory of games, and described in the next section.

THE GAME THEORY MODEL

The behavior of the two nominating committees —one for each party—may be viewed as a two-person game in which the payoff is the number of candidates elected. The game is essentially zero-sum, in that positions not filled by one party are filled by the other. Each party has three strategies, namely to nominate one, two, or three candidates (the theoretical alternative of nominating none may be eliminated from consideration since it is never optimal and never employed).

Applicability of Model

The theory of games assumes that each player (1) knows all the rules of the game, i.e., the payoff matrix, (2) has a preference ordering of the payoffs, and knows that of his opponents, and (3) expresses his preference ordering in selecting strategies, i.e., he acts to maximize expected utility. As Luce and Raiffa (1957) point out, the third condition may perhaps best be taken as tautological, being simply a description of "preference ordering." The alternative is to determine preference orderings independently and use these to test the postulate.

In this application—unlike many experimental trials of the theory of games—the stakes are so considerable that there is good reason to assume a preference ordering. At the least, electing one candidate is not preferred to electing two candidates, and neither is preferred to electing three candidates. The assumption of a known payoff matrix is also met unusually well in this case, for both the available strategies and the outcomes for different combinations of strategies are completely prescribed and known to both parties. Hence, among situations involving actual political behavior, the case of cumulative voting appears particularly appropriate and specifically overcomes many of the problems raised by Deutsch (1954) in the application of game theory to politics.

Factors in the Committee's Decision

In determining the number of candidates to nominate, the committee acts under uncertainty as to (1) the percentage of the vote their party will receive, (2) how it will be divided among their candidates, and (3) how many candidates the other party will nominate. The division of the vote among the various candidates of the same party,

[1]This research was aided by support from the National Science Foundation and the Social Science Research Council.

however, in practice is usually very even, reflecting perhaps a combination of voter indifference, party discipline, and the effect (in machine voting) of the party lever, which results in an equal division, whereas to split one's vote unevenly requires further manipulation. In any event, given a rational opponent, an equal division of the votes among the candidates is always best. Otherwise, as has happened in a small percentage of elections, a party barely in the majority may find its less preferred candidate running behind both the equally preferred minority party candidates. Consequently, the formulations which follow assume that each party's vote is divided equally among all candidates of that party. (Actually, as will be made apparent later, the most desirable situation—although very difficult to attain—would be the ability to maintain a very small, but highly reliable, difference among the candidates.

The Payoff Matrices

The two remaining uncertainties may be dealt with by considering, for any given distribution of the vote, what the payoffs are in terms of number of candidates elected for any particular combination of strategies. It turns out, in fact, that there are only six different payoff matrices, one for each of the following ranges of the vote for the first party: 0-25%, 25-40%, 40-50%, 50-60%, 60-75%, 75-100%. These six matrices are given in Table 1, with the payoff the number of candidates elected by party A.

Take, for example, the matrix for the case in which A has 50-60% of the vote. If each party nominates only one candidate, both are elected, and the payoff to A is one. (This is an extremely rare event, occurring only when there is a strong third party—a case not covered by the present exposition. Technically, this feature makes the game non-zero-sum, in that the gains of A plus the gains of B do not sum to the constant three for this case. In a two party situation, however, this combination of strategies is never optimal, and, in practice, never occurs.) If A nominates one while B nominates two or three, A, being in the majority, will elect that one, while B will elect the other two. If A nominates two, he will elect both, regardless of how many B nominates. The third strategy of A presents an interesting inversion in outcomes: if B nominates one, he will elect that one and A will elect two; if B nominates two, he will elect both (despite his overall minority, each of his <u>two</u> will have <u>more</u> than 20% of the vote, while each of A's <u>three</u> candidates will have <u>less</u> than 20%); but if B nominates three, he will elect none and A will elect three. This latter case illustrates particularly how it is possible to nominate either too many <u>or</u> too few candidates.

Solution of Game

Taking the same payoff matrix (50-60%) as before, what is the optimal strategy for each party? Examining A's strategies, it is apparent that by nominating two, he can guarantee elcting that many, which is more than can be guaranteed by any other choice of strategy. B, on the other hand, who wishes to minimize the number elected by A, can assure by choosing to nominate two, that A will elect at most two. The same assurance is given B is he chooses to nominate only one, but

Table 1

Number Elected by Party A, by Percentage of Vote

<u>0-25% A</u>

A Nominates	B Nominates 1	2	3
1	1	1	0
2	2	1	0
3	2	1	0

<u>50-60% A</u>

A Nominates	B Nominates 1	2	3
1	1	1	1
2	2	2	2
3	2	1	3

<u>25-40% A</u>

A Nominates	B Nominates 1	2	3
1	1	1	1
2	2	1	0
3	2	1	0

<u>60-75% A</u>

A Nominates	B Nominates 1	2	3
1	1	1	1
2	2	2	2
3	2	3	3

<u>40-50% A</u>

A Nominates	B Nominates 1	2	3
1	1	1	1
2	2	1	2
3	2	1	0

<u>75-100% A</u>

A Nominates	B Nominates 1	2	3
1	1	1	1
2	2	2	2
3	3	3	3

this strategy is dominated, since the outcome resulting from its selection is in no case better, and sometimes worse than that from nominating two. Thus the minimax loss solution for this game matrix calls for each party to nominate two, in which case the outcome is that A selects two.

In the 60-75% game, B can, by nominating only one, guarantee himself electing that one, while for A, nominating either two or three guarantees his electing two. Nominating two, however, is consistent with his minimax strategy in the 50-60% case, and provides for the event that the vote should actually fall below 60% (which is much more likely, on the basis of the distribution of the vote, than that it will exceed 75%).

In the 75-100% game, A's best strategy clearly is to nominate three, which guarantees his electing all of them, while B can by no choice guarantee himself even one. It might seem that B's best choice were to nominate two or three, in the hope that A might grossly err by nominating only one. It seems more reasonable, to assume, however, that B in this case acts as if he knew the a priori distribution of A's choices to have zero probability for the strategy of nominating one. (In the nearly 1500 district elections since 1902, this has indeed been the case: with 75% of the vote, the majority party has never nominated less than two.) With this assumption, B's three strategies given identical outcomes and leave no choice. However, given even the slightest positive probability (which, in practice, there always is) of the percentage of the vote to party A being not 75-100 but between 60% and 75%, then B should nominate only one, as a hedge against this possibility. In addition, other considerations enter, such as the cost of candidacy, the effect of running superfluous candidates in the face of certain defeat, etc. It might be supposed in this situation that B should really choose to run no candidates. Again, however, there is always some positive probability that the 75-100% matrix is not the one which applies, but rather 60-75%, in which case a candidate should always be nominated. In addition, there may be a positive benefit to party morale to run one rather than none.

Solutions for the three remaining matrices are symmetrical to the ones already obtained, and, in summary, the hypothesized behavior is that the number of nominations by each party will conform to the table below:

% of the Vote to Party A	Nomination Pattern
0-25	1-3
25-40	1-2
40-50	2-2
50-60	2-2
60-75	2-1
75-100	3-1

Rationale for minimax solution. The conservative minimax loss criterion—providing a sure minimum rather than a chance of greater gain or loss—appears particularly compatible with a stable political system. Using the minimax loss solution assumes that parties do not try to completely demolish the opposition, even at considerable risk to themselves. One would not expect, on the other hand, such a consensus-promoting model to apply to revolutionary parties, who might be more likely to risk all for total victory.

Alternative Solutions. One might consider the minimax regret criterion, and ask, given a particular choice of strategy by the opponent, what the regret would be at one's own choice of strategy compared with one's best choice given that particular strategy of the opponent. While much post mortem speculation is carried on over election results, it is not clear that this neurotic criterion is invoked before the choice is made. One can compare, however, the solutions obtained by minimaxing loss with those obtained by minimaxing regret. In the latter case, of course, a non-zero-sum game results, since A's regret for a particular outcome depends upon the other entries in that row. In the case, however, of these small matrices and limited range of payoffs, it happens that precisely the same solutions result from minimaxing either regret or loss.

In general, solution of a game by the minimax principle assumes a rational opponent who will behave in the same conservative fashion, i.e., to assure himself a guaranteed himimum. Such behavior may not be optimal against an opponent who behaves in some other fashion, although it will still guarantee one the stated minimum. It may be more profitable yet, though, to utilize some more risky strategy. In particular, if some a priori distribution can be ascribed to the strategies of the opponent (as was done in a limited sense in the 0-25% matrix), then a Bayesian solution, given this particular distribution, can be made. A party nominating committee could act on such a basis, of course; the problem is in specifying the a priori distribution. The number of previous election upon which to base experience is small, and in most districts the distribution of the vote has varied considerably, affecting the choice of strategy. It seems unreasonable, therefore, to assume constant probabilities, unchanging over time. Yet if separate distributions among strategies are posited for different distributions of the vote, the number of previous elections upon which to base such a distribution becomes very small. In addition, committee membership, and perhaps party philosophy, is continually changing. Considering all these contingencies, minimax loss seems the most appropriate model for this situation.

Table 2

Distribution of Illinois General Assembly Elections 1902-1954, by Number Nominated and % Democratic

Number Nominated		% Democratic for the Same Election						
Dem	Rep	0-25	25-40	40-50	50-60	60-75	75-100	Total
1	3	27%	5%	1%				3%
1	2	69%	81%	26%	6%	2%		36%
2	2	4%	13%	68%	67%	10%		40½%
2	1		1%	5%	27%	87%	91%	20%
3	1					1%	9%	½%
Total		77	371	419	295	145	45	1353

Since the proportion of the vote received by party A is a variable, imperfectly estimated, it is not entirely certain, of course, that the vote will be within the range applicable to a particular matrix. Hence, as a first approximation to a stochastic model, the solutions just described incorporate the following feature: in the case of a matrix for which a party has two possible minimax strategies, that strategy is preferred which is consistent with the minimax strategy of one of the adjacent matrices (the one closer to the middle of the distribution). A further step might involve attaching a probability to the vote being in each of the six ranges, and solving the composite game, based upon the relations among the six probabilities. However, the present examination of empirical data, described in the next section, is based upon the simpler model.

ANALYSIS OF ELECTION DATA

To examine the fit of the model, data were obtained from official records (Illinois Secretary of State, 1902-1954) for the 1377 biennial elections (27 in each of Illinois' 51 districts) for the years 1902-1954. (Fortunately for research purposes —although in direct disregard of the constitution

Table 3

Partial Regression Coefficients for Number Nominated at Time t

Variable	Democrats	Republicans
Number nominated at time t-1	.462	.519
% Democratic at time t	.470	.301
% Democratic at time t-1	.012	.044
% Democratic at time t-2	-.085	.027
Minimax outcome at time t-1	.041	-.022
Multiple Correlation	.767	.800

Note: For the Republicans, the second, third, and fourth partial regression coefficients are for % Republican, at times t, t-1, and t-2.

Table 4

Change in Number Nominated vs. Change in % Democratic

Change in Number Nominated	Change (±) in % Dem 0-10	10-50
Democrats		
Same direction as (Dem) vote	9%	26%
No change in number nominated	87%	72%
Opposite direction from vote	4%	2%
Republicans		
Same direction as (Rep) vote	7%	12%
No change in number nominated	91%	84%
Opposite direction from vote	2%	4%
Number of elections	1080	246

—the Illinois Assembly had failed to redistrict during this entire period.)

Of these 1377 elections, 187 resulted in one party receiving less seats than was guaranteed it by following a minimax strategy. (Of the non-minimax outcomes 29 were, however, the result of uneven distribution of the party's vote among its candidates.) In 86 of the non-minimax outcome elections, the Democrats won less seats than guaranteed by minimax strategy; in 101 elections, the Republicans were the losers. The following sections further examine factors involved in rationality of the choice of the number nominated.

Number Nominated and Per Cent Democratic

Table 2 displays, for each of the six ranges of the vote which is Democratic (the same as for the six payoff matrices in Table 1), the distribution of the number nominated: 1-3 (one Democrat; three Republicans), 1-2, 2-2, 2-1, 3-1. Columns add to one hundred per cent. The less than two per cent of the elections (many of them involving strong third parties) which display a nomination pattern other than one of these five are omitted from the table.

From both sets of marginals, it is evident that Illinois has been more Republican than Democratic over this period. The table entries, themselves, however, show a remarkable symmetry between the two parties, for the more than ninety per cent of the elections in which the Democratic vote was between 25% and 75%. The second (25-40% Dem) column (5, 81, 13, 1, 0) corresponds closely to the reverse of the fifth (25-40% Rep) column (1, 87, 10, 2, 0); likewise the third column (1, 26, 68, 5, 0) to the reverse of the fourth (0, 27, 67, 6, 0). Given a vote between 25% and 75%, the parties behave rather similarly with respect to number nominated.

As an indication of the appropriateness of the particular game theory model in this situation, one may examine the distribution of number nominated for each of the six ranges of the vote. The predicted cells (as on the previous page) contain 27%, 81%, 68%, 67%, 87%, and 9% of the elections in their range of the vote, the overall proportion in those cells being 69%.

In over half of all elections, the vote is between 40% and 60% Democratic. In about two-thirds of these elections both parties employ the minimax strategy of nominating two candidates. In another quarter of these elections, only the majority party nominates two candidates, with the result that the election is uncontested, but its outcome is the same as that obtained by use of minimax strategy. For about five per cent of the elections, however, the majority party nominates only one candidate, thereby losing a seat it could have obtained.

The case in which one party has between 60% and 75% of the vote results in an uncontested election about five-sixths of the time. Only rarely (2% for Democrats; 1% for Republicans) does the party with this distinct majority fail to nominate the two candidates which it can surely elect.

With more than 75% of the vote, however, while both parties are considerably reluctant to nominate the three which that proportion enables them to elect, the Democrats have been much less likely to nominate three than the Republicans (9% compared with 27%).

There is, in the foregoing results, a strong implication of non-linear utility; parties never fail to run at least one candidate; 5% of the time when they could elect two, they fail to run the second candidate; but 80% of the time when they could elect three, they fail to nominate a third candidate. Such non-linear utility, of course, implies a non-zero-sum game.

Sensitivity to Change

In examining various aspects of the sensitivity of the behavior of the nominating committee to change, the most striking result is the basic conservatism of both parties. No matter what, by far the most likely number of nominations for any given year is the same as that for the previous election.

Table 5

Correlations over 51 Districts

	Variable	1	2	3	4	5	6	7	8	9	10	11
1.	Number of contested elections	—										
2.	Number of non-minimax losses: Dem	-.43	—									
3.	Number of non-minimax losses: Rep	-.40	-.17	—								
4.	Number of changes in No. Nom'd: Dem	.14	-.08	-.12	—							
5.	Number of changes in No. Nom'd: Rep	.01	.21	.05	-.03	—						
6.	Mean % Democratic	-.04	.67	-.41	-.23	.19	—					
7.	Absolute value of (Mean % Dem - 50%)	-.57	.24	.44	.11	-.07	-.28	—				
8.	Variance (over years) in % Dem	-.27	.43	.15	.04	.23	.10	.11	—			
9.	Change in % Democratic, 1902-1954	-.20	.62	-.06	-.11	.05	.42	.06	.70	—		
10.	Var of 2 yr change in % Democratic	-.42	.58	.20	.00	.19	.29	.17	.64	.59	—	
11.	Var of presid't'l yr change in % Dem	-.32	.41	-.02	.01	.17	.20	.03	.37	.34	.72	—
12.	Var of off-year change in % Dem	-.21	.31	.25	.18	.09	.09	.25	.59	.28	.54	.14

Changes in number nominated often represent changes in the vote of several years prior.

Multiple regression on number nominated. To investigate the association of variables with the choice of the number of candidates, multiple regressions were performed separately for Democrats and Republicans, using the 1275 elections of 1906-1954. Independent variables were number nominated at the previous election; Percentage of the vote at the present election, the previous election, and the next previous election; and the occurrence of a minimax outcome (rather than a non-minimax outcome) at the previous election.

The same general pattern of partial regression coefficients was obtained for both parties (Table 3); nearly all the weight is placed on the two variables, number nominated the previous election, and percentage of the vote for the present election. (The relevance of the latter variable may be taken as reflecting on the committee's use of estimates of the forthcoming vote—which correlates, however, .82 with the vote of the previous election.) Note, though, that while for Democrats, the two variables are weighted about equally, for the Republicans, the number nominated in the previous year has significantly ($p<.01$) more weight, an appropriately more conservative result. The correlations of the number nominated the present election with the number nominated the previous election, the percentage Democratic the present election, the previous election, and the next previous election, are all between .60 and .75, and in decreasing order as named, for both Republicans and Democrats.

Number nominated vs. change in % Democratic. Examination of the behavior of each party in the face of a changing vote gives additional insight into the differential role of conservatism as a correlate of choice. The following implications are drawn from Table 4: (1) change (in the number of nominations from that of the previous election) —for all values of change in the vote—comes slowly for both parties, but moreso for the Republicans, (2) Democrats show a larger reduction, between the small and large vote change situations, in the proportion of elections showing no change in number nominated, than do Republicans, (3) when the vote is changing over 10%, and a party does change its number of nominations, the Democrats are more likely to change in the proper direction, (4) for the Democrats, the change in the number nominated is much more likely to be in the appropriate direction if the change in the vote is large, (5) for the Republicans, however, the change in the number nominated is no more likely to be in the appropriate direction for a large change of the vote than for a small change. (The differences of statements (1) to (4) above are all significant at $p<.01$.)

Examination of particular years of great change in the vote further documents the conservatism. Given below are the total number of changes in number nominated, over all 51 districts for the years of 1920 and 1924 (when the Republicans should have been increasing number nominated) and 1932 (when the Democrats should have been doing the same).

	Democrats			Republicans		
	Up	Down	Same	Up	Down	Same
1920	1	8	42	5	1	45
1924	4	4	43	5	7	39
1932	8	1	42	0	7	44

Some anticipation is shown, but far from adequate: in 1920, the Republicans lost 21 seats they could otherwise have gained, through failing to run sufficient candidates; four years later they lost 14 seats in a similar fashion; in 1932, the Democrats lost 8 seats by failing to run enough candidates.

Table 6

Correlations over 25 years, 1906-1954

Variable	1	2	3	4	5	6	7	8	9	10	11
1. Year	—										
2. Number of contested elections	.42	—									
3. Number of non-minimax losses: Dem	.14	.38	—								
4. Number of non-minimax losses: Rep	-.25	-.07	-.26	—							
5. Number of changes in no. nom'd: Dem	-.36	-.58	-.36	.32	—						
6. Number of changes in no. nom'd: Rep	-.43	.24	.31	-.07	-.42	—					
7. Mean % Democratic	.45	.44	.71	-.69	-.54	.19	—				
8. Mean of 2 yr change in % Democratic	-.06	-.10	.27	-.67	-.34	.37	.44	—			
9. Mean of 4 yr change in % Democratic	-.06	.15	.35	-.40	-.44	.64	.44	.46	—		
10. Variance in % Democratic	.39	.20	-.03	.01	.10	-.17	-.08	.02	-.13	—	
11. Variance of 2 yr change in % Dem	-.74	-.05	-.12	.52	.38	.37	-.53	-.09	-.07	-.02	—
12. Variance of 4 yr change in % Dem	-.70	-.37	-.18	.04	.20	.30	-.31	.38	.00	-.18	.59

Number Nominated vs. Non-minimax outcome. A similar reluctance to change the number nominated can be observed even in the case in which the previous election resulted in a lesser number of seats than the minimax strategy would have guaranteed. For the Democrats, loss of a seat by occurrence of a non-minimax outcome at the previous election increases the probability of change in the number nominated from .15 to .26, while for the Republicans, the corresponding probabilities are .10 and .20. Following a non-minimax outcome, the Democrats and Republicans each have a probability of changing the number of nominations which is only about ten per cent greater than that obtaining if the number elected were at least equal to the number guaranteed by following minimax strategy. Of the changes in the latter situation, two-thirds are in the direction of nominating more candidates, which is appropriate, since practically all of the non-minimax losses result from nominating too few, rather than too many candidates.

Relations over Districts and Years

In the following three sections, relations are examined between a number of variables by (1) correlating them using the 51 districts as the units, (2) correlating them over the 25 years, 1906-1954, and (3) examining in further detail changes over time.

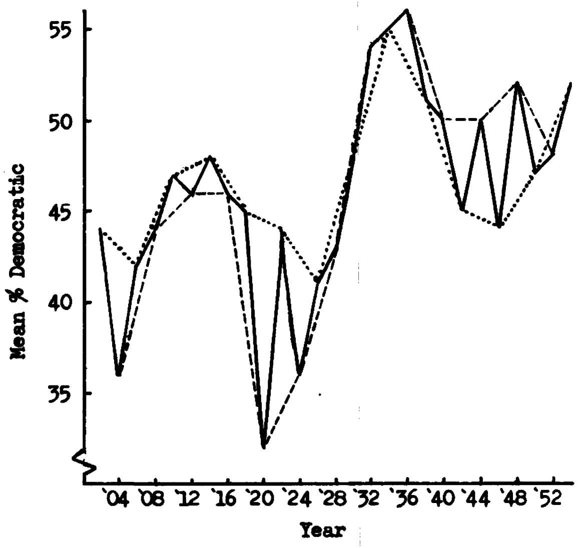
Correlations over districts. Table 5 presents correlations among 12 variables for the 51 districts. The variables represent the experience of each district over the period 1902-1954, and are as follows:

1. Number of elections in which there were a total of more than three candidates for the three seats.
2. Number of elections in which the Democrats elected fewer than guaranteed them by following the minimax strategy.
3. Number of elections in which the Republicans elected fewer than guaranteed them by following the minimax strategy.
4. Number of elections for which the number nominated by the Democrats represents a change from the previous election.
5. Number of elections for which the number nominated by the Republicans represents a change from the previous election.
6. Mean, over all years, of the variable, % Democratic.
7. Absolute value of the difference between 50% and variable (6) above.
8. Variance, over all years, of the variable, % Democratic.
9. Change in % Democratic from 1902 to 1954.
10. Variance of the change in % Democratic from the previous election.
11. Variance of the change in % Democratic from the second previous election, for the 12 presidential election years, 1908-1952.
12. Variance of the change in % Democratic from the second previous election, for the 13 non-presidential election years, 1906-1954.

Notice the relation of non-minimax losses to the other variables. Districts having more contested elections have fewer non-minimax outcomes, although even in these districts, far from all of the elections are contested. Perhaps the willingness to contest is the critical element. Each party has the most non-minimax losses in those districts in which its mean proportion of the vote is highest. This may again reflect the decreasing utility of the third seat. For the Democrats, non-minimax losses are also greater in those districts where the mean vote has been increasing over time. The number of contested elections is much greater

Figure 1

Mean % Democratic of all 51 Districts, 1902-1954



in those districts where the mean vote is close to 50%.

Correlation over years. Correlations were also computed (Table 6) among 12 variables for the 25 years, 1906-1954 (1902 and 1904 could not be used, since some of the variables involved changes from two or four years previous). For each year, the variables below represent all 51 districts. Variables 2-7, and 10 correspond to variables 1-6, and 8, respectively in the correlations over districts. The other variables are

1. Year of the election: 1906, 1908, ..., 1954.
8. Mean, over all districts, of the variable, change in % Democratic from the previous election.
9. Mean, over all districts, of the variable, change in % Democratic from the second previous election.
11. Variance, over all districts, of the variable, change in % Democratic from the previous election.
12. Variance, over all districts, of the variable, change in % Democratic from the second previous election.

The changes associated with time have been considerable in Illinois, and many significant relations emerge. Paralleling the previous finding concerning districts, non-minimax losses are highly related to the party strength that year, as well as to increase in the party's vote from two and four years previous. This is true for both parties. Both parties are also more likely to change the number of nomination from the previous year when their vote is dropping than when it is rising.

A number of variables change with time itself, there being an increase in the number of contested elections; decreases in the number of changes in number nominated from the previous election (for both parties); an increase in the Democratic vote; an increase in the variance among districts, and sharp decreases in the variance of the change from the previous election or the second previous election. Changes over time are further examined in the next section.

Changes in the vote, 1902-1954. Figure 1 gives the mean Democratic vote of the 51 counties for each of the 27 biennial elections, 1902-1954. (The solid line connects adjacent election years; the dashed line connects presidential years; the dotted line connects non-presidential years.)

It can be seen that the variance among the means of the districts for presidential years is much greater than that among off-years, in fact, less than half. However, the variance among districts within years is much less for presidential years than for off-years. Associated with presidential voting, then, is a unifying effect among districts. However, the unity among districts of the presidential years is less stable than the relative diversity among districts of the off-years. If the presidential election could be said to bring the districts together, then it tends to unite them at a different level from election to election, following which the districts return in the off-years to a more stable diversity.

In general, the variance among districts is greater than the variance among years, and the greater variance among districts is associated particularly with the division between Cook county (Chicago and suburbs) and the rest of the state. Moreover, this division has been increasing over time. In 1958, the average Democratic vote in the Cook county districts was 64%, compared to 47% for the remaining districts.

IMPLICATIONS

It is apparent from the preceeding sections that the particular game theory model employed provides a partial, though not a complete description of the behavior of the nominating committee in determining the number of candidates to run. This section examines some factors relevant to the agreement and disagreement of data and model.

Rationality

Game theory may be employed as a model of rational behavior, defined in terms of minimaxing a certain quantity. In this sense, a considerable amount of rationality was demonstrated by the nominating committees in their determination of the number of candidates: in 69% of all elections, both parties employed a minimax strategy.

It is, however, for both parties, a rationality associated with a basic conservatism. Some years often elapse from a change in the vote to the corresponding adjustment in the number nominated. There is some question whether the effective behavior which is shown results from a considered rationality, or simply from the conservatism of not changing the number nominated, which usually turns out to be the best thing to do. Or perhaps, more significantly, the general decision for conservatism has been made on a rational basis, seeing that this is generally effective.

Utility and Job Security

The principal departures from the model occur in the relatively rare cases in which one party has more than seventy-five per cent of the vote. In these cases, however, the departure is very considerable, with parties only nominating three candidates about 20% of the time. As has been previously suggested, there is a strong implication of non-linear utility particularly with respect to the election of third candidates. Further investigation has revealed some possible bases for this non-linear utility.

A relevant consideration seems to be the individual job security of the incumbent (or less often of another strong candidate). While parties may acknowledge that with 75% of the vote, they may elect three, and even if the vote falls somewhat below that, will surely elect two, the critical question to them is "Which two?" The party committee naturally seeks to control who is nominated, as well as the number nominated. Hence, coalitions may form between the two stronger forces within the party and mitigate against adding a third nomination which would reduce the certainty. (The possibility of coalitions of two against one would seem to make job security initiated pressure against an additional nomination more likely against a third candidate when the vote is 75%, than against a second candidate when the vote is 50%.)

In addition, there is evidence that in certain cases, again most likely occurring when one party has 75% of the vote, bi-partisan agreements are reached to allow the minority a seat, in exchange for an appropriate "side-payment." The bi-partisan agreement is also seen to alleviate the job security problem, and perhaps to satisfy both parties, inasmuch as, if the utilities are non-linear, the side payment can be larger than the utility of the majority's third member and less than the utility of the minority's first member.

Minority Representation

In terms of the intent of the law--to provide minority representation--the game theory solution is optimal. If both parties follow a minimax strategy, and divide their votes equally among all their candidates, the result will be to give the minority one seat whenever it has as much as 25% of the vote. This end has been generally achieved, as noted by Blair (1960). (Indeed, as we have seen nearly all of the mal-representation which has occurred has been in giving the minority <u>over</u>-representation.)

In the earlier days of cumulative voting in Illinois, some "good government" groups protested the lack of choice available to the voter in the general election--with many elections having only three candidates, no more than the number of seats --and hence proposed that parties be required to run full slates. This would, of course, negate the basic purpose of this scheme of minority representation. In addition, several writers have implied that a non-contested election represented collusion, but this is not necessarily so. Such a casy may indeed be the best strategy for each party, and may also represent the <u>only</u> way in which nearly proportional representation can be assured.

A modification might be proposed, however, to guarantee more proportional representation by removing the uncertainty which influences committees to unnecessarily (and unprofitably) limit the number of candidates. Allow the primary elections to specify the <u>order</u> of election of candidates, and the number of votes for the party (without regard to candidates) in the general election to determine the <u>number</u> elected from each party, in the already-specified order. Better yet, let one election suffice for both functions, by allowing the votes within a party to determine the ordering of candidates within that party, and the total number of votes cast for each party to determine the number of candidates it elects. Each voter would, as before, have a number of votes equal to the number of candidates to be elected, and would still be able to cross party lines if desired. Thus proportional representation could be achieved to a much greater extent, unhampered by possibly non-optimal committee decisions.

SUMMARY

A two-person zero-sum game theory model is devised for the behavior of committees which determine, for their own three-member districts, the number of candidates (one, two, or three) their party shall enter in the general election for representative to the Illinois General Assembly. Application of this model to the 1377 biennial elections from 1902 to 1954 finds that in 69% of all elections both parties employ minimax strategy, and that 86% of all elections result in a minimax outcome, regardless of strategies. Other results include (1) a drastically decreasing utility for the third seat, possibly related to individual job security, (2) a basic conservatism on the part of both parties, which are considerably reluctant to change the number nominated, and (3) changes in the distribution of the vote over the period 1902-1954.

REFERENCES

Blair, G. S. <u>Cumulative voting: an effective electoral device in Illinois politics</u>. Urbana: University of Illinois, 1960.

Deutsch, K. W. Game theory and politics: some problems of application. <u>Canadian Journal of Economics and Political Science</u>, 1954, 20, 76-83.

Glasser, G. J. Game theory and cumulative voting for corporate directors. <u>Management Science</u>, 1959, 5, 151-156.

Hyneman, C. S., & Morgan, J. D. Cumulative voting in Illinois. <u>Illinois Law Review</u>, 1937, 32, 12-31.

Illinois. Secretary of State. <u>Blue book of the state of Illinois</u>. Springfield: State of Illinois, 1902-1954.

Kaplan, M. <u>System and Process in international politics</u>. New York: Wiley, 1957.

Luce, R. D., & Raiffa, H. <u>Games and decisions</u>. New York: Wiley, 1957.

Luce, R. D., & Rogow, A. A. A game theoretic analysis of congressional power distributions for a stable two-party system. <u>Behavioral Science</u>, 1956, 1, 83-95.

Moore, B. F. The history of cumulative voting and minority representation in Illinois, 1870-1919. <u>University of Illinois Studies in the Social Sciences</u>, 1919, 8(2).

Shapley, L. S., & Shubik, M. A method for evaluating the distribution of power in a committee system. <u>American Political Science Review</u>, 1954, 48, 787-792.

Simon, H. A. A comparison of game and learning theory. <u>Psychometrika</u>, 1956, 21, 267-72.

von Neumann, J., & Morgenstern, O. <u>Theory of games and economic behavior</u>. Princeton: Princeton University, 1944.

V

PLACE OF THE 1960 CENSUS IN THE EVOLUTION OF IMPORTANT CENSUS CONCEPTS

Chairman, Robert W. Burgess, Bureau of the Census

WHERE WE STAND IN THE URBAN-RURAL AND FARM RESIDENCE CONCEPTS

By: Henry D. Sheldon, Bureau of the Census

Census concepts rarely spring fully armed from the deliberations of the census staff, like Athena from the brow of Zeus. Rather they begin with a relatively simple and common sense notion which evolves in response in the conflicting needs of the users of census statistics and the difficulties of application, into a reasonably complicated definition which all too frequently represents an unhappy compromise.

The concept of urban population began quite simply as the aggregate population of cities. In the latter part of the 19th Century and the early 20th Century there were differences from census to census as to size of cities included in the urban population. In the Statistical Atlas of 1874 the urban population was defined as the population of places of 8,000 or more and in 1880 this limit was lowered by implication to 4,000. In 1910 a population of 2,500 was taken as the lower limit for reasons which have been lost to posterity, and this limit has persisted without serious question to the present. Between 1920 and 1930, data from all previous censuses were consolidated on the basis of this limit, thus producing a time series on urban population going back to 1790.

More explicitly, the 1910 definition specified the urban population as the population of incorporated places of 2,500 or more and of a few additional areas in which the standard definition did not appear to apply but which were nevertheless considered urban. The problem of definition in these latter areas will be considered later in this discussion. The selection of incorporated places above a certain size had several advantages. It did isolate areas containing the highest population densities; and, since incorporation generally represents a type of organization which permits the setting up of administrative machinery for coping with the special problems arising out of high density - sewage disposal, the provisions of utilities, and police protection, and the like, it generally identified areas which have these secondary urban characteristics. Finally, since incorporation involves a definite area, the boundaries of the population concentration were automatically established and could be used in enumeration. This advantage may seem somewhat irrelevant, but if one examines costs to the Bureau of setting up such boundaries, as was done in the case of the urban fringe in 1950, its usefulness is apparent.

With minor changes in the treatment of special problem areas the 1910 definition remained intact through the 1940 Census. Although nowhere explicitly stated, and although a level of density which set off urban territory from rural territory was not specified, implicitly urban territory was defined in terms of high density in aggregates of appreciable size. This criterion has been designated in some quarters as naive and superficial - the really discriminating characteristics of the urban population are those relating to the urban way of life such as characteristic occupations, services, attitudes and the like, and with the greater fluidity in the daily movement of population in the area in which people live is to a large in the area in which people live is to a large degree irrelevant. This position appears to me to be somewhat misguided. It calls attention to be sure, to important problems, but if the urban concept means anything it certainly refers to the consequences of the high concentration of population in relatively small areas. As Dr. Eldridge has stated it:

> "For much the same reasons that the definition of urbanization must be restricted, that of cities must also be restricted. Cities have been defined as ways of life, states of mind, collections of traits, types of occupation and the like. Such definitions are bound to get us in trouble sooner or later because none of the attributes named are constants of the city and all of them spill over into other areas. Traits change, occupations change, political organization changes, the economic system changes. The only trait that is constant is that the city is different from what is not the city. The nature of this difference varies. If we say that the city is a collection of traits, we cut ourselves loose from the only solid base on which we can set up definite criteria and find ourselves neck-deep in a sea of difficulties connected with the isolation of urban traits. It means that whatever we find more of in the city is an urban trait. But what is the city? Why, it is a collection of urban traits. How do we identify these traits? By their high intensity in the city. What is the city? A collection of traits. And so forth. The only way to break this deadlock is to go back to population concentration. The criterion must be in terms of population. Then we can study traits, relationships, and characteristics to our heart's content."[1]

In preparing for the 1950 Census, it became apparent that if urban population was the population in areas of high density, the definition used in past censuses would be inadequate. There were on the outskirts of most of our larger cities heavy concentrations of suburban population which, since they were outside the city limits, in unincorporated territory would fall into the rural population. There were also concentrations in the open country such as, for example, Kannapolis, North Carolina and Oak Ridge, Tennessee, which, being unincorporated, would likewise fall into the rural population.

In order to make up for these deficiencies, the staff of the Geography Division of the Bureau, in large part by a field examination, established boundaries for urban fringe areas around all cities with a population of 50,000 or more, as established by the 1940 Census or subsequent special censuses, and outside of urbanized areas, around unincorporated places of 1,000 or more.

The urban fringe was laid out to include (with minor exceptions) all incorporated places adjacent to the central city and all adjacent unincorporated territory with a density of 500 dwelling units per square mile, a density normally the minimum found associated with a closely spaced street pattern. In addition, there were added the following elements: (1) adjacent territory devoted to urban land use, (2) clusters meeting the minimum density requirement with 1-1/2 miles of the main contiguous urbanized part, and (3) other outlying areas meeting the density criterion within one-half mile of the initial noncontiguous clusters.

The boundaries of the unincorporated places included a definite nucleus of residences and as far as feasible all surrounding closely settled territory.

The net effect of this change in definition was to increase the size of the urban population by about 9 percent over what it would have been under the 1940 definition. A detailed classification of the components of this change is presented in table 1.

The areas which have not lent themselves to the conventional treatment in terms of incorporated places are the New England towns and to a lesser degree the townships of Pennsylvania and New Jersey.

The problem arises from the unique character of the structure of local government in New England in contrast to the rest of the country. Counties in the United States are composed of minor civil divisions which in their totality account for the entire area of the county. They are known variously as townships in the Middle West, towns in New England and New York, and as judicial districts, election districts, militia districts, and the like in other parts of the country. Outside New England these minor civil divisions are spotted with small areas of high population density which typically are places incorporated as municipalities. In some States they are subordinate to the minor civil division in which they are located and in other States they are independent - in effect a minor civil division in their own right. In the New England States the same points of population concentration exist but typically they have no formal boundaries nor are they recognized as legal entities. Cities are grown up or "filled up" towns, and towns characteristically - like the Greek city State, have centers of population concentration and a considerable rural hinterland.

As the result of this situation, towns in New England were of necessity included or excluded from the urban population on an all or none basis. In 1910 all towns of 2,500 or more were included in the urban population. In later censuses it was felt that this procedure too greatly diluted the urban population, so that a series of special rules were developed which permitted the inclusion of certain New England towns, and elsewhere a relatively small number of townships, in the urban population. In the 1940 Census, New England towns which contained thickly settled areas of 2,500 or more and in which 50 percent or more of the population of the town was to be found in thickly settled areas were recognized as urban. In addition, minor civil divisions with a population of 10,000 or more with a density of 1,000 or more per square mile and containing no incorporated places were classified as urban. The towns recognized as urban under special rule were largely concentrated in Massachusetts, Rhode Island, and New Hampshire and the townships in New Jersey and Pennsylvania, and in the aggregate accounted for a population of nearly 2 million.

In the 1950 Census neither towns nor townships were recognized as urban under special rule on the essentially correct assumption that the really urban parts of these minor civil divisions would either fall in the urban fringe or would be recognized as unincorporated places. About 2.4 million persons were counted in the areas which would have been urban under the 1940 special rules, and about 338,000 or about 14 percent were classified as living in rural areas according to the 1950 rules. Although this procedure met the theoretical specifications of the urban definition with a fair degree of adequacy, it had the net effect of suppressing the tabulation of certain data for a considerable number of New England towns and New Jersey and Pennsylvania townships. Since in both instances these areas are the units of local government, they were deprived of statistics for which there was considerable administrative need.

The definition of the urban population which is to be used in the 1960 Census is essentially the same as that of the 1950 Census. It had been hoped that the delineation of urban fringe would be extended to cities of 25,000 to 50,000 but this forward step could not be accomplished within the budget for the 1960 Census.

In 1960 the procedure for defining urban fringe has been altered. Instead of "on the ground" delineation of the areas prior to the census, areas which are certain to fall in the urban fringe have already been identified and the area so delineated has been surrounded by a band of small enumeration districts. When the population counts for these areas are available they will be included or excluded on the basis of a population density criterion.

Another innovation will be to reinstate selected towns in New England and townships in New Jersey and Pennsylvania as urban. This will be done if they have a population in 1960 of 25,000 or more, or, if they have a population of 2,500 to 25,000 and a density of 1,500 or more per square mile and do not contain a dependent incorporated place within their borders. This latter change, providing as it does for the separate recognition of selected towns and townships, will mean that statistics for administrative users will be available, but at the same time will not materially dilute the urban population with the rural population of the towns and townships in question. Most of the population in these areas would have been urban by virtue of their location in the urban fringe or by the fact that they contained unincorporated places; in terms of 1950 population 95 percent of the aggregate population of the areas which would be urban under this rule, would also be urban in terms of urban fringe and unincorporated places.

Like the urban-rural definition, the definition of the farm population has been an apparently simple one. In the population censuses of 1920 through 1950, the farm population has been defined as the population living on a farm. It included all the persons living in households in which the respondent had answered "yes" to the question "Is this house on a farm?" There were minor variations in the wording from census to census, but substantively it was the same. In 1950, the enumerator was instructed to exclude persons who paid rent for house and yard only as well as persons living in motels, etc. located on farms.

No definition of a farm was given the enumerator, but since he was at the same time filling an agriculture schedule, it was assumed that his classification on the population schedule would be highly correlated with the information collected on the agriculture schedule. There was no reason to assume, however, that the two classifications would be identical, since the final determination of whether a place qualified as a farm was made by an office editing procedure.

As a part of the 1950 program of the Agriculture Census, records from the 1950 Census of Population were matched for a sample of about 11,000 farms and farm operators, and a report, Farms and Farm People, was prepared jointly by the Bureau of the Census and the Department of Agriculture.

This study suggested that approximately 1.7 million or about 8 percent of the rural-farm population, as counted in the Population Census, lived on places which were not classified as farms in the Agriculture Census. On the other hand, about 1 million persons living on farms as defined by the Agriculture Census, were not so classified in the Population Census. The net difference, based on the total rural population, was about 1.3 percent. (table 2).

In view of the rapid changes in the character of agriculture in this decade as indicated by the 1954 Census of Agriculture, the Bureau was considerably concerned about the possibility of a growing discrepancy between the size of the farm population implicit in the two approaches to its measurement. Consequently, in March and April of 1957, the Current Population Survey was used as a means of investigating this problem.

In March of 1957, the Agricultural Economics Division of USDA and the Agriculture Division of the Bureau of the Census sponsored a supplement to the Current Population Survey, which involved an abridged agricultural questionnaire of sufficient detail to make a farm determination according to the Agriculture Census procedure for those households in the CPS sample which had reported themselves as on farms or reported income from the operation of their own farms.

In April 1957, the Agricultural Economics Division of USDA and the Population Division sponsored a survey in an effort to evaluate various criteria of farm residence. Among the questions on the schedule were questions on acreage and global questions on value of products and value of sales. It was therefore possible to approximate the agriculture census farm definition.

In both surveys a farm was defined (as in 1950 and 1954) as a place of 3 or more acres with a value of products of $150 or more or a place of less than 3 acres with sales valued at $150 or more. The results of the special farm determination were in both cases compared with the standard farm-nonfarm classification based on the responses to the question "Is this house located on a farm (or ranch)?" and in both surveys all persons living on farms were included in the farm population regardless of whether or not they were members of farm operator households.

The results of these surveys are summarized in tables 3 and 4. They suggest that, in comparison with 1950, the percentage of the rural population classified differently by the two approaches had increased by a factor of about 2 (5.1 to 11.4, March, or 8.8 April), that the population on agriculture census farms but not in the population census farm population had not changed appreciably; and that the percentage net difference had increased by about a factor of 5 (1.3 to 6.6 March, or 5.6 April).

There are, as tables 3 and 4 indicate, appreciable differences between the levels for some categories in the March and April surveys; but these are not unusual in CPS results relating to farm residence. For the population living on agriculture census farms as defined in 1950 and 1954, there is a surprising agreement between the two surveys in the absolute figures, 15.8 vs. 15.7 million. This agreement, if it is not purely coincidental, suggests that the substitution of summary questions, encapsulating a given farm definition, produces essentially the same results as the more elaborate agriculture census procedure.

In planning for the 1959 Census of Agriculture, the farm definition used in previous censuses have been subjected to a critical review by both the Agriculture Division of the Bureau of the Census and the Department of Agriculture. After due consideration, it was agreed that, for the purposes of the 1959 census of agriculture, a farm would be defined in terms of acreage and value of the sales of agricultural products as follows:

Acreage	Sales
10 or more	$50 or more
Under 10	$250 or more

the net effect of the use of this somewhat more restrictive definition would be to increase further the discrepancy between the size of the population of agriculture farms and the size of the farm population as determined by the 1950 population question.

In view of this problem, and encouraged by the close correspondence of the figures for the population of agriculture census farms in the two surveys, the 2 agencies proposed to include on the 1960 sample schedule, questions relating to acreage and value of sales, which in terms of the response, would permit an identification of farms as defined above.

The exact form of the question is indicated in the material distributed. Admittedly, it may have a number of shortcomings, but these in all likelihood could not be corrected short of an operation similar to that employed by the Agriculture Census. There is neither time nor money for an operation of this complexity in the Population Census. Nor is it possible to obtain a farm residence classification for the Population Census from the Agriculture Census. Such a matching procedure would be costly and would seriously delay the processing of the censuses, and the difficulties of matching across a six month interval would seriously affect the quality of the classification.

The proposed procedure does have the advantage of bringing the Population and Agriculture figure on farm population more closely together and of eliminating from the Population Census farm population a considerable number of persons who would be members of that population only by courtesy or sentiment. Some reduction in the proportion of disagreements (gross error) can be expected from the fact that the entire rural population will be covered, from the gap in time between the two censuses thus allowing time for the conversion of potential sales into actual sales, and from improvements in format and instructions. Confusion in the public mind arising from the former use of two definitions of a farm will be ended.

On the other hand, the count of the farm population in the census and in future surveys, notably the Current Population Survey, will be at a lower level than it has been in the past. Data from the April 1959 survey, in which the questions on farm population were asked in essentially the same form that they will be asked in 1960, indicate a difference between the two series of roughly the same magnitude as that observed in the 1957 Surveys. These data, and those obtained in the coming year in which both approaches will be used, will provide a basis for bridging the gap between the two series.

The Bureau's approach to the definition population has been made consistently in terms of the membership of households located at the seat of some agricultural operation. The same problem could of course be approached in terms of occupation or in terms of income from agricultural operation and employment. For a substantial part of the population in question, any of these approaches would yield the same result, but for the remainder there would be considerable variation depending on the particular approach used.

---

[1] Hope Tisdale Eldridge: "The Process of Urbanization" in Demographic Analysis edited by J. J. Spengler and Otis Dudley Duncan, The Free Press, Glencoe, Illinois, 1956, p. 339.

Table 1.--URBAN AND RURAL POPULATION OF THE UNITED STATES UNDER OLD AND NEW URBAN DEFINITION: 1950

New definition	Total population	Old definition			
		Urban			Rural
		Total	Incorporated places of 2,500 or more	Places urban under special rule	
Total population	150,697,361	88,927,464	86,550,941	2,376,523	61,769,897
Urban	96,467,686	88,589,867	86,550,941	2,038,926	7,877,819
Incorporated places of 2,500 or more	86,550,941	86,550,941	86,550,941	...	...
Unincorporated places of 2,500 or more	1,994,727	320,504	...	320,504	1,674,223
Urban fringe[1]	7,922,018	1,718,422	...	1,718,422	6,203,596
Rural	54,229,675	337,597	...	337,597	53,892,078
Percent distribution by old classification:					
Total population	100.0	59.0	(57.4)	(1.6)	41.0
Urban	100.0	91.8	(89.7)	(2.1)	8.2
Incorporated places of 2,500 or more	100.0	100.0	(100.0)	...	...
Unincorporated places of 2,500 or more	100.0	16.1	...	(16.1)	83.9
Urban fringe[1]	100.0	21.7	...	(21.7)	78.3
Rural	100.0	0.6	...	(0.6)	99.4
Percent distribution by new classification:					
Total population	100.0	100.0	100.0	100.0	100.0
Urban	64.0	99.6	100.0	85.8	12.8
Incorporated places of 2,500 or more	(57.4)	(97.3)	(100.0)	...	...
Unincorporated places of 2,500 or more	(1.3)	(0.4)	...	(13.5)	(2.7)
Urban fringe[1]	(5.3)	(1.9)	...	(72.3)	(10.0)
Rural	36.0	0.4	...	14.2	87.2

[1] Exclusive of incorporated places of 2,500 or more.

Source: Henry D. Sheldon, Changes in the Rural Population, 1940 to 1950, Rural Sociology, June 1952

Table 2.--RURAL POPULATION BY FARM-NONFARM RESIDENCE, POPULATION VERSUS AGRICULTURE CENSUSES: 1950

(In thousands)

Population definition - tabulation of answers to question, "Is this house on a farm?"	Agriculture definition applied by editing agriculture schedule		
	Total	Farm	Nonfarm
Total	54,230	22,325	31,904
Farm	23,048	21,313	1,735
Nonfarm	31,181	1,012	30,169

Table 2a.--DIFFERENCES IN CLASSIFICATION

(In thousands)

Type of difference	Number	Percent
Total rural population.................	54,230	100.0
Net difference.............................	723	1.3
Gross difference...........................	2,747	5.1

Note: Statistics are based in part on the result of a sample matching of agriculture and population schedule and in part on data from the 1950 Census of Population. They have been adjusted to eliminate the urban farm population (about 283,000). Persons who were living on farms but not members of farm operator households have been included in the agriculture census farm population as well as in the population census farm population.

Agriculture farm population: Persons living on places of 3 or more acres with value of products of $150 or more or places of less than 3 acres with sales of agricultural products of $150 or more.

Population farm population: Persons living in households reported by respondents to be on farms.

Source: U. S. Departments of Agriculture and Commerce: Farms and Farm People, Washington, 1953, and U. S. Census of Population: 1950, Vol. II, Characteristics of the Population, Part 1, United States Summary, Washington, 1953.

Table 3.--RURAL POPULATION BY FARM-NONFARM RESIDENCE, POPULATION VERSUS AGRICULTURE DEFINITION: CURRENT POPULATION SURVEY, MARCH 1957

(In thousands)

Population definition - tabulation of answers to question, "Is this house on a farm?"	Agriculture definition applied by editing agriculture schedule		
	Total	Farm	Nonfarm
Total..................................	63,990	17,300	46,690
Farm....................................	21,524	15,772	5,752
Nonfarm.................................	42,466	1,529	40,937

Table 3a.--DIFFERENCES IN CLASSIFICATION

(In thousands)

Type of difference	Number	Percent
Total rural population.................	63,990	100.0
Net difference.............................	4,223	6.6
Gross difference...........................	7,281	11.4

Note: Statistics are based on a survey sponsored by the Agriculture Division covering household classified as farm in the Current Population Survey and other rural household reporting farm income and on estimates of the total rural population derived from the Survey.

Agriculture definition: Persons living on places of 3 or more acres with value of products of $150 or more or places of less than 3 acres with sales of agricultural products of $150 or more.

Population definition: Person living in household classified as "farm" on the CPS control cards which in turn is based on replies to the question, "Is this house on a farm?"

Source: Current Population Survey, March 1957.

Table 4.--RURAL POPULATION BY FARM-NONFARM RESIDENCE, POPULATION VERSUS AGRICULTURE DEFINITION: CURRENT POPULATION SURVEY, APRIL 1957

(In thousands)

Population definition - tabulation of answers to question, "Is this house on a farm?"	Agriculture definition - tabulation of answers to questions on acreage and value of products		
	Total	Farm	Nonfarm
Total	64,362	16,760	47,602
Farm	20,334	15,702	4,632
Nonfarm	44,028	1,058	42,970

Table 4a.--DIFFERENCES IN CLASSIFICATION

(In thousands)

Type of difference	Number	Percent
Total rural population	64,362	100.0
Net difference	3,574	5.6
Gross difference	5,690	8.8

Agriculture definition: Persons living on places of 3 or more acres with value of products of $150 or more or places of less than 3 acres with sales of agricultural products of $150 or more.

Population definition: Person living in household classified as "farm" on the CPS control cards which in turn is based on replies to the question, "Is this house on a farm?"

Source: Current Population Survey, April 1957.

1960 FARM POPULATION QUESTIONS

ON THE HOUSEHOLD QUESTIONNAIRE:
(Self enumeration)

H17 and H18. Is this house:

On a city lot
(or is this an apartment building)?....... ☐

OR

On a place of less than 10 acres?......... ☐ → Last year (1959), did sales of crops, livestock, and other farm products from this place amount to $250 or more?

- $250 or more......... ☐
- Less than $250 or none................ ☐

OR

On a place of 10 or more acres?........... ☐ → Last year (1959), did sales of crops, livestock and other farm products from this place amount to $50 or more?

- $50 or more.......... ☐
- Less than $50 or none. ☐

ON THE SAMPLE FOSDIC SCHEDULE
(Used by enumerator)

H17. Is this house--

On a city lot
(or apt. bldg.)?............ ☐
(Omit H18)

On a place of less than
10 acres?................... ☐ → H18a
(Ask H18a)

On a place of 10 or more
acres?...................... ☐ → H18b
(Ask H18b)

H18a. If occupied--
Last year, 1959, did sales of crops, livestock and other farm products from this place amount to--

- $250 or more?...................... ☐
- Less than $250 (or none)?........... ☐

H18b. If occupied--
Last year, 1959, did sales of crops, livestock and other farm products from this place amount to--

- $50 or more?....................... ☐
- Less than $50 (or none)?............ ☐

LIVING QUARTERS AND HOUSEHOLD CONCEPTS IN THE 1960 CENSUS

By: J. Hugh Rose and Paul C. Glick, Bureau of the Census

Summary

Over the years, users of housing census data have made a number of criticisms of the dwelling unit concept. The weaknesses of this concept have become particularly evident in the course of analysis of urban renewal and slum clearance areas. In these areas particularly, the 1950 concept was subject to varying interpretations by enumerators. For this reason the Bureau of the Census has developed more specific means of delineating each of the separate quarters where people live. As a result, the unit of enumeration has become a "housing unit" instead of a "dwelling unit." In order to define this new unit of enumeration, changes were made in the application of the physical criteria, "access to unit" and "cooking equipment." Since a "household" is defined as the person or group of persons occupying a housing unit, the change affects both housing units and households.

The criteria and the procedures used in 1950 frequently omitted from the universe certain peripheral types of living arrangements. Such quarters should be uniformly included in order to secure a more complete account of housing conditions in America. Thus, consistent with the long range goal of extending coverage, the 1960 Census of Housing will include all living quarters having some degree of "separateness."

The basic criterion of "separateness" used in earlier censuses is not being changed but more criteria are being provided for determining when a particular unit is to be considered as separate.

It is estimated that the new definition will add about one million households or occupied housing units, and perhaps a quarter of a million vacant units. More precise estimates will be available from postcensal evaluation surveys. The impact of wider coverage will be greatest on parts of urban areas, with less impact on rural areas. The change reflects largely the enumeration as separate units those quarters occupied by persons who were formerly counted as members of another household or of "nondwelling unit living quarters."

Nature of the Changes in Definition

Basically, the same physical criteria for determining separateness that were used in 1950—entrance or access and availability of cooking equipment—are being used in 1960. The change is primarily in the application of the criteria. The basic physical definition of a dwelling unit in 1950 was that it must have (a) separate cooking equipment or (b) two or more rooms with a separate entrance. One room in a "regular apartment house" and one room that was the only living quarters in the structure were exceptions to the rule and could be dwelling units. The basic definition of the housing unit for 1960 is that it must have either (1) direct access or (2) kitchen or cooking equipment for exclusive use.

From the data collection standpoint, the change was designed to simplify and systematize the enumerator's task of delineating separate housing units in converted structures, rooming houses and similar places with complex living arrangements. In 1950, the enumerator either divided these places into individual dwelling units or he classified the whole structure as "nondwelling unit quarters" and singly listed the occupants, providing no housing or household information for the structure. There is ample evidence that different courses were taken by different enumerators under the same circumstances. Generally there is no evidence as to why the decision was made to enumerate the structure as individual quarters or as combined quarters. Structures have been identified in New York and other large cities in which virtually identical living arrangements were enumerated as dwelling units by one enumerator and as nondwelling unit quarters by another. In some instances even the same enumerator used different procedures in two separate structures with similar living arrangements, particularly if one of the structures had a sign "Rooms for Rent" and the other had a sign "Apartments for Rent." In each case the quarters had separate access. The only difference was likely to be that cooking equipment was provided by the landlord in the second structure but not in the first.

The burden on the 1950 enumerator was heavy in areas where these marginal units were prevalent. He had to remember the basic definition as well as the exceptions. The exception that was difficult to apply in these marginal cases was that in a "regular apartment house" one room without cooking could become a dwelling unit, whereas if the structure was not a regular apartment house one room had to have cooking to quality as a dwelling unit. The enumerator's manual did not define "regular apartment house." Consequently, this was subject to varying interpretations by enumerators.

After many months of discussion with a working group from advisory committees, the idea evolved that the enumerator should be concerned with only one conceptual task, namely, to ascertain when people live "separately." In essence, the enumerator may accept the respondent's interpretation of separateness, although whereever more than one family is reported at an address he asks if they "live and eat together or have separate quarters." If they have separate quarters, then the more objective criteria of "access" and "cooking equipment" are applied. Pretests in Ithaca, New York and Washington, D. C. indicated that these criteria are applied uniformly only when enumerators must consciously record them on a schedule. No such record was required in 1950.

The application of the physical criteria reduces individual enumerator interpretation by the requirement that when living quarters have either "direct access" or "kitchen or cooking equipment for exclusive use," or both, they are to be classified as housing units. If the quarters have neither of these criteria they are not considered sufficiently separate to be classified as housing units. They are therefore combined with the quarters through which occupants entered even though the respondents in such quarters consider themselves as living separately.

With these new procedures, there are only three aspects of the housing unit determination that are still somewhat subjective.

(1) If the respondent in a converted structure having two families, e.g., parents on the first floor and a son and his family on the second, does not consider that the families live separately, the enumerator records access and cooking only for the whole structure. In such cases, the physical criteria for two units may exist although they are ignored on the basis of the respondent's interpretation of "living together."

(2) The determination of what constitutes direct access may depend in part on the enumerator's interpretation. The definition gives two types, direct from outside the structure and through a common hall. The common hall, vestibule or lobby is defined as one that is not a part of any unit, but is clearly separate from all units in the structure. In application this common hall definition may not be uniformly applied. For example, in certain types of converted houses, a center hall may be used both as a common hall for all families and as a means for one family to go from one of its rooms to another.

(3) The third aspect relates to marginal types of cooking equipment such as hot plates. To be considered as cooking equipment, these appliances must be "used for the regular preparation of meals." The word "regular" may have varying interpretations among the enumerators.

All of these three aspects were also present in the 1950 definition but under less precise definitions. Our pretests have indicated that the more stringent definitions led to more uniformity in interpretation of these rules than in 1950.

A final difference in procedures involves the enumeration of rooming houses and separate living quarters in institutions. In 1950, any living quarters containing ten or more unrelated persons was enumerated as nondwelling unit quarters for which no housing information was obtained. Mechanical editing reduced this maximum to five or more lodgers. In 1960, for rooming houses, the enumerator will first find out whether the living arrangements of the occupants meet the separateness criteria. Each of the quarters that meets the criteria becomes a housing unit. Those that do not are combined. If there are five or more persons unrelated to the person designated as head, the combined quarters are enumerated as "group quarters" and no further housing information is obtained. In institutions for 1950, separate quarters were enumerated as dwelling units only when such quarters were in structures containing no inmates. In 1960, all quarters occupied by staff personnel will be enumerated as housing units if they otherwise qualify. In dormitories, student quarters are excluded from the housing unit universe.

The reasons for the change

The plan to simplify the procedures and improve the coverage of housing units in the 1960 housing universe is largely a direct result of Census users' need for housing characteristics for the entire housing inventory--for all the places where people live in separate quarters.

Housing statistics have been an increasingly important element in economic planning on a number of fronts, such as housing demand, marketing research for consumer goods and many other areas. But in city planning, in urban renewal or in slum clearance programs, the need becomes acute. These programs generally are designed to upgrade economic and social conditions in the decayed parts of large urban areas. It is in these crowded sections of cities that the 1950 Census of Housing was not entirely adequate, since the 1950 procedures did not provide separate housing statistics for the partially self-contained living quarters so common in blighted or slum parts of cities.

The high incidence of partially self-contained living quarters in heavily populated urban areas is plainly evident in 1950 publications for census tracts. In many urban tracts a large number of noninstitutional persons are reported as not living in households. Obviously, these persons were living in quarters that did not meet the 1950 dwelling unit definition, and no housing information was obtained for their quarters. Information on age of structure, condition, plumbing facilities, family income and number of persons per room, influences the selection, and is used to define the boundaries of potential planning, renewal or clearance areas. For planning purposes, these housing data should be available for all housing in an area, not just a part of it.

Other users of housing statistics, such as those concerned with housing market analysis, will find that the new definition will provide information for other types of quarters formerly excluded from the dwelling unit count. Many of the occupants of such quarters may be part of the potential demand for new housing or for consumers' goods. Common examples are young married couples who occupy a room with direct access but no cooking equipment. In economic terms, many persons in quarters classified as "nondwelling unit quarters" in 1950 differed little from those in dwelling units. Nevertheless, no household characteristics, such as age of head or income were available for such persons or groups of persons.

Finally, tabulation of these two new housing characteristics—access and cooking equipment—will provide a much better description of the character of living quarters than has hitherto been available. The term "quality," when applied to housing, previously has depended on condition of structure and availability of plumbing facilities. Data now will be available on condition, type of access, availability of cooking equipment, and plumbing facilities in evaluating quality of housing.

Estimates of the volume and characteristics of units added

In the course of making the change in concept, the Bureau obtained some information on the probable magnitude of the change as well as on the characteristics of the units that might be added. This information was obtained from small scale studies by a task force of housing technicians working with Census staff and from a national sample taken in conjunction with the Current Population Survey in November 1958. It should be emphasized that the information cannot be considered as definitive. Under more favorable circumstandes the national sample would have produced more reliable quantitative statistics. This survey, however, required the same enumerator to apply both sets of enumeration unit definitions, the old and the new, with only a minimum of training. This fact, plus an indication that some enumerators may have suspected that the supplement was a device for checking the accuracy of their work, is believed to have led to an unknown amount of bias.

Keeping in mind that these figures should be considered as approximations at best, some indication of what was found is given in Tables 1 and 2. In Table 1 it is assumed that the number of occupied and vacant housing units in 1960 will amount to about 60.8 million. The November survey indicated that there would be 59.6 million dwelling units as defined by the 1950 procedures. The remaining 1.2 million were found either in quarters classified as nondwelling units in 1950 (500,000 units) or as additional units in quarters classified as one dwelling unit by the 1950 definition (700,000 units). Of the 700,000 units, 100,000 may be assumed to represent a more precise delineation of dwelling units. Since all 700,000 of them have both separate cooking and direct access, they should have been counted separately even under the old definition.

The 60.8 million in this table is also distributed by subcategories of the 1960 housing unit universe. The first group which is tentatively labelled "regular housing units" are those which were found to have both direct access and cooking equipment or kitchen for exclusive use of the occupants. Substantially all of these quarters were dwelling units by the old definition. Even the two hundred thousand not so classified may well represent the sorts of dwelling units not separately identified under the old procedures. As pointed out above, 100,000 were found within quarters defined as one dwelling unit earlier. The balance were quarters that had the access and cooking qualifications of the 1950 dwelling unit but were located in rooming houses, hotels, motels, institutions and the like. The second group, for the purposes of this paper defined as partially self-contained units, had either direct access or cooking equipment or kitchen for exclusive use, but not both. Even though deficient in these criteria for separateness, half of this group were included among the regular dwelling units by the 1950 definition.

Thus from this table, it is possible to draw two conclusions with respect to the probable impact of using the new unit of enumeration:

(1) While the gross increase resulting from the change should be around 1.2 million, one sixth of these units have all of the characteristics of the regular dwelling units and probably should have been included earlier.

(2) Of those units which are lacking in some aspect of complete separateness, half were included under the old procedures.

These conclusions might also be stated in terms of quality of the results:

(1) Coverage of units having either complete separateness or partial separateness will be more uniform.

(2) The user of housing statistics will have more detailed information relating to the adequacy of living quarters.

In Table 2 distributions are shown for three characteristics of partially self-contained units. Again it should be emphasized that these are only approximations of what we may expect to find. Also this group includes both of those that were previously included in the dwelling unit universe (about half) and the additional units included by the new definitions. From these distributions, however, we can conclude that relatively large proportions of these marginal living quarters consist of one furnished room occupied by one person who is a renter.

The qualitative evaluations of the task force mentioned earlier lend support to the distributions for rooms and tenure. There appears to be some question, however, as to whether the number of one-person units should be as high as 70 percent of the partially self-contained units. Observations in New York City and Memphis, Tennessee, indicated relatively large numbers of such quarters contained families.

<u>The question of comparability</u>

Any change or improvement in statistics always raises the problem of comparability. The merits of the improvement must be weighed against the disadvantage of a break in the series. In some respects the break is of a minor nature as far as both housing and population are concerned. It may be no greater than occurred between 1940 and 1950. Only this time we shall be able to quantify the change, whereas previously we have not been able to do so. The obvious advantages of better coverage and better descriptions of adequacy of all living quarters appear to outweigh the disadvantages. Furthermore, even if the old procedures were to be used again, only a rough comparability would exist. The definitions were subject to wide variability of interpretation by individual enumerators.

There is one problem of comparability of housing statistics that has given housing analysts some concern. This stems from the fact that added units which are carved out of the old dwelling units (the 700,000 shown in the first column of Table 1) will give rise to a large proportional increase in the number of units sharing plumbing facilities, although the absolute numbers will still represent a tiny part of the whole inventory. The increase will occur when a unit by the old definition contained one bathroom and is now reported as two or more units. Not only such added units would be shown as sharing these facilities but also that part of the original unit which is left. While it is true that this may result in an apparent increase in substandard units as defined by the housing agencies, it does not represent a change in the true situation. The facilities were in fact shared before the change. It simply was not recognized that there were two separate units using the facilities. Because of the relatively small numbers involved, this increase in sharing will be obvious only in tabulations for some small areas, such as tracts, where these complex living arrangements are prevalent.

The effect of the change in definition may also be acute in the annual series of households and in the anticipated shift of persons from secondary to primary individuals. The change over to the new definition will take place at the time of the Census. Some change in level as well as in characteristics of households will be apparent. On the assumption that 15 to 20 percent of the added units will be vacant, the 1960 household figure from the Current Population Survey may be one million greater than the figure in 1959 solely because of the change in definition. Since about half of these added units result from fragmentation of households in the earlier series, a slight decrease in household size may be expected. Furthermore, there may be roughly 200,000 (or two-thirds) fewer secondary families and up to 700,000 persons whose status may be shifted from secondary to primary individuals. The number of subfamilies living with relatives may be reduced slightly, but not significantly. Finally a larger proportion of one-person households may be found.

Some help in relating the old household series to the new can be expected from the postcensal surveys. At least these surveys will provide a description of the added households and possibly some indication of the source of such households. It is also hoped that a means can be devised for measuring the extent of the change in the annual household series by examining the returns in the Current Population Survey.

The plans for postcensal evaluation of the change

The plans for postcensal evaluation have not yet been formulated in detail. These will, however, include identification of a sample of those units which would not have been counted as separate units in 1960 under the correct application of the 1950 definitions. From this evaluation, reasonably precise estimates of the magnitude of the change can probably be made. Some information will be available on the housing and household characteristics for the new units and for the old units, if any, from which they were derived. These surveys will include information on the source of the added units, that is the former classification of the quarters, as either dwelling unit or nondwelling unit.

Tabulation and publication plans

A simple distribution of all units by whether or not they have direct access, cooking equipment and bath or toilet for exclusive use will be tabulated for all areas. Those units lacking one or more of these characteristics will be tabulated in three groups: (1) with direct access and cooking exclusive but lacking exclusive use of bath or toilet; (2) with direct access but shared or no cooking equipment; and (3) with cooking exclusive but access through another unit. For places of 25,000 or more population, this distribution will also be tabulated by tenure. It is intended to publish these data for all areas containing a significant number of partially self-contained units. Unpublished data for other areas will be available upon request.

In addition to the simple distribution, cross-tabulations of the three groups of units lacking or sharing access, cooking and bath or toilet will be published. The cross-tabulated characteristics will include such items as tenure, rooms, condition, number of units in structure, year built, number of persons, rent, and income. The areas for which these cross-tabulations will be shown include standard metropolitan statistical areas and cities having 100,000 or more inhabitants in 1960.

Table 1.—EFFECT OF CHANGE OF DEFINITION ON ESTIMATED NUMBER OF 1960 UNITS OF ENUMERATION

Living quarters by 1950 definition	Housing units by 1960 definition (in millions)		
	Total	Regular housing units	Partially self-contained units
Total units	60.8	58.8	2.0
Regular dwelling units..	59.6	58.6	1.0
Additional quarters within dwelling units..	0.7	0.1	0.6
Quarters in nondwelling units................	0.5	0.1	0.4

Source: Derived from unpublished data collected by the U. S. Bureau of the Census in the current Population Survey, November 1958.

Table 2.—CHARACTERISTICS OF OCCUPIED PARTIALLY SELF-CONTAINED HOUSING UNITS

Characteristic	Percent Distribution
All partially self-contained housing units................	100
Rooms:	
1..........................	61
2..........................	12
3 or more..................	24
Not reported...............	3
Persons:	
1..........................	70
2..........................	14
3 or more..................	14
Not reported...............	2
Tenure:	
Owner......................	9
Renter.....................	64
Furnished.................	53
Unfurnished...............	10
No cash rent...............	27

Source: Derived from unpublished data collected by the U. S. Bureau of the Census in the Current Population Survey, November 1958.

COMPONENTS OF CHANGE IN THE NATION'S HOUSING INVENTORY IN RELATION TO THE 1960 CENSUS

By: Frank S. Kristof, Bureau of the Census

Introduction

In the short period of 25 years we have observed the transformation of the residential construction industry from a purely private activity of the economy to one of our most publicized, analyzed, and legislated industries. The residential construction industry had constituted an important segment of the Nation's economy before it became clothed with a public interest. However, when it was seized upon, during the 1930's, as both a means of stimulating a lagging economy and of raising the American standard of living, our housing industry moved to the forefront of political and legislative activity. Again, when the unflagging pace of housing construction helped shore up the economy during the recession of 1953-54, the Cinderella-like quality of the industry was further enhanced.[1] Today, with the problem of urban blight plaguing every sizable community in the Nation, housing problems associated with slum clearance and urban renewal have become almost a daily topic of discussion.

It is no coincidence that during this same quarter of a century that statistical knowledge about the housing field has jumped from almost nothing to a fairly well documented aspect of American life. The real property inventories of the 1930's, followed by the 1940 and 1950 Censuses of Housing provided the basis for this major statistical breakthrough. Data provided by the censuses actually preceded the existence of agencies able to exploit fully this information--particularly at the local community level.[2]

During the early 1950's pressures developed from several sources for more sophisticated and useful data on the dynamics of housing than that provided by the Housing Census. The first major recognition of this need was a joint project of the Housing and Home Finance Agency and the Bureau of the Census. The housing experts of these agencies undertook the task of defining concepts to express these needs. Techniques for collecting the necessary data were explored and developed, and the results of the project were reported in a formidable 200 page document.[3] The culmination of this work was a two sentence congressional appropriation in 1956 for $1,000,000 for "expenses necessary for conducting a survey of housing." Shortly thereafter the Bureau of the Census created a separate Housing Division to carry out the 1956 National Housing Inventory--the first systematic attempt to measure changes in the Nation's housing inventory.

The National Housing Inventory was undertaken in the fall of 1956 and preliminary results were released in the fall of 1957; the first of twenty-one reports was issued in May of 1958 and the last appeared in April 1959. Despite the time lapse in producing them, these reports were well received in housing circles, and during the planning for the 1960 Housing Census many requests were received by the Bureau to incorporate a Components of Change survey within the Housing Census program. In response to the urging of the Bureau's Housing Advisory Committees and the Federal Housing Agencies, the Housing Division included within its census program plans for a Components of Change survey.

When the Director of the Bureau of the Census affirmed a decision to undertake a Components of Change survey in the fall of 1959, it marked a milestone in the field of housing statistics in the United States. This decision was the green light for a project which would provide the first complete explanation of the changes in the housing inventory which occur between decennial censuses. Enumeration was undertaken this fall and is nearing completion. With good fortune we should have our results by the end of next year.

The Components of Change program has been designed to utilize the total inventory figures to be obtained from next year's Housing Census to increase the reliability of the new construction estimates. And through the Bureau's Post Census Evaluation Program, we hope to link the 1959 dwelling unit inventory figures with the 1960 Housing Census results obtained from use of the new unit of enumeration--the Housing Unit. The linking process should account for the differences between the inventory figures obtained from the two programs attributable: (1) to the differences in definition of the unit of enumeration, and (2) to the time lapse between the 1959 and 1960 enumeration.

Before further discussing the inter-relationship of the Components of Change and other phases of the 1960 Housing Census program, it would be useful to review the nature and objectives of the Components of Change program as it was developed in 1956.

Major Subjects of Inquiry

What is the character and significance of the changes which were to be measured? These may be classified under the following major categories:

New construction.--What is the rate of new additions to the housing inventory? Despite two housing censuses and Bureau of Labor Statistics monthly housing starts data,[4] no really accurate measure of this statistic existed. BLS housing starts data, even when placed in the most favorable light, have been shown to understate nonfarm new construction when compared with Census results.[5] The "year built" data of both the 1940 and 1950 censuses were widely at variance with the BLS estimates. However, the census "year built" data could not be accepted as conclusive because this information was subject to response error. As a result, users of data have remained dissatisfied with available estimates of new residential construction A corollary question in connection with new additions to the inventory is: What is the rate of "other additions" to the housing inventory, i.e., the shift from nonresidential to residential use? With respect to this question, virtually no information was available.

Conversions and mergers.--These types of changes within the housing inventory long have been the subject of deduced estimates, but of little certain knowledge. During the depression thirties, about a third of the additions to the housing inventory were attributed to conversions--the division of existing dwelling units into smaller units. During the war and immediate postwar forties a somewhat smaller proportion of the net additions was credited to this source. Housing analysts believed that this phenomenon had become a declining influence on the housing inventory by the 1950's, but there were no figures to demonstrate it. Correspondingly, they also suspected that mergers, or the combining of dwelling units into fewer units were becoming an increasingly important source of losses from the inventory. It is, of course, recognized that conversions and mergers add or subtract virtually nothing from total living space, but merely change the existing supply into a larger number of smaller units or smaller number of larger units. Nevertheless, no analysis of the housing inventory would be complete until this phenomenon can be measured.

Demolitions and other losses.--Possibly the most distressing gap in housing knowledge was in data on withdrawals from the housing supply. How much of the Nation's inventory is lost annually through demolition, fire, flood or the shift from housing to nonresidential purposes?

Direct information was of the most fragmentary nature. Some estimates had been made on the basis of the overall changes in the total inventory measured by the 1940 and 1950 housing censuses. But even these were subject to varying assumptions about the other components of change, i.e., new construction, other additions, conversions and mergers.

Qualitative changes in the housing inventory.--In what direction is the quality of the existing housing inventory changing over time? Although new construction tends to improve overall housing quality, what are the characteristics of the housing being removed from the inventory? More important, what is happening to the remaining inventory? And at what rate?

These questions are particularly relevant in any discussion of "housing needs" for America. This subject probably has generated more heat in the housing field than any other as a result of the widely disparate estimates of "housing needs" that have been developed over the years. Glenn Beyer notes that "when the 1949 Housing Act was being considered in Congress, leaders of the home building industry set a range of 600,000 to 900,000 per year" as the annual need for new construction.[6] For the ten year period 1956-1965, Beyer places the need between 1,300,000 and 1,900,000 annually depending upon the assumptions used.[7] At the other extreme, estimates by William Wheaton placed annual need at 2.0 to 2.4 million units annually.[8] Although much of the differences among these estimates stemmed from disagreement among analysts over the rapidity and methods by which the substandard inventory was to be eliminated, a part of the difference is attributable to different guesses about the direction and rate of qualitative changes in the housing inventory.

Household formation.--The annual rate of new household formation is another statistic that eludes accurate measure between censuses. True, the Bureau of the Census now publishes annually a total household figure obtained from its monthly Current Population Survey, and from this has been derived an annual change in number of households. The latter figure, however, is subject to such a large variance that it is of little utility in making year to year estimates of household increase. As long as more precise data on this subject are lacking, analysis of market demand for housing based upon the rate of household formation cannot be undertaken.

Mobility and housing demand.--The mobility of the American population has created problems for housing analysts in estimating the ability of local markets to absorb new construction. Some analysts believe that, by itself, "sheer mobility alone tends to increase the quantity of housing demanded even if net migration is zero."[9] The various types of population mobility further complicates the problems of analysis, e.g.:

(1) The general movement from the central city to the suburbs;

(2) the continued shift from rural to urban areas, including the shift of nonwhites from the rural south to central cities of the north, and the migration of Puerto Ricans to New York City.

(3) The regional movement to Florida and the West Coast (particularly California).

(4) Growth and mergers in American industry have created large industrial aggregates and established thousands of new branches in suburban and rural areas, resulting in the movement of tens of thousands of business managers, engineers and other scientists in a constant cross-current of movement.

One aspect of our population movement which has been well noted is the increase in home ownership from 55 to 60 percent of all households since 1950. Other aspects of this movement are less well known. For example, what are the components of the net shift in tenure? Who are the people that move--are they a typical cross-section of the population or not? Do they have special characteristics which will help predict the impact of future movement on housing demand? Does sheer movement, per se, increase the demand for housing? What effect do household moves have upon the relative expenditures of income for housing?

The Results of the 1956 National Housing Inventory

The information yielded by the 1956 National Housing Inventory provided direct answers to many of the questions raised in the preceding paragraphs. Some of the results and experiences associated with obtaining them are worth reviewing.

NHI results--new construction.--Since new construction represents the largest change in the housing inventory, these figures were of most interest. Staff members of the Housing Division had made national estimates of new construction for the 6-3/4 year period ranging from 9.6 million to 11.5 million or annual rates of 1.4 to 1.7 million.

These figures were higher than any existing published estimates because of a conviction that BLS starts were continuing to understate the volume of new residential construction. The first UNIVAC runs on new construction were dismayingly low. Of the nine SMA's, five actually were slightly below BLS estimates. The national figure was 9.1 million compared to a comparable BLS estimate of 8.1 million for this period. On the surface, the national new construction figure was not unreasonable even though it was below the lowest staff estimate. The extent to which some of the SMA figures were below BLS estimates was more disturbing. This, plus a "red light" warning of a control figure built into the tabulation program, led to a complete review of the processing procedures. Parenthetically, it might be observed that not only was the NHI conceptually a new program, but it was the first major survey to use the new FOSDIC device.[10] Consequently, there were serious fears that something untoward had developed in processing. These fears were realized when a print-out of an SMA output tape was compared with the original schedules and it was discovered that FOSDIC had only intermittently transcribed the new construction mark from the microfilms to the magnetic tape.[11]

Time and cost factors prohibited any consideration of a "refosdicing" the microfilmed data. Since the tracked down difficulty affected only the simple count of new construction, it was decided to obtain these figures for the nine SMA's and the United States by a hand count. The final verified hand counts largely closed the gap between the BLS and NHI figures in the SMA's, nevertheless, it left four SMA's with slightly lower totals than the BLS estimates since the overall count was increased by only four percent. The significant change as a result of the hand count occurred for the U.S. total which was raised from the original UNIVAC figure of 9,100,000 to 10,900,000, an increase of about 20 percent.[12]

Although Housing Division personnel involved in processing NHI spent several anxious weeks before finally obtaining a valid new construction count, the experience carried an important lesson. No matter how marvelous and efficient are the complex electronic devices available to us, they will perform no better than the quality of the personnel responsible for their operation. Although they need not be responsible for operations, subject matter people must understand the workings and monitor the output of our electronic equipment. In this instance, had the subject matter people been unfamiliar with operations, the warning provided by the check incorporated within the UNIVAC program might have been rationalized or even ignored. But with operational knowledge, housing personnel were alerted by the warning, and coupled with their intuitive conviction about the magnitude of new construction figures to be expected, the figures obtained from the original UNIVAC run were emmediately investigated.[13]

It was many months before the significance of the new construction data provided by NHI had any public impact. And when it came, the reactions were understandably adverse. The possibility that the United States had been adding new housing at an average rate of 1,600,000 units annually since 1950 seemed incredible. It is not surprising that the figure was branded as incorrect by many analysts. Many months of checking by staffs of the Bureau and other interested government agencies, however, failed to reveal any significant source of error in the NHI findings.

Only recently, a leading daily newspaper headlined a story that the government was looking for 2 million houses "lost" somewhere between 1950 and 1956--referring, of course, to the BLS-NHI disparity.[14] This time, however, Federal administrative processes moved rapidly. While some readers were learning for the first time in November 1959 that the Nation had built far more homes than it suspected between 1950 and 1956, the responsibility for collecting housing starts data had rested with the Bureau of the Census since July 1, 1959. Henceforth, inconsistencies between housing starts statistics and future components of change data will be the headache of a single statistical family. But with the responsibility for both sets of data so centralized, the pressures for clarifying concepts and improving methods of data collection to close the statistical gap will be multiplied manyfold.

<u>NHI results--conversions and mergers</u>.--The intuitive feeling among housing analysts that conversions had become a declining influence on the Nation's housing inventory was confirmed by NHI. Approximately 700,000 units became 1,400,000 between 1950 and 1956. This average of one unit gained from each one involved in conversion is the lowest that normally might be expected from this process and it signalizes the end of the era of large old single family mansions being cut up into four, six, or ten units. The evidence in table 1

Table 1.--CONVERSIONS AND MERGERS IN THE HOUSING INVENTORY: 1950 TO 1956

(In thousands of dwelling units)

Area	Conversions			Mergers			Net change from conversions and mergers
	From	To	Net Gain	From	To	Net Loss	
United States	668	1,376	708	1,321	649	672	+36
Inside SMA's	422	880	458	765	371	394	+64
Outside SMA's	246	496	250	556	277	279	-29
Northeast	199	412	213	349	170	179	+34
North Central	215	452	237	422	216	206	+31
South	184	378	194	431	204	227	-33
West	71	133	62	120	60	60	+2

indicates that this type of conversion, since 1950, no longer has been an influence of significance in the housing supply.

Almost as interesting is the manner in which mergers have completely nullified the effect of conversions on the housing inventory. The picture presented in table 1 demonstrates that this occurred in each of four regions as well as inside and outside standard metropolitan areas. Again, the average loss of one unit from each two involved in merger is the exact reverse of the results obtained from conversions.

NHI results--demolitions and "other losses."--The loss from all sources of 2-1/2 million units in 6-3/4 years, an average of nearly 400,000 annually, was somewhat higher than most analysts' estimates. That less than half of these were classed as demolitions might seem a little surprising, but the demolition figure probably is a conservative estimate because some units actually demolished were incorrectly reported as "other losses" by enumerators. The "other losses" component of withdrawals from the housing inventory, which numbered 1,400,000, probably is conceptually the weakest figures produced by NHI. A significant proportion of "other losses," about 400,000, consisted of units recorded by enumerators as "abandoned." This proved to be a poor concept in that it offers to less conscientious enumerators an easy disposition for vacant units in rural areas when no one is conveniently nearby to furnish information on the dwelling. Post enumeration investigation of "abandoned" units suggests that some proportion of them should have been classified as vacant. For this reason the term "abandoned" has been discontinued for both the 1960 Housing Census and the 1959 Components of Change program.

Another and equally large component of "other losses"--about 400,000 units--was classified as "moved from site." This is essentially a segment concept rather than an inventory concept, that is, it is associated with the data collection technique. The concept "moved from site" has real meaning to an enumerator seeking a house which has moved out of the way of a road-building project, or trying to explain a trailer which has left the site in the segment. This concept has a counterpart within the category "units added through other sources," and is usually labeled "moved to side."

Although these two concepts literally are examples of housing mobility, they are not, strictly speaking, housing inventory concepts since they really do not represent additions or subtractions from the Nation's housing inventory. However, even this statement must be qualified. For example, a vacation trailer which is sold and moved from a backyard to a lot where it is used as a permanent residence actually is an "other gain" to the housing inventory.[15] A house that is moved out of the way of a road-building project to a "used house" lot or to another site where it is used for nonresidential purposes actually is an "other loss." Some of the houses moved out of a segment may subsequently have been demolished. It is not known to what extent this occurred. For this reason it was decided to record all units moved into, or moved out of a segment as "other gains" or "other losses." Since these two factors essentially are opposite sides of the same coin, they should be self-cancelling except for the type of examples cited above. Actually, about 400,000 units were recorded as moved out of segments and slightly less than 500,000 were recorded as moved into segments.

Of the remaining 600,000 units ascribed to "other losses," nearly 200,000 represented shifts from dwelling units to quasi dwellings, i.e., rooming houses, boarding houses, transient hotels, etc. It is probable that some large old former single family structures have continued to disappear into this category. Another 250,000 units shifted to nonresidential use and the remaining 150,000 were ascribed to fire, wind, storm, and "torn down," i.e., demolitions inaccurately recorded in this category.

It is probable that the rate of demolitions and "other losses" will continue to increase in the years to come. The full impact of urban renewal and our highway construction programs has not yet been felt. Although the future rate of withdrawals from the housing inventory is a matter of speculation, there is little evidence to justify a recent estimate that "new permanent construction for the next decade will have to be 1,250,000 units greater than the net increase in households and vacancies in order to replace those units lost from the inventory."[16] An estimate of twice this size would not be unreasonable.

NHI results--qualitative changes.--The NHI figures on condition and plumbing facilities showed a significant decline in substandard dwellings between 1950 and 1956. About 3,000,000 units, which were recorded in 1950 as dilapidated or lacking some or all plumbing facilities, in 1956 were recorded as not dilapidated, with all plumbing facilities. Upon consideration, this change is not too surprising since the Nation's level of living has moved upward in every respect during the 1950's. According to the definition of substandard housing generally accepted today, two-thirds of the Nation's rural housing, much of which was sound in construction, but which lacked running water and inside plumbing facilities was classified as substandard in 1950. Since the nature of urban life makes a facility like running water a necessity, whereas this is not necessarily the case in homes in rural areas, application of the term substandard to some of our rural housing probably is questionable. Nevertheless, nearly 2,000,000 of the 3,000,000 units which changed from substandard to standard between 1950 and 1956 were outside standard metropolitan areas, where much of the housing is rural.

Nothing in these figures, however, should give cause for any complacency about the problem of substandard housing. It is to be expected that housing should participate in the general rise in our level of living, and presumably only the lack of adequate data on expenditures on private residential repairs and rehabilitation justifies surprise at the extent of improvement in housing quality recorded since 1950.[17] On the other hand, it is reasonable to presume: (1) that major expenditures on improving existing housing most frequently occurs in urban neighborhoods that are not rundown and in rural housing that does not require major rehabilitation; and (2) that the largest proportion of repairs and improvement expenditures was not major rehabilitation but the installation of inside plumbing facilities, particularly running water and water heating equipment--a relatively less expensive means of shifting units from substandard to standard status than major rehabilitation work. This would argue, then, that the rapidity of the improvement in housing quality from 1950 to 1956 is attributable to the completion of the easier and less expensive improvements required to remove housing from substandard status. If this is true, it means that continued improvement in quality of the existing stock during the 1960's will come harder and more slowly. It also means that the hard core of the substandard housing problem--namely homes of low income owners in deteriorated neighborhoods who are unable to improve their housing and rental properties in slum areas whose owners are not financially justified in investing in improvements or unable to obtain financing when they are willing--basically is not touched by this type of upgrading.

Consequently, it may be unwarranted to utilize a projection of the 1950-56 NHI results to justify the conclusion reached by Reinhold Wolff that repair and rehabilitation will upgrade as many as 4,600,000 substandard units during the 1960's, unless there is substantial increase of local and Federal activity in the form of code enforcement, urban renewal, and subsidies to accelerate the removal or improvement of substandard housing.[18]

NHI results--household formation.--In March 1957, the Census Bureau published a CPS household figure of 49,543,000 which, compared to the CPS household figure of March 1950, gives an annual rate of household increase of 856,000 for the seven year period.[19] The 1956 NHI figures for occupied dwelling units (households) was 49.9 million compared with the 1950 Census figure of 42.8 million, or an annual rate of increase of about 1,000,000 for the 6-3/4 years. Although dissatisfaction has been expressed with this variance in household figures emerging from the same agency, it is a danger inherent in the measurement of the same concept by separate surveys with different orientations. The primary task of CPS is to identify employment status of persons in the labor force. True, in the process, households are defined and an attempt is made to obtain their proper definition, but it is not the primary emphasis. in NHI, on the other hand, great stress in the training and induction of enumerators was placed upon the definition of the dwelling unit.[20] As a consequence, it is reasonable to expect that the NHI enumerators did a more thorough job in this area than is done by the CPS. Although it is regrettable that such differences occur, it should be remembered that before the NHI results became available, CPS was the only source of any information on households. Even though Bureau statisticians had reservations about the use that might be made of the annual CPS household figure, it was published upon the urging of many users that some information on households was better than none at all.

NHI results--mobility.--Some aspects of the amazing mobility of American households were measured by NHI. The dimensions of this mobility were familiar, but its characteristics were not well known. For example, it was well established that owner households tended to move less than renter households, but it was not known that over half (57%) of the renter households (as of 1956) had been in their present unit between two and three years while about half of the owner occupants had been in their units 6 years or more. The instability of the renter population was least marked inside central cities--47 percent of the households outside SMA's had moved into their present unit within the past two years; for SMA's, the corresponding figure was 44 percent and for central cities 41 percent.

Information obtained on present tenure, previous tenure, and "year moved into present unit" permits analysis of the increase in owner-occupancy from 55 to 60 percent of all households between 1950 and 1956. About one quarter (14 million) of the Nation's 55.3 million households moved into their dwelling unit during 1955 or 1956. Twelve million of these were households with the same head, of which 9 million were former renters and 3 million were former owners. Although most renter movers remained renters, 2.7 million or 30 percent shifted to owner-occupancy. Only 3/4 million former home owners became renters; however, they constituted 27 percent of all former owners who moved in these two years. During this same period, about 2 million moves involved a change in household composition, a large part of which represented new household formation. Three quarters of these 2 million movers whose previous tenure was not investigated, became renter households.

Not too much can be said from our tabulations about the distance of moves. For example, 10 percent of all households who moved crossed State boundaries. But we do not know what proportion of this group merely crossed nearby borders, such as the Hudson River from New York City to New Jersey or vice versa, and what proportion made cross-country moves. The flow to the suburbs, however, is indicated in the moves of households formerly residing in central cities. About one third or 1.3 million of nearly 4 million central city households who moved in 1955 and 1956 left the city.

The majority of households who moved from one owner occupied single family unit to another paid more for their new homes. The median value of the house they moved from was $10,000 compared to a median value of $13,100 for their new house--a 30 percent increase.[21] There are also significant differences between households which have moved from one owned home to another and all owner households. For example, their median household size was 3.5 compared to 3.1 for all owner households; their median income was $5,400 contrasted to $4,800 reported for all owners.[22] The quality of homes acquired by these movers also improved significantly. Whereas 16 percent of all owner occupied units in the United States were dilapidated or lacking plumbing facilities (substandard), only 9 percent of the units occupied by recent owner to owner movers fell into this category.

In contrast, the characteristics of households who moved from renter-occupied units to other rental units are not very different from all renters. In terms of household size, median number of persons for renter to renter movers was 3.0 compared to 2.8 for all renter households; median incomes were about the same, $3,700 for renter to renter movers compared to $3,600 for all renters. Similarly, there was virtually no change in the quality of housing of this group of movers, although there was a slight but possibly not significant increase in median contract rent from $54 in the previous unit to $58 in the new unit.

A comparison of the average value of homes occupied by former renters with the average rents they used to pay shows a correlation of +.48, but the dispersion is wide. In short, previous rents would be poor predictors of the value of homes presently occupied.[23] The median contract rent of households which shifted from renting to owning in 1955 and 1956 was $61. Although this was higher than the $54 median for renters who moved to other rented units, many of the new home owners came from the low rent brackets. About a sixth previously paid less than $40 a month rent. The wide spread in value of homes purchased by these previously low rent payers is shown in the following illustration:

Table 2.--VALUE OF ONE-DWELLING UNIT NONFARM STRUCTURES FOR RECENT MOVERS FROM THE $30-$39 RENT CATEGORY: 1956

Percent of recent movers from the $30-$39 rent category	Purchased homes valued at:
27	$6,000 or less
27	$6,000 to $10,000
26	$10,000 to $15,000
19	$15,000 or more

Source: Unpublished 1956 NHI tabulations

Only one cross-tabulation was obtained for recent movers who shifted from owner to renter occupancy (table 4). The data show a correlation of +.46 between value of previous residence and the present rent paid by former owner-occupants indicating that the higher the value of the previous residence, the higher will be the present rent paid by former owner-occupants. The median value of their former homes was the same as that for all owner-occupants who moved in 1955 or 1956--$10,000. But the $68 median contract rent of former owner-occupants was $10 higher than that of movers from renter-occupied units and $15 higher than the median contract rent for all renter-occupied units. This fact seems to be significant relative to the demand for rental units. If the shift from owner to renter occupancy increases appreciably, the demand for better quality and higher rent units may be strengthened. Even today, this may be a factor in the recent resurgence in the rental market discussed by Louis Winnick's ACTION report.[24]

<u>NHI results--some conclusions</u>.--The 1956 NHI results answered many questions that have been raised about the housing market in recent years. Even the cursory discussion in the preceding paragraphs, however, indicates the endless possibilities that exist for more intensive mining of these data. Nevertheless, we now have a basis for evaluating an approach commonly used by housing analysts in estimating future "demand" for new residential construction. This method begins by making assumptions or estimates for the period ahead about:

(1) The state of the economy--employment, incomes, etc.

(2) Net new household formation.

(3) Units lost and units added in the existing inventory.

(4) Change in gross vacancies.

The analysis of "(1) The state of the economy..." is used in making estimates of net new household formation, losses and gains in the existing inventory, and the expected direction and magnitude of change in vacancies. These estimates then become factors in the following equation:

> "(2) Net new household formation" + "(3) Units lost and units added in the existing inventory" + "(4) Change in gross vacancies" (i.e., + increase or - decrease in gross vacancies) = "need" for new construction.

The resulting "need" for new construction theoretically can be translated into effective demand if certain conditions about the price and mortgage credit terms of new construction are met.

Table 5b fills in the variables of the above equation from data obtained from NHI for the period 1950-1956.[25] The adjoining table 5a provides the background of absolute change against which these relative changes may be examined. The relationships derived for this period were obtained from an economic climate of full employment and high incomes even though encompassing the recession of 1953-1954. For the United States as a whole, 178 units were provided for each 100 households added. Of this total, 155 were new units and 23 were additions from other sources. Losses in the inventory amounted to 46 units for each 100 households added. The surplus of 32 (178 provided, less 46 lost) represents additions to the vacancy supply--although only 14 of these were in the form of available vacancies. The other 19 units were either held off the market, owned or bought and awaiting occupancy, or dilapidated.

When the national figures are classified by inside and outside standard metropolitan areas, relationships change sharply.[26] The first and most obvious fact is that relative growth inside standard metropolitan areas was twice that outside standard metropolitan areas (table 5a). More important, however, only 146 units were needed to accommodate each 100 households added inside standard metropolitan areas compared with nearly twice this figure (275) outside standard metropolitan areas (table 5b). A similar ratio held for the number of new units constructed for each 100 households added--128 inside standard metropolitan areas and 233 outside standard metropolitan areas.

Table 3.--CONTRACT MONTHLY RENT OF PRESENT UNITS BY VALUE OF PREVIOUS PROPERTIES, FOR UNITS OCCUPIED BY RECENT MOVERS: 1956

(In thousands)

Contract monthly rent--nonfarm renter units occupied by recent movers	Same Head in Present and Previous Unit									
	Value of Previous Property: Owner-occupied nonfarm 1-unit structures									
	Total	Under $6,000	$6,000 to $7,999	$8,000 to $9,999	$10,000 to $11,999	$12,000 to $14,999	$15,000 to $17,999	$18,000 to $19,999	$20,000 or more	NR
Total............	510	124	43	73	57	67	46	14	54	30
Less than $30.............	31	20	4	2	--	2	--	--	--	3
$30 to $39................	31	10	3	4	3	2	4	--	3	2
$40 to $49................	51	18	4	8	8	2	4	--	--	6
$50 to $59................	71	27	9	21	4	1	--	--	10	--
$60 to $69................	73	10	8	10	15	16	7	1	1	6
$70 to $79................	61	17	6	7	9	10	4	3	2	2
$80 to $99................	69	10	5	9	7	14	14	1	7	2
$100 or more..............	97	3	4	7	8	16	14	9	29	8
Not reported..............	25	9	--	4	3	3	--	1	2	2

NOTE: Due to independent rounding, figures may not add to totals.

Source: Unpublished 1956 NHI tabulations.

Table 4.--VALUE OF PREVIOUS CONTRACT MONTHLY RENT BY VALUE OF PRESENT PROPERTY, FOR UNITS OCCUPIED BY RECENT MOVERS: 1956

(In thousands)

Value of present property--Owner-occupied nonfarm 1-unit structures	Contract monthly rent--nonfarm renter units occupied by recent movers									
	Total	Less than $30	$30 to $39	$40 to $49	$50 to $59	$60 to $69	$70 to $79	$80 to $99	$100 or more	NCR or NR
Total.......................	2,214	142	205	293	346	391	242	248	195	152
Less than $6,000................	227	50	56	52	26	20	2	2	--	19
$6,000 to $7,999................	208	21	27	57	34	39	18	5	--	7
$8,000 to $9,999................	273	25	29	47	64	50	15	18	9	16
$10,000 to $11,999...............	291	12	21	37	66	66	37	27	8	18
$12,000 to $14,999...............	501	20	33	63	87	114	82	53	26	23
$15,000 to $17,999...............	312	10	22	15	35	64	35	53	55	22
$18,000 to $19,999...............	117	--	2	3	17	13	31	32	10	8
$20,000 or more..................	257	2	16	16	13	24	22	56	83	25
Not reported.....................	29	2	1	4	3	--	--	2	4	13

Source: Unpublished 1956 NHI tabulations.

It is clear that, inside standard metropolitan areas, the relationship between net new household formation and new construction (between 1950 and 1956) was close enough so that the measure of one would have provided a good basis for predicting the other, whereas this was not true outside standard metropolitan areas.[27]

Since standard metropolitan areas are frequently treated as analyzable housing markets, and since most forecasts of new construction are made for local housing markets, the NHI figures showing the relationship of new construction to increase in number of households for the nine standard metropolitan areas are of particular interest. These relationships range from a low of 105 new dwelling units per 100 new households in the Los Angeles SMA to 144 per 100 in the Dallas SMA. The average for the nine SMA's was 124, close to the figure of 128 recorded for all SMA's of the Nation. This suggests that the variation in the ratio between new construction and net new household formation for individual SMA's is not so wide as to vitiate the usefulness of the average in predicting demand for new construction for individual SMA's. The relevant question is with respect to the stability of this ratio for the near-term future, i.e., can it be used for prediction?

Examination of the composition of this ratio indicates that there is basis for expecting it to be stable over the next decade for SMA's which continue to grow at approximately their 1950 to 1956 rate. Internal change [28] in the housing inventory of these areas are small and largely offsetting relative to net household growth. However,

Table 5a.--NUMBER AND INCREASE IN HOUSEHOLDS: 1950 to 1956

Area	Number of households 1950 (thousands)	Increase in households 1950 to 1956	
		Number (thousands)	Percent
U. S.	42,826	7,048	16.5
Inside SMA's	24,514	5,264	21.5
Outside SMA's	18,312	1,784	9.7
REGIONS			
Northeast	11,228	1,780	15.9
North Central	12,972	1,618	12.5
South	12,633	1,852	14.7
West	5,994	1,797	30.0
SMA's			
Atlanta	191	66	34.6
Boston	646	50	7.7
Chicago	1,607	268	16.7
Dallas	187	52	27.8
Detroit	829	199	24.0
Los Angeles	1,440	600	41.7
N.Y. - N.E.N.J.	3,774	654	17.3
Philadelphia	1,018	185	18.2
Seattle	236	42	17.8

Source: 1956 National Housing Inventory, Volume I.

Table 5b.--CHANGES IN THE HOUSING INVENTORY PER 100 HOUSEHOLDS ADDED: 1950 to 1956

Area	Number of units added			Minus all losses[2]	Equals net additions	Minus increase in gross vacancies	Equals increase in number of households
	Total	By new construction	From existing inventory[1]				
U. S.	178	155	23	46	132	33	100
Inside SMA's	146	128	17	32	114	14	100
Outside SMA's	275	233	42	86	188	89	100
REGIONS							
Northeast	153	133	20	32	121	21	100
North Central	179	150	29	50	130	30	100
South	231	203	28	70	161	61	100
West	149	131	17	30	118	18	100
SMA's							
Atlanta	139	129	11	23	117	15	100
Boston	182	138	44	44	138	40	100
Chicago	134	115	19	27	106	5	100
Dallas	156	144	12	38	117	17	100
Detroit	136	124	12	23	113	14	100
Los Angeles	117	105	12	16	102	2	100
N.Y. - N.E.N.J.	128	113	16	25	104	4	100
Philadelphia	143	127	16	21	122	22	100
Seattle	138	121	17	33	105	5	100

[1]Includes units added through conversion and from other sources.

[2]Includes units lost through merger, demolition and other means.

NOTE: Because of independent rounding of figures, detail may not add to totals.

Source: 1956 National Housing Inventory, Volume III.

for those few standard metropolitan areas which are growing only slowly, internal changes will be large relative to net household change and the relationship between new construction and net household increase will fall apart.[29]

What are the implications of a close relationship between new construction and net increase in households? To the realtor, the builder, and the supplier of mortgage funds, a close relationship between these two variables probably is desirable because it indicates a stable housing market. From a purely social standpoint, a close relationship between these variables is not necessarily desirable. When the housing market starts from a position of tightness--a very low available vacancy rate--as in 1950, a practically one-to-one ratio between new construction and net new household formation precludes any easing of the market and the efficient operation of the filtration process--one means whereby the quality of the housing supply may be improved.[30] Thus, the construction of 281 new units for each additional 100 households in Philadelphia theoretically would mean that improvement in the quality of housing occurred through operation of the filtration process to the extent that new units replaced substandard units which shifted to the available vacant category or completely disappeared from the housing inventory. However, no such conclusion can be deduced. Although about one-quarter of the increase in Philadelphia's vacancies and three-quarters of the units removed from the inventory were substandard, the city's experience was similar to that of SMA's whose new construction-household increase ratio approached one.

Without further evidence, the most that can be said is that the occurrence of a high ratio of new construction to net new household formation in a housing market creates the condition for an improvement in the quality of housing by providing scope for the operation of the filtration process. So long as such a development does not continue to the point where vacancies become a glut on the market and react negatively on construction activity, a high ratio of new construction to net new household formation may be regarded as desirable.

Part of the preceding discussion alluded to the reversibility of the new construction-net new household formation ratio, i.e., one can attempt to estimate new construction if estimates of household formation are available, and occasionally, estimates of net new household formation for the recent past are made on the basis of new construction figures. The NHI data indicate that the foregoing types of estimates could have been made with reasonable accuracy for the early 1950's. The difficult task for individual SMA's is a forecast of net new household increase which, for the nine NHI standard metropolitan areas, ranged from 8 percent for Boston to 42 percent for Los Angeles. It is obvious that it would be difficult to use the new construction-net new household formation relationship to forecast the demand for new construction unless techniques for obtaining good estimates of net new household formation in local areas are developed. If the new construction-net new household formation relationship for the Nation has any stability, it can be useful for forecasting national demand for new construction since aggregate household growth is more easily predicted than that for a single standard metropolitan area.[31]

Extrapolating the NHI Results to 1959

It will soon be possible to check the stability of the relationships shown in table 5b and thus their usefulness for prediction. What kind of figures will the 1959 program yield on the basis of the 1956 relationships? Since economic conditions for 1956-1959 basically were similar to the earlier years of the decade, it is reasonable to assume a rate of net new household formation of 1,000,000 per year for the past three years. This would give a net addition of 4,000,000 units and a total inventory figure of 59,300,000 dwelling units; the figure for new construction would be 4,650,000.[32] As a forecast, the figure for new construction seems reasonable, but the total inventory figure appears high.[33] At any rate, the results of the 1959 Components of Change survey will permit an evaluation of the stability and usefulness of these relationships.

The 1959 Components of Change program and the 1960 Census

The 1959 Components of Change program will mark the climax of a decade of research in the dynamics of housing statistics. Basically, the program is identical with the NHI which has been discussed in some detail in the previous pages. The experience gained from NHI proved valuable in planning the present survey and some of the problems that arose in 1956 have been eliminated. The information obtained for the United States and the nine standard metropolitan areas in the 1956 sample segments has been transcribed to schedules used in the current survey. This information will be compared with the 1959 status of dwelling units in the same segments. In order to obtain measurement of new construction since 1956, an added group of new construction segments have been delineated and will be enumerated throughout the United States.

The 1959 Components of Change enumeration will be carried out independently of the 1960 Housing Census. However, the total housing inventory figure obtained in the 1960 Census will be used for ratio estimating to obtain a more accurate new construction figure, which is the largest single change in the inventory. This will require re-enumeration of the components of change segments immediately after the 1960 Census to obtain the count of housing units--the new unit of enumeration to be used in the 1960 Housing Census--in those segments. The count of housing units in the sample segments, and the total count for the Nation will provide the 1960 factors for the ratio estimating equation.

At an earlier stage it was planned to utilize the 1960 re-enumeration of the Components of Change segments to obtain a complete reconciliation of the 1959 dwelling unit figures with the 1960 housing unit figures--including accounting for changes that occurred in the time lapse between the survey and the census. Time and cost factors have led to a shift of this objective to the Bureau's 1960 Census Evaluation program, which will provide such a reconciliation on a national basis. Information about several housing characteristics and about family income were left off the 1959 schedule because these data will be obtainable from the 1960 Census. Consequently, tabulations for the 1959 program will be essentially limited to those provided in Volume I of the 1956 NHI plus some tabulations for recent movers that appeared in Volume III.

The sample for the Components of Change program was expanded in 1959 to include eight additional standard metropolitan areas--all those of 1,000,000 or more inhabitants in 1950. Budgetary limitations, however, require use of a smaller sample for the newly added standard metropolitan areas. The method of making the 1950-1959 comparison will be essentially the same as that used for the 1956 National Housing Inventory. Enumerators will compare the status of dwelling units in their segments with the information reported on the 1950 Census schedules.

When all phases of the 1959 Components of Change program are complete, components of change statistics in terms of dwelling units will be provided for the 9-2/3 years since the 1950 Census, permitting a virtually complete explanation of all housing changes which have occurred during the intercensal period; estimates will be available to link these dwelling unit data with the housing unit results obtained from the 1960 Census of Housing; and housing and household characteristics from the 1960 Census of Housing will provide the information not available from the Components of Change program. This will make available to analysts the most complete set of housing statistics ever provided--in this Nation or any other.

FOOTNOTES

[1] The favorable terms of the 1954 Housing Act probably contributed materially to the high level of residential construction through this recession.

[2] Major exceptions to this point were some local and regional planning agencies, and local housing authorities which were compelled to undertake serious study of available statistics to justify requests for public housing subsidies. The recent success of the Bureau's contract block statistics program indicates that this pressure still is strong. Some 250 places below the 50,000 population level are under contract for 1960 housing statistics by blocks. The willingness of local governments to expend funds for these statistics, plus the ability to read, understand and carry out the Bureau's meticulous directions for preparation of adequate block maps is not only a sign of growth in statistical sophistication in the Nation's communities, but an indication of the extent to which small communities now are participating in federally subsidized urban renewal programs.

[3] U. S. Bureau of the Census, Intercensal Housing Surveys, (1957).

[4] Collection of accurate data on housing starts was hampered by what we now see were inadequate funds.

[5] Grebler, Blank and Winnick, Capital Formation in Residential Real Estate, Princeton University Press, Princeton, N. J., 1956, p. 373. The authors quote a "Reconciliation of the Net Change in the Nonfarm Housing Inventory, 1940-1950, (reported by Census) and New Construction as reported by the Bureau of Labor Statistics" labeled as a "preliminary report of an interdepartmental committee of federal agencies (to be published)," which indicated that BLS underestimated new construction by 6 percent for this period. The "Reconciliation" quoted was prepared by the Bureau of Labor Statistics members of the interdepartmental committee, but it neither was accepted as official by the committee nor was its publication authorized.

[6] Glenn H. Beyer, Housing: A Factual Analysis, Macmillan, New York, 1958, p. 281.

[7] Ibid., p. 285.

[8] William C. Wheaton, "American Housing Needs, 1956-1970," The Housing Yearbook, 1954, Washington National Housing Conference, p. 11.

[9] Rapkin, Winnick, and Blank, Housing Market Analysis, U. S. Housing and Home Finance Agency, 1953, p. 60.

[10] FOSDIC is a device, developed by Bureau of Standards' engineers, designed to eliminate punch cards and card-to-tape operations by a process of transcribing marks from 16 or 35 millimeter microfilm to magnetic tape. This requires a specially designed enumeration form although it is marked by ordinary pencil. The document is microfilmed and then fed through the FOSDIC device. The magnetic tape output of FOSDIC then is ready as input to UNIVAC.

[11] The difficulty eventually was traced to inexact printing of the schedules where accuracy is measured up to the thousanth of an inch.

[12] The problems in obtaining NHI new construction figures turned out to be the only major processing difficulty the Housing Staff experienced. In retrospect, it can fairly be said that the processing and tabulation job accomplished by FOSDIC and UNIVAC was an amazing accomplishment. In addition, the Bureau's engineering staff gained invaluable experience in preparing the new FOSDIC's for processing the coming 1960 Census.

[13] The memory of this experience has carried over into preparations for processing the 1960 Census. Both operations and subject matter staffs have combined to incorporate into the final program sufficient checks to assist in the detection of processing flaws.

[14] Wall Street Journal, November 2, 1959, p. 1, col. 1.

[15] Trailers comprised about 150,000 units "moved to site" and nearly 100,000 units "moved from site."

[16] U. S. Senate Subcommittee on Housing, Study of Mortgage Credit, Dec. 1958, Sherman J. Maisel, "Importance of Net Replacements in Housebuilding Demand," pp. 32-42. The analysis in the Maisel paper on this point is open to question. Table 16, the core of his argument, is a combination of estimates and extrapolations that can be accepted or rejected depending upon the assumptions one is willing to make. The statement (p. 37) "Table 16 shows a Census estimate that net losses...averaged 210,000 annually from 1950 through 1956" is not accurate--this is Maisel's estimate not a Census estimate. On another point, I would reverse his projection of 40,000 units gained from conversions and mergers in 1961-70 to an annual net loss of nearly the same magnitude. Finally, seasonal units should not be included in table 16 as an offset to losses from the inventory. These in effect are second homes; or they are held as income properties for seasonal use of persons who normally have their own homes elsewhere. An increase in the number of units of this type does not offset losses in the inventory.

[17] It is hoped that work of the Bureau of the Census' new Construction Office will, in the near future, fill the gap in data on expenditures on residential repairs and rehabilitation.

[18] U. S. Senate Subcommittee on Housing, Study of Mortgage Credit, Dec. 1958, Reinhold P. Wolff, "Substandard Units and Their Replacement, 1961-70," pp. 43-58. Wolff overstates the NHI 1950-56 figure for "substandard units rehabilitated and now standard" by 1.1 million units in his table 20. It is not clear how he arrived at this result. This error may have contributed to his projection of 4.6 million units to be shifted from substandard to standard during the 1960's which is, I judge, too optimistic. If this gross figure reaches 3 million the Nation would be fortunate.

[19] U. S. Bureau of the Census, Current Population Reports, P-20, No. 76, table 1.

[20] In an effort to improve the quality of enumeration, Bureau personnel accompanied each new enumerator in the nine SMA's through his first few interviews. NHI was the first major Bureau survey in which a systematic attempt was made to induct new enumerators in the field under supervision.

[21] Value information for recent owner to owner movers probably is the most reliable of the NHI value data collected, since it represents, for the most part, actual prices received and paid for homes in the two years, 1955 and 1956.

[22] The median value for all owner-occupied nonfarm 1-unit structures was $11,400.

[23] Nevertheless, the generalization may be made that the higher previous rent, the higher is the

value of the home subsequently purchased.

[24] Louis Winnick, Rental Housing: Opportunities for Private Investment, McGraw-Hill Book Co., Inc., New York, 1958.

[25] Analysis of the relationships shown in table 5b was suggested by Miss Beulah Washabaugh, Chief of the Housing Division's Occupancy and Utilization Branch.

[26] Although the dichotomy is not at all clear cut, housing inside standard metropolitan areas is essentially urban while virtually all rural (as well as much urban) housing is located outside standard metropolitan areas.

[27] The construction of 233 new units for each 100 households added outside SMA's has virtually no analytical significance because of the diverse composition of this part of the Nation. About half of the housing outside SMA's consists of urban places of 2,500 to 49,999; the other half is classified as rural and includes farm and other isolated housing, small settlements below 2,500 population, and the suburban housing outside of places of 2,500 to 49,999.

[28] Conversions, mergers, demolitions, other losses, additions other than new construction, and changes in vacancies.

[29] Indirect evidence on this point exists from data about this relationship for two NHI cities. In Chicago where the percent increase in number of households from 1950 to 1956 was 4.2 percent, 183 new construction units were provided for each 100 households added. In Philadelphia, with a household increase of 2.7 percent, the ratio was 281 new construction units provided for each 100 households added.

[30] The essence of the filtration process is in the production of a surplus of new housing--a new construction-net new household formation ratio substantially greater than one--permitting poorer quality housing to be released to successively lower levels of demand until the effect reaches the bottom of the market, where the poorest housing will remain vacant or be removed.

[31] This is true because migration into or out of standard metropolitan areas, which causes wide variations in household growth rates of individual SMA's, does not affect such estimates for the Nation.

[32] Column 5 of table 5b gives a relative of 132 for net additions to the housing inventory and column 2 a figure of 155 as the relative for new construction. Thus:

3,000,000 added households X 132 = 4,000,000 net addition of dwelling units

3,000,000 added households X 155 = 4,650,000 newly constructed units

Total inventory = 55,300,000 + 4,000,000=59,300,000

[33] If any modification is to be looked for in the U. S. relationships in table 5b, one might expect a decline in units added from the existing inventory and an increase in all losses, which would tend to drop the figure for net additions. I would estimate that the relative of 132 for net additions thus might decline to about 124, which would yield a total inventory figure of about 59,000,000 dwelling units as of December 1959.

VI

DEMOGRAPHIC FACTS AND INTERNATIONAL POLITICAL REALITIES

Chairman, Philip M. Hauser, University of Chicago

Western Expansion and Ethnic Convergence in the Population of the United States—Irene B. Taeuber, Office of Population Research, Princeton University

Problems of Questions Such as Religion and Origin in the Canadian Census—J. T. Marshall, Dominion Bureau of Statistics

Population and Welfare in South Asia—Nathan Keyfitz, University of Toronto

WESTERN EXPANSION AND ETHNIC CONVERGENCE IN THE POPULATION OF THE UNITED STATES

By: Irene B. Taeuber, Office of Population Research, Princeton University

The census volumes provide materials for decennial pictures of the westward expansion of the population of the United States. The problem is the focus of analysis. Usually it has been from within the United States. Some have seen the trek of settlers across an almost empty continent, with the influence of the frontier as the unifying theme. Others have thought within the framework of the manifest destiny of a dynamic and crusading people. Here we look eastward across the Pacific and ask the questions that Asians might ask about the growth of this westward moving population. The first question concerns the role of colonialism in population growth, for we were once colonial subjects of European powers, and we absorbed many areas that we labelled territories. The second and related question is whether the ethnic groups in the expanding area of the United States maintained the demographic behavior of their parent populations or became assimilated to the larger culture and displayed its population characteristics and trends.

The statistical record is long and relatively abundant, and the laboratory of American history is complex. If the focus is Asian, the area of major interest extends in an east-west direction from the Mexican border to the Philippines; in the other direction, it extends from Alaska in the north to the Micronesian islands in the south. This analysis is limited to the comparative numbers and the convergent characteristics of the populations of the former Spanish areas from Mexico to California, together with Alaska and Hawaii. Five groups of people and their descendants are distinguished: The pre-Columbian inhabitants; immigrants from Asia; immigrants from Africa; those whose associations were with Spain and Mexico; and other European immigrants. These groups are designated in terms of their origins as indigenous, Asian, African, Spanish surname, and other European.

A Century of Growth

In 1840, some 17 million people were distributed sparsely over the 1.8 million square miles of the United States. Mexico held 800 thousand square miles to the southwest, but she lacked either the manpower for peasant settlement or the resources for industrial development. The Russian - American Company was based in Alaska, but people and supplies had to come from St. Petersberg. In China and Japan, large populations lived inadequately on limited land. Technologies were backward, capital sparse, and political organization antiquated. In this Pacific confrontation of empty areas on the one side, massive populations on the other, it was social and economic factors rather than demographic ones that were determinative. The lands were added to the United States, and their settlement came through western expansion from the Atlantic rather than eastern expansion across the Pacific.

Barriers of distance, topography, rainfall, and soil retarded settlement in the southwestern regions, Alaska, and Hawaii. There were some migrations for the gold of California and Alaska, the seals of the Pribloff Islands, and the whales of the South Seas, but great movements and rapid developments were products of the science and technology of the last century. These were frontiers of the industrial economy and the city rather than agriculture and the trade center.

The population growth of the mainland areas from Texas to California is a familiar story (Table 1). Total numbers were 378 thousand in 1850, 21 million in 1950. The population is characteristically American. The predominant numbers are European in origin, with increasing proportions of the native born and decreasing proportions of the foreign born (Table 2). The migrations of Africans follow the regularities of distance and origin that characterize their movements elsewhere.

Despite their further distances and their differences in climate and resources, the population histories of Hawaii and Alaska are variants of those of the southwestern mainland areas. The Asian labor imported into the Hawaiian Islands was proportionately larger than that brought to California, but recent migrants are predominantly European in origin. Alaska's growth on a continuing basis required the technologies of the mid-twentieth century. In recent decades, its major population increases have come from the older states to the south, and they have been European in origin. Here, also, the pattern of migration and labor force utilization for Africans is an extension of the national one.

The populations of the areas that became parts of the United States in the last century are American in ethnic composition and in growth. There are differences in the balances of the components, but these differences are related to economic and demographic developments in the nation as a whole over the last century. They are not inherent characteristics of the early histories or the contained economies of the respective areas.

Thus our first question is answered in the negative. The furthest westward migrations and the growth of the most westward populations were regional aspects of growth and redistribution in an industrializing nation. Concepts of areas as territories or of the population growth as colonial are not relevant.

Origins and Characteristics

Our second question on colonial demography within the United States concerned stability and change among peoples of diverse origins, ethnic affiliations, and cultures. Given residence in

Table 1. —Population of the former Mexican areas, Alaska, and Hawaii, 1850-1950.

(Numbers in '000)

State	1850	1860	1890	1920	1950
Former Mexican areas:					
Texas	213	604	2,236	4,663	7,711
Utah	11	40	211	449	689
New Mexico	62	94	160	360	681
Arizona	-	-	88	334	750
Nevada	-	7	47	77	160
California	93	380	1,218	3,427	10,586
Alaska	-	-	32	55	129
Hawaii	84	70	90	256	500

Source: U. S. Bureau of the Census. U. S. Census of population: 1950. Vol. 1. Number of inhabitants. Tables 6 and 7.

Table 2. — Percentage composition of the populations of the former Spanish area, by origin, 1880 to 1950.

Group	1880	1890	1900	1910	1920	1930 a/	1940	1950
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
European	81.6	83.5	85.3	87.9	89.4	90.5	90.9	91.1
Native	66.9	69.9	73.9	75.4	76.3	78.9	82.9	84.5
Foreign	14.7	13.6	11.4	12.5	13.1	11.7	8.0	6.6
African b/	14.2	12.7	12.3	9.9	8.5	7.3	7.1	7.2
Indian c/	1.2	1.8	1.2	1.0	0.9	0.8	0.8	0.7
Chinese	2.9	2.0	1.0	0.5	0.3	0.3	0.3	0.3
Japanese	0.0	0.0	0.2	0.6	0.8	0.8	0.7	0.4
Other	0.0	0.0	0.0	0.0	0.1	0.3	0.2	0.2

a/ Mexicans re-classified as Europeans. b/ For California, the percentages of Africans in the successive enumerations from 1880 to 1950 were 0.7, 0.9, 0.7, 0.9, 1.1, 1.4, 1.8, and 4.4. c/ American Indians.

Source of data: U.S. Bureau of the Census. U.S. Census of population: 1950. Vol. II. Characteristics of the population. Table 14 in state reports: 3. Arizona. 5. California. 28. Nevada. 31. New Mexico. 43. Texas. 44. Utah. 51. Alaska. 52. Hawaii.

areas of rapid social and economic development, were there distinctive processes of demographic adjustment that were related to origin, race, or culture? The approach to an answer involves analysis of the status of the ethnic and cultural minorities in 1950, utilizing primarily educational levels, the extent and the adequacy of the economic activity, the proportions of youth, and the ratios of children to women. Given simple measures of status by age, changes over time may be inferred. Given measures for comparable age groups, the changes associated with urbanization and the impact of regional cultural and economic factors may be inferred. Nativity may be introduced as a further variable. Comparable classifications of the European population of the United States and the western region provide standards for comparisons of the levels and the differences in characteristics by age, urban or rural residence, and nativity.

Europeans

The population of European origin is usually designated as white, occasionally as Caucasian. The overall characteristics and their changes over time need not be summarized here. It is a diverse population, with intricate patterns of differences that reflect relative participation in the developing society and economy. The proportion completing the fourth grade or less in school may be taken as an index for retardation, the proportion completing the fourth year of high school or more as an index of adjustment. The former is inversely, the latter positively correlated with the proportion of the labor force in professional, technical, and related occupations, and with income.

In the total population of European origin, or in any major segment of it, the reverse movement from older to younger age cohorts is a movement to increasing educational achievement; to higher proportions of the labor force in managerial, technical, and professional employment; and to higher income levels or income potentials. Each oncoming age groups has had greater opportunities than the preceding group, and has made greater economic progress as measured by employment, occupation, and income. Demographic characteristics are related directly to social and economic characteristics, with high proportions of youth and high child-woman ratios persisting among the groups that are educationally and economically retarded.

These differences by age exist among the native born of native parentage, the native born of foreign or mixed parentage, and the foreign born. In each instance, the accommodation to the advanced society is greater among the native than among the foreign born. If the variables are examined for nativity groups in urban, rural nonfarm, and rural farm areas, the age and nativity patterns persist, with urban populations at the highest levels, rural farm populations at the lowest levels. If the relationships are examined for the regions of the country, the patterns remain, but there are differences in levels. In general, whatever the age, nativity, or residence category, social and economic levels and demographic structure are most traditional in the South, most altered in the West.

These are the relations that existed in 1950. Analysis of the age differences in 1950 and the relationships in earlier decades indicate rapid reductions in the differences among the groups, whether classified by nativity, residence, or region. The forward movements among the younger people in once retarded groups may be regarded as convergence, or the existing differences may be regarded as time lags. In either view, there are substantial but declining differences in accommodations to opportunities and responsibilities within the modal European population.

Persons with Spanish surnames

The largest recognized minority in the southwestern states consists of Mexican immigrants and the descendants of such immigrants, along with the descendants of the early Spanish settlers. In 1950, the percentage of the population with Spanish surnames was 6 in California, 11 in Texas, 14 in Arizona and 35 in New Mexico. Numerically, there were 758 thousand in California and 1.0 million in Texas.

The differences by age within nativity, residence, and regional groups and the differences within comparable age groups by nativity, residence, and region were similar to those among all people of European origin. The proportion with minimum education advanced with age. At a specific age, it was lowest among the native born of native parentage, highest among the foreign born, intermediate among the native born of foreign or mixed parentage. This relationship held in urban, rural nonfarm, and rural farm areas in all states. Labor force participation, occupational structure, and income level showed the same regularities. So also did age structures and ratios of children to women.

Persons of Spanish surname were retarded in comparison with the general European population in the nation or in the western region, but there were major differences within the group itself. These differences presumably reflected opportunities and responsibilities in the areas of residence and selectivities among the migrants from Mexico and to the cities. In Texas, almost one-half of the men aged 25 to 44 who were native born of native parentage had 4 years or less of school. The comparable proportion in California was one in eight. Among the foreign-born men aged 25 to 44, almost two-thirds of those in California had not completed the fourth grade.

The retardation of the people with Spanish surnames was major, but so also was the upward mobility. If we take the native-born men of native parentage who were aged 45 and over and living in the rural-farm areas of New Mexico in 1950 as the closest approximation to the original culture and economy, we find a median education of 3.8 years. The median education of the foreign-born men in this same age group in New Mexico was 2.0 years. For the native born of native parentage in urban California, the median education of the men aged 25 to 44 was 10.4 years.

For the persons with Spanish surnames, advance was associated with younger age, native birth, and migration from the rural areas of Texas, Arizona, or New Mexico to urban areas in the same states or in California. Retardation was associated with recent migration from Mexico or with continued residence in the rural areas of states that were traditionally Spanish-American centers.

Africans

The response of Africans to the opportunities of the larger economy proceeded in major part through movement to cities within or outside the South. The areas from the Mexican border to the Pacific include the Texas Africans, who are southern in economic characteristics and demographic structures, and the California Africans, who are even more highly selected in migration and more advanced in characteristics than the metropolitan Africans in the industrial region from the Great Lakes to New York.

The patterned differences of social and economic characteristics by age within residential and regional groups are found among the Africans, as among the Europeans. For any given group in any specific setting, however, the attainment of the Africans is less than that of the Europeans.

Retardation in the area of origin and upward mobility among out-migrants and their descendants are characteristic of Africans, as of Spanish Americans. Africans in the rural South are somewhat more educated than persons of Spanish surname in rural New Mexico; Africans in the urban West are somewhat more educated than the native-born Spanish Americans in the same area. In occupation and income, though, the barriers to African advance seem to be substantially greater than those for Europeans of Spanish surname.

Asians

The Africans came from diverse cultures in Africa, but the practices of slavery replaced native languages, religions, and behavior with appropriate modifications of the English language, the Christian religion, and the mores of the South. Asians also came from diverse cultures, but they came in successive time periods as contract laborers, and their associations were within their own groups or with members of the dominant European culture.

The Chinese and the Japanese who came to Hawaii or to the West Coast were lowly members of their respective societies. Among those aged 65 or over in 1950, less than one per cent had school attendance beyond the fourth grade level. Among those aged 25 to 44, the median years of school completed were equal to those of the Europeans. This was true for the West Coast and for Hawaii. However, the pursuit of education was most assiduous among the Japanese. In the urban areas of the West, less than two per cent of the men aged 25 to 44 had four years or less of schooling, while more than three-fourths had four years of high school or more. The Chinese had higher proportions with low education, lesser proportions with higher education, but substantially the same median level.

Among Chinese and Japanese on the mainland, eight or nine per cent of the men aged 25 to 44 who were in the labor force reported professional, technical, and related occupations. In Hawaii, the percentage of the male labor force in professional and related occupations was 5.4 for Japanese, 10.7 for Chinese.

Both in California and in Hawaii, the ethnic identities of the Chinese and the Japanese were being preserved, but accommodation had gone so far that differences among Chinese, Japanese, and Europeans were not great.

In Hawaii and in the West, the Filipinos were more recent migrants. In 1950, the proportions who were laborers were high, the proportions who had been born and educated in urban areas low. On the mainland, educational levels, unemployment, occupational structure, and income placed the Filipinos below any other ethnic or linquistic minority. The relative retardation was also severe in Hawaii. Presumably the responsible factors included the lateness of the migrations and the brief periods in the new areas, though the differences among the adjustments of Chinese and Japanese in Hawaii and in California suggest that cultural factors may also be involved.

Indigenous peoples

The indigenous populations are ethnically and culturally diverse, and their contacts with the in-migrants to their areas have differed in type, extent, and duration. Here we shall consider separately the Indians of the mainland area from the Mexican border to California, the native peoples of Alaska, and the Hawaiians.

If the Indians of the United States or the West are considered as groups, the familiar patterns of the upwardly mobile populations appear. The younger are more educated than the older. For any age group, the educational levels and the types of economic activity are more advanced for the urban than for the rural population. Proportions of youth are less in the urban areas, and ratios of children to women are lower. In relative terms, the Indian population is less advanced than the African. In all residence areas of the West, the proportion of men aged 25 to 44 who are poorly educated is higher. The characteristics of men aged 15 to 24 as contrasted with those of men aged 25 to 44 suggest substantial upward mobility among the Africans, relatively little among the Indians. Moreover, in all residential areas the proportions of youth and the ratios of children to women are higher among the Indians than among the Africans.

The reservation Indians show less accommodation to the industrial economy than any other group recognizable in the data for the southwestern area. As of 1950, 52 per cent of the Navajo men aged 25 and over had not completed a single grade of school. Almost three-tenths of the men aged 14 and over reported that they had neither worked nor sought work during the week preceding the census. More than half of those with income reported amounts below $500, while less than two per cent reported incomes of $3000 or more. The median income was $471. Forty-six per cent of all the Navajo were below age 15, and there were 829 children below age 5 for each 1,000 women aged 15 to 44.

The indigenous peoples of Alaska — Aleut, Eskimo, and Indian — have had only late and limited contact with the constructive aspects of modern society and economy. As a group, their retardation is greater than that of the general Indian population of the West, but less than that of the Navajo. Among men aged 25 and over in 1950, one-fourth had not completed a single grade of school and one-half had not completed the fourth grade. There was mobility here, though, for the percentage of men with less than a grade of school was 67 per cent at ages 75 and over but only 14 per cent at ages 25 to 29. Four-fifths of the boys and girls aged 7 to 15 were enrolled in school.

The residences and the economic activities of the indigenous groups in Alaska were distinct from those of the Europeans. Three-fifths of the employed civilian labor force were in hunting and trapping, fishing, and manufacturing, mainly the canning and preserving of fish. However, 29 per cent of the men aged 14 and above reported that they were neither working nor seeking work. Sixteen per cent of those who reported themselves as in the civilian labor force were unemployed. Some 13 per cent of the men reported themselves as without income, while the median income of those with income was less than one-third that of the Europeans. High proportions of youth and high ratios of children to women further attested the material retardation of the Aleuts, the Eskimos, and the Indians.

The isolation of the indigenous population is greatest in Alaska, the assimilation most advanced in Hawaii. In 1849, 98 per cent of the people in the islands were native born of Hawaiian parentage.[1] In 1950, two per cent were pure Hawaiian, 15 per cent part Hawaiian. In numerical terms unmixed Hawaiians declined from 79 thousand in 1840 to 12 thousand in 1950, while part Hawaiians increased from less than 500 in 1849 to 74 thousand in 1950. By the latter year, 86 per cent of the Hawaiians were classified as mixed.

In 1950, the Hawaiians were an intermediate population, below the Europeans, the Chinese, and the Japanese on the one hand, above the Filipinos on the other. This relationship existed in educational level, labor force participation, occupational structure, and income. Within the Hawaiian group itself, the usual patterns of differences prevailed. Advance in any characteristic was greater among the younger than the older people, greater among the mixed than the pure ethnic stock, and greater in metropolitan Oahu than in the other islands.

Interpretation of the significance of formal tabulations of data are difficult in Hawaii because of the intricacy of the relations among the groups that compose the population. Straightforward demographic analysis is seldom possible. Persons of European origin lose to all other groups, for in mixtures of white and nonwhite, the classification of the children is that of the nonwhite parent. Persons of Asian and African origin gain from the Europeans but lose to the Hawaiians, for any person with either parent Hawaiian or part-Hawaiian is Hawaiian by definition. Vital statistics compound the difficulties, for the data on births are tabulated by the ethnic classification of the infants. Biological accretions to the Hawaiian group give it a high component of youth and a high ratio of children to women. In 1950, more than 46 per cent were below age 15, while there were 860 children age 5 for each 1,000 women aged 15 to 44.

The difficulties in classification and interpretation are products of the distinctive characteristics of the Hawaiians, their cultural assimilation and their biological intermixture. Indexes of dissimilarity for the residential distribution in the census tracts of metropolitan Hawaii in 1950 show maximum segregation among the Europeans, minimum segregation among Hawaiians (Table 3). Among the non-European groups, the maximum segregation is that of each group with the Filipinos, while the minimum segregation is that among Hawaiians, Chinese, and Japanese.

Convergence and Persistence

In the area from the Mexican border to California, in Alaska, and in Hawaii, the relative status of the population groups reflects the nearness to and the accessibility of the opportunities of the advanced economy. Neither area of origin, native culture, nor race are determinants of absolute or relative status. However, in no group measurable in census tabulations are origin, culture, or race irrelevant to levels of achievement and presumably to status.

The recognizable groups that we have considered are the dominant peoples of European origin, the Africans, the people with Spanish surnames, Chinese, Japanese, Filipinos, Indians, Aleuts, Eskimos, and Hawaiians. In all the major groups there are patterns of differences that persist over time. Advance to the modal characteristics of the Europeans occurs primarily through the adjustments and the achievements of youth. It is thus a continuing process. For any given age groups at any given time, however, advance has been greater for the native than for the foreign born, for the urban than for the rural farm population, for the West than for the South or the total population, for areas of in-migration rather than for areas of original settlement. These relationships exist for educational level; the adequacy of economic activity as measured by labor force participation rates and the unemployment of those in the labor force; the proportion of men in the labor force who are in professional, technical, and related occupations; and the level and structure of the income distribution. Demographic structure is related and has the same intricate patterns of differences. Here the specific variables examined were the proportion of youth in the total population and the ratios of children to women.

Among all groups in all areas there was movement toward the levels of the European population within the same area. Migration and redistribution were placing increasing proportions of all groups in the urban areas of the more favored regions and so producing overall convergence in regional or national figures. However, no group in any area had achieved the full stature of the comparable group of European origin. In general, retardation was greatest where the movement required for identity was most difficult. Backwardness was most persistent among groups who continued to live in their areas of habitual residence where traditional social structures and economic discriminations precluded advance in place and discouraged movement. These included the persons of Spanish surname and the Africans in the rural farm areas of Texas, the reservation Indians of the southwest states, and the indigenous peoples of Alaska. Not even in these groups, however, was there the relative fixity of residence and characteristics that existed among Asian

Table 3 — Indexes of residential segregation, census tracts of Metropolitan Honolulu, 1950.[a]

Group	Hawaiian	European	Chinese	Japanese	Filipino
Hawaiian	-	.462	.361	.314	.415
European	-	-	.527	.492	.596
Chinese	-	-	-	.321	.546
Japanese	-	-	-	-	.392
Filipino	-	-	-	-	-

a. If X_i and Y_i are uncumulated percentage distributions for census tracts, Δ is the sum of the positive differences between the two percentage distributions. For algebraic and geometric derivations, see: Duncan, Otis D. "The measurement of population distribution." Population Studies 11 (1):27 - 45. 1957. See especially pp. 29-32.

Source of data: U.S. Bureau of the Census. U.S. Census of population: 1950. Vol. III. Census tract statistics. Chapter 62. Honolulu, T. H. census tracts. Table 1.

peasants in the century or so prior to the Second World War.

The question of our hypothetical Asian colleague concerning colonial demography among populations within the United States can be answered in the negative. Little in the American experience is directly relevant to the evaluation of the future of Asian populations. The social and economic advance of minority groups here occurred largely in association with movements to areas and regions of greater opportunity. The advance of massive Asian populations requires overall transformation. Problems cannot be solved by quick removal of practically all the peasants to industrial employment in metropolitan areas. Hence transfer of findings on adaptation and transformation within the United States to Asian areas can be made only with extreme caution.

The present characteristics and the future problems of the Navajo, together with those of the Aleuts, the Eskimo, and the Indians of Alaska, may seem Asian in nature and magnitude. If only present demographic structures and vital rates are considered, the similarities are striking. In Alaska, populations living in traditional ways with inadequate and declining resources have the age structures and the high birth rates appropriate to their ways of life. Public health activities have reduced or are reducing death rates to low levels. However, the differences between these American problems and the population problems of Asia are more striking than the similarities. Here, the capital, the scientific knowledge, and the techniques are available for alternative ways of life within the native areas. Migration from these areas is quite feasible, provided there are psychological, educational, and cultural preparations for life in different areas at different occupations.

The major area for research is no longer the convergence of the groups but the persistence of group differences. Here our quantitative data are seriously inadequate. If adjustment is associated with migration, and group identification is basically a social definition, those who are assimilated disappear in some unknown proportion from the parent group. Moreover, there may be major artificialities in conclusions derived from the study of groups recognized as minorities. Further analysis of the formation of a population from peoples of diverse origins requires coordinate analysis for the descendants of the nationality groups of European origin along with the African, Asian, Spanish, and indigenous minorities. For this, ethnic or cultural statistics must be extended to the total population rather than limited to the problem minorities.

1/ Schmitt, Robert C. "A census comparison of Hawaii's citizens." Paradise of the Pacific 65 (6):28-29. June, 1953.

Sources

U.S. Bureau of the Census. Census of population: 1950. Vol. 1. Number of inhabitants. Vol. II. Characteristics of the population. 1. United States Summary. 3. Arizona. 5. California. 28. Nevada. 31. New Mexico. 43. Texas. 44. Utah. 51. Alaska. 52. Hawaii. Vol. IV. Special reports. Part 3, Chapter A. Nativity and parentage. Chapter B. Nonwhite population by race. Chapter C. Persons of Spanish Surname.

PROBLEMS OF QUESTIONS SUCH AS RELIGION AND ORIGIN IN THE CANADIAN CENSUS

By: J. T. Marshall, Assistant Dominion Statistician, Canada

Today we are to speak about problems encountered in asking such questions as religion and origin in the Canadian Decennial Censuses. You will have surmised that our problems do not arise primarily in the areas of tabulating or compiling the data as, aside from the editing involved in preparing origin statistics, nothing could be more straight-forward than the mechanical aspects of putting the answers together in tables. Actually, such problems as we have arise in the field and stem largely from two causes. We must find ways to phrase each question in such a way that the populace will understand what it is that the census-taker is attempting to find out. Secondly, questions such as those on "religion" and "origin" may have an emotional impact on some people and may lead to some distortion in the answers hence affecting the relative accuracy and usefulness of the data. But I am getting ahead of myself and had best drop back to cover some of the background of the Canadian Census and review how some of the questions came to be asked and how they are interpreted.

In 1870 an Act was passed by the Canadian Parliament to provide for "The first census in Canada to be taken in the year 1871 --- to ascertain and show, with the utmost accuracy possible --- all statistical information which can conveniently be obtained and stated in tabular form touching population and the classification thereof, as regards age, sex, social conditions, religion, education, race, occupation, and otherwise ---".1/

Here then, we see that from the very beginning the law of Canada provided for inclusion of questions on "religion" and "race" and, as a result, we have asked these questions continuously since 1871. "Race" in recent years has fallen into discard being replaced by "origin". It is under the authority of the Census Act that such questions came to be asked and have continued from census to census.

Provision being made for collection of information on "race" and "religion", the question arose as to the interpretation of these terms. Dealing with the question of definition first we find that the interpretation of the term "religion" in the Census has been approached rather broadly in that people are asked what religious denomination they "profess" or "prefer" - Roman Catholic, Anglican, Judaism, etc. - being given the opportunity to state "none" or "atheist" and the like, this being their wish.

This approach we believe accounts for the relatively few problems of any kind found in the collection of statistics on religion. Because religion is associated with a denominational "preference" and the classification "none" is admissible, no complications regarding definition are apparent.

In the case of "race" it has always been the intent to determine "ethnic" origin. This concept has not varied throughout the history of the Censuses but, unfortunately, it has not been easy to define and has been difficult to explain concisely and briefly on a questionnaire. It has not been possible to strike upon any single factor that will adequately identify the "origin" of all persons. Changing political boundaries and intermarriage are two of the factors that make it difficult for many to say with assurance that "my origin is Scottish, Syrian, Greek or Polish". At some time in the progress of Canadian Censuses, there took place a switch in the concept of "origin". The initial emphasis on geographic or national origin in the definition of "origin" gave way to an increasing concern about cultural associations.

Consequently, the approach to collecting the information about "origin" has shifted from census to census in an attempt to make the census-taker's requirements clear to the respondent. One approach that has been tried is to state the question in terms of political-geographic background and the second to link it to linguistic affiliations.

Before 1941, the birthplace concept predominated because, aside from selected ethnic groups such as the Jews, Poles and Ukrainians, the country of origin is as good an indicator of most persons' ethnic background as one can readily find. Thus, in the 1931 Census, the statement on the meaning of origin contained: "--- For the remaining elements of the population (whites), those namely which derive originally from Europe --- the question as to origin usually elicits the original place of residence and implied cultural surroundings of the family before transfer to the North American continent. In most cases, therefore, the replies to the census questions indicate the country or section of Europe from which the family originally came ---".2/

Through time, however, the feeling grew that linguistic affiliations were the most important influence in most cases in determining the cultural background of our people rather than the geographic or political affiliations. Therefore by 1941, the relevance of birthplace as the primary determinant of "origin" decreased in importance and in 1951 origin was defined for the purpose of census collection as follows: "You will first try to establish a person's origin by asking the language spoken by the person (if he is an immigrant), or by his paternal ancestor when he first came to this continent".3/ However, the 1951 Census has shown that this linguistic affiliation is not sufficient to clearly distinguish backgrounds in the case of origin. The 1951

Census of Canada recorded, for instance, that there were:

2,569 Jews of British Origin
429 Jews of French Origin
7,600 Jews of Polish Origin
9,118 Jews of Russian Origin

in the Canadian population.

It is apparent that aside from difficulties in defining the concepts in understandable terms one might have difficulty because the answers, if given under emotional stress, could be distorted or untrue. Some questions provoke relatively little emotional feeling on the part of respondents. Thus, for example, the questions directed towards establishing a person's sex has, in general, more chance of eliciting a straight factual answer. However, as one attempts to penetrate more deeply into the characteristics of the population and uncover "facts" about its social, religious and economic circumstances, the questions asked may produce emotional responses to some degree. For example, some persons are sensitive concerning their ages, while some men would rather be known as "sanitary engineers" than as "street cleaners".

In the same way, questions on religion and origin could provoke emotional reactions that might lead to erroneous replies. In the case of religion there is apparently less difficulty; the reason being that the Census does not try to determine the "degree" of a person's religious participation. Actually all that is attempted is to obtain a statement of "profession" or "preference", rather than "affiliation", with the opportunity supplied to give the answer "none" if desired. Enumerators are, in fact, cautioned against "forcing" a person to state a denomination if the person has no religious association. And, although some people may feel obliged to state a connection with some denomination when their association with that religion is remote in the extreme, the concept behind the question is substantially filled because by definition we are looking for a "preference" and are not striving in any degree to determine the extent of religious participation. Thus, census totals for a given denomination exceed those shown on church rolls and this excess is, in itself, of interest to religious authorities.

Questions pertaining to origin may be subject to similar emotional reaction. Many people of some, we think a relatively small number, ethnic and language groups seem to feel it expedient to claim affiliation with another group rather than admitting their true origin. For instance, people of Russian origin may feel that they will be more readily accepted if they claim to be Polish or Ukrainian. During the two World Wars, persons of German extraction often laid claim to Dutch ancestry. Such distortions are not too difficult to detect because the answers received will vary with the political and social climate existing at the times that censuses are carried out.

This being true, it becomes necessary to attempt some appraisal of the extent that such problems distort the accuracy of the statistics. That errors and distortions exist would not be denied. But in spite of some distortion valuable use can be made of the figures.

Some effort was made by Norman Ryder to appraise the accuracy of the "origin" series, the results of which were published in the Canadian Journal of Economics and Political Science.[4] Those interested in this subject would be well paid to turn up his article.

Briefly put, however, Mr. Ryder attempted to establish the quality of origin statistics "viz., the extent to which the aggregates established by answers to this question actually correspond with the language groups".[5] After an examination of data relating to people of European extraction and reporting other than English or French as their mother tongue, Mr. Ryder concluded that the Census statistics, useful figures though they may be, contain errors for certain specific groups.

For instance, he demonstrated that certain origin groups contain a number of persons that judged on the basis of language spoken should belong to other groups. Typical of these are the Netherlands, Russians and Poles. On the other hand, there is evidence that some of those who speak German and Ukrainian find their way into the wrong categories leaving their own origin groups under-enumerated. Some of these errors can be traced to the fact that immigrants speaking a given language will come from an area that has changed hands and now belongs to another country. Thus, language and birthplace do not apparently match up. Such complications are difficult to sort out. In addition, these errors vary between Censuses apparently depending on how the public opinion of the time regards the various groups.

If then, questions like these create problems in enumeration and response that are liable to distort the results, one might ask: why carry them on? In the first place these figures, rough though they are, contribute a good guide to the numbers of people of various ethnic groups and religious denominations settled in Canada. In a world where rumours and ignorance concerning the size and composition of minority groups feed suspicion and often persecution, even rough data bearing some relation to facts, can prove a useful counter.

Aside from this, a surprising number of individuals and groups ask for material of this nature. They seem to feel that although the statistics are rough, they still supply a guide in their work and they would prefer to have them, approximate though they may be, rather than have no data at all. Among the users of origin statistics are found the various national organizations, such as the Canadian Jewish Congress, the Ukrainian Association of Canada, St. Jean Baptiste Society, and the Lebanese and Syrian Society. Business firms and marketing research agencies are also heavy users of this material. As for religious statistics, church authorities use them for a good many purposes and all denominations are insistent upon the question being retained in the Census schedule. This, of course, gives them some idea of how numerous their supporters will be in the various areas of Canada. Demographers can also use statistics on religious denominations to attempt to trace patterns of family size and other characteristics for Canada as a whole and for smaller geographical areas.

Canada, like the U.S.A., is a nation built out of immigrant peoples. Since the early 1600's successive waves of people born in foreign lands have come ashore to settle and develop the country. Figures on birthplace, language spoken, religion, and origin, taken together with the questions on education, occupation, etc. should and could be of great value to anyone engaged in studying the impact that Englishmen, Jews, Germans, Methodists, Roman Catholics and Mormons have had on the building up of a culture in Canada. Those interested in the social and cultural development of our country - in some respects unique with its strange association of two distinct main cultural groups of people - will find material of this nature important.

Government departments dealing with aspects of citizenship and immigration use the figures extensively. They have an important bearing on the study of immigration for they show the extent that the newer peoples are mixing with the basic stock of the country and adapting to Canadian institutions. Certain classes of immigrants adapt readily to the Canadian way of life, intermarrying with native British and French stock, and are easily assimilated. Others are less successful in adapting to the society and institutions of Canada or because of recent arrival are comparatively unassimilated. The statistics from the Census on the relationship between official language, mother tongue and the language indicated by the origin reported can also help to determine the extent to which immigrant groups maintain their individual identities.

This, we think, illustrates that an appreciable demand exists for statistics of this type. Merely because, to date, figures gathered have contained errors seems a poor excuse not to continue the search for more reliable methods of collecting accurate data.

In doing this, some may say that it is not right to "pry" into another's religion or his origin, but we submit that this does not constitute any more "prying" than asking a man his age or occupation. Some may say that these are personal matters and attempting to uncover them will create difficulties. We can only say that we have not found that the people of Canada in general regard the questions either as "impertinent" or "prying".

In the taking of a census some individuals for one reason or another refuse to answer some or all of the questions. We can say, however, that no exceptional trouble has been caused by our questions on religion, language, origin, etc. People generally do not resent them or refuse to answer them, but generally speaking it is lack of knowledge and problems of definition, consistant with concept, that cause the Canadian census-takers the most headaches.

In 1951 it was necessary for the Canadian Census to accept "Canadian" and "American" for those respondents who were borne out of families having several generations of birth on this continent. Particularly is this true in those families in which the admixture prevents them from making a clear-cut selection of their origin. Too wide a use, however, of these two terms would ultimately defeat the purpose of the census questions. This has not happened yet, however, as only one per cent of the total population reported "Canadian", "American" and "unknown" in 1951.

Similar problems are encountered, of course, in other areas of the Census. For instance, questions on education and occupation are hard to define. In the case of education, Canada's principal statistics are compiled in the Education Division of DBS, being supplied from the institutions and the provinces who have complete jurisdiction in the field of education. Although it has been suggested that the Census be used to obtain extensive and precise information on education, the net result of the extensive questions necessary would be to make the Census unwieldy. Consequently, in 1961 our Census will be confined to two questions - "What was the highest grade or year attended" and "In the past school year has the person attended a school or university".

Questions on occupation raise difficulties centering around the concepts of employment. Particularly serious is the problem as to whether or not the enumerator can obtain an accurate listing of a person's occupation if the enumerator talks to someone other than the person concerned. In 1961 we hope to further improve our census data on employment subjects through such means as more intensive publicity, stepping up the cooperation of employers who provide each employee with a more precise definition of his job (a very useful effort in 1951), better wording of the questions and improvements in the Classification of occupation and industry.

Historically speaking, the difficulties that have arisen in securing answers to census questions in Canada, have not been due to resistance or resentment on the part of the public, but rather to a lack of knowledge of, or an understanding of, definitive answers on the part of the respondent. In the future, therefore, we must continue to look for improved methods of securing our data in such a way that the relative usefulness of the statistics will not be impaired. In order to do this we must be very clear concerning what we seek. If we determine exactly what we want, then by thought and trial we can develop the particular combination of questions that will give us the answers we need, adding to the accuracy and the usefulness of the Census of Canada.

In conclusion, it is apparent why my US colleagues requested this paper and that it stems out of the problems that they encountered particularly on the inclusion of the question of "religion" in 1960. We have been told by an emminent authority that one reason such questions are practical in our Census may be sought in some fundamental characteristics of our constitution which one is not competent to discuss in this paper. Census-wise, however, as has been pointed out by Dr. Omer Lemieux on many occasions, and as was implied in the first part of this paper, "we have asked these questions continuously since 1871 and Canadians, generally speaking, around Census time go all out to provide the best information in their possession."

Nevertheless, we must admit that were we trying to ask such questions for the first time for inclusion in 1961 one could almost bet that we should face greater problems because of public reaction. "Familiarity" has made the Canadian Census respondent more willing to supply the answers to census questions. His main problem is in trying to sort out just where he belongs in the census classifications and hence in the census tables. "Familiarity" in this case has not bred contempt because first and foremost the Canadian wants a good census.

The writer is indebted to Messrs. O. A. Lemieux, A. H. LeNeveu, D. L. Ralston and A. C. Mikel of the Dominion Bureau of Statistics for their interest in the preparation of this paper.

Bibliography

1/ 1941 Census, Administrative Report of the Dominion Statistician.
2/ 1931 Census, 1, 45.
3/ 1951 Census, Enumeration Manual, 44.
4/ Canadian Journal of Economics and Political Science, Vol. XXI, November, 1955.
5/ Ibid, Page 469.

POPULATION AND WELFARE IN SOUTH ASIA

By: Nathan Keyfitz

University of Toronto

The preceding papers of this session are concerned with population increase among peoples within range of North America, for whose problem a solution is to enter the broader economic life of this continent. This paper turns to an area where no such simple solution is possible.

That Asian population is large and growing is well known. It is equally well known that the growth is a handicap to development, since each year the provision of capital to new workers takes up a large portion of such saving as occurs; only saving in excess of this contributes to raising income per head. The image commonly used is that of the Red Queen who had to keep running in order to stay in the same place. These pertinent facts of the overall situation have been documented by Kingsley Davis[1], Coale and Hoover[2], and Thompson[3], whose excellent work requires no sponsorship from me.

This paper explores a further aspect of the problem: that in the framework of the institutions of South Asia the handicap constituted by population growth is much greater than appears in the overall struggle to raise income per head. Instead of the reference to the Red Queen I shall ask you to think of a runner who has to carry a burden weighing five pounds. This is a handicap in the race that he is running, and it may well stand for the effect of population growth when considered in aggregate figures. But when one takes into account the unevenness of local distribution of population, and particularly the institutions of dividing and sharing that operate in South Asia, then one is forced to consider the load of five pounds as tied to the legs of the runner and so constituting a much more severe handicap than its mere weight. For it appears that the effect of rapid growth is to depress constantly the technical level of agriculture; to make impractical the adequate collection of agricultural taxes; to split countries apart politically into regions of high density which are poor and regions of lower density which are better off but disinclined to support their poorer fellow citizens; to make it increasingly difficult for employers to hire and pay industrial labour. In short the peculiar way in which population is growing in South Asia makes for a steady deterioration in the quality of economic effort under a free market system.

First a brief review of the statistics on the present level of population and its rate of increase. There are in South Asia : one country of about 400,000,000 (India: 397,540,000); two countries edging towards 90,000,000 (Indonesia: 86,900,000, and Pakistan: 85,635,000); three countries of about 20,000,000 (Philippines: 24,010,000, Thailand: 21,474,000, and Burma: 20,255,000); four countries around 10,000,000 (VietNam: 12,800,000, Ceylon: 9,388,000, Nepal: 8,910,000, and Malaya: 6,515,000); and several smaller countries. The figures in parentheses are those given by the United Nations for mid-1958, and the total for that year is approximately 685,000,000 for the area which it is convenient to designate as South Asia. The United Nations does not pretend that these figures are exact and complete, but they are good enough for our purpose. During the year and a half that has passed since the date to which they apply the total has undoubtedly passed the 700,000,000 mark. This is slightly more than the reported population of mainland China.

It is not these levels of population which have attracted the attention of students of Asian development so much as the rates of change. Rates are not known precisely for the whole of Asia, but two countries do keep good vital statistics: Ceylon officially shows 36.6 per thousand in 1957 and Malaya 43.2 per thousand, both based on registrations which may be incomplete but which are unlikely to contain any duplication. The Indian censuses of 1941 and 1951 imply rates of 39.9 per thousand. Indonesia officially estimates a round 40 per thousand for the period since 1950, and this figure is as good a guess as any for the whole of South Asia at the present time. These rates have probably not changed greatly during the past century.

On the other hand death rates have changed drastically. During most of the past hundred years they must have averaged some 25 to 30 per thousand; no direct information is at hand, but deaths of this order along with births of 40 per thousand would give rise to the increases of population which actually occurred. Since World War II there has been a sudden drop in deaths. The figures for Ceylon show 9.9 per thousand in 1957, for Malaya 11.0.

The 1957 balance for Ceylon was 334,000 births against 93,000 deaths, a net increase of 241,000 persons, or 2.7 per cent. Malaya's growth reckoned in the same way was 3.2 per cent. It is true that Ceylon and Malaya are the richest countries in Asia and have had the benefit of the most thoroughgoing attacks on malaria and other diseases, but it is safe to assume that what is true of them today will apply to the rest of Asia tomorrow.

It is not difficult to foresee what the population of South Asia will be if these rates become general. Arithmetic shows that money or people compounding at 3 per cent double in about 24 years and at 2.5 per cent in about 29 years. Thus between 1984 and 1989 South Asia's 700,000,000 will have become 1,400,000,000. This is roughly in accord with the projections of the United Nations, which give for the year 2000 some 3,870,000,000 people in all of non-Soviet Asia, including South Asia, China, and Japan.

Interpretation of these figures remains a problem. To show how difficult it is I should like to contrast two theories of growth. The first is due to Malthus, and for our purpose it may be condensed down to the following statement:

Population tends to grow in geometric progression, and hence will increase faster than any growth of the means to feed it, which it is assumed will be arithmetic at most. Malthus used as an example the rate of increase of the American colonies which doubled in 25 years (about as Asia is now doing), but his theory is valid for any fixed rate of geometric increase however slow. The conclusion is that growth of population, if left to itself, is a menace that will sooner or later cancel out the benefits of any striving for improvement in the conditions of human life.

Now let me cite an alternative theory of growth: Birth rates are variable and there is a tendency for them to adapt themselves to death rates. During most of human history deaths were heavy, and the groups that survived were the fertile ones; they used every ideological, material, and institutional incentive to keep birth rates high. When death rates decline there is a tendency for birth rates to follow them downwards after a lag of two or three generations, which is the length of time that it seems to take institutions to adapt to new requirements. The evidence to support this theory is that a constant difference of one tenth of one per cent--say a birth rate of 40 per thousand and a death rate of 39 per thousand--would suffice to account for the present population of the earth if it originated from a single couple less than 25,000 years ago.

These theories are supported by somewhat similar arguments involving the properties of a geometric progression, but nonetheless they are as far apart as can be conceived in their consequences. One is radically pessimistic, the other optimistic. Malthus' pessimism is not simply that things are hard here and now, but that no amount of invention and no possible bounty of nature that may be discovered in the future will provide more than temporary relief. Each invention that produces more food constitutes a niche but all such niches will be promptly filled. Given Malthus' starting point his conclusion is inescapable, but his starting point must be thought of as selected from a whole barrel full of possibilities, of which I have indicated one other, each of which could be suitably dressed in arithmetic to constitute an equally inexorable law. Out of this collection Malthus took the one that would support the institutions of property that seemed necessary for Britain's development. The Essay followed the Wealth of Nations by only 22 years. Malthus feared that population growth would swallow the gains resulting from a market that was being rapidly freed. One sometimes feels in reading his text that it was not the theory of population that gave rise to his pessimism, but his pessimism that gave rise to--or at least selected--the theory.

Malthus is at his best not on the elaboraration of his arithmetic of doom but on the institutions necessary for progress. He describes them in the allegory of Nature's Feast. This feast is rather like an English dinner-party and has nothing in common with Asian hospitality. There are just so many places set, and those who through their own improvidence or that of their parents have no places do not eat. If any of those who do have places were to reach out some food to the hungry ones behind him instead of leaving them to starve, this intended kindness would have a weakening effect on basic institutions. With enough of such kindness society itself would dissolve as the disinherited fought for the table. Malthus wanted to jar people out of that short-sighted kindness which fits well in a static society but is not consistent with rapid economic growth. No more skilful argument has ever been devised to promote a social system that would lead to hard work and responsible behavior.

To see the application of Malthus's thought to South Asia it seems well to take a point from his friend and opponent David Ricardo. In England the owner and cultivator were usually different people; Ricardo described a process by which all cultivators had the same income for the same effort, and good land yielded a rent to the landlord equal to excess of its product over that of marginal land. The cultivator had either to adjust the number of his children to the share of crop which he received, or send them to the city where they pressed for jobs and so reduced labour cost in manufacturing. But if owner and cultivator are the same person, his family will consume whatever it needs and then sell what is left. On this very different arrangement a steadily smaller surplus of food comes off the land as population grows. In Java owner-cultivation is common; in other areas landlords still exist, but the pressure for nationalization is so strong that they are likely to disappear. From many points of view this will be an advantage, for the tenancy system has not in practice given either the landlord or the tenant an incentive to make improvements. No advantage whatever was taken of the possibilities of large-scale tilling. But disappearance of the landlord will entail the disadvantage that food which was sold to pay rent will now remain in the village. One might say that in those cases where the owner and cultivator are the same person Ricardo's law still holds; income received can be thought of as partly rent and partly wages, each following its own laws. But this formal statement is of no help in tracing the motivation of the person, for which we must consider how he treats his total income.

Taxes like rent bring food off the land, and their objective ought to be to induce the peasant to grow something he would not otherwise have grown, rather than to depress him by taking from him something that he has grown for his own use. But what becomes of this principle where population is dense? The peasant has plenty of time to grow crops other than those he needs for his own food, but he has no land. The tax then becomes the removal of part of the food that would be eaten by the cultivator and his family, or which would be exchanged for essential clothing or kerosene. Such a tax seems unfair, a squeezing of the holder rather than an incentive to more work, and in practice taxes have been largely abandoned in South Asia since World War II.

Institutional systems may be distinguished by who gets the food first. In those of post-colonial Asia it is the local population which pre-empts it before the tax collector or the landlord. We shall see also that it gets to the food before the seller of factory-made articles from the city.

The discussion so far has been in terms of aggregates, as though the new population was the same mixture of men, women, and children as the old. In fact at the present moment, just following a sudden drop in the death rate, the major increase in population consists of children who would formerly have died. I have heard Ceylonese statisticians say that though national income shows an annual increase of less than the 2.7 per cent by which population is rising, people feel that they are better off. The added income is not consumed by an extra child or two. This exemplifies the fact that the drop in the death rate will show itself in different forms during the next generation. First it appears as more babies surviving, then in crowding of the schools and the need to organize the building of new ones; its most dramatic consequence will occur when the new people reach working age and begin to demand jobs. The political effect of this is becoming apparent in Ceylon today. The gradual, virtually unseen onset of the consequences of population growth is one aspect of its treacherous character.

If those who reach manhood cannot be given either land or capital, what are they worth to the economy? What is the value of the lumber a man can saw in a day, less the cost of the logs? What is the value of the yarn he can spin, less the cost of the raw cotton? What is the value of the service of transporting goods on his back? I am anxious to obtain figures on these as they appear at prices determined by the competition of electric power. It will be surprising if they are sufficient to buy the rice or wheat that will keep the man working, let alone housed and clad. A human being is an inefficient machine in most kinds of routine physical work.

To move on to another of the ways in which the growth of population hinders a country aiming at development, consider the fact that population is unevenly distributed throughout Asia. Leaving for more leisurely study the historical reasons for differences in density, I shall here raise the question of the consequences within a single country. In Pakistan, East Bengal is densely populated compared with West Pakistan, even taking account of the richer soil of East Bengal. In India, Kerala shows great density and misery, while parts of the Punjab are relatively wealthy. But the most striking instance is Indonesia: Java, with 48,000 square miles, has a population of some 55,000,000, while Sumatra has five times the area and less than one fifth the population. Wages in Sumatra are much higher than in Java and there is some spontaneous movement of people; in addition there has been some assisted movement to government-sponsored settlements. But the best these have attained has been far less than 10 per cent of Java's natural increase of nearly one million people a year. The degree of organization needed to move people in the numbers required to keep Java's population from growing further is not present with all the other problems of an underdeveloped and crowded country.

Under the pressure of this growing population the fragile tissue of nationhood can easily be torn apart. The densely-settled portion of the country converts its land from plantations of sugar cane and other export crops to food and still runs short. The lightly-settled portion--Sumatra and Celebes in the case of Indonesia--maintains a brisk export trade. The patriotism of the exporting areas is strained; they do not see why the foreign exchange that they earn should go to the support of the denser areas which cannot possibly produce anything in return. One Asian official asked me how it was that Canada could tolerate differences in income among its different areas; did this not mean that we lack a sense of nationhood? I replied that some Dominion-Provincial payments are made to redress the balance, but no one considers or desires that these be large enough to prevent migration; areas such as Prince Edward Island have been supplying people to the rest of Canada during the whole of the present century. But for Asians no feasible movement of people could balance off standards of living. Hence they consider that the redistribution of income is up to governments, which ought to be able to accomplish this by some means easier than moving large numbers of people.

Within the crowded areas of the countries of Asia there is an interaction between population and landscape. In Java, for instance. forests of teak and other wood, mostly of commercial value, are maintained by the state partly with the objective of preventing soil erosion. But squatters take up plots in the forest, cut down the trees, and put the land to the plough. A democratic government does not easily expel people who have no other source of livelihood; the consequence is that population pressure increases erosion, and the erosion in turn diminishes the crops that can be obtained from previously settled land, and so increases population pressure.

But far more drastic in its operation than the foregoing is a mechanism which is day by day lowering the technological levels of agriculture in the more crowded parts of Asia, as well as effectively discouraging the entry of people into industry. It appears especially in the irrigated rice fields whose construction so greatly increased the carrying capacity of land in each of the areas of Asia to which it has spread since its introduction. Our model may be described in terms of a typical fixed village occupying a few hundred acres. Once a tank and irrigation channels have been built, the lowest land can be cropped two or three times a year; higher land is capable of producing one rice crop when the terraces in which it is arranged are filled with rain; higher land yet is used for dry gardens which depend entirely on the monsoon and for houses.

The mechanism that is operating in the rice-surplus villages may be expressed in the words of a population-wise Javanese headman. He explained that the people of his village each year produced about half as much rice again as they needed for themselves, and thus one third of their output was sold outside the village. He plainly had in mind a sort of village balance of payments in which the rice eaten by the producers was considered to disappear along with that traded to others for services within the village, and attention focussed on the net surplus actually moved outside the village. It was sold for money in outside markets, and the money was used to buy roofing tiles, bicycles, clothing, baskets, and other goods; the headman was sophisticated enough to see that money was only a means of exchange, and essentially the rice was being traded for these things, so that there would be more or less of these outside goods entering the village according as there was more or less rice surplus.

He also had some historic sense. He acknowledged that the population of the village had been increasing in recent years, that there was noticeably less rice available for export than there had been when he was a young man, and that in the years ahead there was likely to be less yet. He was farsighted enough to be worried about this, not on the grounds that any of his villagers would be hungry but rather because they would have fewer of the goods that make life pleasant after one has eaten. The consequences for people in the city that produced the inedible goods traded for rice would be more serious--they would see their markets declining, and they would have correspondingly less to eat.

One need not look very far into the future to see population climbing up on its food supply until village after village drops out of rice export. The amount of outside goods bought by the villagers diminishes in exact proportion to the decline in rice sent out, and a time will come when there is no rice for export and no village market for city-produced clothing and other goods. The clothing factories have gradually been reducing their output during this period complaining that the fall of purchasing power is throttling them. On the assumptions we have adopted, the growing population need have no effect on agricultural production, but acts exclusively to choke off town industry and to reduce the town population to unemployment and starvation, unless it can sell its goods in foreign markets where there is an agricultural surplus.

The operator of the clothing factory sees his difficulties steadily increasing. His workers insist on spending about 70 per cent of their wages on food; only beyond this fixed amount for themselves and their families will they take their salaries in clothing. In wage disputes the price of food is always a key item. If the clothing workers are to spend 70 per cent of their wages on rice, then 70 per cent of their clothing output must be sold to rice farmers--disregarding profits and sales abroad for foreign food. The result of surplus money in the city (brought about by an inflationary policy by which money is printed to defray government expenses) is that selling effort is directed to the wrong market. Competitive effort to sell in the city market does bring results to individual manufacturers, but collectively the manufacturers can sell clothing only up to the amount of the surplus of rice that is elicited from the countryside. When they aim the kind and design of goods at city buyers aggregate sales are less than if they were using the same effort to find rural customers.

Thus the only consequence of selling effort directed at city populations is to alter the share of individual manufacturers in a fixed total amount of food wage goods. Unfortunately, if there is a good deal of money in the city and none in the countryside, there is no use exhorting the producer of shirts to seek markets in the countryside--he will reply that he must try to sell where the money is to be found.

All this is to demonstrate how much harder the rapid growth of population makes development. It seems to lead to a shared misery on a scale which has never been seen before, which cumulatively prevents the application of practical methods for its correction. On the other hand, once a degree of the education and discipline that come with development has appeared there seems to be no difficulty in handling the population problem.

Japan presents an example of how easy it is to control population when one has at one's disposal the means of communication and control that have been built up in the course of industrialization--after one has given up aspirations to world conquest. Japan's birth rate fell from about 33 per thousand of population in 1946 to about 17 per thousand 12 years later. Japan's methods included large-scale abortion, but it is expected that this will gradually give place to more widely acceptable methods. Even with the drastic fall in births the population is still increasing by 900,000 persons each year, but this will be further reduced if it seems necessary. The main point is that she has educated her population to understand an incentive system and has then applied it sharply. The incentive system is that wages are low enough and life sufficiently hard that people find it desirable not to have more family than they can afford. In other words, the national shortage of food is transmitted through an incentive system to the individual. The opposite condition applies in South Asia where large areas within nations are faced with starvation, but food subsidies prevent the urban dweller from being warned of this; and peasants on rich lands consume the surplus that was once removed by the landlord, the tax-collector, and the attraction of city-produced goods.

[1]Kingsley Davis, The Population of India and Pakistan (Princeton N.J.: Princeton University Press, 1951).

[2]Ansley J. Coale and Edgar M. Hoover, Population Growth and Economic Development in Low-Income Countries (Princeton N.J.: Princeton University Press, 1958).

[3]Warren S. Thompson, Population and Progress in the Far East (Chicago: University of Chicago Press, 1959).

VII

EDUCATIONAL RESEARCH THROUGH APPLIED STATISTICAL METHODS

Chairman, Herbert S. Conrad, U. S. Office of Education

The Identification, Development, and Utilization of Human Talents—John T. Dailey, University of Pittsburgh

An Illustrative Projection of Debt and Debt Service for Public School Construction in 1970—Louis H. Conger, Jr., U. S. Office of Education

An Experimental Study of Several Factors Affecting Rate and Nature of Response to a Mail Survey of Beginning Teachers—Ward S. Mason, and Robert J. Dressel, U. S. Office of Education and Robert K. Bain, Purdue University

THE IDENTIFICATION, DEVELOPMENT, AND UTILIZATION OF HUMAN TALENTS

By: John T. Dailey, University of Pittsburgh

For at least the past 25 years, there has been serious discussion of the need for a program of research which would, on a nationwide basis, yield factual information about the nature and distribution of human talents and how best to assist individuals to identify, develop, and use them. There has been an urgent need for a comprehensive survey including aptitude, interest, motivational, and background factors, to be accompanied by intermediate and long-range follow-ups of the individuals studied in the survey. Such a survey, Project Talent, is now underway.

Through the cooperative assistance of the U. S. Office of Education, the National Institute of Mental Health, the National Science Foundation, and the Office of Naval Research, such a research study was planned, and the testing was carried out during the school year 1959-1960.

It was planned that the two-day battery of tests be representative of the best current types of aptitude and achievement tests that have been proven to be useful and important in selection, measurement, and guidance of young people at the secondary school level. There were also questionnaires to obtain important background information regarding the previous experiences of the students: their neighborhood, family, community, interests, activities, etc. There were also measures of their motivations, levels of vocational and economic aspirations, and personal plans. The tests were designed to measure the functions covered by most of the standardized tests now widely used with high school students.

The tests that have been used in the project were prepared especially for this research following detailed specifications developed by the staff and advisory panel following a year's intensive study. No test was an exact parallel of any existing commercial test. The tests will be withheld from subsequent publication and will be reserved for research use only.

In the planning phase of Project Talent, one of the first tasks in developing the test battery was to decide on a number of types of tests to be developed. The Staff were assisted actively in this by the Test Panel and the Executive Committee. In September 1958, following the Test Panel meeting, a number of test types were selected which were believed to cover most of the important variance in aptitude and achievement tests in current use and which were suitable for administration by teachers in public high schools and other secondary schools under the logistic limitations of the Project. In the early spring of 1959, detailed rationales and specifications were prepared by Staff for each of these types of tests. These specifications gave a detailed description for the type of test to be developed in the battery, giving its purpose, what it was supposed to measure, the number of items with their difficulty levels, and the like. After these had been reviewed and approved and following the Test Panel meeting in the spring of 1959, the Staff prepared two forms of each type of test and they were administered in May 1959 to approximately 6000 students in eleven schools. The purpose of this administration was to obtain item analysis data and test inter-relationships data to guide the staff in devising the final form of the Project Talent battery.

One of the most important considerations in the design of the Talent battery was conserving testing time to the utmost. Although ten hours of testing time may seem to be a great deal, when it is necessary to include in this time as many different measures as were strongly recommended by the Panels and the Committee, there is not a minute to spare. Also, for any given test to be included, it must meet the very strong competition of all the other tests very strongly recommended by various individuals within the Panels and Committees. Virtually all the test types recommended also had been widely used and had proven themselves by demonstrating value for use by counselors and research personnel. Accordingly, it was necessary that each test demonstrate empirically that it had important variance to contribute over and above the variance covered by all the other tests in the battery. In all, fifty-four measures were scored and included in the analysis.

It proved not to be possible to accomplish all of the scoring of such a large battery during the limited time available with any known existing equipment. Also, any known existing equipment would have been extremely expensive, since it was necessary to score each test separately for two separately-timed halves and for total score. It was also necessary to score rights and wrongs separately and combine them into formula scores. Further, some of the tests were on more than one answer sheet, and some of them consisted of items scattered throughout a 350-item booklet. The scoring for each individual was equivalent to 750 rights scorings. To score the test conventionally would have meant obtaining these 750 scores and then combining them into the relevant scores, total scores, etc.

The scoring and item analysis were accomplished simultaneously on an IBM 704 computer. A special program for this purpose was developed. This scoring and item analysis program scores all the tests in the battery, computes the half-scores separately plus the total score, is flexible to score by formula (rights minus wrongs, rights minus one-fourth wrongs, etc.), and it computes a point biserial correlation coefficient between each item choice and the test score it appears in. Later, it prints out these results in roster form for sending back to the participating schools. Existing programs were used to compute comparable half reliabilities, inter-correlations, the multiple correlations between each test versus all the other tests in the battery, and the partial correlations between each pair of tests in the battery with all other tests held constant. The regression program also computes the inverse of the matrix and the regression weights. This program can also use any one variable as the dependent variable and call in each of the other tests successively, one at a time, in the order of its partial correlation with the dependent variable, and compute the multiple correlation and raw score regression weight for each successive combination of tests.

The final step in the analysis of the pre-test data was to compute the uniqueness coefficient which consists of the reliability coefficient minus the square of the multiple correlation coefficient for the test versus all the other tests in the battery. This uniqueness coefficient was very valuable in helping to evaluate the tests in the experimental battery. Most of the tests turned out to have some useful unique variance. This is not surprising since each test tried out was believed by those sponsoring it to have an important unique contribution. After all of the statistical data had been completed and interpreted, several tests were eliminated from the battery.

It has been found that, once the items of a test are on tape, it is about as cheap to score and analyze simultaneously as it is to analyze alone after the tests have been scored in a separate operation. It is believed that this has profound implications for the scoring and analysis of tests. Efforts are under way to follow through and develop a combination scoring and analysis routine, known as Test Assay, to capitalize on these features. Test Assay may be used as a description of the total process of determining the communality inter-relationships and also the uniqueness of any specific test included in a given battery. This, in a sense, assays a test both in terms of its overlap with the other tests tried out simultaneously with it, and in terms of its unique potential contribution. Of course, the uniqueness of a test is specific to the particular combination of tests it is tried out in. This makes it important to use a standard, known, highly comprehensive and non-redundant battery for this purpose. It is believed that the Talent battery should be very useful for this purpose. Fortunately, it is easy and economical to iterate this process by cycling through it several times and changing the combination of tests as indicated.

It is also important that Test Assay be carried out on at least moderately comparable samples since it is well known that reliabilities and correlations vary a great deal from one kind of population to another. However, one of the most interesting aspects of the uniqueness coefficient is that is appears to be highly stable from sample to sample as compared with the stability of reliabilities and correlation coefficients. Apparently, if because of restricted range or some other factor, the reliabilities tend to be lower on one sample than another, the multiple correlations will be correspondingly lower and thus the uniqueness coefficient will tend to be stable.

The key to the feasibility and economy of this approach is the adequacy of the sensing mechanism for putting the answer sheets on tape. In the pre-test tryout, the answer sheets were run through a type of document reader that automatically sensed and punched cards at a price only slightly cheaper than hand-punching. The Measurement Research Center at the University of Iowa now has a document reader punch which will punch one card per answer sheet insertion at a speed of about 6000 per hour. It will get about 120 items on each card by means of punching two items per column. It uses a special answer sheet similar to the regular type multiple choice answer sheet. It is highly flexible and alpha-numerical in character. If desired, test items can be printed directly on the answer sheet. It can also be used as a separate answer sheet.

The next step in the development of the art in this area will be the development of a document reader which can put an entire answer sheet of 500 multiple choice items directly on magnetic tape at a speed of 6000 per hour or greater, and without going through the intermediary of a punched card. This would have the important advantage of eliminating the necessity of dealing with large numbers of cards and having to sort together several cards for each subject. It is reported that developments of this kind are under way by several groups.

AN ILLUSTRATIVE PROJECTION OF DEBT AND DEBT SERVICE FOR PUBLIC SCHOOL CONSTRUCTION IN 1970

By: Louis H. Conger, Jr.
Educational Statistics Branch, U. S. Office of Education

The viewpoints and projections presented in this paper are my own and are not to be taken as official announcements of the U. S. Office of Education.

Considering the problems of getting a current figure on debt service for public school purposes, it requires a certain measure of temerity to undertake to peer into the future. Nevertheless, I am encouraged to do so for several reasons. In the first place, the statistic is an important one, providing an index to the actual revenue that must be raised to finance capital outlay programs. Secondly, even when painting with crude, broad strokes, a picture emerges because of the large scale of the changes involved. And finally, a share of future debt service is already settled and independent of future developments, since the obligation has already been incurred.

The term "illustrative projection" is highly appropriate here, and may be taken to mean that the assumptions and methods used, while appearing reasonable to me for the purposes at hand, are not the only possibilities for the future. A basic principle of the projections is that present practices continue, particularly present methods of financing capital outlay. There are many alternative possibilities here. The future costs are developed in terms of constant dollars. This is the usual procedure in economic projections, and does not constitute an assumption that price levels will not actually change.

We begin by postulating a need for 610,000 classrooms to be built in the 10 year period 1959-60 through 1968-69 in order to accommodate increased enrollment, to allow for abandonment and replacement of facilities becoming unsatisfactory, and to eliminate the shortage of classrooms that now exists. It should be clearly understood that this figure represents the needs and is not a projection of the expected volume of construction.

These classrooms are judged to entail a capital outlay for construction, site, equipment, and associated facilities of $40,000 apiece, or a total of $24.4 billion (in constant dollars) in the 10 year period.

We now come to the conversion of this capital outlay expenditure into debt and debt service. The first question is the amount of capital outlay to be financed by the issuance of long-term debt. The percent of capital outlay financed by long-term debt appears to have been 85% in the first 7 years of the 1950's, but there are indications that a lower percentage, say 70%, may be more typical today. In order to assess the effect of these two possibilities, they are both projected, under the labels of Model A and Model B respectively.

In converting capital outlay into debt service it is assumed that the bonds will be issued equally in each of the 10 years. An interest rate of 4% is adopted, and a maturity of 25 years. Both of these figures correspond to current conditions for newly issued public school bonds in the third quarter of 1959.

By application of these rates, the debt service arising from construction in the period 1959-60 through 1968-69 is projected to 1969-70 (columns 5 and 6 of table below).

There is also debt service in 1969-70 on debt incurred before 1959-60. Since the debt is already in existence, the projection is less open to alternatives than in the case of debt arising from future activities.

There was $12,488 million long-term debt for public schools outstanding at the end of fiscal year 1957; the maturation of this debt during the time period projected is based on redemption schedules known for local government debt as a whole.

Debt assumed in 1957-58 and in 1958-59 is taken from bond sales for which both the amount and the interest rates are known. A maturity schedule of 1/25 th of the total each year following the year of issue is adopted, corresponding to current conditions as to maturity.

The debt and debt service in 1969-70 derived from debt assumed before 1959-60 is shown in column 4 of the table below.

The components of debt are brought together in columns 2 and 3 to show the total projected debt service in 1969-70; similar figures for 1959-60 are given for comparison.

It is noteworthy that, under the assumptions adopted, debt service doubles from 1960 to 1970 in the case of model A. It is even more striking that the actual amount of debt service in 1970, $2.4 billion, happens to be equal to the average capital outlay over the 10 year period projected. This is not a temporary extreme, since the large volume of debt in recent years ensures that reductions in debt service after 1970 will come slowly even if no new debt is issued.

Under the B model, with long-term debt providing for 70% of the capital outlay expenditures in the decade ahead, the 1970 debt service costs are ameliorated to an increase of 65% over the decade. Of course, this amelioration has been purchased with double infusions of current funds into construction costs, as compared to model A (30% from current funds in model B; 15% from current funds in model A). If these cash costs were added to the debt service figures, model A and B would be brought closer together in 1970.

We are now in a better position to assess the fact that the projected need of 610,000 classrooms averages out to 61,000 a year, which is somewhat less than the number actually built in recent years. Although physical volume averaged over the decade ahead may not need to exceed the 1959-60 level, the actual cost of maintaining this level is shown to be rising rapidly.

In closing, several qualifications need to be specified as to the limitations of the present projections, relative to the problem of providing the needed classrooms. A National projection takes no account of the distribution of the needed construction into the States and localities where it is needed, but the attainment of this outcome requires the solution of substantial problems of distribution not brought out in the National model. A similar matter is the distribution of the construction in time. The existence of a mass of unsatisfactory facilities at the present time argues for an increased pace in order to clean up the backlog before a whole 10 years has gone by. In the long run, the debt service paid in the aggregate is not greatly affected by the timing of the construction, but there are practical problems in mustering sufficient resources to accomplish the necessary stepping up of construction in the next few years. It should be noted that total construction costs are not given here; only the long-term debt component.

Estimated future status of existing long-term debt for public school purposes, and long-term debt for public school classrooms needed to be built in the 10 years 1959-60 through 1968-69 (all State and local government): 48 States and D. C.

(Millions of actual dollars for debt incurred prior to 1959-60; millions of 1959 dollars for debt incurred 1959-60 through 1968-69)

Item	Total long-term debt		Long-term debt incurred prior to 1959-60	Long-term debt incurred 1959-60 through 1968-69 for public schools needed	
	MODEL A	MODEL B		MODEL A	MODEL B
(1)	(2)	(3)	(4)	(5)	(6)
DEBT OUTSTANDING, END OF 1958-59	$15,385	$15,385	$15,385	0	0
DEBT SERVICE, 1959-60:					
Interest	471	471	471	0	0
Redemption	824	824	824	0	0
Total debt service ...	1,295	1,295	1,295	0	0
DEBT OUTSTANDING, END OF 1968-69	24,963	21,962	7,956	17,007	14,006
DEBT SERVICE, 1969-70:					
Interest	925	805	245	680	560
Redemption	1,480	1,333	650	830	683
Total debt service ...	2,405	2,138	895	1,510	1,243

NOTE: Capital outlay cost paid from sources other than long-term debt is not included above.

Model A assumes that long-term debt is issued for 85% of the capital outlay for public schools in the period 1959-60 through 1968-69.

Model B assumes a rate of 70%.

AN EXPERIMENTAL STUDY OF SEVERAL FACTORS AFFECTING RATE AND NATURE OF RESPONSE TO A MAIL SURVEY OF BEGINNING TEACHERS

By: Ward S. Mason, and Robert J. Dressel, U. S. Office of Education
and Robert K. Bain, Purdue University

It has long been known that many features of questionnaire design and mailing procedure can influence the response rate in mail surveys. Length, color of paper, format, and type of stamp are a few of those which have been studied. 1/ Some of these effects seem to hold generally for different populations, while others seem to be specific, or else evidence is lacking on the generality of the effect. While such matters as format and color of paper may seem trivial at first, anything which affects rate of response is an important factor. Probably the major problem of mail surveys is that of obtaining a high and representative response. 2/ Yet mail surveys have many advantages, and if this problem can be overcome there is much to recommend them. 3/

Because rate of response is related to the validity of survey results, it is important that data regarding mechanical features of questionnaires which affect response rate continue to be accumulated. It is with this in mind that the present paper was written.

In the spring of 1957 the U. S. Office of Education was planning a national survey of beginning public school classroom teachers. 4/ A sample was developed in which each beginning teacher in the country had a ten percent chance of being included. A pretest of the form was to be conducted. Given the importance of obtaining a very high rate of response, there were two aspects of questionnaire design which it was desired to test in terms of their effect on response rate: length, and mode of addressing.

Length. There is already considerable evidence that length of questionnaire may have an effect on response rate, although some findings suggest that interest and other factors may be more important. 5/ Of course, if one takes extreme examples, it is highly probable that a relationship will hold between length and response rate; it is hard to imagine the conditions under which one could obtain the same rate of response in a given population with a one-page questionnaire and with a fifty-page questionnaire. But the practical question facing the survey director is not "Is there a statistically significant relationship between the abstractions of questionnaire length and response rate?" but rather "For this population and a survey covering this topic, will there be a 'practical' difference in response rate between a questionnaire of length A and one of length B?" "Practical" is a word of varied meanings, but it relates here to such topics as the precision desired in the results and the relative costs of processing questionnaires of different length. In the present instance, the question was whether there would be an appreciable difference in response rate for a form of six pages and a form of eight pages in a survey of beginning public school teachers. The six-page form had 62 items of information, the eight-page form, 92 items.

Mode of Addressing. In the conduct of mail surveys there are two standards which are seemingly contradictory. On the one hand, other things being equal, it is usually thought desirable to have the respondents participate on an anonymous basis. A number of studies have shown that different responses are sometimes obtained under anonymous and non-anonymous conditions. 6/ In investigating some topics, such as sexual behavior, this problem may be deemed to be of sufficient importance to warrant taking elaborate precautions to conserve anonymity while providing a means of matching data. 7/

The other consideration is that it is usually desirable to be able to identify respondents, either for purposes of data-matching, for follow-up studies, or for being able to mail follow-up requests in order to increase the response.

Several methods have been devised of "eating your cake and having it, too." 8/ Such methods should be carefully considered if it is felt that the anonymity factor will really make a difference in the research in question.

There is one method which is intermediate between complete anonymity and having the respondents' name on the questionnaire: that is to put a code number on the questionnaire and have an office file in which the code numbers are matched to names and addresses. Such a procedure has the appearance of anonymity, although little thought is required for the respondent to realize that such a file stands behind the code number appearing on the form, particularly after he has received a follow-up request. Still, it may be reassuring to some people to know that their name does not appear on the form itself, where anyone who works on processing the data may make the connection between respondent and reply. The chief difficulty with this method is in controlling mailings. Extreme care must be exercised to see that forms with code numbers are matched correctly with envelopes having names and addresses. A mistake can be disastrous, and the matching process is an extremely inefficient operation which becomes quite impractical when the sample N is large. Since the N for the survey of beginning teachers was to be over 10,000, this factor was a serious matter. If the teacher's name and address appear on the form, then the mailing operation is facilitated because the form can be placed in a window envelope by machine methods in what is a fast and efficient operation. At the same time sample control is facilitated because respondents can be readily identified.

It was therefore determined that one of the factors to be tested in the pretest was the mode of addressing the form. Half the pretest sample members were to receive forms on which were found the teacher's name and school address plus a series of code numbers, while the other half was to receive a form which was identified only by a set of code numbers. (The case of complete anonymity was not tested.) For forms with code numbers only, the respondent's name and address appeared on the cover letter (which was separate from the form). Both forms contained assurances that the forms would be seen only by a few survey staff members and that only summary results would be published. In addition, those having their name on the form were told that their name and address appeared on the form only in order to facilitate mailing.

In order to test both factors--length and mode of addressing--four forms were designed as follows:

Form A: long form-name and address on form
Form B: long form-code number on form
Form C: short form-name and address on form
Form D: short form-code number on form

A judgment sample was developed for the pretest by asking a series of school districts to submit the names and school addresses of their beginning teachers. The sample of 741 teachers obtained in this way was drawn from various sections of the country and from school districts of various sizes, though it was weighted rather heavily in favor of large districts. The four forms were assigned systematically, sending form A to the first name on the list, form B to the second, etc. Because of the time pressures under which the study was conducted, it was necessary to make the initial mailing in three installments, as responses were received from school districts. Nonrespondents in the first group of 278 teachers received two follow-up requests; the second group of 51 received two follow-ups, and the final group of 411 received only one follow-up.

Before presenting the results it would be well to describe one further feature of the sample. We had requested the names and addresses of beginning teachers from the school districts, but we could not be sure that the names submitted did not include some nonbeginning teachers. Consequently, it was necessary to include on the teacher questionnaire several screening questions to determine whether the teacher belonged in the population being sampled. Those whose answers indicated that they did not belong were asked to return the questionnaire without completing the other questions. The forms received are therefore of two types: those for members of the sample and those for nonmembers of the sample. The task of the nonmember was obviously much simpler in that he had to fill out only three items. On the other hand, the nonmember might be less inclined to return the questionnaire, thinking that a reply was not needed. It is therefore important to look at the results for the two groups of teachers separately.

Rate of Response. The results of the experiment are shown in table 1.

Table 1--Mailout and receipt of four forms of a questionnaire for beginning teachers

Mailout and receipt	All forms	Form A long addressed	Form B long coded	Form C short addressed	Form D short coded
Total mailed	741	186	185	185	185
Received	613	153	150	154	156
Members of sample ...	534	129	131	141	133
Nonmembers of sample.	79	24	19	13	23
Not received	128	33	35	31	29

The null hypothesis that there were no differences in response to the four forms was tested with the use of chi square, comparing total received with the not-received group. No significant differences in response were found either among the four forms, between the two long forms and the two short forms, or between the two addressed forms and the two coded forms. Furthermore, even if such differences had been found to be statistically significant through the use of a much larger sample, they are so small as to be of little practical significance. In percentage terms, the difference between the two long forms and the two short forms is only 2 percent, and there is no difference between the two addressed forms and the two with code numbers.

Because of the fairly large number of teachers who disqualified themselves from membership in the population, a further check is necessary to be sure that this result holds also for the subgroup composed of those who do belong in the population being studied. Of course, we can distinguish between sample members and nonmembers only among those who responded, and consequently cannot do a test comparing respondents and nonrespondents among sample members only. However, because of the way in which the forms were assigned in the sample, it is possible to assume that nonmembers were randomly distributed by type of form. Therefore a chi square was computed for which the actual frequencies were the number of sample members who responded to each form, and the expected frequencies were derived from the total mailout figures. Again, the results were not statistically significant, either for all four forms or when combining them by length or type of address. However, similar calculations for nonmembers of the sample yielded a chi square significant at the .05 level. Thus, it appears that in this particular case there may be a relationship between form design and response among teachers for whom the form is not intended, even though there is no evidence of such a relationship among those who belong in the population being studied.

In view of the fact that no statistically significant differences were found for respondent sample members, the decision was made in the final study to use a form comparable to form B, i.e., eight-page form with the questions from form B as revised in the light of the pretest results, with the name and address of the beginning teacher appearing on the form. This form was mailed to 10,012 beginning teachers throughout the country on March 28, 1957. Questionnaires were received over a three-month period until the end of June, and some teachers received as many as five follow-up requests. Replies were received from 89 percent of the teachers on the mailing list, a highly satisfactory result which minimized the problem of nonresponse bias and one which justified the use of the form chosen on the basis of the pretest experiment.

Comparability of Response. So far, we have been concerned only with response rates: what proportion of sample members replied among those receiving each form. Another question is whether the form design is related to the characteristics or replies of the respondents. Such differences could arise in either of two ways: either the subsample of individuals who reply to one form is different, at least in part, from the subsample which would reply to another; or the same individuals reply to each form, but their replies are affected by the form design. Of course, in a given instance, both factors could be operating simultaneously.

In order to explore this question, respondents to the four forms were compared on a number of characteristics. These included size of school, age of respondent, salary, education, education of father, career plans, life goals, feeling of adequacy in the human relations aspects of the job, and five different satisfactions items. None of the differences by either length or mode of addressing proved to be statistically significant except one: on the average, those returning short forms had a gross annual salary $143 higher than those returning long forms. Such a difference would be expected on the basis of chance alone fewer than two times in a hundred. However, when mode of addressing is controlled, the finding is replicated only for coded forms; the difference in salary found for those returning long and short forms is not statistically significant for those having forms containing their name and address.

Although this result creates a suspicion that there may be a difference in the mean salaries of these two groups, such a result would have to be duplicated in an independent study before it could be entertained seriously. For among all the tests of significance computed in connection with this paper (with the exception of the relation of form design to response for those who were not members of the sample), it was the only result which achieved statistical significance, and we would expect some tests to be significant by chance alone.

Summary and Conclusions. In a pretest of a study of beginning public school teachers an experiment was conducted to determine the effect of two factors, length of form and mode of addressing, on the rate of response and type of response. It was concluded that for this population an eight-page questionnaire with name and address of the respondent appearing on the form would produce a response not significantly different in rate or type from that of a six-page form identified by only a code number, or any other combination of these factors. There was some evidence that when forms were inadvertently sent to individuals who did not belong in the population being studied, rate of response is related in slight degree to these factors of form design.

The findings of this experiment were quite favorable relative to the survey of beginning teachers. On the one hand, the ability to place the respondent's name and address directly on the questionnaire without influencing the results greatly facilitated the mailing and processing of questionnaires. On the other hand, the use of an eight-page rather than a six-page form meant that it was possible to obtain approximately 33 percent more information. More information was important not only for its own sake but because it allows a much more thorough analysis of the interrelationships among variables.

It is worthwhile to raise the question of whether these findings can be generalized to other studies involving mail questionnaires, for it may be that they are limited by some of the features of this particular survey. For example, the beginning teachers studied had a high educational attainment and were relatively homogeneous in this respect. Although it is not known what image the U. S. Office of Education has among teachers, it may be rather safely assumed that, on the whole, sponsorship by this agency exerted a favorable influence. The study was conducted in the interest of the teaching profession, and thus in the self-interest of the respondents. Questionnaires were returned in franked envelopes addressed to the Commissioner of Education. All of these factors undoubtedly contributed to the relatively high response rate of 83 percent, 9/ and a high response rate tends to minimize the possibility of bias in rate of response, although not necessarily in type of response. A final factor, the influence of which is problematic, is the restriction of the study to a population of young adults. If any of these conditions had been different, it is, of course, possible that the results would have turned out otherwise. It is hoped that other similar studies may be undertaken with different values for each of these variables as conditions, which will enable us to accumulate evidence on the generalizability of these findings.

1/ Mildred Parten, Surveys, Polls, and Samples: Practical Procedures, Harper & Brothers, New York, 1950, p. 383 ff.

2/ Morris H. Hansen and William N. Hurwitz, "The Problem of Nonresponse in Sample Surveys," Journal of the American Statistical Association, Vol. XLI (1946), p. 517-529.

3/ Lawrence E. Benson, "Mail Surveys Can be Valuable," Public Opinion Quarterly, Vol. 10, p. 234-241.

4/ This study was made as part of the regular research program of the U. S. Office of Education. Acknowledgement is made to Dr. Herbert S. Conrad, former Director of the Education Statistics Branch, for review of the manuscript and administrative aid.

5/ Parten, op. cit., p. 385-386.

6/ However, it is not self-evident that data obtained under the anonymous condition are more valid. Cf. Jack Elinson and Valerie T. Haines, "Role of Anonymity in Attitude Surveys," American Psychologist, 5 (July, 1950), p. 315; W. C. Olsen, "The Waiver of Signature in Personal Reports," Journal of Applied Psychology, 20, (1936), p. 442-450.

7/ Erik Manniche and Donald P. Hayes, "Respondent Anonymity and Data-Matching," Public Opinion Quarterly,21 (Fall, 1957), p. 384-388.

8/ Ibid.; Don Cahalan, "Effectiveness of a Mail Questionnaire Technique in the Army," Public Opinion Quarterly, 15 (Fall, 1951), p.575-578; Kenneth Bradt, "The Usefulness of a Post Card Technique in a Mail Questionnaire Survey," Public Opinion Quarterly, 19 (Summer, 1955), p. 218-222.

9/ The response rate in the final survey was 89 percent.

VIII

SOME METHODOLOGICAL PROBLEMS IN SURVEYS

Chairman, Frederick F. Stephan, Princeton University

The Use of Rotating Samples in the Census Bureau's Monthly Surveys—Ralph S. Woodruff, Bureau of the Census

Some Sampling Techniques for Continuing Survey Operations—Leslie Kish and Irene Hess, Survey Research Center, University of Michigan

THE USE OF ROTATING SAMPLES IN THE CENSUS BUREAU'S MONTHLY SURVEYS

By: Ralph S. Woodruff, Bureau of the Census

Rotating panels are used on several of the monthly surveys of the Bureau of the Census. Examples are the Current Population Survey, the Monthly Retail Trade Survey, the Monthly Accounts Receivable Survey, and the Monthly Wholesale Survey. Rotation is used in monthly (and other repetitive) surveys because of one or more of the following advantages:

1. Rotation spreads the burden of reporting among more respondents.

2. Rotation permits the use of data from past samples to improve the current monthly estimate. This is done by means of the composite estimation procedure.

3. Rotation may afford an unbiased solution to the problem of large observations which occur in the sample.

The advantages of rotation may be so great, that I believe that the possibility of rotation should be considered for every monthly survey. This is especially true if the data being surveyed are of such a nature that they are expected to have a high month-to-month correlation.

The first advantage, that of spreading the burden of reporting on a sample survey among more respondents, may be very important from the standpoint of maintaining the rate of response. However, this advantage will not be discussed further in this paper. The nature of gains from the composite estimation and large observation procedures will be discussed, however, and their effect on variances of estimates obtained from the survey will be approximated. Examples of the use of these principles will be given in terms of the Monthly Retail Trade Report since this is the survey with which I am most familiar and since this is probably the survey where the greatest gains from rotation have been realized. The principles illustrated in these examples, however, can be applied to any repetitive survey concerned with any subject matter.

I shall describe briefly the sample for the Monthly Retail Trade Survey to serve as a background for the illustrations. This sample provides information on retail sales for individual kinds of business and all kinds of business combined. The sample can be divided into two main categories--the list sample and the area sample. The list sample consists of multiunit organizations and individual establishments which have been identified from previous Censuses and which are large enough to justify their inclusion in a nonrotating sample to be surveyed each month. We shall not further concern ourselves with this portion of the sample since it does not involve rotation. All remaining establishments are represented by the area sample. This area sample consists of a 2-stage sample, the first stage of which is the selection of 230 primary sampling units (counties or groups of counties) from 230 strata which account for the entire United States. In effect, each of these 230 primary sampling units is completely subdivided into area sample segments with definable boundaries and containing on the average about four retail stores each but varying considerably around this average. A sample of these segments equivalent to an over-all rate of 6 percent was drawn. This sample was divided into 12 equal panels each representing a $\frac{1}{2}$ percent sample of all the retail stores in the United States. Each of the 12 panels is assigned to a particular month and this panel is enumerated for that month each year. Two months of data (current and previous) are obtained from each respondent during each enumeration which is made by personal visit of the enumerator each year. This is the rotating sample which is used for illustration of principles in the remainder of the paper.

A. Use of the Composite Estimation Procedure With Rotating Panels

We shall first discuss the use of the composite estimation procedure with rotating panels. I should like to go back to the fundamental principle behind this use of rotating panels. This principle is that in order to develop the most efficient estimate possible, a search for correlated data should always be made. In order to be useful, these correlated data must be either universe data or based on a sample different from that used for the estimate. Then a means of linking these correlated data to the desired estimate must be found. This linking is usually done through a sample survey where data on both the estimate and the correlated item are obtained for an identical sample. There are many ways of using correlated data which may already be available or developing such data when they are not available.

The use of rotating sample is a direct application of this principle. If we consider the estimates which can be made from a rotating sample for the month of June, we can of course obtain the simple estimate for the month of June from the June panel. However, we can also obtain an estimate for the month of June from the May panel by applying the ratio of the June-May results from the June panel to the estimate for May obtained from the May panel. Progressively less reliable estimates for the month of June can be produced from the April, March, etc., panels by using products of the month-to-month ratios which can be developed from the sample. Now, instead of a single estimate for the month of June, we have a number of estimates for June at practically no additional cost and by proper weighting of these estimates can produce a much more reliable composite estimate than the single simple unbiased estimate.

At this point, it should be noted that there are two different systems of rotation which can be used for developing the data necessary for producing these estimates. The system used in the retail trade report is to completely rotate the sample from one month to the next and to obtain from the entire panel two months of data. Another system often used, for example in the Current Population Survey, is to obtain only one month's data at each enumeration and to rotate only part of the panel, retaining part of the panel to provide information from identicals.

The rotation scheme used in the retail trade survey is more efficient in terms of variance per report because an entirely new panel is available in successive months and because the entire panel is used for the identical links. Therefore, it is to be preferred over the alternative method if it is possible to obtain two months of data (current and previous) with usable accuracy in a single enumeration. This is often not the case. Even in the case of retail sales, which are largely a matter of record, we have had some difficulty with the "previous" sales being regularly reported below the level of the "current" sales. Investigation showed that this was due to a general tendency on the part of enumerators to ignore stores in business the "previous" but not the "current" month. When they were urged to take special care in accounting for such establishments, the differences between the current and previous reports dropped to a much lower level. In many surveys based on on-the-spot observation or upon memory, it may be impractical to attempt to obtain data for two periods in one enumeration. In this case, the less efficient form of rotation must be used. While this will result in different optimum constants and different percentages of gain over the nonrotating system than those presented for the retail system of rotation, the principle is the same and the gains will be well worthwhile if the month-to-month correlations are high.

In the Retail Trade Survey, we use a composite estimate (equation 1) which uses information available from all panels through the ith panel to make an estimate for the ith month. This estimate was used as the only estimate in the Monthly Retail Survey through 1959 and since then as a preliminary estimate. It has the following form:

$$(1)\quad X_i''' = (1-W)X_i' + W\left(\frac{X_i'}{X_{i-1}''}\right)\left(X_{i-1}'''\right)$$

In the above equation:

X_i''' = the composite estimate for the current (ith) month (note that the composite estimate for the previous month, X_{i-1}''' is used in the estimate for the current month).

X_i' = the simple unbiased estimate for the current month.

X_{i-1}'' = the simple unbiased estimate for the previous (i-1) month.

Note that X_i' and X_{i-1}'' are from the same panel and that $\frac{X_i'}{X_{i-1}''}$ is therefore the month-to-month ratio for an identical panel.

W = a constant having a value less than one (.8 is the value of W now used in the Monthly Retail Trade Survey).

If this composite estimate is used for an indefinite number of months, it can be reduced to a series of the following form:

$$(2)\quad X_i''' = (1-W)X_i' + W(1-W)R_i'X_{i-1}' + W^2(1-W)R_i'R_{i-1}'X_{i-2}' + \ldots W^m(1-W)R'R_{i-1}' \ldots\ldots R_{i-m+1}'X_{i-m}'$$

In the above equation:

R_i' = the ratio of the current to the previous estimate for the ith panel $\left(\frac{X_i'}{X_{i-1}''}\right)$

This form clearly shows that the estimate uses all estimates available from past panels. A weight of $W^n(1-W)$ is placed on the estimate derived from the panel n months prior to the month being estimated. It can be shown using the type of reasoning used by Patterson in his comprehensive article on this subject (2) that the weights used on each term of the series yields optimum results provided the following conditions are met:

1. The relvariances are equal for all months.
2. The month-to-month correlations are equal for all months.
3. There are no correlations between the results from different panels.
4. The constant W is chosen in optimum fashion (equation 5).

These conditions differ somewhat from those used by Patterson because a different plan of rotation and a different form of estimate is used. It is believed that the above conditions are approximately met in the Retail Trade Survey.

It has been stated that the preliminary composite estimates shown in equations (1) and (2) make approximate optimum use of all data available from all panels through the ith panel. However, data become available from the i plus one panel which can be used to improve the estimate for the ith month. This of course requires a revision of the preliminary estimate but we have recently decided that this revision is worthwhile because it results in some striking variance gains particularly in the ratio between the estimates for the two most recent months. The form of this final estimate (issued one month after the publication of the preliminary estimate) is:

$$(3)\quad X_i'''' = K\left(X_i'''\right) + (1-K)X_i''$$

In the above equation:

X_i'''' = final composite estimate.

K = a constant having a value less than one. (.83 is the value of K proposed to be used in the Monthly Retail Trade Survey.)

Note that the new piece of information available from the i plus one panel namely the simple estimate for the previous month (X''_i) is averaged with the preliminary composite estimate (X'''_i) with weights (1-K) and K.

Two remarks should be made about these composite estimates before we proceed to determining the optimum constants and examining the gains from the estimate. The first is that the estimates are of the ratio form. Regression or difference estimates could have been used instead. The regression estimate would result in a gain in reliability if regression coefficients were properly computed. However, computation of the regression coefficients involve considerable labor and where correlations are very high and relvariances are roughly the same from month-to-month (as in the case of retail trade) regression and ratio estimates yield very similar results. Actually, although the estimates used are ratio estimates at the United States level they are put in a linear form similar to that of the regression or difference estimate at the primary sampling unit level in order to facilitate the computation of variances.

The second remark which should be made is that there is additional information which theoretically could have been used in the composite estimate. In the first place, since the rotation scheme provides identical panels for the same month each year a ratio to the year-ago composite estimate could be used. However, the composite estimate a year earlier is closely correlated with existing terms in the estimate so that it yields very little additional information. The weight indicated for this term and the resulting variance gain appeared to be too small to justify the considerable complications which result from its use.

Another source of information which could be used to improve the estimate is the information available from panels succeeding the month being estimated. We have already indicated that we have decided to revise the estimate on the basis of information available from the panel following the month being estimated. Theoretically all following panels could also be used but this would require successive revisions and it is our opinion that the resulting variance gains do not justify the cost and confusion resulting from this procedure.

If we accept the preliminary and final composite estimates (equations 1 and 3) as the ones we are to use, the next problem is to optimize the constants K and W which determine the weights on the various parts of the composite estimate. This is done by expressing the variance of the composite estimates and then minimizing these variances with respect to the constants. The variance of the preliminary composite estimate (equation 1) may be expressed as:

$$(4) \quad V^2_{x'''} = V^2_{x'} \left\{ \frac{1+W^2-2W\rho}{1-W^2} \right\}$$

In the above formula:

$V^2_{x'''}$ = the relvariance of the preliminary composite estimate.

$V^2_{x'}$ = the relvariance of the simple unbiased estimate from a single panel.

ρ = the month-to-month correlation between the current and previous estimate from the same panel.

If the above variance is minimized with respect to W then:

$$(5) \quad W = \frac{1-\sqrt{1-\rho^2}}{\rho}$$

The variance of the final composite estimate (equation 3) may be expressed as:

$$(6) \quad V^2_{x''''} = K^2 V^2_{x'''} + (1-K)^2 V^2_{x'}$$

In the above formula:

$V^2_{x''''}$ = the relvariance of the final composite estimate.

If the above variance is minimized with respect to K then:

$$(7) \quad K = \frac{V^2_{x'}}{V^2_{x'''} + V^2_{x'}}$$

or if optimum W is used in $X''' = \frac{1-\sqrt{1-\rho^2}}{\rho^2}$

The above relationships are subject to the conditions previously mentioned (i.e., that the relvariances ($V^2_{x'}$) and month-to-month correlations (ρ) be equal for all months and that there be no correlation between the results for different panels). As previously mentioned these conditions are approximated by the Monthly Retail Trade Survey. The relvariances and month-to-month correlations are roughly equal although not exactly so. There are slight correlations among the various panels due to the fact that the sampling was done without replacement. Also there are year-to-year correlations since the same panel is used each year for a given month. These latter correlations are associated with powers of K and W of 12 or greater and should not affect the variance significantly. The fact that the conditions are not exactly met does not bias the results but means that the constants may not be precisely optimum and that variances may be slightly greater than those indicated by theoretical computations based on the assumptions.

The constants actually used in the Monthly Retail Trade Survey are W = .8 and K = .83. The variance results obtained from the use of these constants is compared with the results obtained from optimum constants in table 1. The results are obtained for month-to-month correlation (ρ) = .98, .99 and .95. The correlation of .98 is roughly that obtained for all kinds of business combined and is approximately the median of the correlations for individual kinds of business. The correlations of .99 and .95 are relatively high and low respectively among those obtained for individual kinds of business. The loss over optimum constants is not great even for the high and low correlations. All the relvariances are expressed as multiples of the relvariance of the simple estimate, or in other words the variance which would be obtained from a nonrotating sample. Note that all relvariances whether or not optimum constants are used are a third or less of the variance from a nonrotating sample.

Table 1: OPTIMUM CONSTANTS UNDER HIGH, MEDIUM AND LOW CORRELATION ASSUMPTIONS AND COMPARISON OF VARIANCES OBTAINED USING THESE CONSTANTS WITH VARIANCES OBTAINED WITH SPECIFIED CONSTANTS

Correlation assumption	Optimum constants		Relvariances of composite estimates of level for a single month				Ratio of variance obtained with specified constants to variances obtained with optimum constants	
			With optimum constants		With specified constants[1]			
ρ =	W	K	Prelim-inary	Final	Prelim-inary	Final	Prelim-inary	Final
			(As multiples of $V_{x'}^2$)[2]					
.99 (high)	.868	.876	.141	.124	.156	.136	1.11	1.10
.98 (med.)	.817	.834	.199	.166	.200	.167	1.01	1.01
.95 (low)	.724	.762	.312	.238	.333	.258	1.07	1.08

[1] W = .8, K = .83, (the constants used).

[2] Since $V_{x'}^2$ is equal to the relvariance of the simple estimate obtained from a nonrotating panel the relvariances shown are in the form of ratios of variances of composite estimates to variances of estimates from nonrotating panels of the same size.

Computations are theoretical-based on stated correlations and other stated assumptions.

The optimum values of the constants K and W shown in table 1 are those needed to produce the most efficient estimate of level for a single month. Of course, the statistics may be used in many other relationships, for example, to obtain month-to-month trends, month-to-year ago trends or annual totals to name a few of the most common relationships. These relationships may be more important to the user than the level itself. The optimum constants for any one of these relationships will in general not be the optimum constants for a single month's level. However, we adopted the approximate optimum constants for a single month's level on the philosophy that the estimates can and will be used in a very large number of relationships and the only way to insure that all these relationships will have a reasonably low variance is to produce a good level estimate.

It should be noted at this point that there is one very important exception to the general rule that constants which are optimum for level are not optimum for other relationships. If constants necessary to obtain optimum results for the level of the preliminary composite estimate and the level of the final composite estimate are used, optimum results will also be obtained for the ratio between the two most recent months.

Table 2 shows the variance results obtained from composite estimates for the month-to-month ratio, the month-to-year ago ratio and the annual totals. These are compared with the results obtained from a nonrotating sample. All results in this table are theoretical based on certain correlation assumptions and other assumptions. All computations for composite estimates have been made with the constant W = .8 and constant K = .83 which are the constants used rather than the optimum constants. On the month-to-month and month-to-year ago relationships, two results are shown, one for the ratio between a preliminary and final estimate which would be the ratio available when the month in the numerator is first published. The ratio between two final composite estimates which would not be available until one month later is also shown.

Some generalizations can be made from this table. The variance of the preliminary month-to-month ratio from composite estimates is equal to or slightly smaller than the variance of the month-to-month ratio which can be obtained from a nonrotating (identical) sample of the same size. The variance of the month-to-month ratio between two final composite estimates is substantially higher. However, primary interest is probably centered on this relationship when it first appears. On the month-to-year ago relationship the variance of the final ratio is somewhat smaller than the variance of the preliminary ratio. In this case, however, the variance of both the preliminary and final ratios are smaller than those obtained from an identical nonrotating sample. On annual totals the variances of the sum of 12 final composite estimates is far below the result which would be obtained from a nonrotating sample. For this particular statistic, however, an even lower variance would be obtained from a sum of the 12 simple results from the rotating sample.

Table 2: RELVARIANCES OF MONTH-TO-MONTH RATIOS, MONTH-TO-YEAR-AGO RATIOS AND ANNUAL TOTALS FROM COMPOSITE ESTIMATES COMPARED WITH RELVARIANCES OF ESTIMATES FROM A NONROTATING SAMPLE OF THE SAME SIZE

Correlation assumptions	Month-to-month ratios			Month-to-year-ago ratios			Annual totals		
	Non-rotating sample	Preliminary composite[4]	Final composite[5]	Non-rotating sample	Preliminary composite[4]	Final composite[5]	Nonrotating sample[6]	Sum of final composite estimates	Sum of simple estimates from rotating sample
	(All relvariances expressed as multiples of V_x^2)								
High[1]......	.020	.020	.047	.200	.126	.082	.908 - .991	.093	.083
Medium[2].....	.040	.040	.062	.300	.210	.154	.863 - .982	.105	.083
Low[3].......	.100	.098	.107	.500	.454	.359	.771 - .954	.143	.083

[1] Month-to-month correlation = .99, Year-to-year correlation = .90.
[2] Month-to-month correlation = .98, Year-to-year correlation = .85.
[3] Month-to-month correlation = .95, Year-to-year correlation = .75.
[4] Variance of ratio between preliminary composite estimate for most recent month and final composite estimate for earlier month. This is the ratio available when the current month is first published.
[5] Variance of ratio between two final composite estimates (note that this ratio is not available until one month after the data are first published).
[6] The relvariance of the annual estimate from a nonrotating sample is dependent on the average correlation between estimates one to twelve months apart. We have assumptions for only the two extremes. The computations are made assuming the average correlation is at these two extremes. The actual variances are within the ranges shown.

All computations on a theoretical basis. See text for assumptions made. In composite estimates the constants used were W = .8, K = .83.

B. Use of Rotation to Establish a Panel of Large Observations Which can be Sampled at Heavier Rates

The occurrence of large observations is one of the principal problems in sampling. If the sample is nonrotating one is usually confronted with the unhappy choice of accepting the considerable increase in variance they create or of taking a bias by arbitrarily reducing their weight. In the rotating system these large observations can be placed in a special panel which can be sampled at heavier than normal rates, thus permitting the weights to be reduced without biasing the results.

The principle of this procedure is simple. Identify in n-1 previous panels these large observations and survey them in the current panel. The weight of these observations (including all like them in the current panel) is then divided by n which may drastically reduce their effect on the estimate and the variance of the estimate. While the principle is simple, it is sometimes difficult to put into effect because it is required if the estimate is to remain unbiased, that the definition of "large" that is used be applied equally to all of the n panels. This is difficult because data obtained in each of the rotating panels is usually for different months so that there is no common statistic which is readily available for all the n panels. However, this difficulty can often be overcome as will be illustrated in the Monthly Retail Trade Survey.

While the illustrations which will be given concern large retail establishments which appear in the area sample of the retail survey, the principle can be applied to any survey using rotating panels. For example, a similar feature termed the "rare event universe" has been established in the Current Population Survey for those area sample segments which contain an unusually large number of households.

A number of large establishments appear in the area sample of the Monthly Retail Trade Survey in spite of the existence of the list sample of large establishments and firms taken from the most recent Census. These large establishments in the area sample may be establishments which were born or have become large since the latest Census or they may be establishments too small to put on the certainty list but large enough to cause considerable variance in the area sample. The problem is to create what we call the large area sample panel for such establishments.

In the case of the Retail Trade Survey, we faced the previously mentioned difficulty of having no common statistic available in the 12 panels to use for the definition of "large" since we had obtained only two months of data from each respondent. A criterion suggested by Max Bershad was used in this case. This criterion was that each large area sample panel member equal or exceed a certain sales cutoff for each month of the year. By looking at the particular months we had, we could determine if it was possible for the establishment to meet the criterion. Where the establishment equalled or exceeded the criteria in the months we had, it was placed on a "potential" large area sample list. At the end of 12 months, all members of the "potential" large area sample were surveyed for their sales in each of the 12 months. Those that qualified (about half of the potential list) were placed on the "permanent" large area sample list.

The above procedure requires some time after the end of the criterion period to determine the members of the permanent large area sample panel. For this reason, it is probably practical only in cases where the rotation is periodically repeated (as in the monthly retail trade sample). However, with other systems of rotation other procedures can be used, some completely unbiased and others with biases much smaller than those resulting from an arbitrary reduction in weight.

While the establishment of the permanent large area sample panel made an important reduction in variances, we found that large observations were still appearing in the area sample. These establishments were those which had appeared since the permanent large area sample panel was last brought up to date or which had failed the most recent large area sample test because some months were low. To reduce the variance from such cases, we set up what we call a temporary large area sample panel. Each adjacent pair of panels contains a common month of data (due to the fact that two months of data are obtained from each panel). If, for example, we consider the adjacent panels of May and June--we have information from both panels for the month of May. We then set a criterion for the month of May. Any establishments which equal or exceed this criterion for the current month for the May panel are also surveyed for June and placed in this tabulation at half weight. At the same time, the weight of all establishments in the June panel whose previous month's sales exceed the criterion are halved. In this fashion, the weight of all "large" establishments (except those "large" in the current month only) are halved. The "temporary" and "permanent" large area sample procedures are integrated by using the same cutoffs for both. Those used in the "temporary" large area sample procedure therefore constitute the "potential" large area sample which is surveyed for the permanent large area sample panel after the end of 12 months.

The optimization problem involved in the use of a large area sample panel is to decide that large area sample cutoff which will produce the most efficient results. All establishments with sales equal to or greater than this cutoff are placed on the large area sample list while those with sales smaller than this cutoff are left in the regular area sample.

The large area sample cutoff was determined empirically by fixing a cost and then approximating the variance from the combined large area sample and regular area sample strata that would be obtained with various cutoffs.

The formula for the variance obtained from the combined strata for a fixed cost and stated cutoff may be expressed as follows:

$$(8)\quad \frac{\dfrac{M^2\sigma_y^2}{12} + M^2\sigma_x^2\left\{(.83)^2\left[\dfrac{1+.8^2-2\rho_x(.8)}{1-.8^2}\right]+(.17)^2\right\}}{\dfrac{C}{12C_y\bar{N}_y + C_x\bar{N}_x}}$$

The numerator of the above fraction represents the combined variance from the area sample and large sample strata per area sample segment drawn. A final-type composite estimate is assumed with the values of W and K which are proposed to be used.

The denominator of the fraction represents the number of area sample segments which can be drawn for the fixed cost and designated cutoff.

The meaning of the individual symbols in the numerator as follows:

M = the number of area sample segments in the universe.

σ_y^2 = the variance per area sample segment for the large area sample stratum.

$= \dfrac{\sum_{i=1}^{M}(Y_i-\bar{Y})^2}{M}$ where Y_i is the segment total of those defined to be in the large area sample universe by the cutoff.

σ_x^2 = the variance per area sample segment for the area sample stratum (of similar form as the variance shown above but with X, or area sample values, substituted for Y values).

ρ_x = the month-to-month correlation for the segment totals of the X values.

Note that the variance of the large area sample universe is divided by 12 (since 12 area sample panels are included) while the area sample portion is multiplied by the reduction factor achieved through the use of the composite estimate. Note also that the values of σ_y^2, σ_x^2 and ρ_x are dependent on the cutoff selected while the other values in the numerator are constants. This variance form assumes that there is no correlation between the X and Y values. (In a small scale study made in the New York primary sampling unit these covariances had a negligible effect on the results).

In the denominator the individual symbols have the following meaning:

C = total resources available for expenditure on the combined area sample and large area sample strata.

C_y = the cost per unit per month of obtaining and processing a large area sample report. (Note that this constant is multiplied by 12 because large area sample establishments from all 12 panels are surveyed each month.)

$\bar{N}_y$ = the average number of large sample establishments per segment.

$C_x, \bar{N}_x$ are similar values relating to the area sample universe.

Note that $\bar{N}_x$ and $\bar{N}_y$ are dependent on the cutoff selected but their sum is constant ($\frac{N}{M}$). All other values in the denominator are constants.

In the above formula, the variance per segment draw declines as does the number of segments which can be afforded for a given cost. The problem is to find that design which gives the lowest value for the fraction. This was done empirically by designating various large area sample cutoffs and then computing the variances for these cutoffs using the above formula. The variances in table 3 are for an estimate of sales of all kinds of business in the New York Metropolitan District while the variances in table 4 are for an estimate of sales of proprietary stores in the United States. Two sets of per unit cost assumptions are used. In the first set (those shown in columns 2 and 4 of each table) it is assumed that the per report cost of an area sample case is four times that of a large area sample case. This approximates the conditions of the retail survey where the area sample reports are collected by personal enumeration and the large area sample reports by mail. In columns 3 and 5 it is assumed that the method of collection and the per unit costs are the same for both strata. The optimum large area sample cutoff in this case is somewhat higher.

Only limited empirical evidence relating to optimum large area sample cutoffs is available because of the labor of computation. However, those data available point to the same general conclusion as column 4 of tables 3 and 4 namely, that there is a broad range for the cutoff centered around 3 to 5 times the average sales value per establishment where losses over the optimum cutoff are apparently small. This is true both for optimums computed from the standpoint of estimates for individual kinds of business and for those computed from the standpoint of the total estimate for all kinds of business combined. The large area sample cutoff for the Retail Trade Survey has been set at about three times the average sales value per establishment. Since only one cutoff could be used for each kind of business, arbitrary compromises were made where the average sales for a particular kind of business differed significantly from the average sales for all kinds of business combined.

Formula (8) and the computations in tables 3 and 4 are designed to produce optimum results for the final composite estimate. Optimum cutoffs could have been computed also for the preliminary composite estimate, the month-to-month change or other relationships.

Table 3: EMPIRICAL COMPUTATION OF APPROXIMATE OPTIMUM CUTOFF BETWEEN AREA SAMPLE AND LARGE AREA SAMPLE STRATA--ALL KINDS OF BUSINESS IN THE NEW YORK METROPOLITAN DISTRICT

Cutoff (monthly sales)	Variance per area sample segment drawn ($x\ 10^{12}$)	Number segments drawn with fixed cost[3]		Variance obtained with stated cutoff	
		With $C_x = 4C_y$	With $C_x = C_y$	With $C_x = 4C_y$ (1) + (2) ($x\ 10^{12}$)	With $C_x = C_y$ (1) + (3) ($x\ 10^{12}$)
	(1)	(2)	(3)	(4)	(5)
(00) [1]	187,678	199.6	278.5	940	674
$100,000 ($20\bar{x}$) [2]	136,391	198.5	271.0	687	503
40,000 ($8\bar{x}$) [2]	116,531	196.1	253.7	594	459
25,000 ($5\bar{x}$) [2]	93,126	191.3	225.1	487	414
15,000 ($3\bar{x}$) [2]	89,078	182.0	182.0	489	489
10,000 ($2\bar{x}$) [2]	92,689	169.8	141.9	546	653
5,000 ($\bar{x}$) [2]	94,145	129.1	69.6	729	1,353

[1] In other words, all cases in area sample stratum.

[2] $\bar{x}$ is the average sales per establishment in the combined strata ($5,025).

[3] Cost is fixed at level needed for 182 segments with cutoff = 3 x (approximate present design) but note that the optimum is independent of the size of the fixed cost. As indicated in Equation 8, Cy is the per unit cost of collecting and tabulating a large area sample report while Cx is the same cost for an area sample report. The first cost assumption (Cx = 4 Cy) is approximately the relationship in the Monthly Retail Trade Report because area sample reports are collected personally while the large area sample reports are collected by mail. The other assumption is used to show the variance relationships if the two types of reports were collected by the same method.

Table 4: EMPIRICAL COMPUTATION OF APPROXIMATE OPTIMUM CUTOFF BETWEEN AREA SAMPLE AND LARGE AREA SAMPLE UNIVERSES--PROPRIETARY STORES

Cutoff (monthly sales)	Variance per area sample segment drawn ($x\ 10^{12}$)	Number segments drawn with fixed cost[3]		Variance obtained with stated cutoff	
		With $C_x = 4C_y$	With $C_x = C_y$	With $C_x = 4C_y$ (1) + (2) ($x\ 10^9$)	With $C_x = C_y$ (1) + (3) ($x\ 10^9$)
	(1)	(2)	(3)	(4)	(5)
(00) [1]	13,495	2,061	2,785	6,548	4,846
\$33,048 ($8\bar{x}$) [2]	12,476	2,043	2,656	6,107	4,697
20,655 ($5\bar{x}$) [2]	11,348	1,923	2,000	5,901	5,674
12,393 ($3\bar{x}$) [2]	11,211	1,900	1,900	5,901	5,901
8,262 ($2\bar{x}$) [2]	10,951	1,797	1,540	6,094	7,111
4,131 ($\bar{x}$) [2]	10,297	1,284	643	8,019	16,014

[1] In other words all cases in area sample stratum.

[2] $\bar{x}$ is the average sales per establishment in the combined strata (\$4,131).

[3] Cost is fixed at level needed for 1,900 segments with cutoff = $3\bar{x}$ (approximate present design). See note [3], table 3 for per unit cost assumptions used.

C. Summary of Effect of Composite Estimation Procedure, and Large Observation Procedure (both temporary and permanent) on Variances of the Monthly Retail Trade Survey

The actual combined effect of the composite estimation procedure and the large observation procedures (both permanent and temporary) is shown by comparing the variances of the estimates obtained using these procedures with the variance of a nonrotating sample. These variances have been computed for all kinds of business combined in the United States for May and June 1959 in table 5. The comparisons shown are the variances of the preliminary and final composite estimates versus the variance of the simple estimate and the variance of the preliminary-to-final month-to-month ratio versus the variance of the ratio of the simple estimates from a nonrotating panel.

These are total variances and a between primary sampling unit contribution from the entire sample is included in the variances of both the simple estimates and the composite estimates. This contribution has not been affected by any of the devices discussed in this paper, therefore percentage of gain is not as large as for the within-primary sampling unit variance alone which is the component affected by the procedures discussed. In spite of this, it appears that the variance of level has been reduced to about 30% of that which would have been obtained from a nonrotating sample while the variance of the ratio between the two most recent months has been reduced to about 45% of that obtained from a nonrotating sample. These results are only approximate because actual conditions are different for each estimate considered and because these variance results are themselves subject to variance. The cost of the rotating sample used may be between 10 and 15% larger than the cost of a nonrotating sample due principally to the cost of the large area sample procedure.

Table 5: COMPARISON OF VARIANCES OF ESTIMATES OBTAINED USING COMPOSITE ESTIMATE AND TEMPORARY AND PERMANENT LARGE AREA SAMPLE PROCEDURES WITH VARIANCES OF THE SIMPLE ESTIMATES FROM NONROTATING PANEL; ALL KINDS OF BUSINESS: MAY AND JUNE 1959

Estimate	Relvariances of simple estimate	Relvariances of composite estimate	Ratio of variances of composite estimate to variances of simple estimate
May 1959 level.......	.000688	[2] .000197	.29
June 1959 level......	.000685	[3] .000202	.29
June-May 1959 ratio..	[1] .000033	[4] .000015	.45

[1] Same panel used for both May and June.

[2] Final composite estimate (relvariance of preliminary composite estimate for May was .000230).

[3] Preliminary composite estimate.

[4] Ratio preliminary to final composite estimate.

Bibliography:

(1) Hansen, M.H., Hurwitz, W.N., Madow, W. G., (1953). Sample survey methods and theory. John Wiley and Sons, Inc., New York.

(2) Patterson, H.D., (1950). Sampling on successive occasions with partial replacement of units. Jour. Roy. Stat. Soc. Series B, 12, 241-255.

(3) Jessen, R.J., (1942). Statistical investigation of a sample survey for obtaining farm facts. Iowa Agr. Exp. Sta. Res. Bull. 304

(4) Yates, F. (1949). Sampling methods for censuses and surveys. Charles Griffin and Co., London

(5) Cochran, W.G., (1953). Sampling techniques. John Wiley and Sons, Inc., New York

SOME SAMPLING TECHNIQUES FOR CONTINUING SURVEY OPERATIONS

By: Leslie Kish and Irene Hess
Survey Research Center, University of Michigan

Sample survey methods have been developed largely in terms of individual sampling operations. However, many organizations carry on their survey operations on a rather continuous basis. The special techniques of "rotating" samples devoted to collecting periodically a specific set of data have been treated by Hansen, Hurwitz and Madow, Cochran, Yates, Albert Eckler, and others [1]. We deal here instead with some techniques generally applicable to any repeated use of the same selection frames with similar survey methods. We shall not dwell on the obvious advantages of an adequate staff working as a team experienced in their special problems. Nor shall we discuss the obvious advantages of prorating the costs of expensive sampling materials over many surveys, as was done in the Master Sample of 1945 [3].

We shall present briefly several simple but useful ideas. Probably most of them will appear obvious to statisticians engaged in similar continuing sampling operations. Yet several of these we had to invent -- or more likely reinvent -- for ourselves. This experience and the impressions gained in teaching these methods to others indicate that at least some of these ideas will prove both interesting and useful to some of our readers.

1. The actual sample for each new survey is selected not from the entire population but from a frame of segments and dwellings selected earlier with a predetermined rate from the entire population. The preparation of this frame needs specialized work, done separately from the surveys, preferably during slack periods both in the office and field schedules. We try to do the field work in the months with favorable weather when travelling around the primary areas is easier and somewhat cheaper. In each place we design a frame which can supply us with sampling materials for a period of one to three years. The period is varied and controlled for efficiency considerations; for example, we prepare a smaller frame in places where we expect rapid growth. From this frame we subsample as needed in such a way that the product of the two probabilities of selection -- first, into the frame and second, out of the frame -- equals the probability desired for a particular survey. In the simple situation of uniform monthly surveys (a kind we do not have at the Survey Research Center) the frame can be compiled for a year and then subsampled, one-twelfth for each month.

In addition to the economy and the speed of obtaining samples for specific surveys, subsampling from a frame offers opportunities for greater statistical efficiency through various devices. Of these, we shall describe four briefly. Some of these introduce additional selection stages or phases for obtaining the necessary information for creating good segment boundaries, for assigning measures of size, or for stratification. For example, our interviewers cheaply obtain economic ratings of dwellings, and we have used these for stratified allocation schemes.

a. Instead of sampling area segments directly from a map, we ask the interviewer to create segments out of a "chunk." As viewed on the office work map, the rural chunk is an area which seems to have good boundaries and contains about 30 dwellings; we send the interviewer a chunk sketch and ask her to revise it to reflect the current situation, to indicate the location of each dwelling, and to add any internal features that may be used to divide the chunk into segments of about four dwellings each. In urban areas the chunk may be all or part of a city block, subsampled after a preliminary visit to the block by the interviewer who reports dwelling unit counts and locations, number of floors and apartments in apartment buildings, and related information. This permits us to send the interviewer, for her survey work, segments with boundaries that are familiar to her and with recent measures of size. From the chunk we sample segments without replacement and continue until virtually every segment has been selected for some survey.

b. When we select tracts or blocks from Census listings, we use double sampling; first we select a relatively large sample, stratify it, and then draw from it as needed. It costs scarcely more to select from the Census lists a larger initial sample than a smaller one; thus the cost of drawing the sample can be divided by the number of times it is used. Furthermore, we obtain for the actual sample the gains of stratifying (by geography and by income indicators) the larger initial sample.

c. Currently, we use city directory listings of addresses wherever these are practicable; here, too, we select a larger frame and then subsample repeatedly. Generally, we first select clusters of ten or twenty directory lines, then subdivide these into clusters of about four dwellings for a survey sample. We estimate the probable number of dwellings at each address and group nearby addresses to create clusters of about four dwellings.

But for a three-year frame of the City of Detroit, we selected clusters of three lines and then subselected single lines for each year's sample; thus obtaining, for each year's sample, individually selected addresses throughout the city.

d. We also use double sampling in the area supplement designed for picking up new growth and missed dwellings. This supplement is particularly necessary to correct selections from city directories. First, we may select an initial sample of blocks eight times as large as we need for a single survey and obtain for these, in the field, rough estimates of size. Then we are free to subselect either one-eighth of the blocks with little growth, or one-eighth of the dwellings from the blocks with much growth.

2. Practitioners of survey sampling know the painful feeling of surprise when they find 20 or even 200 dwellings in a small segment where they expected about four. Area samplers are in a continuing race with home builders. This problem is likely to occur six months or a year after the compilation of the frame of chunks and segments, when the sampler can no longer distribute the surprise building over all subsamples from the frame. For the sake of equal probabilities the sampler might accept all of the surprise dwellings -- and the corresponding increase in variance and cost. At some point, however, he may decide to cut the sample take, thus accepting some bias, but probably with a lower mean square error than the unbiased procedure.

This source of bias could be reduced over the long run of continuing operations by averaging these events in a surprise stratum. The population expansion from a surprise of x_g dwellings (or other elements), from the g-th survey taken with the over-all sampling rate of f_g, is x_g / f_g. The average over G surveys is

$$[\sum^{G} w_g \, x_g / f_g] \,/\, \sum^{G} w_g \,,$$

where w_g is the weight given to the g-th survey. In many situations w_g should be made proportional to f_g; then the average becomes

$$\sum^{G} x_g \,/\, \sum^{G} f_g \quad .$$

The sample "take" from the surprise stratum for the last survey should be, then,

$$f_g \sum^{G} x_g \,/\, \sum^{G} f_g \,.$$

The summation G would be over a period (perhaps two years) large enough to provide a "long run" for averaging but not so large as to cause bias by obsolescense. The current survey can be included as the last survey, denoted by G.

We have not used this method since we had only one surprise over the past several years, thanks to an elaborate system of information on growth from the sample counties. This includes interviewers' reports, during surveys, about perceived growth in the visited chunks. But now we intend to lower the criteria for "surprises" and, over several studies, to establish a "surprise stratum." We invite your suggestions and reports of your experiences with this problem.

3. When facing the problem of changing the selection probabilities of a set of sampling units, the sampler may want to use a method that will minimize the number of sampling units that must be changed because changing them is expensive. In particular, we have in mind primary sampling areas, counties or metropolitan areas, each representing an investment of hundreds of dollars in interviewer training and in sampling materials. Nathan Keyfitz has described the problem and a procedure for introducing new population sizes [2].

We may represent the original probabilities used for the selection of sampling units from a stratum as

$$\sum^{J} i_j + \sum^{K} d_k + \sum^{M} s_m = 1 \quad .$$

Similarly, for the same sampling units the new probabilities to which we want to change are

$$\sum^{J} I_j + \sum^{K} D_k + \sum^{M} S_m = 1 \,.$$

Here i_j and I_j denote the original and the new probabilities of the same sampling unit; there are J + K + M sampling units in the stratum. We use the letters i and I, d and D, s and S to denote, respectively, sampling units with increase, decrease or the same probability from the original to the new measure. These three subsets of sampling units have the relationships

$$I_j > i_j \,, \qquad D_k < d_k \,, \qquad S_m = s_m \,.$$

Also we have $\sum^{M} S_m = \sum^{M} s_m$;

hence, $\sum^{J}(I_j - i_j) = \sum^{K}(d_k - D_k)$.

That is, the sum of the probability increases must equal the sum of the probability decreases.

The procedure for changing probabilities is as follows:

a. If the originally selected sample unit shows either an increase or no change in probability, it remains in the sample with the new probability I_j or S_m .

b. If a sample unit decreases in selection probability, then its probability of remaining is made D_k / d_k and its probability of being dropped is made $1 - \frac{D_k}{d_k}$. If we decide (by resorting to a table of random numbers) that the unit remains, the compound probability of original selection and remaining is

$$d_k \times \frac{D_k}{d_k} = D_k .$$

c. If a unit is dropped from the sample, we select a replacement from among the increased units with probabilities proportional to the increases, the probability of selection for the j-th unit being

$$\frac{I_j - i_j}{\Sigma(I_j - i_j)} .$$

Thus, the total selection probability for a unit that increased is

$$i_j + \Sigma(d_k - D_k) \times \frac{I_j - i_j}{\Sigma(I_j - i_j)} = I_j .$$

Our methods represent generalizations of the Keyfitz technique in three directions. First, we introduced considerations of statistical efficiency into the problem, knowing that it is neither necessary nor possible to have precise measures of size. It is necessary and sufficient that the sum total of net changes be zero. Within that requirement we can adjust the probabilities of selection to satisfy some criteria of change sufficiently large to be recognized as "important." We noted that for many of the sampling units the change in probabilities was small and unimportant. To these units we reassigned the old probabilities and they became the S units. This procedure reduces the probability of having to switch sampling units; it also eliminates the task of having to revise office records for the sample units with no change in probabilities. In choosing criteria, we tried to balance the increased costs involved in changing primary sampling units (psu's) against the increase in survey variances due to the increased variation in the sizes of sample clusters arising from the small distortions in the probabilities of selection. Of course, we had only crude measures for these criteria, but we believe "anything worth doing at all is worth doing badly." We decided on the following procedure: (a) define important increase as 10 per cent or more and add all such increases over the entire stratum; (b) then add enough decreases from an ordered set of decreases and adjust balance <u>exactly</u> the increases; (c) consider all other sampling units as not having changed. We might have defined a minimum critical decrease, but this we did not consider necessary. Incidentally, rather than merely accepting a specific amount of change from one Census period to another, the rate of change can be projected into the middle of the period of the use of the frame. In other words, the California counties which have increased from 1950 to 1950 will tend to increase through the 1960's; and the sampler making the adjustment in 1960 may take this into account in designing his sample for the '60's.

Second, we also introduced controlled selection into the changes of probabilities. Faced with rather small probabilities of change $(1 - \frac{D_k}{d_k})$ in each of 54 strata (zero in many), instead of drawing independently in each stratum, we cumulated the expected fractions of change from one stratum to another and applied an interval of one, after a random start. Thus, the actual number of changes was controlled within a fraction of the expected number of changes.

To illustrate the application of both the "strict" and the "flexible" assignments of probabilities we display their results for the 54 primary sampling units selected from as many strata in the Survey Research Center's national sample:

Classification of Sample psu's	Number
Total primary sampling units	54
A. Increase with both strict and flexible plans	6
B. Increase with strict, but remains same with flexible	20
C. Decrease with strict, but remains same with flexible	12
D. Decrease with both strict and flexible plans	16

This saves all changes in 32 units, among which 12 were also exposed to being dropped under the "strict" plan. The 6 units with increases were retained in the sample and assigned their new probabilities, with all their records relating to probabilities of selection corrected. Using a controlled selection procedure for dropping psu's resulted in changing 3 of the possible 16. The other 13 required record changes only to convert the old probabilities to the newly adjusted ones.

The illustration in Table 1 of one stratum may clarify some of the details of the flexible procedure. Subset A of psu's includes all with probability increases of 10 per cent or more (see column 4); the new probabilities for these psu's are the same with either plan (columns 3 and 7). Subset B includes psu's with increases of less than 10 per cent (in column 4); these psu's are assigned probability changes of zero in column 6, and the new probabilities in column 7 are identical with the original probabilities (column 2). Next, we must define "decrease" in such a manner that the net change in probabilities will be zero over the entire stratum.

To define "decrease" we order the psu's with respect to the ratio of probabilities (that is, the order of column 4). Beginning with the lowest, proceed up column 5 cumulating the decreases until their absolute values just exceed the sum of the increases, .03640; each decrease (the last five entries in column 5) was adjusted proportionately so that the decreases of column 6 equal the increases in absolute value. The remaining psu's, subgroup C, are considered to have no change in probabilities.

Third, we want to add that this method may also be used to adjust the probabilities from the original population to some other population. For example, suppose that the psu's were selected with probabilities proportional to numbers of persons and that we now want to sample another population which is distributed somewhat (but not very) differently from the original population -- for example, a population of physicians, or farmers, college students, or Boy Scouts. The same techniques may be used to adjust the probabilities of selection from the original to the newly desired population.

To summarize our modifications of the Keyfitz method: first, we reduce the expected number of necessary changes; second, we reduce the variability of that number; third, we generalize the applicability of the method.

4. The estimators of the variance of survey results are often subject to large variations. This is particularly true for models which use few primary selections (approximate degrees of freedom). For similar items in several surveys, greater precision may be obtained by averaging computations over several succeeding periods. We are conducting investigations of this problem.

5. Another technique available to organizations conducting surveys at intervals with similar methods is the reduction of the effects of nonresponse by simulating an increase in the number of recalls [4]. This technique consists in adding to the addresses of current surveys the nonresponse addresses from similar recent surveys. The replacement addresses should be chosen from surveys using similar respondent units because not-at-homes and refusals among some respondent units may differ from those among others. Refusals may also depend to some extent on survey objectives and questions. The effect of the procedure is about equal to that obtained by doubling the number of recalls, but without the corresponding increase in expense and trouble.

6. The accumulation of evidence on response rates and coverage rates is another advantage of continuing operations. This permits better control of the sample size through the accumulated knowledge of field results. Furthermore, by studying factors associated with varying rates of response, the researcher can learn something about the sources of nonresponse and how to cope with them. Of course, this knowledge does not accumulate automatically but only with planning and labor.

7. One result of continuing operations is the presence of inertia in different parts of the design. For example, in designing some modest size studies, we often find it cheaper and easier to utilize our standard 66 primary sampling areas, or perhaps half of them, than to design and staff six cities, let us say, for the study. This results from having trained interviewers and sampling frames and materials available in our regularly used primary areas. It seems to contradict the usual rule that it is cheaper to use fewer sampling units.

8. Continuing operations also result in a certain conservatism of methods. Some of this is the justifiable result of having available certain good, economic and reliable methods, tested with long experience; and the preceding example can serve as an illustration. However, we suspect that there must also exist many less justifiable types of conservatism, because one naturally thinks first of methods that seem to have worked well enough, that is, without noticeable catastrophes. It is difficult to view new problems with fresh, unbiased eyes. But one should always strive for that fresh point of view and question his familiar methods, trying to separate the seasoned timber from dead wood.

REFERENCES

[1] Hansen, M.H., Hurwitz, W.N., Madow, W.G., Sample Survey Methods and Theory, Vol. I., New York: John Wiley and Sons, Inc., 1953, pp 491-493.

Cochran, W.G., Sampling Techniques, New York: John Wiley and Sons, Inc., 1953, pp 282-290.

Yates, Frank, Sampling Methods for Censuses and Surveys, New York: Hafner Publishing Company, 1953, Second edition, pp 175-182, 233-235, 260-262.

Eckler, Albert R., "Rotation Sampling", The Annals of Mathematical Statistics, Vol. 26, No. 4 (December, 1955), pp 664-685.

[2] Keyfitz, Nathan, "Sampling with Probabilities Proportional to Size", Journal of the American Statistical Association, 46 (March, 1951), pp 105-109.

[3] King, A.J., Jessen, R.J., "The Master Sample of Agriculture", Journal of the American Statistical Association, Vol. 40, No. 229 (March, 1945), pp 38-56.

[4] Kish, Leslie and Hess, Irene, "A 'Replacement' Procedure for Reducing the Bias of Nonresponse", The American Statistician, Vol. 13, No. 4, (October, 1959), pp 17-19.

TABLE 1

COMPARISON OF STRICT WITH FLEXIBLE PLAN FOR CHANGING SELECTION PROBABILITIES OF SAMPLING UNITS IN ONE STRATUM

Item	Original probability	New probability for strict plan	Ratio of probabilities (Col. 3 ÷ Col. 2)	Change in prob. for strict plan (Col. 3-Col. 2)	Flexible plan: Change* in probability	Flexible plan: New probability (Col. 2 + Col. 6)
1	2	3	4	5	6	7
PSU Classification						
A. Increase with both strict and flexible plans 1.	.07307	.09000	1.232	+ .01693	+ .01693	.09000
2.	.09407	.11354	1.207	+ .01947	+ .01947	.11354
B. Increase with strict, same with flexible 1.	.03317	.03636	1.096	+ .00319	.00000	.03317
2.	.04381	.04718	1.077	+ .00337	.00000	.04381
3.	.08915	.09142	1.025	+ .00227	.00000	.08915
4.	.09073	.09083	1.001	+ .00010	.00000	.09073
C. Decrease with strict, same with flexible 1.	.05500	.05433	.988	- .00067	.00000	.05500
2.	.04258	.04204	.987	- .00054	.00000	.04258
3.	.07297	.07138	.978	- .00159	.00000	.07297
4.	.09719	.09317	.959	- .00402	.00000	.09719
D. Decrease with both strict and flexible plans 1.	.05059	.04807	.950	- .00252	- .00238	.04821
2.	.02478	.02343	.946	- .00135	- .00128	.02350
3.	.05611	.04894	.872	- .00717	- .00678	.04933
4.	.09263	.07871	.850	- .01392	- .01315	.07948
5.	.08415	.07060	.839	- .01355	- .01281	.07134
Total	1.00000	1.00000	- - -	- - -	- - -	1.00000
Summation by subset A.	.16714	.20354	- - -	+ .03640	+ .03640	.20354
B.	.25686	.26579	- - -	+ .00893	.00000	.25686
C.	.26774	.26092	- - -	- .00682	.00000	.26774
D.	.30826	.26975	- - -	- .03851	- .03640	.27186

* Col. 5 entries for the five psu's of class D. were adjusted by a factor of .03640/.03851 = .9452 to obtain the corresponding col. 6 entries.

IX

RATIO ESTIMATION

Chairman, H. O. Hartley, Iowa State University

ANALYTIC STUDIES OF SURVEY DATA

By: H. O. Hartley, Iowa State University

1. *The problem of group comparisons in sample surveys*

One of the main objectives of a sample survey is the computation of estimates of (say) means and totals of a number of characteristics attached to the units of a population. More often than not, however, the data are also used for what is known as an 'analytic study' or a 'critical analysis' of a survey. Such an analysis usually involves the comparison of means and totals of certain subgroups of the population. To fix the idea, a farm-economic survey in the State of Iowa may have been primarily planned to obtain estimates of totals of numerous farm-economic items such as annual hog sales or annual bushels of corn sealed for the total population of farms in Iowa. In a subsequent analytic study one may then be concerned with the comparison of some of these items for certain subgroups of farms such as 'owner operators' and 'tenant operators'. Such subgroups of the total population have been termed 'domains of study' by the U. N. Subcommission on Sampling and this term is also used by Yates (1949, 1953) who provides certain formulas for the estimation of their means and variances in the more elementary survey designs.

Although such domains are usually fairly well defined, it will often not be known until after sampling which of the domains any particular unit belongs to. Thus the domains with which an analytic study is concerned are normally not represented in the sample in prescribed fixed sample proportions and the number of sampled units in each domain will itself be a random variable. This is perhaps the most characteristic difference between 'domains' in analytic studies and 'treatment groups' of experiments: but there are others, and the main departures from standard analysis of variance conditions may be summarized under three headings:

1. The number of units in the domains (subgroups) are random variables.
2. The population from which the samples are drawn is finite.
3. Sampling is often not simple random but stratified and/or multistage resulting in correlations of the characteristics of units in the same domain as well as of units in different domains.

To illustrate these points and, at the same time, introduce notation required later we give below three examples of simple survey designs.

Example 1. Simple random sample.

Yates (1949, p. 152) gives data* for a simple random sample of 1/20 Hertfordshire farms. The $n = 125$ sampled farms were, after selection, classified into 7 districts and the number of farms and their total acreages (of crops and grass) are shown in Table 1 below.

TABLE 1

Numbers and total acreages for 125 Hertfordshire farms classified in 7 districts after selection

	District Number							Total (mean)
	1	2	3	4	5	6	7	
No. of farms in sample .	15	8	40	24	4	24	10	125
Total acreage.	1,935	1,385	4,851	4,034	335	2,027	547	15,114
Mean acreage.	129.0	173.1	121.3	168.1	83.8	84.5	54.7	120.9

* Yates (1949) stresses that these data are for illustration only, see his explanations pg. 30-31.

The notation for the entries in Table 1 are shown in Table 1a below.

TABLE 1a

General notation for random sample of n units classified in k domains after sampling

	Domain No.				Total (mean)
	1	2.	j.	k	
No. of units in sample .	${}_1n$	${}_2n$	${}_jn$	${}_kn$	n
Total of character . . .	${}_1y$	${}_2y$	${}_jy$	${}_ky$	y
Mean of character . . .	${}_1\bar{y}$	${}_2\bar{y}$	${}_j\bar{y}={}_jy/{}_jn$.	${}_k\bar{y}$	$\bar{y} = y/n$

If we visualize the whole population as subdivided into the domains we reach the notation set out in Table 1b.

TABLE 1b

General notation for population values of domains from which a random sample was drawn

	Domain No.				Total (mean)
	1	2.	j.	k	
No. of units in pop. domain	${}_1N$	${}_2N$. . . .	${}_jN$	${}_kN$	N
Total of characteristic . .	${}_1Y$	${}_2Y$. . . .	${}_jY$	${}_kY$	Y
Mean of characteristic . .	${}_1\bar{Y}$	${}_2\bar{Y}$. . . .	${}_j\bar{Y}$	${}_k\bar{Y}$	$\bar{Y} = Y/N$

The main purpose of an analytic study would now consist in estimating the domain means and totals ${}_j\bar{Y}$ and ${}_jY$ from the sample and to provide errors for such estimates. In the present simplest case of a random sample it is easy to guess (as will, in fact, be established later) that the domain population means ${}_j\bar{Y}$ may be estimated by the corresponding sample means ${}_j\bar{y}$. However, the computation of the errors of means of the 'single classification' given in Tables 1 and 1a by standard analysis of variance technique would take no account of either the sampling procedure by which the data were collected or of the actual population for which estimates are required and would, in fact, introduce the assumption of an artificial model not necessarily relevant to the data. That faulty inferences* can be drawn from the application of standard analysis of variance procedures is obvious if we consider the special case when sampling is 100%, i.e. $n = N$ and when the ${}_j\bar{Y} \equiv {}_j\bar{y}$ are in fact estimated without error.

Example 2. Simple stratified sampling (Yates, 1949, p. 154).

Only the first two entries in each cell, the number of units ${}_jn_h$ and the totals ${}_jy_h$ are shown in the example Table 2; the means ${}_j\bar{y}_h = {}_jy_h/{}_jn_h$ are not entered.

If we visualize the whole population of units (farms) as likewise classified by strata and domains the population numbers and totals corresponding to those in the sample of Table 2a would be denoted by capital letters, i.e.

$$ {}_jN_h\,,\ {}_jY_h\,,\ {}_j\bar{Y}_h = {}_jY_h/{}_jN_h\,; \quad N_h\,,\ Y_h\,,\ \bar{Y}_h = Y_h/N_h\,; $$

$$ {}_jN,\ {}_jY,\ {}_j\bar{Y} = {}_jY/{}_jN, \quad N,\ Y,\ \bar{Y} = Y/N\,; $$

This simple example shows the occurrence of correlation (say) between units in the same domain: Of the ${}_jn$ units giv-

* Note also discussion in Section 8.

Table 2

Numbers of farms and total wheat acreages for a sample of $n = 135$ Hertfordshire farms stratified by 'size-group' and classified by District after selection

Size group stratum (acres)	Population size N_h	District Number 1	2	3	4	5	6	7	Total
6- 20	519	0 0	1 0	0 0	1 0	0 0	1 0	0 0	3 0
21- 50	357	1 0	1 10	2 0	1 17	0 0	1 0	0 0	6 27
51-150	519	3 36	3 40	5 40	5 65	2 0	5 19	3 14	26 214
151-300	400	4 63	5 213	10 270	8 305	5 112	6 140	2 60	40 1163
301-500	215	4 320	10 1074	11 659	12 989	3 234	2 0	1 16	43 3292
501-	51	1 114	4 487	2 315	9 1937	0 0	1 72	0 0	17 2925
Total	Number	13	24	30	36	10	16	6	135
	Wheat acreage	533	1824	1284	3313	346	231	90	7621

Table 2a

General notation for sample of n units stratified in L strata and classified in k domains after selection

Stratum number		Domain Number 1. . .	j. . .	k	Total (mean)
$h = 1$	No. of units	$_1n_1$	$_jn_1$	$_kn_1$	n_1
	Total.	$_1y_1$	$_jy_1$	$_ky_1$	y_1
	Mean.	$_1\bar{y}_1$	$_j\bar{y}_1$	$_k\bar{y}_1$	$\bar{y}_1$
.		. . . .	. . . .	. . . .	. . . .
h	No. of units	$_1n_h$	$_jn_h$	$_kn_h$	n_h
	Total.	$_1y_h$	$_jy_h$	$_ky_h$	y_h
	Mean.	$_1\bar{y}_h$	$_j\bar{y}_h$	$_k\bar{y}_h$	$\bar{y}_h$
.		. . . .	. . . .	. . . .	. . . .
L	No. of units	$_1n_L$	$_jn_L$	$_kn_L$	n_L
	Total.	$_1y_L$	$_jy_L$	$_ky_L$	y_L
	Mean.	$_1\bar{y}_L$	$_j\bar{y}_L$	$_k\bar{y}_L$	$\bar{y}_L$
Total	No. of units	$_1n$	$_jn$	$_kn$	n
(Mean)	Total.	$_1y$ . .	$_jy$ . .	$_ky$	y
	Mean.	$_1\bar{y}$	$_j\bar{y}$	$_k\bar{y}$	$\bar{y}$

ing rise to the domain mean $_j\bar{y}$ 'a cluster' of $_jn_h$ will be found in stratum h and will usually be positively correlated. The correlation is similar to that found in a two-way table of an analysis of variance with unequal cell frequencies but, of course, cannot here be assumed to have been caused by an additive model.

Example 3. Two stage sample, primaries drawn with replacement and probabilities proportional to size (p. p. s.) and secondaries drawn without replacement and with equal probability.

Table 3 below gives data from a 'consumer preference survey' carried out by the Statistical Laboratory of Iowa State College in the City of Des Moines. The survey was arranged as a stratified two-stage design with 50 city blocks as strata from each of which were sampled 2 'segments' (primaries) each containing an expected number of 5 households (secondaries). After completion of the survey the data in each stratum were classified by segment number $t = 1, t = 2$ and in 6 different Income Groups $j = 1, 2, \ldots, 6$. Actually Table 3 gives the totals for 5 strata-groups combined by pooling the answers for all segments 1 and all segments 2 for the 10 strata in each group. The table shows for each stratum group:

(*a*) the number of households in each classification cell ($_jm_t$, top line)

(*b*) the total number of persons $_jy_t$ in households in the j^{th} income group and in segments $t = 1$ and segments $t = 2$ respectively

(*c*) the average number of persons per household ($_j\bar{y}_t$) for each cell of the two-way classification.

Table 3a shows the notation in the general case of a stratum from which n primaries ($t = 1, 2, \ldots, n$) were drawn with the sample results subdivided into k domains ($j = 1, 2, \ldots, k$). The stratum subscript h has been omitted from all symbols.

Table 3

Number of households (a), number of persons (b) and number of persons per household (c) for 467 households sampled from the City of Des Moines and arranged in 5 'strata groups', 6 'income groups' and 2 segments

Stratum group	Segment		Income group 1 Less than $25 per week	2 $25-$50 per week	3 $50-$75 per week	4 $75-$100 per week	5 $100-$125 per week	6 More than $125 per week	Total
1	1	a	—	7	5	13	14	8	47
		b	—	21	14	42	55	25	157
		c	—	3.00	2.80	3.23	3.93	3.12	3.34
	2	a	2	8	6	12	5	13	46
		b	2	23	16	45	21	50	157
		c	1.00	2.88	2.67	3.75	4.20	3.85	3.41
2	1	a	2	9	13	9	10	13	56
		b	2	20	42	26	37	43	170
		c	1.00	2.22	3.23	2.89	3.70	3.31	3.04
	2	a	2	9	9	13	7	10	50
		b	2	14	20	48	30	34	148
		c	1.00	1.56	2.22	3.69	4.28	3.40	2.96
3	1	a	11	8	15	9	—	4	47
		b	21	24	50	31	—	10	136
		c	1.91	3.00	3.33	3.44	—	2.50	2.89
	2	a	1	8	11	7	5	3	35
		b	1	17	36	26	19	17	116
		c	1.00	2.12	3.27	3.71	3.80	5.67	3.31
4	1	a	4	5	15	5	1	12	42
		b	8	12	47	13	3	36	119
		c	2.00	2.40	3.13	2.60	3.00	3.00	2.83
	2	a	3	2	20	9	2	9	44
		b	6	1	50	26	3	28	115
		c	2.00	2.00	2.50	2.89	1.50	3.11	2.61
5	1	a	8	10	10	9	9	12	58
		b	15	26	37	32	27	45	182
		c	1.88	2.60	3.70	3.56	3.00	3.75	3.14
	2	a	—	7	14	8	8	5	42
		b	—	15	50	34	27	9	135
		c	—	2.14	3.57	4.25	3.38	1.80	3.21

If we visualize the whole population of units as likewise classified by primaries and domains the population numbers and totals corresponding to those in the sample of Table 3a would be denoted by capital letters, i.e. by

$$N; \quad _jM_t\,, {_jY_t}\,, {_j\bar{Y}_t} = {_jY_t}/{_jM_t}\,; \quad M_t\,, Y_t\,, \bar{Y}_t = Y_t/M_t\,;$$

$$_jM, {_jY}, {_j\bar{Y}} = {_jY}/{_jM}\,; \quad M, Y, \bar{Y} = Y/M\,;$$

TABLE 3a

General notation for two stage sample of n primaries $t = 1, 2, \dots, n$ containing respectively m_t sampled secondaries which are classified into k domains after sampling

Primary Number		Domain No. 1...	j...	k	Total
$t = 1$	No. of units	$_1m_1$	$_jm_1$	$_km_1$	m_1
	Total.	$_1y_1$	$_jy_1$	$_ky_1$	y_1
	Mean.	$_1\bar{y}_1$	$_j\bar{y}_1$	$_k\bar{y}_1$	$\bar{y}_1$
.		. . . .	. . . .	. . . .	.
t	No. of units	$_1m_t$	$_jm_t$	$_km_t$	m_t
	Total.	$_1y_t$	$_jy_t$	$_ky_t$	y_t
	Mean.	$_1\bar{y}_t$	$_j\bar{y}_t$	$_k\bar{y}_t$	$\bar{y}_t$
.		. . . .	. . . .	. . . .	.
n	No. of units	$_1m_n$	$_jm_n$	$_km_n$	m_n
	Total.	$_1y_n$	$_jy_n$	$_ky_n$	y_n
	Mean.	$_1\bar{y}_n$	$_j\bar{y}_n$	$_k\bar{y}_n$	$\bar{y}_n$
Total	No. of units	$_1m$	$_jm$	$_km$	m
	Total.	$_1y$	$_jy$	$_ky$	y
	Mean.	$_1\bar{y}$	$_j\bar{y}$	$_k\bar{y}$	$\bar{y}$

The primary index $t = 1, 2, \dots, N$ now runs through the complete population of primary units. If the population consists of L strata the stratum index $h = 1, 2, \dots, L$ will precede the primary index t in all symbols.

2. *Estimation of domain totals, variance formulas and variance estimation*

We begin with the estimation of domain totals leaving that for domain means for the next section. It is clear that any theory of estimating domain totals or means will have to cover, as a special case, the situation when the 'domain' consists of the total population, and this is, of course, the well known theme dealt with in the literature on sample survey methodology. As is well known, this generally accepted theory of estimation is essentially distribution free, does not accept any particular models, is in fact based on the first two moments only and can be roughly described as 'unbiased estimation with optimum variance properties'.

In the following we shall accept this theory of estimation also for the more general problem of estimating domain totals and means.

We recall here certain basic formulas for the estimation of population totals which may be found in most text books on the subject: Appropriate to any particular design there are three basic formulas pertinent to the estimation of the population total. They are in general notation:

(*a*) The estimate of the population total Y: —

$$\hat{Y} = \hat{Y}(y_i) \quad [1]^*$$

(*b*) The population variance formula for $\hat{Y}$: —

$$\text{Var}(\hat{Y}) = V(y_i) \quad [2]^*$$

(*c*) The estimated variance of $\hat{Y}$: —

$$\text{var}(\hat{Y}) = v(y_i) \quad [3]^*$$

The multi-variable functions $\hat{Y}(y_i)$ $V(y_i)$ and $v(y_i)$ will depend on the particular design which has given rise to the sample of y_i *.

For example in stratified sampling (*Example* 2) we would have

$$\hat{Y} \equiv \hat{Y}(y_{hi}) = \sum_{h=1}^{L} N_h \bar{y}_h \quad [4]$$

$$\text{Var}(\hat{Y}) \equiv V(y_{hi}) = \sum_{h=1}^{L} N_h^2 (1 - n_h/N_h)\, n_h^{-1} S_h^2 \quad [5]$$

$$\text{var}(\hat{Y}) \equiv v(y_{hi}) = \sum_{h=1}^{L} N_h^2 (1 - n_h/N_h)\, n_h^{-1} s_h^2 \quad [6]$$

$$\text{where } S_h^2 = (N_h - 1)^{-1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2 \quad [7]$$

$$s_h^2 = (n_h - 1)^{-1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2 \quad [8]$$

and these formulas for $L = 1$ (one stratum) would yield as a special case the corresponding formulas for simple sampling (*Example* 1). The appropriate formulas for the two stage design of *Example* 3 (primaries p.p.s. equal take $\bar{m}$ of secondaries)

$$\hat{Y} \equiv \hat{Y}(y_{ti}) = M\bar{y} \quad [9]$$

$$\text{Var}(\hat{Y}) \equiv V(y_{ti}) = \frac{M}{n} \sum_{t=1}^{N} \frac{M_t}{\bar{m}} (1 - \bar{m}/M_t)\, S_t^2 + \frac{M^2}{n} \sum_{t=1}^{N} \frac{M_t}{M} (\bar{Y}_t - \bar{Y})^2 \quad [10]$$

$$\text{var}(\hat{Y}) \equiv v(y_{ti}) = \frac{M^2}{n(n-1)} \sum_{t=1}^{n} (\bar{y}_t - \bar{y})^2 \quad [11]$$

$$\text{where } S_t^2 = (N_t - 1)^{-1} \sum_{i=1}^{M_t} (y_{ti} - \bar{Y}_t)^2 \quad [12]$$

$$s_t^2 = (n_t - 1)^{-1} \sum_{i=1}^{\bar{m}} (y_{ti} - \bar{y}_t)^2 \quad [13]$$

In general $\hat{Y}$ and $v(\hat{Y})$ will depend on the sampled y_i only, and these are the formulas used in practical evaluation of the estimate and its variance, whilst $V(y_i)$ depends on the entire population of y_i values and is used for the theoretical evaluation of the merits of the estimator. The three formulas are applicable to ***any*** set of characteristics y_i attached to the population units. This enables us to use these same formulas for the estimation of domain totals and their variances: — To this end we introduce the following characteristics $_jy_i$ which we shall attach to ***all*** units in the population.

$$_jy_i = \begin{cases} y_i \text{ if the ith unit belongs to jth group (domain)} \\ 0 \text{ otherwise} \end{cases} \quad [14]$$

Now the group total $_jY$ of our j^{th} group is seen to be the population total of the $_jy_i$ and standard sample survey theory, therefore, provides the following estimators:

(*a*) The estimate of the domain total $_jY$: —

$$_j\hat{Y} = \hat{Y}(_jy_i) \quad [15]$$

(*b*) The population variance formula for $_j\hat{Y}$: —

$$_jV = V(_jy_i) \quad [16]$$

* The arguments y_i denote the characteristics attached to the *individual* units; in stratified sampling they would of course show a double subscript y_{hi}, in multistage sampling a triple subscript y_{hti}, etc.

(c) The estimated variance of ${}_j\hat{Y}$: —

$${}_jv = v\ ({}_jy_i) \qquad [17]$$

together with the assurance of unbiasedness

$$E\ ({}_j\hat{Y}) = {}_jY \qquad [18]$$

$$E\ ({}_jV) = {}_jV_i \qquad [19]$$

The meaning of formulas [15] to [17] is simply that the standard formulas for estimation [4] to [13] be applied to the characteristics [14]. The 'spelling out' of these formulas often results in simplifications: In Example 2 (stratified sampling) we find that formulas [15] to [17] can be written as

(a) The estimate of the domain total:

$${}_j\hat{Y} = \sum_h N_h\, {}_jy_h/n_h \qquad [20]$$

(b) The population variance formula for ${}_j\hat{Y}$

$${}_jV = \sum_h \frac{N_h\,(N_h - n_h)}{n_h\,(N_h - 1)} \left\{ \sum_{i=1}^{{}_jN_h} {}_jy_{hi}^2 - {}_jY_h^2/N_h \right\} \qquad [21]$$

(c) The estimated variance of ${}_j\hat{Y}$

$${}_jv = \sum_h \frac{N_h\,(N_h - n_h)}{n_h\,(n_h - 1)} \left\{ \sum_{i=1}^{{}_jn_h} {}_jy_{hi}^2 - {}_jy_h^2/n_h \right\} \qquad [22]$$

The last formula may be compared with one given in the 2nd edition of Yates' book (Yates 1953, 301 formula [9.3.e]) which corrects an earlier (faulty) formula given in his 1st edition (Yates 1949, 202). To make this comparison we use the sample variance of the units in the h^{th} stratum and in the j^{th} domain, we write

$${}_js_h^2 = \sum_{i=1}^{{}_jn_h} ({}_jy_{hi} - {}_j\bar{y}_h)^2/({}_jn_h - 1)$$

and hence obtain [22] in the form

$${}_jV = \sum_h \frac{N_h\,(N_h - n_h)}{n_h\,(n_h - 1)} \left\{ ({}_jn_h - 1)\,{}_js_h^2 + \left(\frac{1}{{}_jn_h} - \frac{1}{n_h}\right) {}_jy_h^2 \right\} \qquad [23]$$

which agrees with Yates' formula [9.3.e]. Our formulas [22] and [23] have been proved to be unbiased estimates of the exact variance of ${}_j\hat{Y}$. The second term of [23] which is proportional to the square of the domain total ${}_jy_h^2$, is characteristic of estimators of *totals* of a population domain whose size is unknown and this feature will, of course, disappear in the estimation of domain *means*.

In Example 3, (two stage sampling with primaries drawn p.p.s.) if we assume that an equal take of $\bar{m}$ secondaries are sampled from each primary, formulas [15] and [17] spell out as follows:

(a) The estimate of the domain total:

$${}_j\hat{Y} = M\, {}_jy/m = M \left(\frac{{}_jm}{m}\right) {}_j\bar{y} \qquad [24]$$

(c) The estimated variance of ${}_j\hat{Y}$

$${}_jv = \frac{M^2}{n\,(n-1)} \sum_{t=1}^{n} \left(\frac{{}_jy_t}{m_t} - \frac{{}_jy}{m}\right)^2 \qquad [25]$$

The interpretation of [24] is simple: The sample domain mean ${}_j\bar{y}$ estimates the corresponding population mean ${}_j\bar{Y}$ and the fraction ${}_jm/m$ the corresponding population fraction ${}_jM/M$, the unbiasedness of the product being assured by [18].

3. *Estimation of domain means, variance formulas and variance estimation*

In the preceding section we gave estimators ${}_j\hat{Y}$ of the population total ${}_jY$ of the j^{th} domain. We now turn to the estimation of the domain means ${}_j\bar{Y}$ (given by ${}_jY/{}_jN$ in the notation of Example (1 and 2 and Table 1 b, and by ${}_jY_j/M$ in the notation of Example 3). Even if the number of units in the j^{th} domain ${}_jN$ (or ${}_jM$) were known it would usually be unwise to use ${}_j\hat{Y}/{}_jN$ (or in two-stage sampling ${}_j\hat{Y}/{}_jM$) as an estimator of ${}_j\bar{Y}$, since it is well known from experience with such estimators of population means that their variances are large. In any case ${}_jN$ or ${}_jM$ will often be unknown (as are, for example, the number of tenant-operators in Iowa) and in these circumstances we are forced to estimate *both* the unknown numerator ${}_jY$ and unknown denominator ${}_jN$ (or ${}_jM$). This direct approach automatically leads to what is known as the "combined ratio estimator" of ${}_j\bar{Y}$. Other ratio estimators, available in special situations, are briefly discussed in section 6. In order to estimate, then, the denominator ${}_jN$ (or ${}_jM$) by precisely the same method used for the numerator ${}_jY$ we introduce the "count variates"

$${}_ju_i = \begin{cases} 1 \text{ if } i^{th} \text{ unit is in } j^{th} \text{ domain} \\ 0 \text{ otherwise} \end{cases} \qquad [26]$$

and these obtain the estimate of the domain size ${}_jN$ (or ${}_jM$) as

$${}_j\hat{U} = \hat{Y}\ ({}_ju_i)\,. \qquad [27]$$

Provided that ${}_j\hat{U} > 0$, that is provided there is at least one sampled unit in the j^{th} domain, we can then use

(a) the combined ratio estimator of the domain mean ${}_j\bar{Y}$

$${}_j\hat{y} = {}_j\hat{Y}/{}_j\hat{U} \qquad [28]$$

In general this estimator will be biased and conditions as to when this bias is negligible are known from the literature (see e.g. Cochran, 1953, pp. 114-8). Unbiased ratio estimators are briefly discussed in section 6. Standard approximate formulas (see e.g. Cochran, 1953, pp. 114-120) for the variance and estimated variance of the ratio estimator likewise yield

(b) The approximate variance formula for ${}_j\hat{y}$

$$\text{Var}\ ({}_j\hat{y}) \doteq {}_jN^{-2}\ V\ ({}_jy_i - {}_j\bar{Y}\ {}_ju_i) \qquad [29]$$

(c) The approximate estimate of the variance of ${}_j\hat{y}$

$$\text{var}\ ({}_j\hat{y}) = {}_j\hat{U}^{-2}\ v\ ({}_jy_i - {}_j\hat{y}\ {}_ju_i) \qquad [30]$$

The spelling out of these formulas [28], [29] and [30] will show that the domain mean ${}_j\bar{Y}$ is usually estimated by the simple sample domain mean ${}_j\bar{y}$. This will certainly be so in Examples 1 and 3 as indeed with all self-weighting survey designs. Likewise the estimated variance of ${}_j\hat{y}$ will often simplify:

In Example 2 we have the following formulas

$$(a) \qquad {}_j\hat{y} = \sum_h \frac{N_h}{n_h}\, {}_jy_h \Big/ \sum_h \frac{N_h}{n_h}\, {}_jn_h \qquad [31]$$

which, in the case of proportional allocation $N_h/n_h = N/n$ will reduce to

$$(a) \qquad {}_j\hat{y} = {}_j\bar{y}\,. \qquad [32]$$

The estimated variance in this example from [30] and [23] is given by

$$\text{var}\ ({}_j\hat{y}) = {}_j\hat{U}^{-2} \sum_{h=1}^{L} \frac{N_h\,(N_h - n_h)}{n_h\,(n_h - 1)} \left\{ {}_j\Sigma_h + {}_jq_h\, {}_jn_h\, ({}_j\bar{y}_h - {}_j\hat{y})^2 \right\} \qquad [33]$$

where ${}_j\Sigma_h = ({}_jn_h - 1)\,{}_js_h^2$ is the sum of squares of deviations of the y in the j^{th} domain of the h_{th} stratum about their mean $_j\bar{y}_h$ and $_jq_h = (n_h - {_jn_h})/n_h$ the proportion of units sampled outside the j^{th} domain. Formula [33] agrees with Yates' (1953, p. 301 formula [9.3.c]) except that the latter appears to be restricted to a proportional allocation of the sample to strata. The formula for Example 1 (simple sampling) is obtained as a special case of [33] when $L = 1$ and $_j\bar{y}_h = {_j\bar{y}}$ yielding

$$\text{var}\ {_j\hat{y}} = \left(1 - \frac{n}{N}\right)\left(\frac{n}{n-1}\right)\left(\frac{_jn - 1}{_jn}\right){_js^2}/{_jn} \qquad [34]$$

which is, approximately, equal to the familiar variance of a mean of a sample of size $_jn$.

It should be noted that the case of stratified sampling is not derivable from the simple sampling case by summation over strata since this would only yield the first terms inside the }{ of [33]. The second terms allow for the fact that the strata proportions $_jN_h/_jN$ for the j^{th} domain are not known.

A numerical example:

We illustrate the above formulas by evaluating $_j\hat{y}$ and var $(_j\hat{y})$ for Example 2 using the data of Table 2 to evaluate an estimate of the mean wheat acreage for District $j = 1$.

The N_h and n_h are given in the second and last (top entry) columns of Table 2 and the $_1y_h$ and $_1m_h$ in the third column. For the individual wheat acreage values $_1y_{hi}$ of the farms in the first district reference is made to Yates (1949, p. 154). We compute

$$_1\hat{Y} = 3291\,,\quad _1\hat{U} = 183\,,\quad _1\hat{y} = 17.98\quad ,\quad \text{var}\ (_1\hat{y}) = 61.52\,.$$

Finally the spelling out of formulas [30] and [28] for Example 3 yields

$$\text{var}\ (_j\hat{y}) \doteq \frac{n}{n-1}\sum_{i=1}^{n}\left(\frac{_jm_i}{_jm}\right)^2(_j\bar{y}_i - {_j\hat{y}})^2 \qquad [35]$$

where $_j\hat{y} = {_j\bar{y}}$.

4. *Correlation between the estimates of domain totals and means and variances of differences between means*

In the preceding sections we derived formulas for the estimation of domain totals and means and their variances. We now turn to the task of the "comparison" of two estimated domain means, and to this end require estimates of their covariances in order to obtain variance estimates for the *difference of two* estimated domain means.

Again, we first deal with the estimates of domain-totals and afterwards with the domain means. It is convenient to write all formulas for the comparison of two particular domains which, without loss of generality, may be taken as domains $j = 1$ and $j = 2$.

The following formulas are obvious by arguments similar to those of section 2.

$$\text{Var}\ (_1\hat{Y} - {_2\hat{Y}}) = V\,(_1y_i - {_2y_i}) \qquad [36]$$

$$2\ \text{Cov}\ (_1\hat{Y}\,, {_2\hat{Y}}) = V\,(_1y_i) + V\,(_2y_i) - V\,(_1y_i - {_2y_i}) \qquad [37]$$

$$\text{var}\ (_1\hat{Y} - {_2\hat{Y}}) = \text{v}\,(_1y_i - {_2y_i}) \qquad [38]$$

$$2\ \text{cov}\ (_1\hat{Y}\,, {_2\hat{Y}}) = \text{v}\,(_1y_i) + \text{v}\,(_2y_i) - \text{v}\,(_1y_i - {_2y_i}) \qquad [39]$$

Formula [38] gives directly the expression required for estimating the variance of the difference between two estimates $_1\hat{Y}$ and $_2\hat{Y}$ of the two domain totals. Formula [39], however, is required for proving that cov $(_1\hat{Y}, {_2\hat{Y}})$ is given by the bilinear form in the variates $_1y_i$ and $_2y_i$ which corresponds to the symmetric quadratic form v (y_i). The same remarks apply to the variance formulas [37] and [36].

Turning now to the domain means we require an approximate expression for the covariance between two ratios y/u and x/v: This follows on identical lines as the familiar variance formulas and yields:

$$\text{Cov}\left(\frac{y}{u}\,, \frac{x}{v}\right) \doteq (E\,(u)\,E\,(v))^{-1}\,\text{Cov}\left(y - \frac{E\,(y)}{E\,(u)}u\,, x - \frac{E\,(x)}{E\,(v)}v\right) \quad [40]$$

The combination of [36] and [39] can be shown to yield

$$\begin{aligned}\text{Var}\ (_1\hat{y} - {_2\hat{y}}) &\equiv \text{Var}\left(\frac{_1\hat{Y}}{_1\hat{U}} - \frac{_2\hat{Y}}{_2\hat{U}}\right)\\ &\doteq V\left(\frac{1}{_1N}(_1y_i - {_1\bar{Y}}\,{_1u_i}) - \frac{1}{_2N}(_2y_i - {_2\bar{Y}}\,{_2u_i})\right)\end{aligned} \qquad [41]$$

The corresponding approximation for the estimated variance yields

$$\text{var}\ (_1\hat{y} - {_2\hat{y}}) \doteq \text{v}\left(\frac{1}{_1\hat{U}}(_1y_i - {_1\hat{y}}\,{_1u_i}) - \frac{1}{_2\hat{U}}(_2y_i - {_2\hat{y}}\,{_2u_i})\right) \quad [42]$$

where the variables $_jy_i$ and $_ju_i$ are given by [14] and [26] respectively, the estimates $_j\hat{U}$ and $_j\hat{y}$ by [27] and [28] and the multivariable function v is defined in [3]. This function v will of course depend on the particular survey design but there is no difficulty in spelling it out in any particular case. We proceed to do so for our examples 1, 2 and 3 when v (y_i) is given by equations (6) and (11) respectively. For example 2, simple stratified sampling with varying sampling fractions, we obtain the formula

$$\begin{aligned}\text{v}\,(_1\hat{y} - {_2\hat{y}}) = \sum_h a_h \Bigg\{ &\frac{_1\Sigma_h}{_1\hat{U}^2} + \frac{_2\Sigma_h}{_2\hat{U}^2} + 2\,n_h\,{_1p_h}\,{_2p_h}\frac{(_1\bar{y}_h - {_1\hat{y}})}{_1\hat{U}}\frac{(_2\bar{y}_h - {_2\hat{y}})}{_2\hat{U}} + \\ &+ n_h\,{_1p_h}\,{_1q_h}\,(_1\bar{y}_h - {_1\hat{y}})^2 + n_h\,{_2p_h}\,{_2q_h}\frac{(_2\bar{y}_h - {_2\hat{y}})^2}{_2\hat{U}^2}\Bigg\}\end{aligned} \qquad [43]$$

where

$$a_h = \frac{N_h\,(N_h - n_h)}{n_h\,(n_h - 1)}\,,\quad _jp_h = \frac{_jn_h}{n_h}\,,\quad _jq_h = 1 - {_jp_h}$$

$$_j\Sigma_h = \sum_{i=1}^{_jn_h}(_jy_{hi} - {_j\bar{y}_h})^2$$

Formula [43] is in agreement with that for $\hat{\text{cov}}\ (_1\hat{y}\,, {_2\hat{y}})$ given by Yates (1953, p. 301, formula 9.3.d) except that the latter appears to be restricted to stratified sampling with proportional allocation. The first two terms in [43] are the within domain within strata components, whilst the last three terms have the form of multinomial variances and covariances and allow for the fact that the strata proportions $_1N_h/_1N$ and $_2N_h/_2N$ are unknown. These latter terms disappear in the special case of a single stratum $L = 1$ which yields the answer for example 1 in the form

$$V\,(_1\hat{y} - {_2\hat{y}}) = \left(1 - \frac{n}{N}\right)\left(\frac{n}{n-1}\right)\left\{\frac{_1n - 1}{_1n}\frac{_1s^2}{_1n} + \frac{_2n - 1}{_2n}\frac{_2s^2}{_2n}\right\} \quad [44]$$

This expression is approximately equal to $(_1s^2/_1n + {_2s^2}/_1n)$ i.e. the familiar variance formula for the difference of two means based on fixed samples of size $_1n$ and $_2n$.

Finally, the spelling out of formula [42] for Example 3 (two stage sampling primaries drawn p.p.s. equal take of secondaries) yields the surprisingly simple formula

$$\text{v}\,(_1\hat{y} - {_2\hat{y}}) \doteq \frac{n}{n-1}\sum_{i=1}^{n}\left\{\frac{_1m_i}{_1m}(_1\bar{y}_i - {_1\hat{y}}) - \frac{_2m_i}{_2m}(_2\bar{y}_i - {_2\hat{y}})\right\}^2 \quad [45]$$

In the important case of $n = 2$ primaries this formula simplifies further to

$$v\,({}_1y - {}_2y) = 4\,({}_1w\,({}_1\bar{y}_1 - {}_1\bar{y}_2) - {}_2w\,({}_2\bar{y}_1 - {}_2\bar{y}_2))^2 \qquad [46]$$

where

$${}_jw = {}_jm_1\,{}_jm_2/({}_jm_1 + {}_jm_2)^2$$

Certain special cases of formulas [44] and [45] were recently obtained by an independent derivation by L. Kish and Irene Hess (1955).

5. *The domain total expressed as a proportion of the population total.*

In many analytic studies of sample survey data we are interested in the proportion of a measured characteristic which falls into a specified domain. For example in a consumer study we may be interested in the proportion of say the total milk consumption which is attributable to families of a particular income group. Or, again, in a soil survey we may be interested in the proportion of the total farm land which is of a particular soil type. In the preceding sections we have estimated the total ${}_jY$ of the variate values y_i which fall into the j^{th} domain by the estimator ${}_j\hat{Y}$. It is therefore suggested that the proportion ${}_jY/Y$ be estimated by the ratio

$${}_jp = {}_j\hat{Y}/\hat{Y} \qquad [47]$$

The properties of this ratio estimator can be evaluated by a method similar to that used in Section 3: We attach two variates to each unit in the population viz
the numerator variable given (by [14] i.e.) by

$${}_jy_i = \begin{cases} y_i \text{ if the } i^{th} \text{ unit belongs to the } j^{th} \text{ group} \\ 0 \text{ otherwise} \end{cases} \qquad [48]$$

and the denominator variable given by y_i. Since ${}_j\hat{Y}$ and $\hat{Y}$ are, respectively, the estimates of the population totals for the variates ${}_jy_i$ and y_i the estimator ${}_jp$ is seen to be the standard (combined) ratio estimator for ${}_jP = {}_jY/Y$ of the respective population totals. Its variance may therefore be obtained by the standard (approximate) formula viz.

$$\text{Var}\,{}_jp = Y^{-2}\,V\,({}_jy_i - {}_jP\,y_i) \qquad [49]$$

where the variance function $V(y)$ is defined by [2]. Let us spell out the general formula [49] for the particular case of a simple stratified design and in terms of the notation used in Table 2a and that described immediately following Table 2a. We obtain the formula

$$\text{Var}\,{}_jp = Y^{-2}\sum \frac{N_h^2}{n_h\,(N_h - 1)} \sum_{i=1}^{N_h} (d_{hi} - \bar{D}_h)^2 \qquad [50]$$

where $d_{hi} = {}_jy_{hi} - {}_jP\,y_{hi}$ and $\bar{D}_h$ is the stratum mean of the d_{hi}.

Yates (1953) p. 304 only discusses this case of a stratified sample and offers a formula only for the special case when the domains do not cut across the strata. In the general case of domains cutting across the strata Yates gives hints for variance computation and we have not attempted to identify the result of these instructions with our simple formula [50].

For the purpose of variance comparison, an alternative form of [50] may be more useful. After some algebra we reach the formula

$$\text{Var}\,({}_jp) = Y^{-2}\sum_h \frac{N_h^2}{n_h\,(N_h - 1)} \left\{ {}_jQ^2\,{}_j\Sigma_h + {}_jP^2\,{}_J\Sigma_h + N_h^{-1}\,{}_jN_h\,{}_JN_h\,({}_jQ\,{}_j\bar{Y}_h + {}_jP\,{}_J\bar{Y}_h)^2 \right\} \qquad [51]$$

where

${}_jN_h$ is the number of units in the h^{th} stratum falling within the j^{th} domain

${}_JN_h$ is the number of units in the h^{th} stratum not falling within the j^{th} domain

${}_jY_h$ is the y-total for the h^{th} stratum of units in the j^{th} domain

${}_JY_h$ is the y-total for the h^{th} stratum of units not in the j^{th} domain.

So that

$${}_j\bar{Y}_h = {}_jY_h/{}_jN_h \qquad {}_J\bar{Y}_h = {}_JY_h/{}_JN_h$$

are the corresponding means.

Further

$${}_jP = \sum_h {}_jY_h/Y \qquad {}_jQ = \sum_h {}_JY_h/Y = 1 - {}_jP$$

$${}_j\Sigma_h = \sum_{i=1}^{{}_jN_h} ({}_jy_{hi} - {}_j\bar{Y}_h)^2$$

is the sum of squares of deviations of the y values in the h^{th} stratum and j^{th} domain from their mean and

$${}_J\Sigma_h = \sum_{i=1}^{{}_JN_h} ({}_jy_{hi} - {}_J\bar{Y}_h)^2$$

is the corresponding sum of squares for the y-values in the h^{th} stratum and *not* in the j^{th} domain.

To interpret the three terms in [51] we may write the estimator in the form

$$1/{}_jp = 1 + ({}_J\hat{Y}/{}_j\hat{Y}) \qquad [52]$$

where ${}_J\hat{Y} = \hat{Y} - {}_j\hat{Y}$ is an estimate for the y-total *not* in the j^{th} domain. It can be shown that the first two terms of [51] are contributed by variation in y-values only, holding the ratios ${}_jn_h/n_h$ constant at their expected values $\frac{{}_jN_h}{N_h}\,n_h$. The third term allows for the variability in the proportion ${}_jn_h/n_h$ and is seen to be a binomial type of variance being approximately equal to

$$\sum_h \left(\frac{{}_jN_h}{N_h}\right)\left(\frac{{}_JN_h}{N_h}\right)\frac{1}{n_h}\,Y^{-2}\,N_h^2\,({}_J\bar{Y}_h\,{}_jP + {}_j\bar{Y}_h\,{}_jQ)^2\,. \qquad [53]$$

In this equation [53] the term

$$\left(\frac{{}_jN_h}{N_h}\right)\left(\frac{{}_JN_h}{N_h}\right)\frac{1}{n_h}$$

is the binomial variance of ${}_jn_h/n_h$ and the term inside the brackets $\{\,\}$, is the square of the expected value of the coefficient of ${}_jn_h/n_h$ in the expansion of the estimator ${}_j\hat{Y}/\hat{Y}$.

For the computation of an estimate of variance we, again, rely on standard formulas for ratio estimators. The standard estimate (not necessarily unbiased) of the variance given by [51] is computed from

$$\hat{\text{var}}\,({}_jp) = \hat{Y}^{-2}\,v\,({}_jy_i - {}_jp\,y_i) \qquad [54]$$

where the sample variance function is given by [3].

If we spell out this general formula in the particular case of a stratified sample we obtain

$$\text{var}\,{}_jp = \hat{Y}^{-2}\sum_h \frac{N_h^2}{n_h\,(n_h - 1)} \sum_{i=1}^{n_h} (d_{hi} - \bar{d}_h)^2 \qquad [55]$$

where $d_{hi} = {}_jy_{hi} - {}_jp\,y_{hi}$ and $\bar{d}_h$ is their stratum sample means and $\hat{Y}$ is the estimate of the population total i.e. $\hat{Y} = \sum_h N_h\,\bar{y}_h$ and ${}_jp = {}_j\hat{Y}/\hat{Y}$ the estimate of the proportion. An alternative formula which does not require the computation of the d_{hi} can be obtained, after some algebra as follows.

$$\text{var}\ {}_j p = \hat{Y}^{-2} \sum_h \frac{N_h^2}{n_h(n_h - 1)} \{ {}_j q^2\, {}_j T_h + {}_j p^2\, {}_{\bar{j}} T_h + n_h^{-1}\, {}_j n_h\, {}_{\bar{j}} n_h ({}_j q\, {}_j \bar{y}_h + {}_j p\, {}_{\bar{j}} \bar{y}_h)^2 \} . \quad [56]$$

Here

${}_j n_h$ is the number of sampled units in the h^{th} stratum falling within the j^{th} domain.

${}_{\bar{j}} n_h$ is the number of sampled units in the h^{th} stratum not falling within the j^{th} domain.

${}_j y_h$ is the y-total for the h^{th} stratum of units in the j^{th} domain.

${}_{\bar{j}} y_h$ is the y-total for the h^{th} stratum of units not in the j^{th} domain.

So that

$${}_j \bar{y}_h = {}_j y_h / {}_j n_h \qquad {}_{\bar{j}} \bar{y}_h = {}_{\bar{j}} y_h / {}_{\bar{j}} n_h$$

are the corresponding means. Further

$${}_j p = {}_j \hat{Y} / \hat{Y} \quad , \quad {}_j q = 1 - {}_j p$$

$${}_j T_h = \sum_{i=1}^{{}_j n_h} ({}_j y_{hi} - {}_j \bar{y}_h)^2$$

is the sum of squares of deviations of the sampled y-values in the h^{th} stratum and j^{th} domain from their mean and

$${}_{\bar{j}} T_h = \sum_{i=1}^{{}_{\bar{j}} n_h} ({}_{\bar{j}} y_{hi} - {}_{\bar{j}} \bar{y}_h)^2$$

is the corresponding sum of squares for the y-values in the h^{th} stratum and not in the j^{th} domain.

6. *Domain means adjusted for concomitant variables.*

We now turn to the analogue to « analysis of variance and covariance » for domain means. First let us briefly recall the essentials of "analysis of covariance" as it is practiced with the adjustment of (say) two group means from experimental designs. It is usually assumed that:

(*a*) The true (or population) means for the concomitant variable x are identical for the two groups, so that any differences in observed x-group means are due to sampling. In popular language the failure of this condition to be satisfied is sometimes described by the warning that " Analysis of Covariance should not be misused to correct away real treatment differences in the x-means."

(*b*) The true (or population) regression lines for the two groups are parallel.

Condition (*b*) is, of course, not essential, but in the situations in which it is satisfied (or approximately satisfied) it is almost universally invoked and simplifies the analysis. If condition (*a*) is not satisfied a meaningful generalization of analysis of covariance can still be employed if the two population group x-means ${}_j\bar{X}$ are known, and estimates of the y-group means can be obtained as the ordinates of the respective estimated group regression lines evaluated at the respective abscissa values ${}_j\bar{X}$. Now with situations as are usually encountered in " analytic studies " of sample survey data one will hardly ever be able to assume identity (or even approximate identity) of the x-domain means ${}_j\bar{X}$; on the other hand, there are situations when these means are known and we shall therefore here attempt to develop an analogue to analysis of covariance in this case.

We therefore deal with the following situation: A sample survey provides paired data $y_i\, x_i$ for a sample of n units sampled from a population of N units. After sampling the n units are classified into k domains $j = 1, 2, \ldots, k$ for which the population means ${}_j\bar{X}$ are known. It is now required to estimate the (unknown) domain y-means ${}_j\bar{Y}$ utilizing the ${}_j\bar{X}$.

The question then arises as to which estimator of ${}_j\bar{Y}$ should be used. The regression theory in classical analysis of covariance arises as the maximum likelihood estimation and results from the assumption of a linear model. Although the validity of such a model, even for very large finite populations, will often be in doubt, the use of regression estimation may still result in a gain of precision. Nevertheless we shall here *not* employ regression estimators. The reason for this is *not* that we consider regression theory inappropriate, but that this theory for finite populations requires considerable development before it can be applied in the present situation. On the other hand, ratio estimators are easily adapted to the estimation of domain means but in using these we should stress their well known limitation, namely, that they are likely to be effective only if the y and/or x scales can be so chosen that the population regression will intersect near the origin.

If a combined ratio estimator is used to estimate ${}_j\bar{Y}$ with the help of the known x-domain mean ${}_j\bar{X}$ the theory is almost identical with that developed for the ratio estimation of Section 3, the only essential difference is that the concomitant variable x will now take the place of the "count variable" ${}_j u_i$ of [26]. Accordingly we introduce the variate

$${}_j x_i = \begin{cases} x_i \text{ if } i^{th} \text{ unit is in } j^{th} \text{ domain} \\ 0 \quad \text{otherwise} \end{cases} \quad [57]$$

and obtain the ratio estimator of ${}_j\bar{Y}$ in the form:

(*a*) The combined ratio estimator of the domain mean ${}_j\bar{Y}$:

$${}_j\breve{y} = {}_j\bar{X}\, ({}_j\hat{Y} / {}_j\hat{X}) \quad [58]$$

where the ${}_j\bar{X}$ are the fixed and known domain means of x, ${}_j\hat{Y}$ is defined by [15] and ${}_j\hat{X}$ by analogy.

Employing standard results from the ratio estimator theory on similar lines as in section 3 we further obtain

(*b*) The approximate variance formula of ${}_j\breve{y}$

$$\text{Var}\,({}_j\breve{y}) = {}_j N^{-2}\, V\, ({}_j y_i - {}_j\bar{Y}\, {}_j x_i / {}_j\bar{X}) \quad [59]$$

(*c*) The approximate estimate of the variance of ${}_j\breve{y}$

$$\text{var}\,({}_j\breve{y}) = {}_j\hat{U}^{-2}\, \text{v}\, ({}_j y_i - {}_j\hat{Y}\, {}_j x_i / {}_j\hat{X}) \quad (60)$$

where the multivariable functions V and v are defined by [2] and [3] the variates ${}_j y_i$ and ${}_j x_i$ by [14] and [57], ${}_j\hat{Y}$ and ${}_j\hat{U}$ by [15] and [27] and ${}_j\hat{X}$ by analogy.

Formulas [59] and [60] show the familiar result that the variance of ${}_j\breve{y}$ depends on the residuals ${}_j y_i - ({}_j\bar{Y}/{}_j\bar{X})\, {}_j x_i$ of a " regression " with slope ${}_j\bar{Y}/{}_j\bar{X}$ passing through the origin. For the majority of units in the population ${}_j y_i = {}_j x_i = 0$ and no contribution is made to the residuals; for the units in the j^{th} domain the above residuals will be small only if the y, x data of this domain alone satisfy the usual conditions required for the effectiveness of ratio estimators, namely that the y, x correlation should be high and that the y, x regression should intersect near the origin. If the latter condition is not satisfied little is gained by using regression estimators. For if the y, x data in the j^{th} domain have a regression which does not intersect near the origin then the addition of the large number of units (outside the j^{th} domain) for which ${}_jy_i = {}_jx_i = 0$ would generate large residuals from the best regression fitted to all N pairs of ${}_jy_i$, ${}_jx_i$.

The spelling out of formula (60) follows on the same lines as in section 3. For example 1 (random sampling) we find in analogy to (34)

$$\text{var}\,(\widehat{{}_j\bar{y}}) \doteq \left(1 - \frac{n}{N}\right) \frac{n}{{}_jn^2\,(n-1)} \sum_{i=1}^{{}_jn} \left({}_jy_i - \frac{{}_jy}{{}_jx}\,{}_jx_i\right)^2 \qquad [61]$$

In a similar manner the formulas for the variance of a difference of two ratio estimators can be obtained following on the lines of the arguments in section 4. We reach the general formula

$$\text{var}\,(\widehat{{}_1\bar{y}} - \widehat{{}_2\bar{y}}) \doteq \text{v}\left(\frac{1}{{}_1\hat{U}}\left({}_1y_i - \frac{{}_1\hat{Y}}{{}_1\hat{X}}\,{}_1x_i\right) - \frac{1}{{}_2\hat{U}}\left({}_2y_i - \frac{{}_2\hat{Y}}{{}_2\hat{X}}\,{}_2x_i\right)\right) \quad [62]$$

which, in the case of example 1 (simple sampling) spells out as

$$\text{var}\,(\widehat{{}_1\bar{y}} - \widehat{{}_2\bar{y}}) \doteq \left(1 - \frac{n}{N}\right)\left(\frac{n}{n-1}\right)\left\{\frac{{}_1\Sigma}{{}_1n^2} + \frac{{}_2\Sigma}{{}_2n^2}\right\} \qquad [63]$$

where

$${}_j\Sigma = \sum_{i=1}^{{}_jn}\left({}_jy_i - \frac{{}_jy}{{}_jx}\,{}_jx_i\right)^2$$

7. *Unbiased ratio estimators for domain means and exact variance formulas.*

In sections 3-6 we have freely used analogies to the well known combined ratio estimators and their approximate variance formulas. For a discussion of the magnitude of the bias and the precision of the approximate variance formulas we must refer to the literature. Since there may well be cases in which these approximations are too inaccurate we wish to put here on record a method which avoids these disadvantages under certain circumstances.

The situations in which our exact formulas will apply are characterized by the following conditions:

(*a*) The number of units in the j^{th} domain, ${}_jN$, must be known.

(*b*) Primary sampling units (or, if the design is single stage, the sampling units) must have been drawn *with* replacements.

Two remarks concerning these conditions may be pertinent: Concerning (*b*), although there are many surveys in which primaries are actually drawn with replacement, this condition becomes of little importance when the sampling fraction of primaries is small as then there is practically no difference between drawing with — and drawing without — replacement. Small primary sampling fractions are very common in survey designs.

Concerning the very important condition (*a*), if the ${}_jN$ are known it is of course possible to obtain unbiased estimates of ${}_j\bar{Y}$ by dividing the unbiased estimates of the ${}_jY$ (developed in the preceding sections) by the known ${}_jN$. It is, however, well known from experience with such estimates that for most survey designs and populations they would have large variances. Confirmation of this may be sought by comparing for the examples quoted in the preceding sections ${}_jN^{-2}$ Var ${}_j\hat{Y}$ with Var $(\widehat{{}_j\bar{y}})$. Roughly speaking the former depends on the variation of ${}_jy$ — totals whilst the latter depends on the variation of ${}_j\bar{y}$ — means.

Turning, therefore, again to the device of ratio estimation we will describe first the unbiased ratio estimators for population means translating them into estimators of domain means later.

Let us consider a survey design consisting of L strata ($h = 1, 2, \ldots, L$) with population proportions P_h and let us assume that n_h primaries $t = 1, 2, \ldots, n_h$ have been drawn with replacement from the h^{th} stratum. Denote by y_{hti} a character + attached to the i^{th} unit in the t^{th} primary of the h^{th} stratum and by $\hat{y}_{ht}$ the unbiased estimate of $\bar{Y}_h$ from the t^{th} primary only. The functional form of $\hat{y}_{ht}$ will depend on the design details. Denote by

$$\bar{y}_h = n_h^{-1} \sum_{t=1}^{n_h} \hat{y}_{ht} \qquad [54]$$

and by

$$\bar{y}_{st} = \sum_{h=1}^{L} P_h\,\bar{y}_h \qquad [55]$$

the stratum means of the $\hat{y}_{ht}$ and the unbiased estimator of $\bar{Y}$.

Denote further by $\hat{x}_{ht}$, $\bar{x}_h$ and $\bar{x}_{st}$ the corresponding means for a second characteristic x for which the population mean $\bar{X}$ is known and for which we assume that either $x_{hti} \geq 0$ or $x_{hti} = y_{hti} = 0$.

Introduce now the ratio variates

$$\hat{r}_{ht} = \begin{cases} \hat{y}_{ht}/\hat{x}_{ht} & \text{if} \quad \hat{x}_{ht} > 0 \\ r^* & \text{if} \quad \hat{x}_{ht} = \hat{y}_{ht} = 0 \end{cases} \qquad [56]$$

where r^* is a constant conveniently chosen as discussed below. Finally introduce for these ratios the means $\bar{r}_h$ and $\bar{r}_{st}$ as above. The survey design can now be thought of as generating in each stratum an infinite joint population of $\hat{y}_{ht}$, $\hat{x}_{ht}$, $\hat{r}_{ht}$ by the infinitely repeatable process of

(*a*) drawing a single primary in accordance with the design,

(*b*) drawing from this primary the prescribed number of secondaries, tertiaries . . . in accordance with the design,

(*c*) computing $\hat{y}_{ht}$ $\hat{x}_{ht}$ $\hat{r}_{ht}$ for the drawn sample,

(*d*) replacing all units.

Following the lines of Hartley-Ross (1954) and Goodman-Hartley (1956) we can now construct the following estimators:

The unbiased ratio estimate of $\bar{Y}/\bar{X}$

$$r' = \bar{r}_{st} + \bar{X}^{-1}\,\hat{c} \qquad [67]$$

The unbiased estimate of $\bar{Y}$:

$$y' = \bar{X}\,\bar{r}_{st} + \hat{c} \qquad [68]$$

where

$$\hat{c} = (\bar{y}_{st} - \bar{x}_{st}\,\bar{r}_{st}) + \sum_h P_h^2\,(n_h - 1)^{-1}\,\{\bar{y}_h - \bar{x}_h\,\bar{r}_h\} \qquad [69]$$

is an unbiased estimate, computed from the stratified sample, of the population covariance of $\hat{x}_{ht}$ and $\hat{r}_{ht}$ defined by

$$C = \sum_h P_h\,E\,(\hat{x}_{ht} - \bar{X})\,(\hat{r}_{ht} - \bar{R}) \qquad [70]$$

in which E refers to the infinite populations described above.

The estimates r' and y' as given by [67] and [68] are unbiased no matter what value was chosen for the constant r^*. The choice of r^* will, however, affect the variance of these estimates. In the practical situations in which these estimators will be applied the chance of $\hat{x}_{ht} = \hat{y}_{ht} = 0$ to arise will be very small and hence any uncertainty in the choice of r^* will have a small effect. The best choice is to take r^* equal to the best estimate

+ The notation applies directly to two stage designs, the index i denoting the secondary unit. In three or higher stage designs i must be replaced by a multiple subscript.

of $\bar{R}$ available in advance of sampling and must not be altered subsequently.

In order to apply these results to the estimation of the j^{th} domain mean we must use for the numerator variable

$$_jy_{hti} = \begin{cases} y_{hti} \text{ if the } hti \text{ unit is in the } j^{th} \text{ domain} \\ 0 \text{ otherwise} \end{cases}$$

and for the denominator the count variable

$$_ju_{hti} = \begin{cases} 1 \text{ if the } hti \text{ unit is in the } j^{th} \text{ domain} \\ 0 \text{ otherwise} \end{cases}$$

and substitute these in [66], [67] and [69]. It will be seen that for all selfweighting designs the leading term of [57] $\bar{r}_{st}$ will be of the form

$$\bar{r}_{st} = \sum_h P_h n_h^{-1} \sum_t {_j\bar{y}_{ht}}$$

where $_j\bar{y}_{ht}$ is defined, as before, as the y- mean of the units (if any) in the j^{th} domain of the t^{th} primary of the h^{th} stratum and by r^* if there are no such units. The condition mentioned above, that the chance of $_j\hat{y}_{ht} = {_j\hat{u}_{ht}} = 0$ should be small means that the present method of estimation should only be used if there is a reasonable chance that all domains are represented in the sample from each primary.

The unbiased estimation of $_j\bar{Y}$ utilizing the knowledge of the domain mean $_j\bar{X}$ of a concomitant variable will be based on [68] when it should be noted that y' will estimate $_jN_j\bar{Y}/N$, that $\bar{X}$ will be $_jN\,_j\bar{X}/N$ so that the estimator of $_j\bar{Y}$ will be

$$\frac{N}{_jN} y' = {_j\bar{X}}\, \bar{r}_{st} + \frac{N}{_jN} \hat{c}$$

where $\hat{c}$ is computed from [69] using the variates $_jy_{hti}$, $_jx_{hti}$ and the ratios $_j\hat{r}_{ht}$. In this case, therefore, both $_j\bar{X}$ and $_jN$ must be known.

The exact variance formulas for both r' and y' have recently been derived by Goodman-Hartley in the special case of simple sampling from one stratum ($L = 1$). Their results indicate that the variance of y' is of a similar order of magnitude (sometimes larger, sometimes smaller) than that of the combined ratio estimator. For the stratified case results are not as yet available.

8. *Inferences based on the variance formulas.*

Two questions concerning inferences arise:

The variance formulas of the preceding sections are based on finite population sampling and, in particular, have the property to become zero when sampling is 100 %. The obvious logic of this property is that when all units of the population are sampled all means of all domains in such a population are known without error so that all non-zero contrasts between domain means are "significant". It must not be forgotten, however, that such inferences can only apply to the particular finite population under investigation and are not of any wider significance. For instance, in the example of Table 3, had the population of Des Moines been sampled 100 % definite statements about differences in the number of persons / household of different Income Groups could have been made "without error" but such inferences would then only apply to the City of Des Moines at the time of the survey. If data from such a survey are to be used for (say) comparing "Income Groups" in "Cities like Des Moines" in "times like the present" in a more general sense, the finite population sampling theory must not be used. Whether the survey sample can be (artificially) regarded as one drawn from a wider populaton (in space or time) is a matter requiring special investigation and the onus of such an investigation falls upon those wishing to draw such wide inferences.

A second question concerns inferences to be drawn for the domains of the particular finite population for which the survey was planned:

In the preceding sections we have given estimates of domain means, their variances and estimates of these variances. We have also given variance formulas and estimates for the differences between estimated domain means. The question arises as to how these variance estimates are to be used for inferences to be drawn from the data. Even if it is not intended to carry out "tests of significance" there would remain the question of the computation of confidence intervals for the domain means and their contrasts. This question is, of course, not confined to the present issues but is an ever present difficulty in the theory of sample surveys. The customary procedure here is to employ normal theory approximations and to appeal to the respective central limit theorems for finite populations and to the relatively large sample sizes which are available in these situations. The results which are available on these problems are very restricted and require development and it is clearly not possible in this context to dwell upon this issue of much wider impact.

References

COCHRAN, W. G., (1953). — *Sampling Techniques.* John Wiley and Sons, Inc. New York.

GOODMAN, L., HARTLEY, H. O., (1956). (Unpublished manuscript).

HARTLEY, H. O., ROSS, A. — *Unbiased Ratio Estimators,* 1954. Nature, 174, 270.

KISH, L. and HESS, IRENE, (1955). — *On Variances of Ratios and Their Differences in Multistage Sampling.* (Paper presented at the New York Meeting of A.S.A. Dec. 1955).

STEPHAN, F. F., (1941). — *Stratification in Representative Sampling.* J. of Marketing (1941), 38.

YATES, F., 1949 (2nd edition 1953). — *Sampling Methods for Censuses.* Surveys, London, Charles Griffin & Co.

ACKNOWLEDGMENT

This paper has appeared as a contribution to a volume published by the Institute of Statistics in honour of Professor Corrado Gini, upon the occasion of his retirement from the Chair of Statistics at the University of Rome. It is reproduced here through the kind permission of the Institute of Statistics and of its Director, Professor Vittorio Castellano. Professor Hartley presented a summary of this paper at the Annual Meeting.

UNBIASED COMPONENT-WISE RATIO ESTIMATION[1]

By: D. S. Robson and Chitra Vithayasai, Cornell University

INTRODUCTION

The precision of a ratio-type estimator such as $\bar{y}\bar{X}/\bar{x}$ can sometimes be substantially increased if the correlated variables y and x can be expressed as sums of more highly correlated components, $y=y_1+\cdots+y_k$ and $x=x_1+\cdots+x_k$. An empirical example of this arises in the ratio estimation of total dry matter yield of corn in field plot experiments; when both the green weight x and oven-dry weight y are measured and estimated separately for ears and the vegetative parts of the plant the efficiency of estimation of plot total dry weight is increased by approximately 70%. An example from general sample survey methodology is the case of cluster sampling with unequal size clusters when the elements in each randomly selected cluster are stratified into k strata; the usual mean per cluster ratio estimate is then replaced by the sum of k such ratio estimates for the individual strata.

In this paper we are concerned primarily with the Hartley-Ross [1]type of unbiased component-wise ratio estimator, for which we present the exact variance formula and an unbiased estimator of the variance. The efficiency of component-wise ratio estimation is then examined empirically with the data from 39 corn plots of 10 hills each, and the bias of the conventional ratio estimate and its variance formula are evaluated.

THE VARIANCE OF THE HARTLEY-ROSS TYPE OF COMPONENT-WISE RATIO ESTIMATOR

The Hartley-Ross unbiased ratio estimator of the population total for a single component y takes the form

$$Y'=X\bar{r}+\frac{n(N-1)}{n-1}(\bar{y}-\overline{xr})$$

where $\bar{r}$ is the mean ratio of y to x in a random sample of size n from a population of size N, and X is the population total for x. Goodman and Hartley [2] gave the limiting form of the variance of this estimator as

$$\lim_{N\to\infty}\frac{1}{N^2}\operatorname{var}(Y')=[\frac{1}{n}\,\sigma_y^2+\bar{R}^2\sigma_x^2-2\bar{R}\sigma_{x,y}+\frac{1}{n-1}(\sigma_r^2\sigma_x^2-\sigma_{r,x}^2)]$$

where $\bar{R}$ is the population mean of the ratio $r=(y/x)$.

The exact variance for finite N was given by Robson [3] in terms of Tukey's [4] multivariate <u>polykays</u> and may be most conveniently expressed in the notation of Tukey's symbolic, dot-multiplication as

$$\operatorname{var}(Y')=\frac{N(N-n)}{n}[\sigma_y^2+\bar{R}\cdot\bar{R}\cdot\sigma_x^2-2\bar{R}\cdot\sigma_{x,y}+\frac{1}{n-1}(\frac{N-1}{N}\,\sigma_r^2\cdot\sigma_x^2+\frac{N-n}{N}\,\sigma_{r,x}\cdot\sigma_{r,x})]$$

All variances and covariances appearing in this formula are understood to be defined in the usual manner for finite populations; for example

$$\sigma_{r,x}=\frac{1}{N-1}\,(\sum_{i=1}^{N}r_ix_i-N\bar{R}\bar{X})$$

This definition, which arises naturally in the algebraic treatment of moments and cumulants of a finite population, also serves to illustrate what is meant by dot-multiplication, since

$$\sigma_{r,x}=\frac{1}{N}\sum_{i=1}^{N}r_ix_i-\frac{1}{N(N-1)}\sum_{i\neq j}^{N}r_ix_j$$
$$=\overline{RX}-\bar{R}\cdot\bar{X}$$

thus, the dot-product of two means is the mean of all possible crossproducts. The same is true for the dot-product of more than two means; for example,

$$\bar{R}\cdot\sigma_{x,y}=\bar{R}\cdot(\overline{XY}-\bar{X}\cdot\bar{Y})$$
$$=\bar{R}\cdot\overline{XY}-\bar{R}\cdot\bar{X}\cdot\bar{Y}$$
$$=\frac{1}{N(N-1)}\sum_{i\neq j}^{N}r_ix_jy_j$$
$$-\frac{1}{N(N-1)(N-2)}\sum_{i\neq j\neq k}^{N}r_ix_jy_k$$

and, similarly,

$$\bar{R}\cdot\bar{R}\cdot\sigma_x^2=\bar{R}\cdot\bar{R}\cdot(\overline{XX}-\bar{X}\cdot\bar{X})$$
$$=\bar{R}\cdot\bar{R}\cdot\overline{XX}-\bar{R}\cdot\bar{R}\cdot\bar{X}\cdot\bar{X}$$
$$=\frac{1}{N(N-1)(N-2)}\sum_{i\neq j\neq k}^{N}r_ir_jx_k^2$$
$$-\frac{1}{N(N-1)(N-2)(N-3)}\sum_{i\neq j\neq k\neq h}^{N}r_ir_jx_kx_h$$

As N gets large, of course, the dot-product of two or more moments approaches the ordinary product of the moments, provided the latter approach a limit, and so the limiting var(Y') of Goodman and Hartley is obtained.

A minimum variance unbiased estimator of var($\hat{Y}$) is easily constructed using the fact that polykays, or dot-products of sample cumulants, are minimum variance unbiased estimators of the corresponding polykays of the finite population. Thus, for example, the minimum variance unbiased estimator of $\bar{R}\cdot\sigma_{x,y}$ is

$$\bar{r}\cdot s_{x,y}=\bar{r}\cdot\overline{xy}-\bar{r}\cdot\bar{x}\cdot\bar{y}$$

$$=\frac{1}{n(n-1)}\sum_{i\neq j}^{n} r_i x_j y_j-\frac{1}{n(n-1)(n-2)}\sum_{i\neq j\neq k}^{n} r_i x_j y_k$$

or, expressed in more convenient computational form,

$$\bar{r}\cdot s_{x,y}=\frac{1}{n(n-1)(n-2)}\Big[(n-1)\sum_1^n x_i y_i \sum_1^n r_i - n\sum_1^n y_i^2+\sum_1^n x_i \sum_1^n y_i r_i$$

$$+(\sum_1^n y_i)^2+\sum_1^n x_i \sum_1^n y_i \sum_1^n r_i\Big]$$

The other components of var(Y') are similarly estimated; computing formulas for the estimates are given by Robson [5] and will follow as special cases of the more general formulas given next for the component-wise ratio estimator.

The general case we wish to consider is

$$Y'=\sum_{i=1}^{k} Y_i'=\sum_{i=1}^{k}\left[X_i\bar{r}_i+\frac{n(N-1)}{n-1}(\bar{y}_i-\bar{x}_i\bar{r}_i)\right]$$

where now

$$\mathrm{var}(Y')=\sum_{i=1}^{k}\mathrm{var}(Y_i')+2\sum_{i<j}^{k}\mathrm{cov}(Y_i',Y_j')$$

Since the individual terms $\mathrm{var}(Y_i')$ take the form indicated earlier for a single component estimator, the only new algebraic problem is the computation of $\mathrm{cov}(Y_i',Y_j')$, and by the same methods used earlier this may be shown to take the analogous form

$$\mathrm{cov}(Y_i',Y_j')=\frac{N(N-n)}{n}\Big[\sigma_{y_i,y_j}+\bar{R}_i\cdot\bar{R}_j\cdot\sigma_{x_i,x_j}-\bar{R}_i\cdot\sigma_{x_i,y_j}$$

$$-R_j\cdot\sigma_{y_i,x_j}+\frac{1}{n-1}\Big(\frac{N-1}{N}\sigma_{r_i,r_j}\cdot\sigma_{x_i,x_j}$$

$$+\frac{N-n}{N}\sigma_{r_i,x_j}\cdot\sigma_{x_i,r_j}\Big)\Big]$$

Computing formulas for the minimum variance unbiased estimators of the terms in this covariance formula are shown in Table 1 for the case i=1, j=2; sample means are expressed in the manner indicated earlier as, for example,

$$\overline{y_1y_2}=\frac{1}{n}\sum_{i=1}^{n}y_{1i}y_{2i}$$

and all products represent ordinary products, as

$$\bar{y}_1\bar{y}_2=\frac{1}{n^2}(\sum_1^n y_{1i})(\sum_1^n y_{2i})$$

In addition, the abbreviation $(n)_m$ is used for $n(n-1)\cdots(n-m+1)$. Computing formulas for estimating the components of $\mathrm{var}(Y_i')$ may be obtained from putting $(r_1,x_1,y_1)=(r_2,x_2,y_2)=(r_i,x_i,y_i)$.

AN EMPIRICAL EVALUATION OF COMPONENT-WISE RATIO ESTIMATION OF CORN PLOT TOTAL DRY WEIGHT

Crop yield in agronomic experiments with silage corn is ordinarily measured in terms of total dry matter production per plot. Dry weight can be measured accurately only by drying the harvested plant material in ovens and there are, of course, distinct limitations on the amount of material which can be handled in this manner. Green, or fresh weight of the production from a plot, however, can be measured directly in the field as the material is harvested, and since green and dry weight are highly correlated the total dry weight for the plot can be accurately estimated by determining the dry matter percentage in a sample from the plot and applying this sample dry matter per cent to the measured total green weight. For the purpose of measuring the sampling error in this method of estimation, green and dry weight determinations were made on 390 individual hills of corn in an experiment containing an early, medium, and a late maturing variety arranged in plots of 10 hills.* These weight determinations were made separately for the ears and stovers of each hill (stovers =husks+stalks+leaves), thus providing an opportunity also to examine the efficiency of a component-wise estimator of plot total dry weight. The separate and combined components of hill green and dry weights are summarized graphically in Figure 1, showing that a somewhat higher green weight-dry weight correlation exists for the separate components, ears and stovers, than for the total, ears + stovers. Average within-plot correlations between green and dry weight of ears, stovers, and ears + stovers were .953, .932, and .824, respectively.

For each plot the efficiency of the unbiased component-wise ratio estimator $Y'=Y'_{stover}+Y'_{ear}$ relative to the unbiased combined ratio estimator $Y'_{stover+ear}$ was computed for samples of n hills, n=2,3,···,9. These efficiencies, in the form of a variance ratio $\mathrm{var}(Y'_{s+e})/\mathrm{var}(Y'_s+Y'_e)$, were relatively constant for all n, and the average efficiencies over all 39 plots as shown in Table 2. The two components of the estimator, Y'_{stover} and Y'_{ear}, were correlated in this experiment, but to a much lesser degree than the green and dry weights within each component (Table 3).

The variances $\mathrm{var}(Y'_{s+e})$, $\mathrm{var}(Y'_s+Y'_e)$, $\mathrm{var}(Y'_s)$, $\mathrm{var}(Y'_e)$ and $\mathrm{cov}(Y'_s,Y'_e)$ employed in the above evaluation were computed directly from the formulas given earlier. In addition to this evaluation, however, the data provided an opportunity to compare the sampling error of the unbiased ratio estimator with the error mean square of the more conventional, but biased, ratio estimator $\hat{Y}=\bar{y}X/\bar{x}$. This was accomplished by enumerating all possible samples of size n for each plot of N=10 hills, computing the conventional ratio estimate for each such sample, and then averaging the squared error, (estimate-known plot dry weight)^2, over all $\binom{N}{n}$ samples. Averaged over all 39 plots, the error mean squares (EMS) for the three estimators $\hat{Y}_{s+e}$, $\hat{Y}_s$, $\hat{Y}_e$ compared to the variances of the corresponding unbiased ratio estimators as shown in Table 4.

The bias of the conventional estimator is negligible in this case, even for small samples, and its sampling error is the same as that of the unbiased ratio estimator. In practice, of course, the conventional estimator offers the advantage that individual hills in the sample need not be weighed and dried separately but may be handled in bulk.

Finally, the actual error mean square of the biased estimator $\hat{Y}=\bar{y}X/\bar{x}$ can be compared to variance approximation

$$\text{var}(\hat{Y}) \doteq \frac{N(N-n)}{n}\,\bar{Y}^2\left[\frac{\sigma_y^2}{\bar{Y}^2}+\frac{\sigma_x^2}{\bar{X}^2}-\frac{2\sigma_{x,y}}{\bar{X}\,\bar{Y}}\right]$$

This comparison is shown graphically in Figure 2. A tendency for this approximation to underestimate the true error mean square decreases as sample size increases since the actual error mean square decreases at a faster rate than the function N-n/n.

Table 1. Computing formulas for the estimation of $\text{cov}(Y_1', Y_2')$

$$s_{y_1,y_2} = [\overline{y_1y_2} - \bar{y}_1\bar{y}_2]\,n^2/(n)_2$$

$$\begin{aligned}\bar{r}_1\cdot\bar{r}_2\cdot s_{x_1,x_2} = \{ & n^2[2(\overline{y_1y_2} - \bar{x}_1\overline{r_1y_2} - \bar{x}_2\overline{y_1r_2}) - (n-1)(\bar{r}_1\overline{x_1y_2} + \bar{r}_2\overline{y_1x_2}) \\ & - (n-2)(\overline{x_1x_2}\,\overline{r_1r_2}) - (\bar{y}_1\bar{y}_2 + \overline{r_1y_2}\,\overline{y_1r_2})] + n^3[(n-2)\bar{r}_1\bar{r}_2\overline{x_1x_2} + \bar{r}_1\bar{x}_1\bar{y}_2 \\ & + \bar{r}_1\bar{x}_2\overline{x_1r_2} + \bar{r}_2\bar{x}_1\overline{r_1x_2} + \bar{r}_2\bar{x}_2\bar{y}_1 + \bar{x}_1\bar{x}_2\overline{r_1r_2}] - n^4\bar{r}_1\bar{r}_2\bar{x}_1\bar{x}_2\}/(n)_4\end{aligned}$$

$$\bar{r}_1\cdot s_{x_1,y_2} = \{n^2[(n-1)\bar{r}_1\overline{x_1y_2} - \overline{y_1y_2} + \bar{x}_1\overline{r_1y_2} + \bar{y}_1\bar{y}_2] - n^3\bar{r}_1\bar{x}_1\bar{y}_2\}/(n)_3$$

$$\bar{r}_2\cdot s_{y_1,x_2} = \{n^2[(n-1)\bar{r}_2\overline{y_1x_2} - \overline{y_1y_2} + \bar{x}_2\overline{y_1r_2} + \bar{y}_1\bar{y}_2] - n^3\bar{r}_2\bar{y}_1\bar{x}_2\}/(n)_3$$

$$\begin{aligned}s_{r_1,r_2}\cdot s_{x_1,x_2} = \{ & n^2[(n^2-3n+1)\overline{r_1r_2}\,\overline{x_1x_2} + (n-1)(\bar{x}_1\overline{r_1y_2} - \overline{y_1y_2} + \bar{x}_2\overline{y_1r_2} + \bar{r}_1\overline{x_1y_2} \\ & + \bar{r}_2\overline{y_1x_2} + \bar{y}_1\bar{y}_2 + \overline{r_1x_2}\,\overline{x_1r_2}] - n_3[(n-2)(\overline{r_1r_2}\,\bar{x}_1\bar{x}_2 + \bar{r}_1\bar{r}_2\overline{x_1x_2}) \\ & + \bar{r}_1\bar{x}_1\bar{y}_2 + \bar{r}_1\bar{x}_2\overline{x_1r_2} + \bar{r}_2\bar{x}_1\overline{r_1x_2} + \bar{r}_2\bar{x}_2\bar{y}_1] + n^4\bar{r}_1\bar{r}_2\bar{x}_1\bar{x}_2\}/(n)_4\end{aligned}$$

$$\begin{aligned}s_{r_1,x_2}\cdot s_{x_1,r_2} = \{ & n^2[(n^2-3n+1)\overline{r_1x_2}\,\overline{x_1r_2} + (n-1)(\bar{x}_1\overline{r_1y_2} - \overline{y_1y_2} + \bar{r}_2\overline{y_1x_2} + \bar{r}_1\overline{x_1y_2} \\ & + \bar{x}_2\overline{y_1r_2}) + \bar{y}_1\bar{y}_2 + \overline{r_1r_2}\,\overline{x_1x_2}] - n^3[(n-2)(\overline{r_1x_2}\,\bar{x}_1\bar{r}_2 + \bar{r}_1\bar{x}_2\overline{x_1r_2}) \\ & + \bar{r}_1\bar{x}_1\bar{y}_2 + \bar{r}_1\bar{r}_2\overline{x_1x_2} + \bar{x}_1\bar{x}_2\overline{r_1r_2} + \bar{r}_2\bar{x}_2\bar{y}_1] + n^4\bar{r}_1\bar{r}_2\bar{x}_1\bar{x}_2\}/(n)_4\end{aligned}$$

Table 2. Average relative efficiency of the unbiased component-wise estimator

n=2	n=3	n=4	n=5	n=6	n=7	n=8	n=9
1.67	1.68	1.69	1.69	1.69	1.69	1.69	1.69

Table 3. Average correlation between the two components of the estimator

n=2	n=3	n=4	n=5	n=6	n=7	n=8	n=9
.237	.240	.241	.241	.242	.242	.242	.241

Figure 1. Summary of data on 390 hills of corn

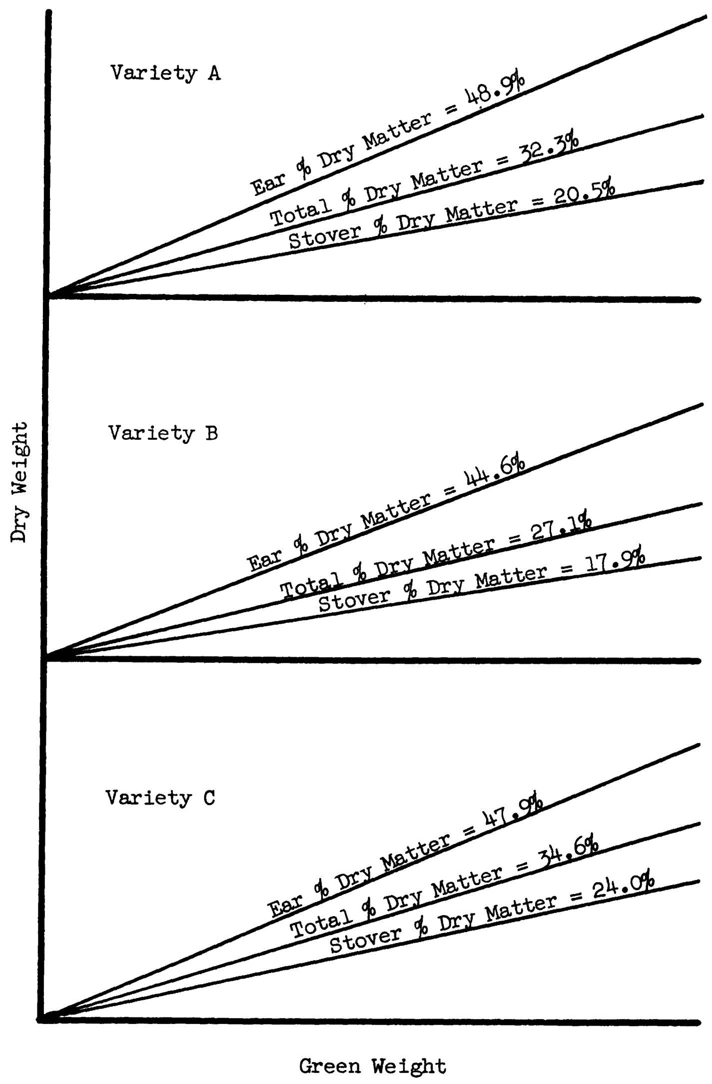


Figure 2. Bias of variance approximation

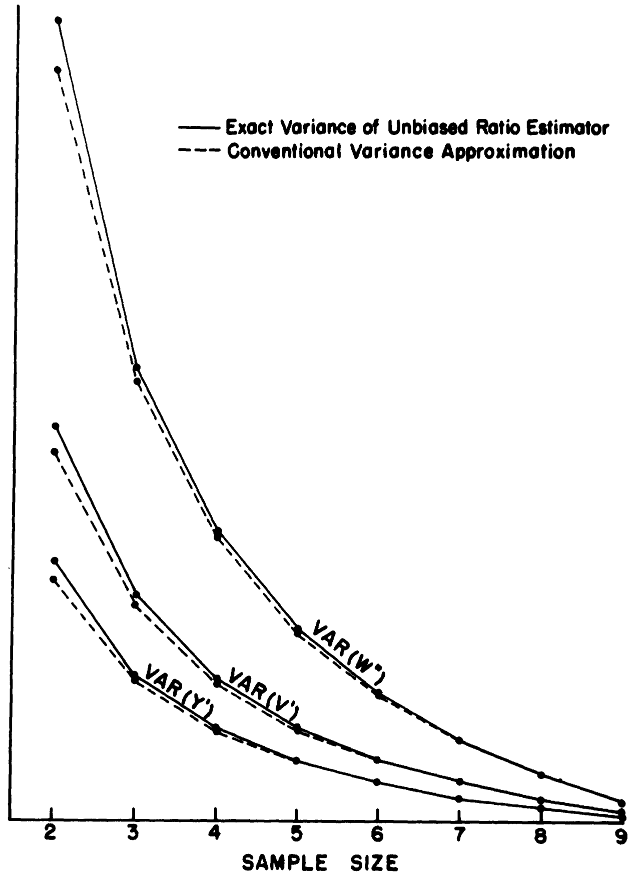


Table 4. Comparison of error mean squares of biased and unbiased ratio estimators

	n=2	n=3	n=4	n=5	n=6	n=7	n=8	n=9
EMS($\hat{Y}_{s+e}$)	45,603	26,102	16,665	--	7,352	4,718	2,748	1,220
var(Y'_{s+e})	46,233	26,275	16,744	11,113	7,389	4,742	2,762	1,226
EMS($\hat{Y}_s$)	9,612	5,362	3,382	--	1,477	945	549	243
var(Y'_s)	9,466	5,317	3,374	2,235	1,484	952	554	246
EMS($\hat{Y}_e$)	14,222	8,102	5,169	--	2,286	1,468	856	380
var(Y'_e)	14,458	8,197	5,217	3,461	2,300	1,476	860	382

[1] Prepared in connection with research sponsored by the National Science Foundation

REFERENCES

1) Hartley, H. O. and A. Ross. Unbiased ratio estimators. Nature 174:270, 1954.

2) Goodman, L. A. and H. O. Hartley. The precision of unbiased ratio-type estimators. Jour. Amer. Stat. Assoc. 53:451-509, 1958.

3) Robson, D. S. Application of multivariate polykays to the theory of unbiased ratio-type estimation. Jour. Amer. Stat. Assoc. 52:511-522, 1957.

4) Tukey, J. W. Keeping moment-like sampling computations simple. Annals Math. Stat. 27: 37-54, 1956.

5) Robson, D. S. An unbiased sampling and estimation procedure for creel censuses of fishermen. Biometrics 16, No. 2, 1960.

SAMPLING EXPERIMENTS ON RATIO ESTIMATORS*

By: Alan Ross, University of Kentucky Medical Center

1. Introduction. An increasing variety of estimation formulas involving auxiliary variables is appearing in the statistical literature and elsewhere [4, 5, 6, 7, 8, 9]. A general tack in this development is a search for estimators that eliminate or suppress bias which is present in the standard ratio and regression estimators used in sample surveys. Small samples are the principal source of difficulty when one attempts a formal evaluation of the characteristics of these estimators, since formulas for the variance and bias have only been obtained as approximations valid for large samples. The purpose of this paper is to present some empirical evidence bearing on the performance of the ordinary ratio of means estimator for samples of size 2, 3, and 4, and to compare this estimator against an unbiased ratio-type estimator.

Let y denote the characteristic whose population mean $\bar{Y}$ is to be estimated, and let x be a variable with known mean $\bar{X}$. The formula

$$\tilde{y} = \bar{y}\bar{X}/\bar{x}, \qquad (1)$$

where $\bar{y}$ and $\bar{x}$ are sample means, is the standard ratio of means estimator in its elementary form appropriate for simple random sampling. The ratio estimator $\tilde{y}$ is known to be biased with the bias decreasing for increasing sample size. In survey designs involving small samples from many strata it is possible that the combined bias in $\tilde{y}$ may assume serious proportions. For it can be shown that if the bias in $\tilde{y}$ for each stratum has the same sign, the bias in the estimated population mean will be approximately constant and equal to the average bias for individual strata, whereas the standard deviation of the overall estimate decreases by a factor of $1/\sqrt{L}$ (where L is the total number of strata) [2].

The large sample variance formula generally used for $\tilde{y}$ is

$$\mathrm{Var}(\tilde{y}) = \frac{N-n}{N}\frac{\bar{Y}^2}{n}\left\{\frac{S_y^2}{\bar{Y}^2} + \frac{S_x^2}{\bar{X}^2} - 2\frac{S_{xy}}{\bar{Y}\bar{X}}\right\}, \qquad (2)$$

where the S^2's are mean squares and S_{xy} is the mean product of x and y. A consequence of (2) is that the ratio estimate $\tilde{y}$ will have smaller variance than the simple mean $\bar{y}$ if

$$\rho_{xy} > \frac{1}{2}\left(\frac{S_x}{\bar{X}}\right)\Big/\left(\frac{S_y}{\bar{Y}}\right),$$

where ρ_{xy} is the correlation between x and y. (See, e.g., [2].)

A fairly simple unbiased estimator based on ratios was proposed in [4]. The estimator for simple random sampling is written

$$y' = \bar{X}\bar{r} + \frac{(N-1)n}{N(n-1)}(\bar{y} - \bar{r}\bar{x})$$

where $r_i = y_i/x_i$, $\bar{r} = \frac{1}{n}\sum r_i$, N is the number of units in the population, and n is the sample size. Robson [4] gave an exact formula for the variance of y' in terms of multivariate symmetric means. In terms of ordinary population mean squares and products the variance of y' is given in [8].

Let

$$\bar{R} = \frac{1}{N}\sum r_i$$

$$S_r^2 = \frac{1}{N-1}\sum (r_i - \bar{R})^2$$

$$S_{xr} = \frac{1}{N-1}\sum (x_i - \bar{X})(r_i - \bar{R})$$

with parallel notation for the mean squares and products involving y's and x's. Then the variance of y' can be written

$$\frac{Nn(n-1)}{(N-1)^2}V(y') = aS_y^2 + (N-1)bS_x^2S_r^2 + a\bar{R}^2S_x^2 + c\bar{X}^2S_r^2 + 2d\bar{X}S_{yr} + 2e\bar{R}S_{xy} + \frac{(N-1)}{N}fS_{xr}^2 + g\bar{X}\bar{R}S_{xr},$$

where a, b, ..., g are constant coefficients as follows:

$$a = \frac{n-1}{N} - \frac{(N-1)(n-2)(n-3)}{N(N-2)(N-3)}$$

$$b = \frac{1}{N^2} - \frac{2(n-2)}{N^2(N-2)} + \frac{(n-2)(n-3)}{(N-2)(N-3)}$$

$$c = \frac{(n-1)}{N} - \frac{(N-1)(n-2)(n-3)}{N(N-2)(N-3)} - \frac{(n-1)}{(N-1)} - \frac{(n-1)^2}{(N-1)^2} + \frac{2(n-1)(n-2)}{(N-1)(N-2)}$$

$$d = \frac{(n-2)(n-3)}{N(N-2)} - \frac{(n-1)(n-3)}{N(N-1)} + \frac{2(n-2)(n-3)}{N(N-2)(N-3)} - \frac{2(n-1)(n-2)}{N(N-1)(N-2)}$$

$$e = \frac{(n-2)(n-3)}{N(N-2)} - \frac{(n-1)}{N} + \frac{2(n-2)(n-3)}{N(N-2)(N-3)}$$

$$f = \frac{1}{N} + \frac{(n-1)^2}{N} + \frac{2(n-2)^2}{N(N-2)} + \frac{2(n-2)(n-3)}{N(N-2)(N-3)} + \frac{(n-1)}{N} - \frac{(N-1)(n-2)(n-3)}{N(N-2)(N-3)} - \frac{Nn(n-1)}{N(N-1)}$$

$$g = \frac{2}{N} + \frac{(n-1)(n-2)}{N} - \frac{n(n-1)}{(N-1)} + \frac{(n-2)(n-3)}{N(N-2)} + \frac{N(n-1)(n-2)}{(N-1)(N-2)} - \frac{(N-1)^2(n-2)(n-3)}{N(N-2)(N-3)}$$

2. The Sampling Experiments. It was indicated above that there is as yet no simple formulation of the characteristics of the distribution of the estimator for arbitrary (finite) populations in terms of population moments or equivalents. In order to gather some evidence on how effective $\tilde{y}$ is for small samples, two universes of 50 elements each were constructed, random samples (without replacement) of size 2, 3, and 4 were drawn, and the distributions of $\tilde{y}$ were estimated on the basis of the repeated sampling.

One universe was composed of the first 50 families included in a sample survey of the University of Kentucky faculty and staff [1]. For that survey data were recorded for each family on the number of persons in the family, number of physician visits in 1957 (home, office, University Health Service Dispensary), total charges to the family by physicians for these outpatient visits, and a host of other material.

In example A we regard the average number of physician visits as the quantity to be estimated from a sample, and we assume that the average number of persons per family is known for the universe. In formula (1) $\bar{X}$ is then the population mean number of persons per family (2.94 for this example), and $\bar{y}$ and $\bar{x}$ denote sample means of physician visits and number of persons per family, respectively.

The same universe of 50 families was used for Example B with physician charges (dollars) forming the y population and number of physician visits was the x population or concomitant variable.

The University of Kentucky Department of Agricultural Economics provided corn acreage data for 1952 and 1956 on a number of farms scattered throughout Kentucky. The first 50 of these farms were used as a universe for Example C. In this situation we shall estimate the average number of acres planted in corn in 1956 (the y variable) with assistance of the population mean acreage planted in 1952 (the x variable).

Sampling was accomplished by loading the x and y variables for a universe onto the drum of an IBM Type 650 Data Processing Machine and then feeding random numbers (between 1 and 50) to designate (x,y) pairs for a sample. The random numbers were introduced four at a time--the first two designated a sample of size 2, the first three numbers designated a sample of size 3, and the four numbers together designated a sample of size 4. The numbers in each set of four were distinct so that the method of selection was random sampling without replacement. No control was exercised over repeated pairs, triplets, or quadruplets of random numbers so that sets of two, three, or four elements were selected at random with replacement from populations consisting of pairs, triplets, and quadruplets of (x,y) sets.

The estimates $\tilde{y}$ and y' were computed for each sample designated as given above. Totals of $\tilde{y}$, $\tilde{y}^2$, $\tilde{y}^3$, $\tilde{y}^4$, y', y'^2, y'^3, and y'^4 were accumulated as the sampling proceeded to allow for computation of moments of the distributions of $\tilde{y}$ and y'. The frequencies were also accumulated in eleven equally spaced intervals for each distribution.

Example A--Physician Visits. Some of the properties of the populations are given in Table 1. Since the relation $\rho_{xy} > C_x/2C_y$ holds we should expect that $\tilde{y}$ will be more efficient than the simple mean $\bar{y}$, at least for large samples.

One thousand samples of size 2, size 3, and size 4 were drawn from this universe. The results of the samplings are summarized in Table 2. The entries for $\tilde{y}$ and y' were calculated from the samplings, while the variances of the simple mean $\bar{y}$ was computed from formula. The variance formula for y' was not used since there was some evidence that rounding error introduced by the machine inflated the estimated variance of both $\tilde{y}$ and y'.

The variance of $\bar{y}$ is given in the first line as a reference point. The estimated bias of $\tilde{y}$, shown in line 2, is seen to be negligible, although the bias demonstrated no tendency to decrease as the sample size increased from 2 to 4. The bias relative to the standard deviation of $\tilde{y}$, given in line 6, showed a tendency to increase. Line 8 indicates an increase in efficiency of the unbiased estimator y' relative to $\tilde{y}$ with increasing sample size.

Example B--Physician Charges. The populations and summary of results for this example are presented in Table 3 and Table 4. As before the characteristics of the distributions were estimated from 1000 samples of each size.

Table 1. Properties of the populations used in Example A.

y = number of physician visits
$\bar{Y} = 18.42$
$S_y^2 = 243.06$
$C_y = .85$
x = number of persons in family
$\bar{X} = 2.94$
$S_x^2 = 2.43$
$C_x = .53$
$_{xy} = .47$
$C_x/2C_y = .31$

Table 2. Summary of Example A

	Sample Size		
	n=2	n=3	n=4
1. var($\bar{y}$)	116.67	76.14	55.90
2. bias($\tilde{y}$)	.09 (0.5%)	.27 (1.5%)	.30 (1.6%)
3. m.s.e.($\tilde{y}$)	75.05	53.28	41.06
4. var($\tilde{y}$)	75.04	53.21	40.97
5. $\sqrt{var(\tilde{y})}$	8.66	7.29	6.40
6. bias($\tilde{y}$)/$\sqrt{var(\tilde{y})}$	1.1%	3.7%	4.7%
7. var(y')	111.93	63.47	47.65
8. m.s.e.($\tilde{y}$)/var(y')	67.0%	83.8%	86.2%

In Example B the bias in $\tilde{y}$ was more pronounced than in the previous example. The estimated biases for the three sample sizes were significantly different from zero with probabilities less than .01. Line 8 shows increasing efficiency of y' relative to $\tilde{y}$ as the sample size increased from 2 to 4.

Example C--Acres in Corn. Tables 5 and 6 give the information pertaining to this example. In this example characteristics of the distributions were estimated from 500 samples of each size.

The estimated biases for $\tilde{y}$ in Table 6 are significantly different from zero at the 1% level for samples of size 2 and size 3 and at the 5% level for samples of size 4.

Table 3. Properties of the populations used in Example B.

y = physician charges
$\bar{Y}$ = 84.92
S_y^2 = 6659.71
C_y = .96
x = number physician visits
$\bar{X}$ = 18.42
S_x^2 = 243.06
C_x = .85
ρ_{xy} = .74
$C_x/2C_y$ = .44

Table 4. Summary of Example B

	Sample Size		
	n=2	n=3	n=4
1. var($\bar{y}$)	3196.66	2086.71	1531.73
2. bias($\tilde{y}$)	4.63 (5.5%)	3.25 (3.8%)	4.45 (5.2%)
3. m.s.e.($\tilde{y}$)	2078.03	1318.37	1022.94
4. var($\tilde{y}$)	2056.60	1307.78	1003.16
5. $\sqrt{var(\tilde{y})}$	45.35	36.16	31.68
6. bias($\tilde{y}$)/$\sqrt{var(\tilde{y})}$	10.2%	9.1%	14.0%
7. var(y')	2833.48	1427.33	1026.56
8. m.s.e.($\tilde{y}$)/var(y')	73.3%	92.4%	99.6%

Table 5. Properties of the populations used in Example C.

y = acres in corn (1956)	
$\bar{Y}$	= 25.28
S_y^2	= 237.31
C_y	= .61
x = acres in corn (1952)	
$\bar{X}$	= 36.56
S_x^2	= 399.37
C_x	= .55
ρ_{xy}	= .70
$C_x/2C_y$	= .45

3. Discussion. Perhaps the most consistent finding for the three examples was the tendency for the efficiency of the unbiased estimator y' to increase relative to $\tilde{y}$ as sample size was increased from 2 to 4. The bias in $\tilde{y}$ was persistent but not overwhelming in two of the examples when compared to the standard error (line 6 in Tables 2, 4, and 6). Judging from the examples presented here it appears that when a ratio estimate is appropriate according to the ρ_{xy} criterion the $\tilde{y}$ estimator maintains its efficiency relative to $\bar{y}$ even for samples of size 2. The unbiased estimator y' compared favorably with $\tilde{y}$ on accuracy for samples of size 4, and we note that an unbiased estimate of the variance of y' can be computed from a sample of four or more elements; but a further result of the sampling experiments indicates that a variance estimate for $\tilde{y}$ for samples of size 4 computed by substituting sample values in (2) may be an underestimate by 15% to 35%.

Table 6. Summary of Example C

	Sample Size		
	n=2	n=3	n=4
1. var($\bar{y}$)	113.91	74.36	54.58
2. bias($\tilde{y}$)	1.33(5.3%)	.98 (3.9%)	.59 (2.3%)
3. m.s.e.($\tilde{y}$)	92.39	51.91	36.31
4. var($\tilde{y}$)	90.62	50.95	35.96
5. $\sqrt{var(\tilde{y})}$	9.52	7.13	5.99
6. bias($\tilde{y}$)/$\sqrt{var(\tilde{y})}$	14.0%	13.8%	9.9%
7. var(y')	115.62	57.11	40.79
8. m.s.e.($\tilde{y}$)/var(y')	79.9%	90.9%	89.0%

Table 7. Coefficients of skewness and kurtosis for the distributions of $\tilde{y}$ and y'.

		Sample Size					
		n=2		n=3		n=4	
		g_1	g_2	g_1	g_2	g_1	g_2
Physician Visits	$\tilde{y}$	.676	3.412	.552	.241	.425	-.066
	y'	.192	1.56	.471	.236	.447	.052
Physician Charges	$\tilde{y}$	1.167	2.358	1.92	1.311	.700	.921
	y'	-.176	4.094	.286	1.215	.342	.370
Acres in Corn	$\tilde{y}$	.716	1.574	.241	.355	.169	.323
	y'	-.087	.936	.131	1.120	.146	.659

An additional set of descriptive measures for the sampling distributions of $\tilde{y}$ and y' is shown in Table 7 where estimates of γ_1 and γ_2 are given. The estimates of all 18 sampling distributions were unimodal with varying degrees of asymmetry and kurtosis. The pair of histograms in Figure 1 is typical and illustrative.

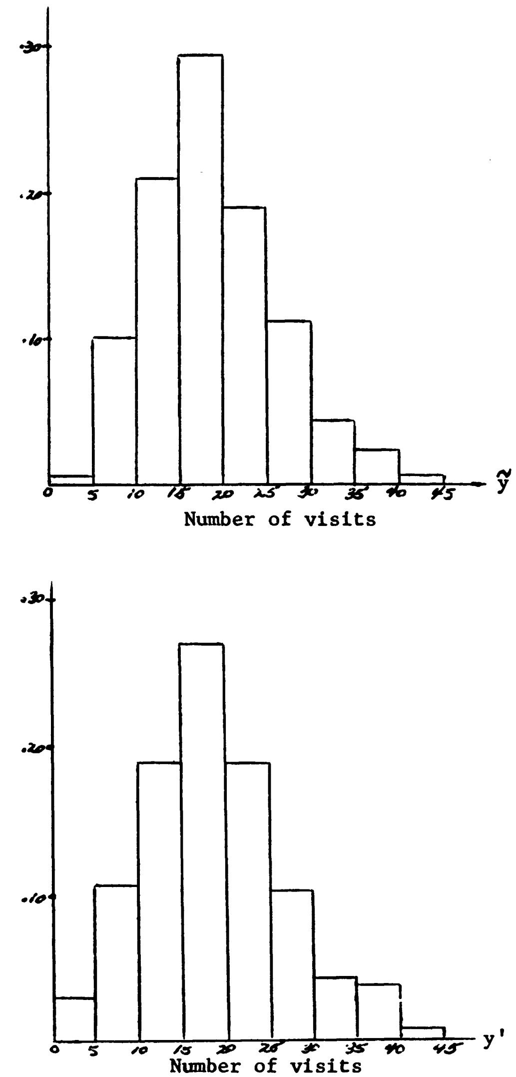


Figure 1. Histograms of distributions of $\tilde{y}$ and y' for physician visits for samples of size 3.

*The courtesy of the University of Kentucky Computing Center, whose facilities were used for the computations in this report, is gratefully acknowledged.

References

[1] Bost, H. L. and Ross, A. University of Kentucky employees and their families. Report no. 1 and report no. 3.(in preparation) on the health and insurance study. Lexington, Kentucky, University of Kentucky, Medical Center. 1959.

[2] Cochran, W. G. Sampling techniques. New York, John Wiley and Sons. 1953.

[3] Goodman, L. A. and Hartley, H. O. The precision of unbiased ratio-type estimators. Jour. Amer. Stat. Assoc. 53:491-508. 1958.

[4] Hartley, H. O. and Ross, A. Unbiased ratio estimators. Nature. 174:270-271. 1954.

[5] Mickey, M. R. Some finite population unbiased ratio and regression estimators. Jour. Amer. Stat. Assoc. 54:594-612. 1959.

[6] Nieto de Pascual, J. Unbiased ratio estimators in stratified sampling. Unpublished report. Ames, Iowa. 1959.

[7] Robson, D. S. Applications of multivariate polykays to the theory of unbiased ratio-type estimation. Jour. Amer. Stat. Assoc. 52:511-522. 1957.

[8] Ross, A. On two problems in sampling theory: unbiased ratio estimators and variance estimates in optimum sampling designs. Unpublished Ph.D. thesis. Ames, Iowa, Iowa State University Library. 1960.

[9] Williams, W. H. Unbiased regression estimators. Unpublished report. Ames, Iowa. 1959.

UNBIASED ESTIMATION

By: W. H. Williams, McMaster University

1. INTRODUCTION

A favourite method in sampling theory of increasing the precision of estimates is the utilization of auxiliary information. Analytically, we have a random sample of n pairs (y_i, x_i) drawn from a population of size N and the problem is to estimate the population mean μ_Y relative to the assumption that the population mean μ_X is known exactly. Ratio and regression estimators have been designed for this problem and they are described in detail along with illustrations in the textbooks on the subject, see for example Cochran [1953]. Additional contributions have been made by Hartley and Ross [1954], Nieto [1958] and Robson [1957].

First, we specify that an estimator is of the regression type if it is invariant under location and scale changes in x and undergoes the same location and scale changes as the y variate. A ratio estimator has these properties for scale changes only.

The two common ratio estimators are known to be biased. These estimators are the ratio of means estimator $\hat{y} = \bar{y}\mu_X/\bar{x}$ and the mean of ratios estimator $\hat{y} = \mu_X \sum_{i=1}^{n} r_i/n$, where $\bar{y}$ and $\bar{x}$ are sample means and $r_i = y_i/x_i$. The classical regression estimator is obtained by evaluating the least squares line of best fit at the point μ_X giving $\hat{y}_b = \bar{y} + b(\mu_X - \bar{x})$ as a regression estimator of μ_Y. This estimator is biased if the assumption of a linear model is not valid.

Some exactly unbiased ratio and regression estimators are presented in this paper.

2. A PROCEDURE FOR UNBIASED ESTIMATION

The following sampling procedure can be used to derive unbiased estimators. The scheme consists of two steps. First select with equal probability one of all possible splits of the population in mutually exclusive groups of size n/k. Let s be the number of groups and assume that n/k divides N so that s is an integer. At the second step select randomly without replacement k of the groups from the total number of groups (s) obtained in step one. Thus a sample of size n is obtained.

Next consider the conditional distribution for a fixed set of s groups. These groups have y and x means which are denoted $\bar{y}^{(i)}$ and $\bar{x}^{(i)}$ (i) = 1, 2 . . . s; also let $b^{(i)}$ denote an as yet unspecified function of the y and x of that group. For a given split and a random selection of groups the expectations of $\bar{y}^i$ and $\bar{x}^i$ i = 1, 2, 3 . . . k are μ_Y and μ_X respectively; furthermore

$$\left(1 - \frac{n}{N}\right) \frac{1}{k(k-1)} \sum_{i=1}^{k} (b^i - \bar{b})(\bar{x}^i - \bar{x}) \tag{1}$$

is an unbiased estimator of $\text{Cov}(\bar{b}, \bar{x})$ where $\bar{b} = \sum_{i=1}^{k} b^i/k$.

Hence $Eg = \mu_Y - \text{Cov}(\bar{b}, \bar{x})$ where $g = \bar{y} + \bar{b}(\mu_X - \bar{x})$ showing that

$$T_k = \bar{y} + \bar{b}(\mu_X - \bar{x}) + \left(1 - \frac{n}{N}\right)\frac{1}{k(k-1)}\sum_{i=1}^{k}(b^i - \bar{b})(\bar{x}^i - \bar{x}) \tag{2}$$

is a conditionally unbiased estimator of μ_Y. It is then unbiased unconditionally.

T_k remains unbiased for any defined form of the coefficients $b^{(i)}$. It is classified as a regression estimator if $b^{(i)}$ has a form which is invariant under linear x and y transformation (for example least squares form). If $b^{(i)} = \bar{y}^{(i)}/\bar{x}^{(i)}$ (say) then T_k falls into the class of a ratio estimator.

3. SOME ILLUSTRATIONS

It is natural to consider $b^{(i)}$ in the least squares slope form shown in equation (3).

$$b^{(i)} = \frac{\sum_{j=1}^{n/k}(y_j - \bar{y}^{(i)})(x_j - \bar{x}^{(i)})}{\sum_{j=1}^{n/k}(x_j - \bar{x}^{(i)})^2} \tag{3}$$

$$(i) = 1, 2 \ . \ . \ . \ s.$$

In this case T_k is similar to $\hat{y}_b$ but possesses an additional term which compensates for possible bias in $\hat{y}_b$. One also wonders about the efficiency of T_k when compared with $\hat{y}_b$, for when the linear model assumption is valid $\hat{y}_b$ possesses certain optimum variance properties. However, $\hat{y}_b$ is then also unbiased and the advantage of T_k is unbiasedness in situations in which $\hat{y}_b$ is not unbiased. It would be desirable that the variance of T_k compare favourably with $\hat{y}_b$ even under the assumption of a linear model. A discussion of this case showing that the loss in efficiency is $O(n^{-1})$ can be found in Williams [1958].

Another possible choice is $b^{(i)} = \sum_{j=1}^{n/k} y_j x_j / \sum_{j=1}^{n/k} x_j^2$. In this form T_k is a ratio estimator and it is unbiased even if the linear relationship of y and x does not pass through the origin. But characteristically the variance will be inflated by such a relationship.

Next, if $b^{(i)} = \bar{y}^{(i)}/\bar{x}^{(i)} = \bar{r}^{(i)}$, T_k will reduce to the form

$$T_k = \bar{r}\mu_X + \frac{Nk-n}{N(k-1)}(\bar{y} - \bar{r}\bar{x}) \tag{4}$$

where $\bar{b}$ is denoted $\bar{r}$. It will be noted that when k=n, T_k=y', the unbiased ratio estimator presented by Hartley and Ross [1954].

Finally, consider $b^i = r^i = \frac{k}{n}\sum_{j=1}^{n/k} r^i_j$, $r_j = y^i_j/x^i_j$ then $\bar{b} = \bar{r} = \sum_{j=1}^{n} r_j/n$ which does not depend upon the particular split of the population. Now if, after substitution of this form into T_k, the estimator is averaged over all possible splits of the sample into groups of size n/k, it will be found that the result is again the Hartley-Ross unbiased ratio estimator.

Other forms could be considered.

4. EXTENSION TO STRATIFIED SAMPLING

Unbiased estimators are important in stratified sampling as a bias may be magnified relative to the standard deviation. Their separate use 'within strata' requires exact knowledge of the population strata means but is straightforward. A 'combined' stratified estimator can also be developed.

Suppose that there are L strata with N_t units in the t^{th} stratum, $t = 1, 2 \ldots L$ with $\sum_{t=1}^{L} N_t = N$. The sampling is again done in two stages. First select with equal probability one of all possible splits of each stratum into s groups of size n_t/k. Then $N_t = sn_t/k$. At the second stage select k groups with equal probability and without replacement from each of the strata, giving a sample of size n_t in the t^{th} stratum, $\sum_{t=1}^{L} n_t = n$.

For a given split and a random selection of groups, μ_Y and μ_X are estimated unbiasedly by $\bar{y}^i_{st} = \sum_{t=1}^{L} N_t\bar{y}^i_t/N$ and $\bar{x}^i_{st} = \sum_{t=1}^{L} N_t\bar{x}^i_t/N$ where $\bar{y}^i_t$ and $\bar{x}^i_t$ denote means of the i^{th} group in the t^{th} stratum. Also we can consider a coefficient $b^{(i)}_{st}$ which is as yet unspecified in form but utilizes the set of elements in the i^{th} group of all strata. For example

$$b^{(i)}_{st} = \frac{\sum_{t=1}^{L}\sum_{j=1}^{n_t/k}(y_{tj} - \bar{y}^{(i)}_t)(x_{tj} - \bar{x}^{(i)}_t)}{\sum_{t=1}^{L}\sum_{j=1}^{n_t/k}(x_{tj} - \bar{x}^{(i)}_t)^2} \qquad (5)$$

$$(i) = 1, 2 \ldots s$$

is an overall slope estimator.

Next, notice that

$$\bar{y}_{st} = \sum_{t=1}^{L} N_t\bar{y}_t/N = \sum_{i=1}^{k}\bar{y}^i_{st}/k \qquad (6)$$

where $\bar{y}_t$ is the mean of the n_t observations in the t^{th} stratum (similarly for $\bar{x}_{st}$) and that a conditionally unbiased estimator of $\mathrm{Cov}(\bar{b}_{st}, \bar{x}_{st})$ is given by $(1 - \frac{n}{N})\frac{1}{k(k-1)}\sum_{i=1}^{k}(\bar{x}^i_{st} - \bar{x}_{st})(b^i_{st} - \bar{b}_{st})$. Therefore, if $g = \bar{y}_{st} + \bar{b}_{st}(\mu_X - \bar{x}_{st})$ then $Eg = \mu_Y - \mathrm{Cov}(\bar{b}_{st}, \bar{x}_{st})$ and

$$T_{k(st)} = \bar{y}_{st} + \bar{b}(\mu_X - \bar{x}_{st}) + (1 - \frac{n}{N})\frac{1}{k(k-1)}\sum_{i=1}^{k}(\bar{x}^i_{st} - \bar{x}_{st})(b^i_{st} - \bar{b}_{st}) \qquad (7)$$

is a combined stratified unbiased estimator of μ_Y. Note that $N_t = sn_t/k$ implies $k/s = n/N$.

Again the generalization to p auxiliary variates is straightforward.

To illustrate the stratified estimator take first $b^{(i)}_{st} = \bar{y}^{(i)}_{st}/\bar{x}^{(i)}_{st} = r^{(i)}_{st}$. Then $T_{k(st)}$ reduces to

$$T_{k(st)} = \bar{r}_{st}\mu_X + \frac{Nk-n}{N(k-1)}(\bar{y}_{st} - \bar{r}_{st}\bar{x}_{st}). \qquad (8)$$

When $N_t = \bar{N}$, $n_t = \bar{n}$ for all t and $k = \bar{n}$, $s = \bar{N}$ then

$$T_{k(st)} = \bar{r}_{st}\mu_X + \frac{(\bar{N}-1)\bar{n}}{(\bar{n}-1)\bar{N}}(\bar{y}_{st} - \bar{r}_{st}\bar{x}_{st}) \qquad (9)$$

which is a generalized Hartley-Ross estimator.

Finally, we again consider an averaging of T_k over all possible splits of the sample into groups of size n_t/k, $t = 1, 2 \ldots L$. For this, the coefficient is taken in the form $b^i_{st} = r^i_{st} = \sum_{t=1}^{L}\frac{N_t}{N}r^i_t$ where $r^i_t = \frac{k}{n_t}\sum_{j=1}^{n_t/k}\frac{y^i_j}{x^i_j}$. Therefore, $\bar{r}_{st} = \sum_{i=1}^{k} r^i_{st}/k = \sum_{t=1}^{L}\frac{N_t}{N}\sum_{j=1}^{n_t}\frac{y_j}{x_j}/n_t = \sum_{t=1}^{L}\frac{N_t}{N}\bar{r}_t$ and some algebraic reduction will show that $T_{k(st)}$ averaged over all possible splits is equal to

$$T^*_{k(st)} = \bar{r}_{st}\mu_X + (\bar{y}_{st} - \bar{r}_{st}\bar{x}_{st}) + (1 - \frac{n}{N})\sum_{t=1}^{L}\frac{N^2_t}{N^2}\frac{(\bar{y}_t - \bar{r}_t\bar{x}_t)}{(n_t - 1)} \qquad (10)$$

which does not quite reduce to a form similar in appearance to equation (8) and the Hartley-Ross estimator.

As before other selections of coefficients will yield other unbiased estimators.

5. MULTISTAGE SAMPLING

We consider a population with N primaries of equal size $\bar{M}$ and the following sampling scheme. First select n primaries from the N available with equal probability with or without replacement. Then select with equal probability one of the splits of each of the primaries into s groups of size $\bar{M}/k$. Then with equal probability and without replacement draw k of the groups so that the sample size is $\bar{m}$ in each selected primary.

Consider now the conditional distribution for a fixed set of primaries and a fixed split of the primaries into s groups each. Then by section 4, equation (11) is an unbiased estimator of $\bar{\bar{Y}}_n$, the population mean of the n selected primaries.

$$T_{k(M)} = \bar{\bar{y}} + \bar{b}(\mu_X - \bar{\bar{x}}) + \left(1 - \frac{\bar{\bar{m}}}{\bar{M}}\right)\frac{1}{k(k-1)}\sum_{i=1}^{k} (b^i - \bar{b})(\bar{\bar{x}}^i - \bar{\bar{x}}) \quad (11)$$

where $\bar{\bar{y}}^i = \frac{k}{n\bar{\bar{m}}} \sum_{t=1}^{n} \sum_{j=1}^{m/k} y^i_{tj}$, $\bar{\bar{y}} = \frac{1}{k} \sum_{i=1}^{k} \bar{\bar{y}}^i = \frac{1}{n\bar{\bar{m}}} \sum_{t=1}^{n} \sum_{j=1}^{\bar{\bar{m}}} y_{tj}$ and similarly for x. The coefficient b^i is again arbitrary in form.

Finally, the expectation of $T_{k(M)}$ over all possible primary selections is the average of $\bar{\bar{Y}}_n$ over all possible primary selections; this is μ_Y and $T_{k(M)}$ is unbiased in multistage sampling.

Again the selection of the coefficients yields estimators of different types. For example, an unbiased ratio estimator of the Hartley-Ross type generalized to multistage sampling can be obtained.

This delivered paper was submitted to Biometrics in February, 1960.

REFERENCES

Cochran, W. G. [1953]. Sampling Techniques. New York: John Wiley and Sons.

Hartley, H. O. and Ross, A. [1954]. Unbiased ratio estimators. Nature 174, 270.

Nieto, J. [1958]. Unbiased ratio estimators in stratified sampling. Contributed paper, Inst. of Math. Stat., Ames, Iowa.

Robson, D. S. [1957]. Application of multivariate polykays to the theory of unbiased ratio-type estimation. J. Amer. Stat. Assoc. 52: 511.

Williams, W. H. [1958]. Unbiased regression estimators. Contributed paper, Inst. of Math. Stat., Ames, Iowa.

II

X

STATISTICAL DEFINITIONS, CONCEPTS AND CATEGORIES

Chairman, Dudley Kirk, Population Council

RELATIONS OF SOME SOCIAL SCIENCE CONCEPTS TO STATISTICAL DATA

By: Frederick F. Stephan, Princeton University

The problem I wish to discuss with you is the alignment of statistical data and social science concepts. By this I mean the matching of a concept and a set of data related to it in such a way that one can move from data to concepts and back again in a thoroughly logical and valid manner. Unless we can do that, social theory will develop without the invigorating effect of close association with systematic observation and statistical enterprises will fail to realize their greatest potential usefulness.

What we lose if we fail to bring concepts and data into fruitful relation to each other in the social sciences may not seem very serious if we only consider past relationships. A substantial amount of social science research has utilized statistical methods and data to good effect and in turn has stimulated statistical activities. But we are only beginning. The need for statistical data in the social sciences will increase very rapidly in the future. We see a great expansion of activity in the natural sciences today, accompanied by phenomenal increases in their demands for data of great accuracy. As the social sciences mature they, too, will increase their rate of expansion and their requirements for data not only on new subjects and, with more exacting specifications, on old subjects.

Need for a New Systematic Review

Twenty-five years ago the Committee on Government Statistics and Information Services studied the problems that beset the Federal statistical agencies as they strove to meet the needs for statistical data arising out of the Great Depression and the New Deal programs. They studied these needs and made recommendations about meeting them through suitable changes in the organization of government statistical activities and the manner in which the work was performed. Ten years later, further recommendations about what should be done, who should do it, and how they should do it were made to the Hoover Commission by F. C. Mills and C. D. Long. I believe the time is approaching when there should be another systematic review of statistical activities. This time it should be directed to the relations of these activities to the research of social scientists. At the same time it should examine the extent to which the theories and concepts of social science have been developed in a manner that is conducive to the use of statistical data for their verification and further development.

This suggestion may seem to place an unnecessary burden on those of us who are struggling to sustain the standards gradually established in the past and to get work done that is needed to meet demands already recognized as most urgent. It is true that great progress has been made in organizing and supporting statistical work, in establishing standard methods and classifications, in improving comparability, in sampling, and in processing data. It is true that there hardly seems room for additions to the program of work now being done. Nevertheless, it would be short-sighted to confine our thinking to what is practical today. If we do that, we will never find what is practical tomorrow.

Some among us will see reason to question the reasonableness of this suggestion because of the present state of the social sciences (with the possible exception of economics). It is true that there are social scientists who have no liking for statistical research. The concepts and theory that dominate social science tend to be vague, at least so they seem to a hard-headed statistician. Social scientists tend to over-simplify matters as if abstract ideas were the only reality and all the variations and exceptions we encounter when we collect data were only errors and imperfections. It is also true that they often refuse to discipline themselves to the painstaking methods we use to obtain and assemble statistics. Yet our frustrations over these shortcomings should not stop us from looking for opportunities to achieve a more effective relationship to them and their work. We have a great stake in what they do.

On our part, we should not overlook our own shortcomings. We have paid a price for the progress we have achieved in statistics. It is a certain rigidity resulting from the solution of many of our problems by somewhat arbitrary decisions and by expediency. We need not continue to limit ourselves to the solutions of problems which we adopted in the past or to confine ourselves to traditional programs, ignoring new opportunities to advance into valuable regions of work which we have neglected heretofore. I would like to make some general observations about the ways in which we have accepted restrictions on our work and relate them to some of the concepts which are important in the present and prospective development of social science.

1. The Use of Opportunistic or Pragmatic Categories

The classifications which are used in statistical work are often those which are convenient for the collection of data though they may be very inconvenient for the purposes of theoretical analysis. Many researchers accept the data gratefully since they see no chance of getting any data at all if they do not accept them in this form. For example, the concepts and theory of urban and rural differences stress the differences in values, patterns of behavior, social organization and communication, and personal development which characterize these two modes of life. As statisticians, we are baffled by these complex and elusive concepts and we take as a practical substitute a classification of political areas by the size of their populations or by population density. Obviously our ideas and theirs are not aligned in such a way that we can establish a close or precise correspondence between them. There is an important error of translation every time we pass from one to the other. We can't continue to progress without reducing this error.

We make use of a classification of persons by socio-economic status, based on the occupation of the person or the head of his household. Social scientists attempt to make use of this classification but they can not progress until they obtain data more closely related to their concepts of stratification and class structure. They in turn must improve their concepts before they can succeed in linking them effectively to even an improved statistical version of socio-economic status.

We tend to define a child as a person under, say, eighteen years of age (or twenty-one if we are tinged with the legal definition) while social scientists look on childhood as a status established by custom which usually is defined less by age than by social attitudes. We define as a person's occupation the kind of work he spent most time doing during a week or the work at which he earned the most money. For the social scientist an occupation is a role, a part in the division of labor which is not determined by such superficial, though clear-cut, rules of classification.

We should not lightly abandon definitions which make it possible to collect data economically and accurately. The accuracy we attain, however, is spurious if it is gained at the expense of a large error of translation when the data are linked in research with social science concepts of another and incongruent character. We should free ourselves from the limitations of our previous compromises. This we can often do in special studies even when we can not do it in the regular collection of standard statistical data.

2. Dependence on Cross-sections and Current Activity

The practical exigencies of collecting data also tend to force us to take data in a brief period of time, or even of a point of time, to the neglect of changes that are going on and more gradual developments of great significance for social science theory. Even in our time series, our data often lack historical depth for the persons or groups which are important units of analysis. Thus we take the activity of a week as the basis for determining whether or not a person is in the labor force. The approach of a social scientist would be less definitive but more meaningful in taking as the basis of classifying workers and non-workers a pattern activities extending over a longer period of time. Likewise, our data on migration tend to seize on a period of time rather than a sequence of changes. If we are to weld together the ideas of statisticians and social scientists for these kinds of human activities, we will have to find ways of revising our definitions to meet their revised concepts within the bounds of practical procedure.

3. Emphasis on Residence and the Place of Data Collection

Social scientists are greatly interested in the relations of parents and children, neighbors, relatives, school-mates and work-mates. They are concerned about community life and community institutions. Our statistics tend to ignore most interpersonal relations except some of the formal relations useful to identify individuals and a few traditional relations centered in the household. The emphasis on residence is strong. The emphasis on kinship and social interaction is weak. There are exceptions, of course, such as the collection and analysis of data for "spending units." No one can deny the great importance of the dwelling unit as a matrix of social relationships. Still we should not accept it as a sufficient substitute for the direct recording of the social relationships and activities it is intended to symbolize. Only in this way will we be able to join statistical definitions and social science concepts effectively.

These examples indicate, though they do so very inadequately, the disjuncture between the classifications which have emerged from the struggles of statisticians to improve their data and operate economically, on the one hand, and the concepts and theories developed by social scientists, on the other. Now of course some social scientists have included statistical definitions in their menu of concepts but they have not succeeded in digesting them and transforming them into the living tissues of scientific theory. We statisticians in turn have failed to assimilate into our data the more essentially social aspects of human life. Until we do, our data will fall far short of their possible usefulness for predicting and understanding, as well as for measuring, human activities and experience.

There are practical problems which warrant the expenditure of far greater sums for the collection of data than are now being spent, but only if the data fit in to the scientific processes of analysis which are needed to make them serviceable in the alleviation or solution of these problems. It is not too soon to investigate what must be done to achieve a better match of statistic and concept and to find out how it can be accomplished.

THE EVALUATION AND RESEARCH PROGRAM OF THE 1960 CENSUSES

By: Morris H. Hansen, Leon Pritzker and Joseph Steinberg
Bureau of the Census

I. INTRODUCTION

Our task is to describe the Evaluation and Research Program of the 1960 Censuses of Population and Housing. This program is directed at two broad classes of objectives:

First, the identification of sources of error in census results, the development of improved methods of measurement and control, and the evaluation of the accuracy and cost of alternative census methods and of alternative methods of evaluation. These are objectives for producers of census and survey statistics.

Second, the measurement of both the variance and the bias of statistics of the 1960 Censuses. These are objectives for users of Census data. Such measurement should guide users in the appropriate application of the statistics and should help guide the Bureau of the Census in planning future census programs to meet the needs of users more adequately.

To make clear how we propose to try to achieve these objectives, we will first have to describe briefly how the 1960 Censuses will be taken. There are two main census-taking programs, the "single-stage" census designed for the sparsely populated areas, especially in the West, and for certain sections of the South, and the "two-stage" census designed for the enumeration of the rest of the country, which includes more than 80 percent of the population. (In the "single-stage" census areas, the sample data will be collected at the same time that the complete canvass is made.) We shall limit our description to the "two-stage" census. The following are some of its main features:

1. The complete two-stage census will include a listing of the population and of housing units, and the collection of certain basic data for them. Much of the data collected in the 1960 Censuses will be obtained from a 25 percent sample of households - designated for the sample as the complete listing is prepared. The first stage of the census will be the complete canvass and sample designation. The second stage will have as its goal the collection of the sample data.

2. During the week preceding April 1, the Post Office will distribute "Advance Census Reports" to all households receiving mail. This "ACR," which asks for all the first-stage census information, is to be filled out by the household and held for the census enumerator. The "ACR" asks for a complete listing of the persons in the household - visitors present overnight on March 31 as well as residents whether present or absent - and for data on the characteristics of these persons and of their housing arrangements. The characteristics are limited in number - relationship to head of household, sex, race, month and year of birth, and marital status, and a small number of housing characteristics. These identify the so-called "100 percent" or "nonsample" data.

3. During the first part of April, after being given nine hours of training, three hours a day on three separate days, a corps of about 150,000 enumerators will personally canvass the households in the districts assigned to them. They will search for all places where people live and will seek to obtain a complete listing of people and of housing units. They will visit each household to record the nonsample data on "Stage I FOSDIC schedules." They will transcribe the data from the ACR's that the respondents fill out, but if an ACR is not available or is inadequately filled out, they will conduct the interviews to obtain the data. (Entries are made on the Stage I FOSDIC schedules by position marking - by filling little circles with ordinary black lead pencils. The data on these Stage I FOSDIC schedules will require no coding. The schedules will be microfilmed, and a Film Optical Sensing Device for Input to Computers - hence FOSDIC - will convert the microfilm images to magnetized spots on computer tapes.)

4. During this Stage I enumeration, there will be a formal quality control program with uniform procedures and acceptance standards. The Stage I FOSDIC schedules and the canvassing activity of the enumerators will be checked for completeness and consistency. Mistakes brought to light by this program will be the basis for remedial action.

5. Also during the Stage I enumeration, the enumerators will leave "Household Questionnaires" at every fourth household. The household will be asked to fill out this questionnaire and mail it within three days to the local census office. The household questionnaire asks for considerably more data on personal characteristics such as education and income and on housing arrangements and facilities.

6. The work of the Stage I enumerator will end with the transcription of the nonsample data from the Stage I FOSDIC schedules to the Stage II FOSDIC schedules, which will finally contain all the data collected - nonsample as well as

sample - for the 25 percent sample of households.

7. About one third of the Stage I enumerators will be employed on Stage II. After taking eight additional hours of training, four hours per day on two separate days, the Stage II enumerators will be given the Household Questionnaires - the ones mailed in to the local census offices by the sample households. They will transcribe the sample data from the questionnaires to the Stage II FOSDIC schedules - to the pages on which the nonsample data were transcribed by the Stage I enumerators.

8. The Stage II enumerators will follow-up households from which questionnaires either were not obtained or were incomplete or identifiably defective in some manner. They will complete the Stage II FOSDIC schedules by telephone-interview for partially missing or defective information or by visiting the households from whom no questionnaires were returned.

9. Again, for Stage II, there will be a formal quality control program with uniform procedures and acceptance standards throughout the country.

10. Some of the population data on the Stage II FOSDIC schedules will have to be coded - for example, birthplace and occupation - before the schedules go through the microfilm-FOSDIC-computer process.

This description has left out many important elements of the methods to be employed next April. It will suffice, however, for the presentation of the main features of the Evaluation and Research Program.

II. PROJECT DESIGNED TO MEASURE NONSAMPLING VARIABILITY

Conceptually, we view a census or a sample survey as a repeatable process of measurement. Thus, we will regard the enumeration in April and May 1960 as one trial of a possible set of trials. The mathematical model underlying this view of a census is contained in a paper by Hansen, Hurwitz, and Bershad (1).

We regard the observed value of a particular census or survey statistic - say the number of white males 25 years old and over who have completed high school and who reside in Montgomery County, Maryland - as a value from a potential population of such values, conceived as arising from a set of possible independent repetitions of the census or survey in the given area. Such a census number is thus regarded as a chance variable with an expected value (of which the result actually to be obtained in the 1960 enumeration is an estimate) and a variance. The objective of this evaluation project is to obtain estimates of such variances and their major components for a probability sample of areas of the United States.

The following is an oversimplified account of the mathematical model and of the specific experiments that are being designed.

1. The model identifies separately the variability that arises from independent repetitions of a census inquiry on the same unit of enumeration. This variability, we label the "simple" or "uncorrelated" response variance. Independent replication of the observations would be required to make unbiased or consistent estimates of this variance. This is not achievable in survey work, where the first response must be regarded as, in general, "conditioning" all subsequent responses - although in some cases perhaps to only a minor degree. Consequently, we attempt to achieve approximations.

   We hope to obtain satisfactory approximations to the simple response variance by two experiments. In the first experiment, we will mail household questionnaires to a probability sample of 1,000 sample households. These are households from which, except for some nonrespondents, we have already obtained sample data either from a household questionnaire or from an interview by the Stage II enumerator with the FOSDIC schedules. We will ask these selected households to report to us again. We will then follow the same procedures that are to be employed in Stage II. We will turn over the list of households (and the questionnaires that have been mailed back) to a small group of Stage II enumerators. They will be given the same instructions for transcription to FOSDIC schedules and for follow-up that were employed in Stage II of the Census.

   In the second experiment, a probability sample of 5,000 sample households will be drawn. A small group of Stage II enumerators will be told to reinterview these households with the Stage II FOSDIC schedules. In both experiments, the enumerators will be supplied with the Stage I data for the households, but not the Stage II data.

   A few remarks about these experiments: First we shall also investigate the simple nonsampling variance that arises from independent replication of coding of the same data. In this case, our experiment will conform quite closely to the demands of our model by providing for independent coding operations. Second, for attribute data - 0-1 variates - the upper limit of the simple response variance of a proportion (P) is

PQ/n, where n is the total number of persons on which the statistic is based. Thus the simple response variance can be appreciable only for statistics of small tabulation cells.

2. The mathematical model also identifies the variability that arises from the correlated errors introduced into the census process by (or associated with) enumerators and their supervisors, by coders, and by personnel engaged in other operations. These correlated errors can make a far more important contribution to response variability than the simple or uncorrelated response errors. We shall indicate briefly what we mean by "correlated response errors" and how we propose to measure them.

   The model postulates a sampling process (in general not under our control). In a given trial, i.e., census or sample survey, the particular interviewer assigned to a district, his particular supervisor, the particular coder assigned to work with the schedules obtained from the district, are regarded as having been random selections from a population of interviewers, crew leaders, coders, etc., eligible to work on the particular district. We postulate an expected value of a response for each unit of enumeration taken over all trials (and, therefore, among other things over-all possible selections and assignments of interviewers, coders, etc.) On any given trial, the difference between the response obtained and its expected value is a response deviation. Any tendency on the part of a particular interviewer, coder, etc., consistently to introduce or be associated with the introduction of a systematic error into the response deviations of each unit of enumeration (i.e., person or housing unit) in his assignment will show up as a correlation between the response deviations.

   There is an analogy here with cluster sampling. The correlation has a multiplicative effect on the simple response variance, and what is a trivial effect for the average person or household may become an appreciable effect for the average enumeration district, crew leader district, tract, small town, or other relatively small tabulation area.

A large-scale experiment was conducted as part of the 1950 Census in four areas in Ohio and Michigan - 24 counties and more than 700 enumerators in all. This "Enumerator Variability Study" (or EVS) provided estimates of the correlated response variances associated with the enumerator. The impact of supervisory personnel and of coders was not studied. These estimates were used in constructing estimates of total non-sampling variability - neglecting supervisors, coders, and other sources of correlated errors - of census statistics for small areas. In particular, estimates were produced for a set of areas containing an average of 6,500 population enumerated by 7 enumerators.

The EVS estimates turned out to be substantial and consequently were an important factor influencing the design of the 1960 Censuses. The use of the so-called "self-enumerative" forms - The Advance Census Report and the Household Questionnaire - is meant to reduce the "enumerator effect" on the statistics, as well as the response bias. This will be discussed later.

The 1950 EVS involved the random assignment of two enumeration districts to an enumerator included in a stratum averaging 7 enumerators and 14 districts. This permitted estimates of the between-enumerator variances and evaluations of the statistical significance of the estimates. Some of these results are presented in papers by Eckler and Hurwitz (2) and by Hanson and Marks (3). The estimates of the between-enumerator variances in the 1950 EVS were estimates of the correlated response variances. These estimates were themselves subject to considerable variability because the unit of randomization was an enumeration district - a very large and variable unit.

Now we are ready to discuss the experiment we are designing to estimate the effect of correlated response errors in 1960. We have recognize the importance of this experiment by allocating about $350,000 to it. It is designed to provide estimates of response variance in statistics compiled from the second stage of the Census for areas of various sizes. We should also like to learn whether or not the quality of the statistics will have been improved by the change in the 1960 methods of enumeration, particularly the use of self-enumeration. This will not be measured directly. The principal reason is that considerations of cost and feasibility prohibit any direct comparisons of the self-enumerative approach with a completely non-self-enumerative approach on the same population at the same time.

We have been developing a much more sophisticated experiment for 1960 than the EVS of 1950. Unlike the EVS, the 1960 experiment will have the following properties:

1. It will be based on a probability sample of the entire area of the United States included in the two-stage census.

2. It will provide estimates of the correlated response variances associated with crew leaders and coders as well as enumerators.

3. It will provide more efficient estimates because the unit of randomization will be the household rather than the Enumeration District.

Unlike the two replication experiments described above, this "interpenetration" experiment will constitute the second stage of the census - in the areas in which the experiment is carried out. In these areas, pairs of enumerators will be purposively assigned to geographic clusters of enumeration districts in much the same manner that they would be in nonexperimental areas. Within each cluster, the households included in the 25 percent census sample will be divided into two random groups, one group assigned at random to one enumerator and the other group assigned to the other enumerator. In half the clusters, one of the enumerators will be selected at random to be supervised by a crew leader of a neighboring crew leader district and the other enumerator will be supervised by the regularly assigned crew leader. Altogether the experiment will include 100 crew leaders, 1,600 enumerators, and 320,000 households.

The results of the replication experiments and the interpenetration experiment will be combined to provide estimates of total nonsampling variability of census statistics for average tabulation areas of various sizes (in terms of number of people) and types (urban-rural, tract, city, county, etc.), from about 1,000 to 16,000 population.

III. PROJECTS DESIGNED TO STUDY POPULATION AND HOUSING COVERAGE

We differentiate between coverage and content errors: Coverage errors are the errors in counting people and housing units. These gross errors of under - and overenumeration affect the unclassified population and housing unit totals for the United States and its subdivisions. The net effects of these errors - the net coverage errors - have generally been net deficiencies in the counts.

Content errors are the errors in the tabulated results that arise from errors in classifying people or housing units with respect to the characteristics for which data are collected. Included are errors arising in reporting, recording, transcribing, coding, and tabulating the data. With respect to any class of any characteristic (say, the age class, 10-14, or the tenure class, owner-occupied), a content error results from an improper transfer out of or omission from one class and a corresponding improper transfer to or inclusion in another class. Thus the net content error for a particular class may be a deficiency or a surplus. Over all classes, however, the gross content errors are equal - what is an omission from one class is an erroneous inclusion in another. The algebraic sum of the net content errors class by class is zero.

The gross content error of a given class (the sum of the erroneous omissions and exclusions) corresponds closely under certain conditions to what we have called the simple response variance. The net content error, however, reflects the bias of the census-taking method.

A general discussion of the meeting of bias and the methods of measuring it is beyond the scope of this paper.

But let us now return to the measurement of coverage error. The best available evidence regarding the 1950 Census of Population and Housing indicates that the total population count was deficient by about 2-3 percent, or perhaps more. The problem of coverage is more serious than indicated by this average error rate because we have evidence that some significant differentials existed in the 1950 Census. Young children, nonwhites, young adult males, persons in rural nonfarm dwelling units, all these and some others had substantially higher risks of being undercounted than the general population.

Some of the innovations introduced in the 1960 Censuses represent efforts to reduce the over-all coverage error rates as well as the differentials. The division of the enumeration into two stages is designed to place emphasis in training and in supervision on obtaining complete coverage in the first stage. The Advance Census Report is designed to pin down the enumeration to a particular date and thus help avoid the loss in coverage that is associated with an enumeration extended over time. The enumeration of visitors is designed to cut down on the coverage losses among people with tenuous attachments to households. The enumeration of the transient hotel population on the night of March 31-April 1 will avoid having these people claim they have already been enumerated elsewhere.

Let us return now to the general problem of evaluating coverage. This problem was one of the main concerns of the evaluation program of the 1950 Census -- the Post-enumeration Survey ("PES"). The 1950 PES was the first major attempt to measure the errors of a decennial census. Emphasis was placed on enumerative methods themselves for measuring the errors of the census enumeration. As far as studying coverage error is concerned, the PES was a more intensive enumeration than the Census. The best available evidence is that the PES found about half of the underenumeration of the population in the 1950 Census. The coverage of the PES was especially deficient among those groups for which the risk of underenumeration in the 1950 Census was highest.

This state of affairs has impelled us to seek new methods of studying coverage error and also to examine ways by which the enumerative methods of the PES might be improved. This has led to the development of three evaluation projects for studying coverage error, in addition to a fourth project designed to study directly a method for improving coverage - the Post Office check:

1. "Reverse Record Check" of a General Sample of the Population of the United States

   The objective is to construct an independent sample of the population of the United States as of April 1, 1960, and to determine how many persons appearing in that sample were not enumerated in the Census.

The independent sample will be drawn from four main components:

a. Persons enumerated in the 1950 Census.
b. Aliens who entered the United States after April 1950 and who are registered in January 1960.
c. Children born during the intercensal period and whose births are registered.
d. Persons omitted from the 1950 Census but detected by the 1950 Post-enumeration Survey.

This combined list or "frame" will be incomplete in a number of respects. It will not include, for example, non-registered births during the intercensal period, citizens (particularly Puerto Ricans) who entered the country after April 1950, and persons missed in the 1950 Census who were not detected by the 1950 PES. The original list will cover about 98 percent of the population.

An attempt will be made to determine the current address of each person selected in the sample and then to determine whether or not the person was enumerated in the 1960 Census.

A pretest to determine the feasibility of the method has just been completed and indicates that we can expect to locate the correct addresses of more than 80 percent of the sample selected from the 1950 Census. We hope to be able to increase this by about 5 or 10 percent.

This method has been developed in an attempt to deal with the situation found in 1950, namely that the PES tended to miss some of the same kinds of people who were missed in the Census itself. The independent sample has the virtue of identifying explicitly some persons who in 1960 will be members of population groups which we believe will have the greatest risk of underenumeration. For example, the sample will identify from the 1950 Census a group of persons who were 8-14 years in 1950 - a group where the risk of underenumeration was relatively low - but who are now 18-24 - a group for which the risk of underenumeration is relatively high.

2. Reverse Record Check of Special Samples of the Population of the United States

This project is directed primarily at evaluating the enumeration status and age reports of three special population groups-aged social security beneficiaries, selective service registrants, and students enrolled in colleges and universities in February 1960. The method is essentially the same as that contemplated for the first project. A current address of each person selected in the sample will be obtained. There will be a match against the census records for the area containing the current address to determine whether or not the person was enumerated, and if enumerated, to evaluate the reported age and perhaps other characteristics. The study of selective service registrants will provide us with data for evaluating the coverage of one of the hardest-to-enumerate population groups - young adult males. A sample of college and university students will also provide data on this point, and in addition, will enable us to evaluate the special census rules for enumerating college and university students wherever they are staying while going to school.

3. Re-enumerative Studies of Coverage Error

The major objective of this project is to obtain estimates - by enumerative methods - of the net and gross errors in counting the population and the housing units in the United States. The studies of coverage error that are contemplated will require the use of specially trained enumerators to return to either area samples or samples of enumerated housing units in a search for errors of both omission and inclusion, i.e., missed persons and housing units and erroneously included persons and housing units.

In comparison to the 1950 PES, what may be regarded as significant improvements are being introduced into this project. First, it is planned to investigate the completeness of coverage of the population during the early part of May 1960, a little more than a month following the beginning of the Census. (In 1950, this investigation was delayed until August and September.) Second, unlike the 1950 PES, the investigation of coverage error is being separated from the investigation of content error. This will provide the opportunity for more intensive training of the enumerators used in this study on a more limited number of subjects.

The study of the omission of persons in enumerated living quarters will be carried out in two ways. First, there will be a *de facto* enumeration of persons present in a sample of enumerated living quarters during the early part of May. We will seek to obtain all the possible addresses at which these persons might have been enumerated and to determine whether or not they were enumerated at any of these addresses. Second, there will be the attempt to reconstruct a list of residents of these living quarters as of April 1, 1960, and a subsequent match against the Census to determine if they were enumerated. (Only the second approach was used in the 1950 PES.)

The re-enumerative study of omitted housing units , and consequently, of omitted persons residing in the omitted housing units will also be accomplished in two ways:

First, the Survey of Change and Residential Financing (SCARF) will provide a basis for evaluating the coverage of housing units and households. As a part of this program, a fairly intensive canvass has been virtually completed of a large sample of small areas (segments) in the United States. Lists of living quarters located in these segments have been prepared. These lists will not be used in the Census, nor will the location of the segments be known by the Census enumerators. The area of the segments will, however, be canvassed in the normal course of the Census. A subsample of about 4,000 segments included in SCARF will be re-enumerated during the summer of 1960 in a search for housing units omitted from the Census or included in error. The enumerators participating in the search will have available to them not only the 1960 Census data but also lists of dwellings enumerated in the SCARF program. Their job will be to reconcile the SCARF and Census enumerations and to find the housing units omitted from either canvass.

The second approach is what has been termed the "successor-predecessor" approach and will be carried out in conjunction with the study of omitted persons in enumerated living quarters described above. This canvass will take place early in May 1960. In addition to visiting a sample of housing units enumerated in the Census as described above, enumerators will locate the housing units that immediately precede and follow the enumerated unit. Rules for providing a unique ordering of housing units will be given to enumerators. If, after matching against the Census records, it turns out that either a successor or a predecessor unit has been missed, another enumerator will be sent out to continue a chain of canvass in the indicated direction until an enumerated housing unit is located.

Some special attempts, in addition to the above, will be made to gauge the adequacy of enumeration of the transient population who are to be enumerated in hotels, motels, and other transient quarters.

It should be noted that no one method will be relied on exclusively for providing estimates of coverage error. An important objective of the program is to evaluate alternative methods. Every effort will be made in the analysis to incorporate data from all the methods employed to achieve estimates of coverage error of maximum reliability and validity.

Also we are developing further, but are not yet ready to state firmly our methods for studying overenumeration - duplicate or factitious enumeration. Provision is being made for such study through re-enumeration of samples of enumerated persons and housing units. Several methods will be used.

4. Post Office Coverage Improvement Study

This project stems from a major innovation proposed for the 1960 Censuses that was not adopted because of the lack of funds. This was the use of Post Office personnel to identify households erroneously omitted from the enumeration. The program was budgeted at about 4 million dollars. Our pre-census experimentation indicated that the deficiency in coverage might have been reduced by about one percentage point by this program. We regard this as a worthwhile goal, especially because our evidence suggested that the reduction by this procedure in the over-all coverage error would be greatest in groups more seriously undercovered in 1950.

Thus, although we are unable to employ this procedure at full effectiveness, one of our evaluation and research projects is directed toward additional study of the feasibility and effectiveness of the use of Post Office personnel to improve coverage. The project has the following main features:

a. Within each of the 15 Postal Regions into which the United States is divided, we shall select - by probability sampling methods - an area containing 10,000-15,000 households. The urban part of the sample will consist of areas that are either parts of postal zones or entire postal zones. The rural part of the sample will consist of areas served by several post offices.

b. We shall identify the census enumeration districts - about 50 on the average - which make up each of the selected areas. We shall instruct the enumerators to make up cards showing the name and address of every enumerated household.

c. After withholding a small sample of the cards to provide a basis for controlling the quality of the work performed by the postal carriers, we shall turn the cards over to the local post offices, where the cards will be sorted by carrier route. The postal carriers will then be asked to sort the cards in delivery order and to make up new cards for any households that appear to be missing.

d. The cards for the apparently missed households will be checked against the Stage I FOSDIC schedules. This will provide evidence of the effectiveness of the Post Office in improving census coverage.

IV. PROJECTS DESIGNED TO STUDY CONTENT BIASES

Response variance is an important factor only for small frequencies or small-area statistics. Response bias is the prime cause of error in the important summary measures. The new methods for collecting and compiling data in the 1960 Censuses impose a special responsibility for studying this type of bias. In the content area, there is a strong presumption and theoretical basis for expecting that the innovations in method will reduce the variance of small-area and small-cell statistics. We believe that self-enumeration with enumerator follow-up will reduce the biases as compared with the usual straight enumerator method - primarily because of the time and opportunity self-enumeration allows for members of a household to consult one another, to consult records, and to give more considered responses. Our task, however, is to substitute objective appraisal for opinion and belief.

In the PES of the 1950 Censuses, two approaches to the measurement of content bias were employed - the re-enumerative check and the record check. In the re-enumerative check, a sample of census households was reinterviewed by a small group of carefully selected and specially trained enumerators. The enumerators were supplied with special questionnaires designed to facilitate more careful questioning, and obtained the information from the best respondents instead of any responsible member of the household. These enumerators were more closely supervised. They were paid on an hourly basis rather than on a piece-rate basis as was the case in the original 1950 Census enumeration. The PES enumerators were provided with transcriptions of the original data so that they might, after the initial interview, reconcile discrepancies between the reinterview responses and the original responses. In short, we instituted an improved method of enumeration, which we were willing to regard, before the fact, as being capable of providing us with estimates of bias in the census enumeration. The results of the re-enumerative check indicated, however, that in general the expected values of our improved method were not much different from the expected values of the 1950 Census, in spite of evidence from other sources that, for at least some items, substantial biases existed. The estimates of net error tended to be quite small, although there were some noteworthy exceptions.

The record checks conducted as part of the 1950 PES involved comparisons of 1950 Census data with data on birth certificates, records of the 1920 Census, income tax returns, social security records, alien and naturalization records of the Immigration and Naturalization Service and in the files of the Veterans Administration. These checks were somewhat disappointing because, in general, we were able to find check data for only about 50-80 percent of the persons in the samples we investigated. Our judgment is, however, that, if we had been able to produce unbiased estimates of net error from the record checks, they would have been greater than the estimates obtained from the re-enumerative check.

Our experience with the record checks led us to the view that much more developmental work is required before they can be used as evaluative instruments. We have also been inclined to the view that current records, such as those maintained by employers for persons now working or by schools for persons now attending school, are more promising than historical records, such as birth certificates. We are now considering comparisons with employer records to evaluate census data on occupation, industry, weeks worked in 1959, and similar items.

We are still placing reliance on enumerative methods for measures of bias arising from errors in obtaining data from respondents. Two of our studies are concerned with this. We also plan to do much more than we did in 1950 in the investigation of processing errors. Let us now turn to the specific projects:

1. Measurement of Content Biases in Data Collection

Two studies are contemplated. The first has some features in common with the 1950 PES. Intensive reinterviews will be conducted at 5,000 households included in the 25 percent Census sample. Specially trained enumerators will be employed to probe intensively for the best possible answers regarding the population and housing characteristics of the sample persons and housing units. Some of the enumerators will not be furnished with the original Census schedules. For these enumerators, the data collected in the intensive interviews will be matched with the data collected in the original Census enumeration. Discrepant cases will be sent back to the field for reconciliation. Other enumerators may be given the original data so that reconciliations may be attempted on the spot. The data obtained in the intensive interviews will be coded by specially trained coders so that the results obtained reflect the best that we are capable of accomplishing by enumerative means.

The second study of content error is a match between the data of the Current Population Survey and the 1960 Census of Population. The Current Population Survey (CPS) is conducted monthly on a sample of 35,000 households by the Bureau of the Census and is the primary source of current data on the labor force as well as periodic reports on other demographic characteristics.

The matching will provide additional estimates of bias in selected Census statistics as well as some data on the extent of gross error and on the causes of error.

The Current Population Survey-Census match conducted in 1950 provided useful information for the planning of the 1960 Censuses. It is proposed to improve the method over that employed in 1950 by carrying out field follow-ups on discrepant cases to obtain some information on the causes of discrepancies. The CPS-Census match will be limited to those households included in the CPS sample that are also included in the 25 percent Census sample (about 8,000 households).

2. Studies of Processing Error

Three aspects of the processing of the data collected in the 1960 Censuses will be studied:

a. The transcription that takes place in the field. In the first stage of the two-stage Census, there are two types of transcription - copying data from the Advance Census Report to the Stage I FOSDIC schedule and copying data from the Stage I to the Stage II FOSDIC schedules. In the second stage of the Census, a key element in the enumeration is that of copying the sample data from the household questionnaires to the Stage II FOSDIC schedules. An item-by-item review will be made of a sample of Advance Census Reports, household questionnaires, and Stage I and Stage II FOSDIC schedules to determine the extent to which transcription errors contribute to the net and gross errors and to the "correlated response deviations."

b. Coding of sample data. Two large-scale coding operations are being established to deal with the data on the Stage II FOSDIC schedules. The first - general coding - requires clerks to enter the codes for detailed family relationship, place of birth, migration status, place of work, and income on the FOSDIC schedules. The second operation is the specialized occupation and industry coding. It is proposed to provide estimates of the contribution to net and gross errors in Census statistics arising from coding error, as well as the contribution to response variability arising from correlated response deviations in coding. This will be accomplished by recoding the data obtained in the Census enumeration for a sample of households and comparing the recodes with the original codes. Some of the recoding will be carried out by regular coders and some by coders selected for special expertness.

c. The microfilm-FOSDIC-computer operation. A comprehensive program of quality control of the microfilm-FOSDIC-computer operation is now being developed. This program plus the inherently high reliability of the electronic equipment should assure tabulations that are subject to far fewer errors than those produced by conventional punched card equipment. The computers will undertake the job formerly performed by editing clerks and conventional equipment. Thus the editing of the data will be done uniformly, in accordance with the rules given to the computer. There is a need, therefore, to evaluate the editing rules - particularly the rules for imputing missing data. A study has been established for this evaluation as well as for an overall evaluation of the entire processing system.

V. SOME ADDITIONAL REMARKS

The following features of this research and evaluation program are worthy of discussion in a meeting of statisticians:

1. We recognize that we are checking our own work. This is a grave responsibility to which we have not found any practicable alternative. We shall strive, however, as conscientiously as possible to make available full descriptions of our methods and their limitations as we have been able to determine them, as well as of our results. The data of our evaluation program can, of course, serve as material for independent appraisal of the 1960 Censuses by other analysts. Indeed, this was the case in 1950 where a very interesting analysis of the completeness of coverage of the census was made by Ansley Coale (4), to whom we made available the relevant data on the 1950 Post-enumeration Survey.

2. We shall use various types of samples in the conduct of most of our evaluation studies. It perhaps no longer requires stating that a sample can be used to check on the accuracy of a complete count. This, of course, has been recognized in the quality-control field.

3. We recognize that we have much to learn about methods of evaluation. It turns out, however, that the best time for developing methods of evaluation is during an evaluation program itself - because of the availability of funds, personnel, and census data.

4. In our judgment, evaluations of the methods and results of the 1960 Censuses are not likely to come directly from the results of any single study or project. It is for this reason that we have designated as one of our evaluation and research projects a series of analytical studies.

This project has been identified in order to reserve funds for analyses of the results of both the 1960 Censuses and the evaluation studies. In these studies, we shall examine the strong points and limitations of the measurements of error made in the evaluation studies and we shall undertake some intensive demographic analyses as well as statistical comparisons with results obtained from other sources. The final appraisals of the quality of the 1960 Censuses and of our evaluation efforts will come from these analytical studies.

REFERENCES

1. Hansen, Morris H., Hurwitz, William N., and Bershad, Max A., Measurement Errors in Censuses and Surveys. Paper to be presented at the annual meeting of the International Statistics Institute in Tokyo, Japan, June 1960. U. S. Bureau of the Census (processed February 8, 1960).

2. Eckler, A. Ross and Hurwitz, William N., "Response Variance and Biases in Censuses and Surveys," Bulletin of the International Statistical Institute, Vol. 36, Part 2, pp. 12-35, (Stockholm, 1958).

3. Hanson, Robert H. and Marks, Eli S., "Influence of the Interviewer on the Accuracy of Survey Results," Journal of the American Statistical Association, Vol. 53 (September 1958), pp. 635-655.

4. Coale, Ansley J., "The Population of the United States in 1950 Classified by Age, Sex, and Color - A Revision of Census Figures," Journal of the American Statistical Association, Vol. 50, (March 1955), pp. 16-54.

SOME PRINCIPLES OF DEFINITION IN STATISTICS

By: L. E. Rowebottom and D. H. Steinthorson, Dominion Bureau of Statistics

The process of measurement has been defined as "an expression of the difference in distinguishable qualities or characteristics".1/ The first portion of this definition -- "expression of the difference" -- points to the "how" of the measurement process; the latter -- "distinguishable qualities or characteristics" -- draws attention to the "what" of measurement. Thus, defining what is to be measured is an essential part of measurement. This principle is well known and has important implications for the role of the statistician -- implications which have received considerable recognition in statistical literature and practice. The thesis of our paper is the proposition that in the field of economic and social statistics these implications deserve further consideration and still more emphatic recognition in the planning of surveys and the publication of survey results.

In fact, we would go so far as to claim that a significant gain in comprehension would follow from introducing a concept called "definitional error", analogous to sampling error and observational error. If it is reasonable to take "statistical error" in its broadest sense to mean the degree to which statistics fail to serve their scientific purposes, then it may be useful to consider choice of definition as a significant contributor to statistical error, worthy of theoretically equal status with other types of error.

The explanation of what we mean by definitional error is inherent in the proposition that definitions ought to be suited to the uses of statistics. It follows from this, that to the extent that the definitions employed in a particular statistic are not appropriate to an intended use, their use for the envisaged purpose involves error. Thus, if a particular series of economic or social statistics fell into disuse, both sampling and non-sampling error would continue to exist and statisticians could still debate, perhaps profitably, the nature and extent of these errors of measurement. But in such circumstances definitional error would be non-existent since there would be no possibility of mistaken application. It is through use that definitional error comes into existence. Since our concept of definitional error is so closely allied with use, it could be argued that a more appropriate title would be error of application. However, we rejected this title because it seems to place the onus for controlling it on the user, whereas in our view statisticians have a definite responsibility for minimizing this type of error.

We realize, of course, that operational definition, although limited by what it is possible to measure, should be chosen in such a way as to ensure that what is in fact measured conforms as precisely as possible to the definition of what is to be measured. In fact, definitional error straddles the whole process of economic and social measurement from determination of concept, to formulation of questions, through tabulation and publication. However, we have considered the problems of definition involved in the enumerative process as a part of observational error outside of the scope of this paper. We will later comment on what we consider to be some of the definitional aspects of tabulation and publication. Thus, what primarily concerns us here is the relationship between the definition (or concept if you prefer) of what is measured for a particular purpose, and related definitions (or concepts) that might be used for the same or different purposes.

It is fruitful to examine the various processes whereby particular definitions come into use. In textbook parlance, the problem is simple and the procedure clear cut. The user provides a precise definition of what is required, including an attachment of permissible sampling error, and the statistician measures accordingly. It may be likened to the butcher filling an order of three pounds of sirloin steak, with, if you like, the requirement of "well trimmed" left to the statistician's judgment, perhaps subject to the customer's approval. In our experience this story book description varies from the real world in a variety of ways.

I. The user knows what he wants in a vague rather than precise way. He will likely know his purpose with precision, but not what statistics are required to achieve his objective. Occasionally this uncertainty provokes criticism of users whom it is felt should know exactly what they want. Such complaints are based on a misunderstanding of the field of knowledge of most users of economic statistics, who are primarily concerned with hypotheses to be tested or decisions to be made, and on a failure to appreciate the role that the statistician should play in assisting the user to decide what he wants. Users are concerned with such questions as controlling inflation, increasing productivity or sales, reducing unemployment or costs. When in order to make decisions they require measurements of prices, production or employment, it is unrealistic to expect them to know, as a matter of course, exactly what measurements they want, or to appreciate the vital connection between the definition of the requirement and its measurement. Because the statistician is involved in measurement and thoroughly conversant with its problems, he should be in a strategic position to assist the user in clarifying the nature of the measurement required. For example, consider what the relationship should be between the businessman or the administrator who wants to know for some specific purpose, how many fishermen there are in British Columbia, and the statistician familiar with problems of defining occupation. To the user, "How many fishermen?" is likely to be a simple question requiring a simple answer. To the statistician accustomed to the difficulties of abstracting from reality the question is not simple, and before an answer which is not misleading can be provided, he must have a definition which will at least enable him to classify a person engaged in both fishing and farming. At the risk of belabouring the obvious still further, but we suggest not "obvious" to most recent graduates in statistics, it would be a worthwhile exercise to contrast (at greater length than we can do here) the relationship between statisticians and their clients, and that between architects or doctors and their clients. The latter professions do not make assumptions that clients have the knowledge to make decisions. However, we suspect that economic statisticians not infrequently do, without probing to see whether or not the assumptions are justified. We cannot resist making the facetious but perhaps illuminating comment that long ago society recognized that self-medication could be dangerous to the health of individuals, but not so in the case of statistics.

2. Textbooks generally describe the situation where the user is in a position to create the particular statistics he wants. In contrast, most economic statistics are the products of government and it is seldom that users are in a position to have produced the particular statistics they require. This is not because government agencies are unreceptive to the problems of users, but rather because some related statistic already exists and the modification required is impossible, too expensive, or unique to the particular user and not in the general interest to the point where additional expenditure is justified. Existing statistics must be adapted to the problem in hand, and the role of the statistician in evaluating differences between existing definitions and those required may be even more important than in establishing original definitions. It seems fair to generalize that the adaptation of existing statistics offers greater scope for definitional error than does the development of new statistics.

An example of this type of error arises in the use of city rent measurements. We are frequently asked to provide average rents for a particular city, and discover on questioning that the enquirer is concerned with what rent he may have to pay on moving. Quite aside from the question of applying averages to the problem of an individual, which could be partially met by the provision of frequency distributions, the enquirer has understandably failed to distinguish between the definition of rent relevant to his problem, and the definition of rent on which the statistics are based. In the latter case, the definition is rent paid by tenants for occupied accommodation; our friend requires a definition yielding a measurement of rent asked for vacant accommodation. One only needs to reflect on the difference between rent paid and asked in New York City where rent control is still in effect, to appreciate the possible extent of the dis-service involved in unquestioning provision of the statistic requested. In this case, the service of the statistician is to dissuade the enquirer from using the statistics and refer him to data which conform to the required definition -- the classified sections of daily newspapers.

3. Textbooks seldom describe the situation in which the statistician is not in close touch with many users who select statistics from published documents, and is compelled to make decisions on their behalf. While in such circumstances the statistician will, perforce, and with whatever consultation is possible, select his definitions in accordance with what he considers to be the predominant use, he may be almost certain that the statistic will be used to serve a variety of purposes impossible to anticipate. In these circumstances, it is particularly incumbent upon the statistician to provide precise statements of definition, and whenever possible statistics according to several definitions.

4. In final contrast to the textbook is the happier circumstance in which a set of relevant statistics applicable to a continuing problem has been in existence for some time. Here the subject matter expert and the statistician are thoroughly familiar with each other's problems, and the continuing improvement of definition and measurement may be described as a joint undertaking. However, even in this situation it has been our experience that there is room for improvement. As it was put by one of our colleagues who shall remain nameless, "I have seldom participated in an entirely satisfactory discussion with users". We wonder how many users would substitute the words, "with statisticians"?

It is from such diverse relationships between producers and users of economic statistics that definitional error arises. Even in such a variety of circumstances covering many subject matter areas it seems possible to offer some generalizations about the problem of definition in the field of economic and social statistics. First, economic phenomena are such that a variety of definitions is usually possible. Prices, productivity, production, and employment are complex subjects which defy representation in terms of single definitions, and frequently one would prefer not a particular defininition, but a comparison of results obtained from a variety of definitions. Second, within some range of definitions the user will almost always be indifferent, on the assumption that resultant differences in the statistics will be so small as to not affect decisions. Beyond this range he assumes differences to be significant, and will insist on the requirement of a particular definition.

Whether or not in a given case the classification is correct as between important and unimportant definitional differences, must await the statistical evidence. Thus in the field of price statistics a user may be indifferent as to whether or not the population covered by a consumer price index has an upper income cut-off of $6,500 or $7,000, but have strong convictions about the exclusion of farmers. Third, even in those instances where a satisfactorily unique definition is possible, the statistics serve a variety of purposes. While this difficulty may not exist for private agencies producing statistics for individual clients, it is most certainly the case with governmental agencies producing what have not incorrectly become known as multi-purpose statistics. In this case the words themselves suggest a variance with accepted statistical practice which states that the statistics shall be designed to measure a particular phenomenon and intended for a particular use. This cardinal principle is, of course, based on the obvious premise that different definitions which will produce different results, and not, as implied in the words multi-purpose, that a particular statistic will satisfactorily suit a variety of purposes.

These aspects of definition -- the preference for more than one, indifference to some and concern with others, the multiple use of single statistics -- have implications regarding measurement of definitional error. By measurement of definitional error we mean, of course, determination of the differences between the results of statistical surveys which differ in definition but are otherwise the same in all respects. While it is seldom impossible to know beforehand the magnitudes of definitional differences, and consequently definitional error, nevertheless, to the extent that survey and calculation techniques incorporate a number of definitions, measurements of the magnitudes becomes possible. When such measurement is made, improved awareness follows on the part of both the statistician and the user. Statisticians become more aware of the significance of the definition employed, and users of the potential or realized error attributable to definition.

We believe that measurement of definitional error should be facilitated by continuing effort on the part of statisticians to incorporate a variety of definitions in surveys. We are convinced that understanding of economic events is enhanced by measurement according to more than one definition, and that publication of such results may commonly have a more salutory influence on uses, than publication of sampling error and sources of non-sampling error. This is because measurements of definitional differences provide users with alternatives from which it is possible to select the one most appropriate to the problem in hand, whereas statements of sampling and non-sampling error inform the user of particular aspects of the survey results which he cannot escape if he is to use the statistics at all. Of course, because of limitations of funds, techniques or data, it is impossible to incorporate all useful definitions in any particular statistic and choice between alternatives is both inevitable and necessary. As among alternatives, predominant use and practical survey possibilities will determine the choice. However, it is our contention that, as a general rule, the greater the number of definitions which can be incorporated the more useful will be the results, and the greater the discrimination in use.

Take as an example statistics of farm income and size by type of farm. In attempting to define a farm as being of a particular type, say wheat, it becomes immediately apparent that there is no concensus as to what is a wheat farm. Some say that if fifty per cent of the total sales of a farm is composed of wheat, the farm should be so defined, others say seventy per cent, others a still different percentage. In such circumstances the most useful procedure is surely to type farms according to a number of definitions and let the user fit the statistics to his concept and purpose. It is worth noting the statistics themselves are unlikely to have much impact on the improvement of concept and purpose if the user is confined to one definition.

In the above example, farm income and size are likewise subject to a variety of definitions and in a recent Canadian farm income and expenditure survey, income is being defined in at least ten ways. It is our intention to publish results according to a variety of definitions of farm, income, size, and type. Thus we do not propose to publish an estimate of "the" number of wheat farms in Canada, but rather estimates that there are between x thousand and y thousand wheat farms, depending upon definition. Such statistics will force users to recognize the existence of the problem of definitional error, and as between alternative definitions will enable them to select the statistics most appropriate to each purpose.

It is sometimes argued that the publication of statistics in such variety is confusing to users and that the statistics should be left simple. As will be obvious this is a point of view with which we do not sympathize. Economic events are complicated and while statistics cannot reflect all of the complexities, it is our view that the user should not be misled, by over-simplification of statistical presentation, into believing that he is getting a simple answer to a simple question, when in fact he is getting a simple answer to a complex question, and should be getting a variety of answers which illuminate the complexity.

Perhaps understanding the definitional error by both users and statistical practitioners would be furthered by greater attention to this subject in academic courses in statistical methods, which (as is occasionally pointed out) are too frequently designed exclusively or primarily for the research worker who will be doing his own statistical work. How can the implications of definitional error be taught? Since generalization alone cannot carry us very far in the treatment of definitional error, its implications need to be taught by the case method, that is, by selecting a variety of subject matter fields to demonstrate the relations between uses and choice of definition. We applaud, therefore, the remarks of the three speakers at this Association's session last year on "Desiderata for the Basic Course in Economics and Business Statistics", in particular, the proposed undergraduate course outlined by George F. Break.2/ This course opens with a discussion of the concept of personal incomes and its uses, and goes on to relate statistical methods to uses of these statistics and the various available sources of data. Such training, we feel, would be helpful in preventing the occurrence of some of the harm which can come from unrecognized definitional error.

REFERENCES

I. Richard Stone, The Role of Measurement in Economics.

2. American Statistical Association, Proceedings of the Business and Economic Statistics Section, (Session of December 1958), 329.

XI

THE PLANNING AND FINANCING OF SOCIAL RESEARCH

Chairman, Stuart A. Rice, Surveys and Research Corporation

Programming and Financing Social Research—Paul Webbink, Social Science Research Council

Programming and Financing Quantitative Research in the Social Sciences—Henry W. Riecken, National Science Foundation

The University as a Force in Social Research—Philip M. Hauser, University of Chicago

PROGRAMMING AND FINANCING SOCIAL RESEARCH

By: Paul Webbink, Social Science Research Council

The programmers of this session asked that I "provide some grand perspectives and overview on strengths and weaknesses of present planning of social statistical research ... with special reference ... to ways in which present planning and support fall short of producing a maximum additive contribution to knowledge." The assignment implies a value judgment upon the current state of our research with which I find myself compelled to agree. The reasons why we are where we are, and not somewhere else, will probably not be defined definitively this afternoon, but our general topic raises issues which should actively worry all who are concerned with the future of research in the social sciences. Perhaps we can at least air some of the contributing factors in the hope that this will stimulate reflection after we leave.

I suggest that we begin by ruling out one topic of conversation. This concerns the illusion that somehow someone ought to be able to put together a global program that would tell us who should do what research over, say, the next ten or twenty years. Presumably none of you regard this as feasible or desirable. Aspirations toward programming of this kind, however, do recur periodically. They are usually advanced by persons who reason by analogy from an imperfect understanding of how advances have taken place in the natural sciences, or who when faced with allocating newly available large funds assume that someone should be able to tell them how this can be invested quickly and simply so as to yield gratifying results within a comfortably short stretch of time. I am sure that most of us are in agreement that this sort of programming results either in a set of judgments that were superannuated before they were formulated or that represent the lowest common denominator of the thinking of those induced to participate. The reasons why research in the social sciences, and for that matter in any other important field of knowledge, does not lend itself successfully to programming on this level are undoubtedly evident to all of you.

While I know of no effort at all-encompassing programming that has produced appreciable additions to knowledge, a host of more limited attempts have had significant impacts. You are all aware of the many successes that have been gained within particular government agencies or cooperatively through the mediation of the Office of Statistical Standards. Further major contributions will surely come from its new Price Statistics Review Committee. Many of you know, too, of the planning activities of private organizations such as the National Bureau's Conference on Research in Income and Wealth, the committees on the decennial census of the Population Association of America, and the recently organized Committee on Vital and Health Statistics Monographs of the American Public Health Association. The Social Science Research Council has had research planning as a central concern since its very early years. It may be sufficient to remind you of its recent Committee on Historical Statistics, which advised the Bureau of the Census on the revision of _Historical Statistics of the United States_, or of the Committee on Population Census Monographs, which is planning analyses of recent changes in several major social phenomena, or of the planning activities of the former Committee on Labor Market Research, or of the still earlier work of the Committee on Migration Differentials that significantly affected certain phases of the 1940 census.

Apart from such relatively specific examples, a few much broader appraisals of the state of research have been valuable when done competently and with sufficient thought. Thus the labors of the President's Committee on Recent Social Trends nearly three decades ago had a marked influence on research for many years. The review of statistical programs in the federal government made by the Committee on Government Statistics and Information Services aided in reshaping many of these programs and led ultimately to the establishment of the Office of Statistical Standards. It is cheering that today an increasing number of journals in the social sciences are seriously concerning themselves with the publication of substantial review or appraisal articles.

Research programming, or planning as I prefer to call it, has gone forward under a great diversity of auspices and in considerable amount. Yet it is by nature largely transitory, especially if it is successful, and there are grounds for thinking that research planning is now lagging farther and farther behind the needs of the times. The factors producing this lag seem to me to array themselves in the following way.

First, the increasing prosperity and widening public acceptance of the social sciences are unquestionably a major cause and perhaps the most important. Research planning requires time and may have tangible results only long afterwards, if ever. Today time can often be found only by neglecting other obligations such as deadlines under sponsored projects, by limiting one's participation in exciting and remunerative consulting and public service opportunities, and by foregoing intriguing opportunities for foreign travel. Those of you who have been involved in reviewing research proposals or applications for funds know how often these allocate the time of the supposed senior investigator in bewilderingly small fractions. It would be amazing if some of our most competent friends did not occasionally find that the number of tenths of their time that has been committed far outruns ten times ten. The commitments thus amassed, and the desires of universities that their faculties give some attention to teaching (or comparable expectations of other employers) leave little time and energy for planning anything but additional project applications. Planning of a broader kind confers little recognition and less monetary reward, and may lead to no discernible results appreciated by either the donors of research grants or those responsible for promotions.

These are the realities in which most able research men are caught. Yet if basic research planning is neglected in order to deal with immediate problems, the significance of the research that is undertaken is bound to have less and less relation to long-range intellectual objectives. It is clear that determined efforts must be made to obtain more budgetary provisions both for the planning of individual projects and for longer-range research planning. It is also clear that resolute efforts should be made to enhance the pride taken in planning and the prestige accorded to those who do it well. Progress in these directions, however, will be slow and in the meantime one must rely on the consciences of those competent to plan.

The appeal to the consciences of the competent must be pursued earnestly. Research planning can be done well only by those who have competence, independence, imagination, and rigor of mind. They must be able to rise above the orthodoxies of cliques, the distress that hard reasoning may create for existing institutional or personal programs, and the too widely prevailing evil of kindness that leads one to keep silent about the softness of the work of others.

If those who have active intellectual curiosity and who are dissatisfied with the limitations of existing knowledge are content to press forward only with their own research projects, there is little likelihood that the results of their work will be cumulated in a way that will really extend the boundaries of knowledge. Some planning will be done but if it is left to those who have lesser qualifications it will have little significance or influence.

Research planning will not be adequate either if it is undertaken mainly by those who derive their greatest satisfaction out of thinking on behalf of others. Planning must be done principally by those who intend to devote themselves to work related to that being planned, or who at least are willing to mortgage their own credit and reputations to make sure that the research planned will be carried forward competently.

Efforts have too often been made to "plan" research with the aid of lay "experts" from social action or other interest groups whose knowledge and concern about current urgent problems is not accompanied by knowledge of what the social sciences can and cannot do. Their hopes that one more research project will resolve some immediate problem have rarely been fulfilled, but the effort to plan in terms of their interests or convictions has often consumed time and energy that could have been more usefully employed.

The various preoccupations of those most competent to program research in social fields too often induce them to delegate responsibility for planning to others. The years immediately after World War II brought forth hopes that the planning task could be assigned to research organizations, a large number of which had been recently established at universities and especially at universities that had little settled research tradition. Some of these, like some of the research committees with a longer history, have on occasion done excellent planning. On close examination it turns out, however, that the resulting plans were the work of one man or of a small group of dedicated individuals, rather than of an organization as such. It is significant that few such research organizations have effectively survived the outmigration of their key figures. The organizations have survived but principally as funnels--sometimes as very efficient funnels--for funds to be spent with little regard to any specific ideas or purposes.

A few other types of especially transitory planning should perhaps be mentioned. One carries out the notion that bringing together the most respected scholars in some field and giving them opportunity for a day or two of unstructured talk will lead to something important that has not previously suggested itself to any member of the assemblage. The participants often find these occasions most agreeable but they have little other significance. Then there is the sort of planning undertaken to provide a rationale for fund raising. This may produce excellent summary statements but, unless it reflects a much longer and more purely intellectual process, it too is unlikely to raise new questions or improve the formulation of continuing ones.

Research planning of a productive kind, of which we have currently too little, depends ultimately on an attitude of mind--on the willingness of competent individuals to devote time and effort and reputation to asking for what purpose research is to be undertaken; whether with present knowledge, insight and techniques research yielding more than a new verbalization of opinions is feasible; and if it is feasible, how it can be done most directly and efficiently. Since research has become a respectable, moderately well supported, and highly pleasurable activity, it is sometimes difficult to force oneself to make clear and harshly objective choices. This is all the more difficult because of the great variety of potential customers for research results. Within a relatively few years we have seen not only a substantial expansion of governmental and industrial interest in research but also an avid acceptance of research by hospital trustees, school boards, and representatives of nearly every other major social institution. Their interests have spread, furthermore, from a one-time concern with obvious administrative problems to a yearning for enlightenment on a wide range of human relations problems in both their own operations and their relation to the larger community. These believers in social science have hopefully turned to research for solutions, or at least for information and rationalizations, regarding every conceivable sort of social worry or distress, and their hopes have been effectively exploited by a growing number of self-designated expert problem-solvers.

Among these conflicting purposes genuine courage is required if the social scientist planner is to raise doubts about the validity of some of the activity that is going on about him. Our friends tend to be sensitive, and many of them have payrolls to meet. Though the raising of questions does not imply an adverse judgment, it is too often considered just that. But constructive planning of future research cannot be done if it is assumed that no questions can be raised about what is now being done. It is difficult to offer illustrations that will not immediately suggest to you that your work is under attack but I am sure that several will occur quickly to each of you.

To return to the point with which we began, the present planning of social research falls short of producing maximum contributions primarily because so few choose to labor at this with a seriousness comparable to that with which they undertake their other tasks. I am sure that when someone objectively reviews the history of social research over the past three or four decades he will find a remarkable decline in the recent years in the proportion of time given to critical appraisal of research objectives, of what research in given fields and on given problems is really adding to knowledge, and of the validity of the procedures and techniques that are being used.

Any reversal of this tendency will be caused only by a shift in the preoccupations of those who by their example set criteria of what is regarded as more important or less important among all the activities that we have come to lump together under the label of research. It will not be easy to overcome the disposition in some circles to view with equal approval all expenditures of time and money on activities that resemble research. Nor will it be easy to persuade colleagues that at some point nothing more can be learned, or that nothing is being learned. This is not to argue that there should be interference with anyone's judgment that he wishes to spend his days in pursuits that he finds intriguing and satisfying. Those, however, who wish to be assured that they are participating in a process of intellectual progress do need to reserve time to decide whether progress is indeed being achieved.

What has just been said is not an appeal for a campaign to organize a better apparatus for programming. I have tried to emphasize that planning is largely a highly personal matter. Most of us already spend too much of our limited days in instituting new organizations and mechanisms for coordinating those already in being. Rather than entrust the planning of research to some remote authoritarian body, it is well that we insist upon a continuing multiplicity of planning efforts under a multiplicity of auspices. In these circumstances the arrangements for planning may seem untidy, but the prospect that something significant will result here and there, from time to time, is bound to be greater than if the responsibility for planning is delegated to a single bureaucracy.

My assignment called for some discussion, also, of the effect of problems of financial support on the extent to which contributions to knowledge are currently being made. Here, too, I suspect that implicit in the thinking of some who are concerned with this question is a yearning for a tidier world in which the choice of sources to which to turn would be simpler, funds would be more easily available, and everyone would be able to get the financing that he needs at the moment when he could best use it. This would indeed be nice but it isn't going to happen, and might not in the end accelerate real intellectual progress.

This is not to say that improvements in arrangements for the financing of research are not desirable. But I would urge that this is not the most grievous problem that now confronts us and that if more competent attention were given to the programming of research, some of the current financial problems might be much more easily solved. Such attention might help materially in stimulating and maintaining the interest of sources such as the private foundations, whose lay boards of trustees are bound to wish for a sense of progress in some comprehensible directions. These directions will not always coincide with those that we consider most important, and vexatious difficulties in communicating the importance of certain types of research will certainly remain with us. It may be not irrelevant to note that communication is itself a problem with which one must deal in the planning of research. It does happen, regrettably, that the results of planning are sometimes verbalized in ways well characterized in a review that recently came to my attention. It commented: "The contributors, with a few notable exceptions, show the occupational characteristics of academic persons who are maintaining their dignity before their colleagues; their prose is turgid and elaborate, they conceal quite simple ideas in a vast apparatus of long words inconveniently arranged, and they sometimes establish their superiority over their readers and listeners by the unnecessary use of rare words." It is worth keeping in mind that few, if any, private foundations have ever been chartered solely for the support of research, especially for the support of social research. Here, too, a competitive situation exists in which other forms of activity than the financing of a random collection of research projects are bound occasionally to seem more appealing. Many acts of faith have been performed by foundation trustees, but their faith, like that of everyone else, needs fortification from time to time. I suspect that some of the financial problems of social science research would most easily be overcome by nothing more than a few more good pieces of work whose findings are presented in prose that is relatively generally understandable.

PROGRAMMING AND FINANCING QUANTITATIVE RESEARCH IN THE SOCIAL SCIENCES

By: Henry W. Riecken, National Science Foundation

When I was invited to take part in this program, I was asked to represent the point of view of a government foundation on the subject of programming quantitative research in the social sciences. The National Science Foundation has consistently stated its mission and its program to be the assistance and promotion of the development of basic research in the sciences. This view is just as true in the social sciences as elsewhere, and applies equally well to quantitative research as to any other kind. While our program is to assist those well-conceived projects that promised to add to scientific knowledge and competence, we feel that the Foundation staff should leave the choice of topics, as well as methods of research to those best qualified to judge--namely, to the scientists who are actively engaged in research work. In other words, the program of the National Science Foundation in the area of quantitative research in social science is to give as much assistance as possible to the best research projects that social scientists propose to us.

Obviously there are some criteria by which one project is judged more meritorious than another, and one of the most important of these criteria is the extent to which a given project promises to enlarge the amount of verified knowledge, and/or to advance the state of the art. Now the time is not yet at hand when social scientists can dismiss good qualitative, descriptive accounts of social phenomena. There is still plenty of need for non-quantitative research and there is plenty of it being done--more in some social science fields than in others but still a good deal of sound, respectable work that does not involve any quantification beyond simple statements of relative magnitude or relative frequency, i.e., "more than", "fewer than".

The case for the importance of non-quantitative research can be made most vividly, I believe, in ethnological and social anthropological research where there are scores of small, pre-literate societies that are as yet unstudied or inadequately described. Many of these have apparently unusual forms of social organization, special features that seem to be responses to particular environmental conditions or past history, and accordingly are informative of the plasticity of human nature and the ingenuity of man. Furthermore, many of these tiny societies are disappearing under the expansion of Western society and the "development" of underdeveloped countries. The case for non-quantitative studies can also be made in other social sciences, including sociology and political science where systematic and accurate descriptions of the structure and behavior of a variety of organizations are needed and are indeed being done. The so-called "case-study method" of collecting and arranging data of predominantly non-quantitative sort is extremely useful in these fields as a source of sheer information about social phenomena, of insights into normal processes of development, maintenance and change, and, ultimately as sources of hypotheses about human behavior in society and general theories of social processes. Even in so highly quantified a subject as economics, the case study is still a standby--for example, in the investigation of the effect of patented inventions upon technological change, economic development and productivity.

All this is by way of saying that the social sciences, like the biological sciences, still have much to learn about the objects of study and still need sound, systematic verbal descriptions which incorporate little or no quantitative material. For this reason, any program of support for social science research cannot be soundly based on a one-sided policy of giving funds for quantitative research and withholding them from non-quantitative inquiries.

At the same time, it would be just as unwise not to recognize that some of the most exciting work in social science, and some of the most promising developments are taking place in connection with quantification of one sort or another. The growing edge of the social sciences, as I see it, is in the development of quantitative methods for collecting and analyzing data. The course of development is very uneven, with various fields in the social sciences taking part at rather diverse levels of complexity, because some fields have solved or have confronted particular kinds of problems of quantification earlier than others.

I should like to try to illustrate some of the current trends and immediate concerns that various social sciences exhibit in each of three topics: measurement, data collection, and data processing.

Measurement. Perhaps the most persistent and irritating problem in many areas of social science has been how to assign numerical values to various manifestations of phenomena. A good many things that interest social scientists are extremely difficult to measure, such as sensations, opinions, utilities, convictions, aggressiveness, and group cohesiveness. Another way of saying the same thing is that social scientists have chosen to interest themselves in matters that are extremely difficult to measure or unmeasurable, depending on how easily one gives up. The obstacles to measurement have been handled, variously, by substituting the counting of frequencies, by trying to invent mensurational techniques, or by simply ignoring the problem as the bumble-bee has often been said to ignore the laws of aerodynamics.

A good many very serious and able people have not given up trying to measure, however, and their work on basic problems now will, hopefully, provide the basis for new kinds of quantitative research in the future. A major line of development here has been in scaling, where the attempt is to develop mathematical models of the properties and behavior of attributes that are of interest to social scientists--especially such attributes as attitudes, traits and other individual characteristics, as well as attributes of groups and collectivities. Without trying to go into the details of any of the scaling models, I will simply say that the general procedure is to look for an underlying order, usually a multi-dimensional order, in terms of which the relative magnitudes of observable phenomena can be estimated. The mathematical model is usually derived in part from masses of observations and must, of course, be evaluated or tested against further data--both phases requiring a heavy investment in quantitative data.

Economists have sometimes excited jealousy among other kinds of social scientists because economists seemed to have easily defined and readily observable units for measurement or counting, such as dollars, tons, miles, employees, and other neatly quantifiable observables. To be sure, economists have shown a good deal of ingenuity in constructing combinations of quanta, such as index numbers, that have served them well. But the measurement of preferences and utility has been a persistent problem, attracting a good deal of somewhat baffled attention. Lately, some fundamental mathematical work has been done (partly under NSF support) that relates the measurement of such diverse phenomena as sensation and utility. It promises to lead to further developments in quantitative research in economics.

I will not try to discuss the measurement problems encountered in anthropology, sociology and political science except to note that a great deal of the quantitative work in these fields is still necessarily based on simple frequency counts, a fact that may account partly for the popularity of non-parametric methods of analysis among researchers in these disciplines.

Collection of data. The problems of social science in collecting data are of a vastly different order than those of measurement, being both more and less formidable--less formidable because the difficulties today are not as intellectually taxing as measurement problems are, although still requiring considerable ingenuity and imagination, and demanding rigor of procedure as well as clear-headedness. Nevertheless, I think it is fair to say that the collection of data has been reduced to fairly mechanical procedures owing largely to the installation of agencies whose task is to keep track of the occurrences of a vast range of social, economic and political phenomena; and owing also to the enormous refinement of sampling methods and survey techniques. The precision of modern statistical surveys is no longer in doubt (provided that what is being surveyed is measurable!), and this tool is perhaps the greatest single methodological asset in the social sciences.

The formidable problems of data collection are, rather, in the realm of financing and manpower. I shall do no more than allude to the present level of expenditure by the federal government and private agencies for the collection of statistics nor will I try to estimate the number of man-hours given to planning and executing surveys. Both figures are large and I am sure many people feel they could usefully be larger. It does seem, however, that no matter what data are gathered by the recurring surveys, there is always some potential consumer who wants an answer to a question about which data have not been collected. Many of these guilty parties are research social scientists who are interested specifically in new questions--questions that may not have ever been entertained before. Such new questions are likely to be annoying because the existing data don't quite tell what is needed; and also, in the long run, to be expensive because the persistent researcher is going to want a special survey of some sort. It may involve specially trained interviewers, a specially weighted sample, repeated interviews over time with the same respondents, or inquiries from people such as: hard-to-find at homes, bank presidents, blacksmiths, ex-convicts, or other folks with unusual stigmata. In any event, the survey is bound to be troublesome and costly. Yet the results, if obtained, might throw some light on a significant methodological or theoretical problem. On the other hand, they might not and we might just have to write off the cost as part of the risk in doing basic research. In any given case, we cannot be sure a priori, how it will come out (and that may be the best reason for doing it) but if we are serious about supporting basic research in the social sciences, we are going to have to take some risks of this sort.

There may, however, be some ways of reducing the cost of these risks, so we can take more of them within the limits of the research budget. It is not entirely clear whether the existing facilities for collecting data, either routinely or by special surveys, can be more fully and efficiently used for research purposes or how this increased efficiency could be brought about. It is possible that social science research needs new and specialized facilities for conducting surveys. It is conceivable that some research questions can "hitch-hike" or "piggy-back" on existing surveys and samples. Perhaps some collaborative arrangements that now exist can be extended to a wider group of social scientists. These matters need further discussion and exploration.

Data processing. One outstanding characteristic of social science data is the very great quantity that can be (and usually is) collected. A graduate student on his first research project typically comes home staggering under a burden of questionnaires and, for the next weeks or months, struggles to avoid drowning in the details. It is perfectly amazing how many different items or observations appear, beforehand, to be essential, important or just "interesting" to know. (It is this last adjective that seduces the inexperienced researcher). Perhaps this deplorable tendency on the part of some social scientists results from the inadequacy of the theoretical frame of reference that is supposed to be guiding their inquiry and, hence, from insufficiently sharp or pointed problems. Perhaps too it is due to the fact that the social researcher usually gets only one shot at the subjects of his inquiry. Unlike the biologist, physicist or chemist, the social scientist does not have very good control over the experimental material and cannot easily order up a duplicate batch of it if he has failed to notice something important the first time around. In this respect the social scientist is in somewhat the same position as the astronomer, only, unfortunately, he is nearer to his material and can see its details more clearly. Perhaps we ought not to criticize the social scientist for taking his one shot with a shotgun.

Of course, there is presently on hand a machine that takes some of the tedium out of processing these vast quantities of data. High speed computers seem to have captured the attention of most quantitatively-inclined social scientists and there is no doubt that they present many advantages. Besides being useful for fast calculation, for storage and retrieval of information, computers extend the problem-solving range. They permit exponentially larger numbers of permutations of variables, so that the full range of relationships can be explored, and they readily assimilate enormous numbers of observations, when these can be obtained. This latter advantage seems to be especially noteworthy in economic research where complete series rather than samples are employed in analysis of fluctuations or cycles of economic phenomena. Finally, computers can be usefully put to work in simulating elaborate and lengthy sequential processes that, heretofore, either could not be handled directly because of their size and complexity or had to be treated by averages or other approximations.

All of these advantages computers undeniably possess and there is much promise in them for the development of quantitative research. But their very advantages are also threats, for the computer's speed and capacity seem to be a substitute for thought. In this sense, they encourage the investigator to use a shotgun of bigger bore, to dismiss forethought, planning, and the development of a sharply focussed set of questions in favor of collecting all the information he can and letting the machine sort it out. As yet, I think, this threat has not made itself felt to any great degree, but it seems to me too serious and near a possibility that social scientists (among other kinds perhaps) may be tempted to let machines do their thinking for them. It is entirely possible that the development of the social sciences may be impaired if we adopt the easy faith that we can discover truth by running through all the combinations of observations that happened to be made.

Let me turn briefly to another aspect of my assignment and take this opportunity to report some data recently released by the Office of Special Studies of NSF ("Federal Funds for Science", covering fiscal years 1958, 1959 and 1960). In general, these figures speak for themselves and they speak a simple message.

In 1958 the total outlay for research in all fields of science by the agencies of the Federal government was almost one billion thirty-four million dollars. Obligations for research in social sciences in 1958 amounted to nearly forty million or about four percent of the total. The corresponding figures for 1959 were one billion, four hundred and forty-three millions total, forty-eight millions for social sciences; and for 1960 the estimates are one billion six hundred and two millions total, fifty-eight millions for social sciences. This estimate covers all topics in social science, includes both quantitative and non-quantitative work (but excludes routine data-collecting operations, testing, surveying or the collection of general purpose statistics). In the last three years, then, social sciences have had a slow but steady growth in absolute amount of funds available, but have not advanced in terms of proportion of total research funds. One further fact is of considerable significance: namely, that only about 23 percent of the funds available for social science research are being devoted to basic research (defined as: "where the primary arm of the investigator is a fuller knowledge or understanding of the subject rather than a practical application thereof"). This contrasts with 47 of the funds in biological sciences and 57 percent of funds in the physical sciences being devoted to basic research. This state of affairs probably reflects the great demand for practical advice and guidance in the solution of social problems--a demand that, in the view of some of us, is not always a help to the sound development of sciences of human behavior. It would seem reasonable, in such enterprises as economics, sociology, social psychology, anthropology, to devote a larger rather than a smaller share of resources to "a fuller knowledge or understanding of the subject" than is the case in the more advanced disciplines of the physical sciences. Whatever may be the extent of our efforts in this direction, I feel sure that the NSF program will include the support of both quantitative and non-quantitative work, where each type is relevant to the development of a sounder more precise and more powerful set of social sciences.

THE UNIVERSITY AS A FORCE IN SOCIAL RESEARCH

By: Philip M. Hauser, University of Chicago

This is a subject which warrants planning and financing as an important area of research into the programming and administration of social research. But this paper is not the product of such investigation. It is an impressionistic overview of developments based on some three decades of participation in, and observation of, efforts to plan, finance and conduct research in the social sciences.

Permit me to set forth my conclusions first, and, then, to proceed to elaborate upon them:

1. The University is becoming a less, rather than a more, important force in determining the general direction of social science research and in the planning of specific projects.
2. Non-university bodies--government, foundations and business--are becoming more important forces in determining the direction and character of social research.
3. Social research has been channeled increasingly into "social problem" areas --particularly into the areas of concern to sponsors and sources of funds--rather than to basic social science problems.
4. There is great imbalance in social science research problems reflecting more the availability of funds than of research needs.
5. There is need for a careful consideration of alternatives in the planning of social research and the role of the university in the process.
6. There is need for research personnel to take more initiative in the planning of research, if not entirely to take the initiative, at least to reverse the trend and to achieve a better balance than now exists.
7. There is need for an increase in free research grants--grants that will permit social scientists to follow the leads that emerge in the conduct of research rather than the decisions of personnel at some remove from the actual research process.

The University. The university, without question, has been and still is the most important agency in the conduct of social science research. Since the 1920's, however, it has probably been becoming a less important force, compared with fund granting agencies, in determining the direction and the character of the research. Several factors have contributed to this trend.

Among these is the fact that the great universities, and certainly the privately endowed ones, have become relatively impoverished over the years. Increased income has, in general, not kept pace with inflation and increased costs. In consequence, the universities have not been in a position appreciably to increase their own support of social science. Moreover, their relative poverty has forced many institutions of higher learning to seek social science research funds partly to help to maintain or to expand social science faculties.

In the great state universities which have adopted social science research as an important part of their mission, problems of budget balancing have been less acute than among the private universities but funds for social science research have not been as easy to get as funds for new construction or expanded faculty. The latter has resulted in greatly increased demand for social science research funds and has led the state universities to increase their competition with the private institution in the search for research grants.

The net effect of the development of the past three decades or so, has been to make both private and state universities greatly dependent on outside sources of funds, not only for the conduct of research, but, also, in part for the expansion or maintenance of faculty or for the enhancement of the prestige of rapidly expanding faculties faced with enormous student bodies and excessive teaching loads. Motivation in seeking research grants undoubtedly, in part, has therefore contained elements other than those provided by the hot pursuit of exciting leads in the development of science.

Another factor that has contributed to the diminution of the role of the university in the planning of research is the failure of university administrations conceptually to keep up with the changing fiscal requirements of social research. "Social science" in the 19th and early 20th centuries was largely a form of "thinking", soft as well as hard, rather than of empirical investigation. Thinking does not require large budgets. Moreover, tradition and accepted budget standards provided relatively little for social science teaching or research compared with natural science outlays. Since the first quarter of this century, however, the social sciences have become much more a body of empirical research disciplines and much less branches of social philosophy. This transition--one which calls for greatly increased expenditures per student, both for teaching and research purposes, is not yet reflected in university budgets. It has yet to be recognized by university administrators that social data are often more expensive to come by than physical science or biological data--and that social science laboratory requirements may be as expensive as natural science ones. Without question the relative impoverishment of universities, and again, especially of the privately endowed ones, does not accelerate the disappearance of this evidence of cultural lag.

Factors such as these, then, have made the universities more dependent than formerly on "outside" sources of funds for social science research purposes. This development along with changes in the availability of research funds and changes in the administration of such funds have increasingly placed the planning and direction of social research in non-university agencies--in the foundations, in government and in business hands.

The Sources of Funds. The funds available for social science research have, of course, tremendously increased since the 1920's. Foundation funds for social research have been swelled by increased income derived from the boom economy generated by the post-war cold-war climate and, of course, by the advent of the gigantic Ford Foundation. Government funds for social science research have also greatly multiplied and, especially, funds available through the military establishment, the various institutions of health and the National Science Foundation. Finally, business has greatly increased its expenditures for social research both by making such research part of its own activity and, in part, by means of contractual arrangements. The great increase in resources for social science research in government and in business should, at least in part, be gratifying because it reflects in each instance increased recognition of the practical benefits to be derived from such research.

But the broadening of the base of support and increase in magnitude of funds available has not been without its cost. For social science research has more and more reflected the interests of the grantors of research funds than of the interests of the investigators. The foundations, not without justification, have developed "programs," largely in problem areas by means of which in concentrating their largess they have sought to maximize, and to have discernible, impact. Government sources of research funds similarly, have tended to focus on specific agency missions such as defense, health, housing, urban renewal or improving agriculture and rural life, with the notable exception of the National Science Foundation which is beginning to develop sizable social science research support. But even the National Science Foundation affects the selection of social research by necessarily having to shy away from politically "hot " questions or "soft" areas not likely to have relatively early and demonstrable results. Private business is, of course, in business, and social science research supported by this source is generally definitively aimed at company or industry problems.

These developments have tended to erode university influence in the planning and management of research in several ways. First, each of these sources of research funds has, over the years, expanded its staffs and procured knowledgeable social science personnel to help develop and plan research areas, assist in making grants, and, in varying degrees, "monitoring" the actual conduct of research. Second, by focusing on areas of research activity, they have wrested the initiative in selection of research problems from the personnel actually engaged in research--usually university personnel. Third, by having preferences for specific approaches, methods, inter-disciplinary or multidisciplinary projects, not to mention universities or specific personnel, they have increasingly affected the design and conduct of investigations and, therefore, the direction of development of the social sciences.

Social Problem Areas. The influence of the grantors of research funds is perhaps no place more evident than in the selection of problems for research. There has been an amazing increase in the proportion of research energy, for example, devoted to defense projects; health and especially mental health projects; metropolitan area studies; economic development studies, especially those requiring foreign travel; and industrial relations, business organization and consumer markets. The proportion of social science personnel "bought" by funds in these applied areas, must in the post-war situation have increased enormously over previous periods. Is the present allocation of research personnel the optimal one either from the standpoint of social engineering or social science? Does the multiplication of specific research projects in these areas reflect the needs of sociology or political science, or economics, or of any of the social science disciplines, for making effective progress in the development of these various disciplines as sciences?

There is a great need for knowledge which can serve social engineering purposes. It would be absurd to say that applied social research should be discouraged or even curtailed. But it is possible, as in the natural science field, that the knowledge most useful for social engineering purposes may come from basic rather than applied research. Too much concentration on research into social problem areas may retard rather than advance the cause of social engineering as well as of social research.

Imbalance. Not only is social research increasingly concentrated in social problem areas but, also, in selected problem areas reflecting specific interests of trustees of foundations, government agencies or business needs. The result is an allocation of research resources to social problems that is heavily imbalanced by almost any criteria that might be used to determine "need". Many social problems are less significant or less acute, but, rather, because they are too "hot" to handle from the standpoint of present political, class, moral, religious, or ideological conviction. For example, relative to studies now being supported, are there sufficient researches on problems of racial integration, birth control, the integrity of advertising, the role of mass media of communication in elections, the effects of deleterious agricultural and industrial practices on health and mortality, or the consequences of U.S. foreign policy?

Need for Reconsideration. Now, I am not arguing that the influence of research granting agencies is all necessarily deleterious; nor that the grantors do not possess the right, or even the obligation to exert influence on research. But I am contending that there is need for careful evaluation of the effects of such influences on the development of social science, in general, as well as on the quality of specific research undertakings. It is difficult for me to believe that the transfer of the initiative in the selection of research problems in the design of the research, or, in the unrestrained conduct of the research is really conducive to serving the best interest of the fund granting organization, let alone the furthering of social science or the development of the social scientist.

Moreover, it does not necessarily follow that because the research is social problem oriented that it cannot be well designed research that contributes to the development of science. Furthermore, it may be argued with justification, that even imbalance in research into social problem areas may be justified by efforts to achieve breakthroughs by concentrating research resources. But it is unlikely that the path strewn with research grants is necessarily the most direct path to the improvement of the various disciplines as science, or even to the solution of the problems with which trustees of foundations, government agencies, or business leaders may be concerned.

No one has deliberately planned the way in which direction is at present being given to social science research. It would seem appropriate therefore that, with the increasing volume and importance of social research that some rational decision making with respect to the direction of social science research may not too much impair or retard the complex processes by which social science proceeds and, let us hope, advances.

Planning. Although the planning of social science research has been slipping away from the university faculties to the agents of fund granting organizations, there have been some countering forces. Some of them are considered in another paper at this session (Paul Webbink, "Programming and Financing Social Research"). They consist in the work of such bodies as the Social Science Research Council, the professional societies, and various independent research bodies such as the National Bureau of Economic Research. These organizations have increasingly devoted, or increased their inputs of, energy to plan and to give direction to research. Their activities may be regarded as countering forces to the trends described, because they uniformly involve the mobilization of professional research personnel--largely university faculty--thus enabling them more directly to take a hand in the research planning process. These activities provide dispersed and unorganized university personnel with means to make their individual and collective voices heard on the research front. The activities to which Mr. Webbink has referred include notable achievements in giving direction to needed social research.

One corrective, then, to the increasing influence in research planning of fund granting agencies may lie in the increased use of research personnel by the Social Science Research Council and professional societies to provide direction for research. But such activities are, also, dependent on financing, often from the same fund granting agencies. These agencies, and especially the great foundations, might do well deliberately to support activities which help more effectively to organize research personnel to counter their own influence in giving direction to research. Such a course of wisdom is not likely to be inconsistent either with the desire of the foundation to illuminate problem areas in which they are interested nor the advancement of the social disciplines as sciences.

Free Funds. The most important corrective factor to the trend described undoubtedly lies in the increase of "free funds" for research. That is, an increase is needed in funds available to competent research personnel which they are free to use at their own discretion in following leads which emerge as they do their research. Such research activity is likely to be more fruitful in the development of the social sciences than the research directed to social problems.

Some awareness of this need is evident in free grants by foundations to individual scholars, but such grants are pitifully small, at present, relative to the "program" type of grants. Appreciable increase of free research funds for freely conducted research could go a long way toward restoring the balance between universities and fund granting agencies in giving general direction to social science research.

Concluding Observations

During the 1920's, as social science was turning away from its beginnings as social philosophy to empirical research, unrestricted research grants did much to promote both social research and the development of social science. With the impact of World War II and the pressing nature of post-war problems, there has been a great widening of the base of support for social science research in the foundations, in government and in business. But one of the effects of the broader base of support and the increase in funds for social research has been the tendency to transfer the planning of research from those who conduct it to the agencies which finance it. Social science research has, in consequence, become much more problem or social engineering oriented. These trends have been exacerbated by the relative impoverishment over the years of privately endowed universities and the rush to expand social research output on the part of rapidly expanding social science faculties of state universities.

There is need for serious consideration of the most effective means of planning and giving direction to social science research so that, on the one hand, the needs of social engineering are met, and, on the other, the social sciences are stimulated to further development as sciences. More specifically, there is need to achieve a better balance in the planning and management of research between the agencies which finance it and the research personnel which conducts it.

XII

HOW THE 1960 CENSUS WILL BE TAKEN

Chairman, Conrad Taeuber, Bureau of the Census

THE ENUMERATOR RINGS THE BELL

by
Lowell T. Galt
and
Robert B. Voight

To statisticians it is axiomatic that not too much can be done to improve data through refinement or manipulation once they have left the pencil of respondent or recording interviewer. If this be accepted, it follows that those interested in census data will do well to take stock of the enumerator, that front line participant in the census taking process. As she mounts the steps and rings the bell, what does the census enumerator bring to the important task before her?

At one time during the Korean War we were told that it cost American taxpayers $14,000 to put a combat infantryman into action, this including direct expenses of indoctrination, training, equipment, and transportation and the indirect expense of logistic support. While it costs considerably less to prepare an enumerator for her rounds, it does represent a significant part of the Census budget and the elements of this preparation may be worth attention of those who expect to use the statistical product.

First, you should understand that it is not by accident of circumstance that there is an enumerator there to ring the bell. Few citizens point for a career as census takers and many of those approached in 1960 will feel themselves too preoccupied with day to day concerns, and often profitable part-time work, to take on full-time but temporary duties for two weeks or so. You may be sure that a vigorous recruitment campaign will have been necessary to produce that able and businesslike enumerator on your doorstep. Diligent and dedicated effort, often on the part of a number of recruitment sources with appeals to patriotic duty, party loyalty, community and personal self-interest will be needed to produce enough qualified candidates for selection. The economic situation also explains why we refer to the enumerator in the feminine gender. Most of them we expect will be women, since employed heads of families can't afford to take such temporary posts.

You may not be aware that the enumerator who calls on you has passed a written examination designed to test her ability to understand written instructions and to read a map. This latter talent, incidentally, does not necessarily equate with literacy or a high I. Q. Its importance to clean-cut census taking makes it a "must" requirement. You can be assured that she has been sworn to hold in confidence the information which you provide the Census and you may expect her to wear her identification badge and carry her official portfolio. She is likely to be a resident of the area of her first assignment.

Your census enumerator will have had the benefit of a training program which is tailored more to her needs and specific responsibilities than in previous censuses. Two major shifts in procedure have made this possible, the first being separation of the Census of Agriculture by several months in point of time so that enumerators next spring will have only the job of finding and reporting the population and the units which house them. The second change provides, in all but sparsely settled areas, a two-stage procedure of enumeration, the first concentrating on coverage and complete recording of items asked of all households and a later stage involving only a third of the enumerators in completing more detailed inquiries at a sample of units. This separation allows for specialization in training with more time devoted to method of canvassing, listing, recording sample identification, and reporting in Stage I. Similarly in Stage II it is feasible to concentrate on sample question concepts and follow-up procedures, building, of course, on enumerator's experience obtained in working through the initial stage.

Each enumerator in Two-Stage areas will receive nine hours of classroom training before embarking upon her initial assignment. It will be given in three-hour sessions on three successive days, allowing time for absorbtion. It will consist of oral presentation--exposition and reference reading, filmstrips and recorded interviews, class discussion and written practice exercises. In designing the program, the Bureau benefited from the advice of professional trainers on the most effective aids in forwarding the learning process. An attempt has also been made to take full advantage of the experience of prior censuses and pretests.

In addition to instruction on concepts and definition of population questions and housing items, the enumerator learns why we take a census, what her job is like, how to locate every possible housing unit and to assure herself that she hasn't missed persons within any household, how to check Advance Census Reports for complete and consistent information, and how to mark and care for schedules so that they will be properly read by FOSDIC, the electronic equipment which translates the data she has recorded through microfilm print to coded computer tape. She learns how to keep within her assigned territory, systematically canvassing, listing, and designating sample households in order of visitation, how to leave the Household Questionhaire at sample units, and how to transcribe information to FOSDIC schedules which will later be used in second stage operation. In order to make her work easier and more profitable (remember that most enumerators are paid on a piece rate) she is taught how to organize her materials for field work, how to manage callbacks, how to handle difficult respondents, and finally how to close out efficiently and with maximum high response.

She is further informed of the fact that her work will be periodically reviewed by her crew leader or (his assistant) field reviewer and is told what her responsibilities are in connection with each such review. She knows that a further sample check of quality will be made in the field office before her work is certified for payment. She also knows that she may be released if her work is not satisfactory. Finally, she is fortified with knowledge of her rights under the law as a census taker, together with the obligations she has assumed under oath.

As she begins her assignment your enumerator will know that if she does well and finishes her appointed task with dispatch, she is likely to be tapped as a second stage enumerator, or to clean up incomplete assignments, or, if she lives in or near the field headquarters city, as a clerk in the office on editing chores. If she is nominated for second stage enumeration by her crew leader and appointed after satisfactory check-out of her work, she will be given eight more hours of training on the detailed sample inquiries. This again will feature trainee participation. After viewing filmstrips and studying definitions, enumerators will practice reviewing Household Questionnaires which have been mailed in from their assigned area, will learn how to mark acceptable entries on the FOSDIC sample schedules and how to telephone or call at an address in person to complete information which has not been adequately furnished.

If you live in a more sparsely settled part of the country where road condition depends on weather, where levels of literacy and education are not as high and where scarcity of private or limited party telephone service rules out this means of follow-up, you may expect your enumerator to cover both 100% and sample questions (if you are the lucky fourth) in a single visit at your household. For this traditional census procedure, she will have trained four hours on four consecutive days, the content combining that taught in both sessions in the two-stage areas.

So much for your enumerator's personal preparation. Now for her logistic support which in this census has required planning in intricate detail.

She will report to one of 6,000 training locations which will be equipped to provide satisfactory classroom environment with provision for audio-visual aids. This space will have been donated usually by a school system, county or city government. It will be shared by two crews scheduled for morning and afternoon sessions. Here the enumerator will receive her portfolio weighing some 14 pounds fully packed with listing book, FOSDIC schedule books, blank questionnaires for sample household with "extra person" inserts where needed and envelopes addressed to the field office. In addition there are individual census report forms for visitors enumerated away from their usual place of residence, and miscellaneous forms such as those used to give notice of close-out after several unsuccessful attempts at different times on different days. In such case, the enumerator will have secured as much of the 100% information as possible from neighbors, building superintendents, or other knowledgeable persons and will leave the form at the address indicating how many persons have been reported, inviting correction or amplification of fact. A parallel close-out form will be left at sample households transmitting the detailed questionnaire.

Her portfolio will also include her Enumerator's Reference Manual, indexed to provide ready answers for problems not covered in training class or to refresh her memory on items which failed to "stick." A workbook containing practice exercises to be completed during training is also provided, and finally, each enumerator receives a unique map picturing in detail the enumeration districts included in her assignment.

A less material, but nonetheless real, factor in enumerator support is the publicity which will be beamed at all Americans who read newspapers or magazines, listen to the radio, or watch television programs. This will endeavor to promote awareness that a census is being taken, that enumerators may be expected to visit every household and that citizens themselves have particular responsibilities and some advance homework to do in preparation for the enumerator's visit. Although nationwide publicity media in cooperation of the National Advertising Council make their services available to the Bureau, much of the detailed information of interest to local communities is distributed through the district field offices. Your enumerator's name will probably be listed in news stories appearing in the latter part of March, and she may even be included in a group picture or in a news photograph demonstrating a practice interview.

All of this care in selection, training, and preparation of material is back of your enumerator as she rings the bell. She is one of some 160,000 public-spirited Americans who will begin the nation's inventory on April 1, 1960. As a good citizen we expect that you will do your part in the Census by being a cheerful and ready respondent.

In fact, we shall welcome observations from this professional audience based on their family experience with the 1960 enumeration process, including suggestions and criticisms.

PROGRAMMING THE POPULATION CENSUS*

By: Richard A. Hornseth, Bureau of the Census

A brief description of the over-all plans for processing returns from the 1960 Census of Population and Housing will provide a perspective for a discussion of the uses of the computer in processing the Census and the associated tasks of programming. Processing begins with the delivery to our Jeffersonville Office of enumeration books which consist of FOSDIC documents permitting the optical-electronic translation of enumerator entries to pulses on magnetic tape. At Jeffersonville, the books are checked in by enumeration district, or ED, assembled by counties and large cities within States, provided with ED numbers and population and housing field counts in FOSDIC form, and then microfilmed. Books for the 25 percent sample enumeration are handled separately from those for the 100 percent enumeration and require, before microfilming, an additional step of manual coding in FOSDIC form of certain written entries such as occupation. After the microfilm is developed, it is sent to Washington where the translation to magnetic tape is performed by FOSDIC.

The FOSDIC output tapes serve as initial input to the computer where each record for persons and housing units is edited for blank items and inconsistencies and a corrected version of it, for the sample, is put out on tape. The 100 percent record, however, is tallied according to the categories to be shown in publication immediately after the editing for blank and inconsistent entries, and only the accumulated tallies for each ED are put on tape. In the editing process, a diary summarizing and evaluating the quality of the records and comparing field and computer counts for each ED is prepared to determine which ED's require correction outside the computer and what types of correction are called for. Also, in the editing process, a control tape listing all ED numbers for a State for purposes of control is checked off to account for each ED. In the 100 percent operation, final population and housing unit counts for each ED are inserted on the control tape and a quick computer pass of the completed control tape will provide the State population totals required for certification to the President of the United States. In addition, the control tape, when merged with a master ED identification tape providing publication area codes for each ED, permits the preparation of the Population Series P-A publication which shows the number of inhabitants for every area in the United States. In the sample operation, the control tape is posted with ratio estimate counts required for later determination of weights for the sample records.

The output of the first computer pass in the 100 percent operation consists of a set of about 500 counts or tallies for each ED of persons and housing units by the characteristics appearing in all tables for the 100 percent publications: Series P-B for population, block statistics for housing, tract statistics for population and housing, and other reports. A merge of this ED tally output with the master ED identification tape permits the summarization of tallies to every area or type of area shown in the publications. These summaries, when merged with tapes containing place names and historical detail, enable the assembly of tables into publication format for running on the High Speed Printer to provide copy for photo-offset reproduction.

The output of the first computer pass in the sample operation is a partially edited file of records for persons and housing units. Another pass of this file is required to complete the editing and to assign the sample weights to each record. The weight for a person or unit in a ratio estimate category, for example, male renter heads 25 to 44 years, is roughly the ratio of the 100 percent count in that category to the corresponding sample count over a collection of ED's comprising the smallest area of publication. The ratios are determined by the computer in an operation involving the merge of the sample control tape containing the sample ratio estimate counts with a tape containing corresponding 100 percent counts. The completely edited and weighted sample file of individual person and housing records is then passed through the computer as many times as required, using sub-files when necessary, to produce all sample tabulations. Finally, these tabulations are summarized and assembled into publication tables for High Speed Printer photo-offset copy.

The above sketch has indicated or implied the major uses planned for the computer: Editing of records; check-off or control; evaluation of quality; determination of sample weights; tabulation; preparation of publication copy. Several subsidiary functions not mentioned include the preparation of the master ED identification tape, historical data and place name tapes, listings of ED numbers for field and processing control outside the computer, and progress reports.

Mainly, the many uses planned for the computer arise from experienced-based expectations of, and sometimes merely hopes for, gains in quality, time, control, economy, or convenience, and these uses have been arrived at somewhat piecemeal as each processing problem came under consideration. Our British counterparts have not been so exposed to large-scale digital computers as we, and they presently plan a much more modest use for computers in their coming census. In one manner or another, we have been led to use the computer as much as possible and the upshot is that we quite likely may become able theoretically to process completely the 100 percent returns for an average State from original books to publication copy in something under a calendar month with a directly involved staff measuring perhaps less than one hundred man-months. In actual production, a longer calendar time may be required for a particular State since its separate operations must be scheduled among those required for other States. We do expect to complete the bulk of the 100 percent publication program for the 50 States by mid-1961.

This particular achievement in automation stands to be accomplished without it or its implications having been foremost in mind, at least among those concerned with the detailed planning. In fact, at the programming level, the all-absorbing problem is how to accomplish a particular task within the limits of the equipment and experience with it, marvelous though that equipment may be. The day-by-day accretion of experience and skill in the use of equipment at hand provides the basis for accomplishment, not the presence of equipment or the appeal of an automation goal. Thus, programmers are apt to be cautious and tend to resist grandiose push button plans. Each particular accomplishment too painfully has been earned.

A closer look at several of the major functions planned for the computers will illustrate the problems in planning specific computer operations and the requirements of balance among such considerations as limitations of the computer and auxiliary equipment, availability of computer time, supply of experienced programming staff, pressure of deadlines, desires for economy and appeals for extra quality, detail, or information--all weighed on the basis of experience gained since 1950.

Tabulation, of course, is the prime function of the computers and it would appear that tabulation particularly would be facilitated by them. Our present computers, for example, can produce a 5,000 cell tabulation at the rate of about 3,000 persons per minute. The potential tabulations of census material are enormous and, unfortunately, it seems that every possible tabulation can claim a user somewhere. Yet, the content and coverage of publications planned for 1960 represent perhaps only about a 25 percent increase over that for 1950. The major limiting factors are economic, mainly the costs of printing and the availability of computer time within the census period. The Population Series P-D tabulations alone may require something like six months of computer time at the rate of 1,000 available hours per month. This can be construed really as a consequence of a limitation of computer capacity. Series P-D would require about 100,000 counters if it were to be tabulated in one pass of the sample file. However, our computers can accommodate only 5,000 counters easily and up to 15,000 with some difficulty. Hence, something of the order of ten passes of the sample file may be necessary.

Printing costs present a particularly stringent limit to publication plans and thus have initiated considerable exploration and expansion of computer use into what may be called the typesetting field. Preparation of tables for photo-offset reproduction from High Speed Printer copy is planned for practically all publication for both the 100 percent and sample results. Not only reductions in cost but gains in time are expected. Table preparation does not involve significant amounts of computer time, but it imposes extraordinary demands on the ingenuity, skill, and regard for detail of programmers. The relatively limited capacity of the computer and the severe limitations of the High Speed Printer pose problems in programming of such magnitude that computer table preparation is not being attempted for the Population Series P-A publication and the United States Summaries. Series P-A involves quantities of historical and footnote material that can best be prepared by hand. The United States Summaries require table widths not possible on the High Speed Printer.

Our experience has been limited in the field of table preparation by the computer. We are proceeding to some extent on the feeling that the programming time and effort expended on preparing tabulations in the computer in a form permitting ease of posting for a manual preparation of printer copy could be extended without too great an effort to programming full table preparation in the computer. In any event, considerations of publication costs and timing are overriding and impel us to use the computer as much as possible.

A discussion of one more major function of computer processing will suffice. This concerns the editing of individual records for completeness of entry and consistency. The main considerations again are costs and time. A clerical editing operation is costly and time consuming and, in some quarters, has not been regarded as particularly effective. In fact, many feel that a computer supplied with sufficient rules can do a better job of making assignments for, say, unknown age or unknown income. The problem revolves around what constitutes sufficient rules. Rules that anticipate all conceivable possibilities generally exceed the capacity of the computer or tie up more programming talent than can be afforded for computer coding that may apply to only a few cases in the entire census. Consequently, editing rules used in the computer are being pruned to those considered essential for maintaining the quality of the census. For protection from the unexpected and the rare violations of quality, a diary for each ED, produced during the editing pass, will provide for inspection of counts of the number of imputations by type, and, when the counts exceed certain levels, the ED's will be flagged for manual inspection of the original books or their microfilm. In this fashion, it is expected that enumeration books can be processed with a minimum of manual intervention and without a clerical edit other than that provided in the field quality control sample check. The quality of the edit will not suffer. Computer capacity does provide for fairly complicated editing rules, and for handling of an entire household as a unit in both the 100 percent and sample operations. Also, new editing techniques are made possible. For example, income for a person not reporting it can be made the same as that of the last person previously processed having the same characteristics of age, sex, weeks worked, and occupation. Though the main considerations in applying the computer to the problem of editing records have been cost and time, improvement over 1950 in quality of the edit can be assured.

The use of computers in processing returns from the 1960 Census represents an extensive development in the application of experience with large-scale digital computers which the Bureau has had for the past 10 years. The presently planned use of computers ought to produce results equal to or better than those produced from the 1950 Census by more traditional means and certainly much more quickly and economically. There yet remain limitations of equipment, personnel, and experience, which will prevent a full realization of a goal of automatic data processing, if any of us ever had that goal. In the decade after the census, the materials and experience gained during the census should support such a surge of development in automatic data processing that the 1970 Census processing operation may become truly a push button affair. Those who do not consider this a proper goal can take heart in the reasonable expectation that marked improvements in quality, utility, and timeliness generally accompany the attainment of such a goal. They may take comfort in the knowledge that at the detailed planning stages the concern is with computer application to specific problems involving quality and utility. At that level, the computer is not considered a means for short cut but a remarkable device for solving problems.

*Paper presented at the session December 29, 1959, on "How the 1960 Census will be taken" of the Social Statistics Section of the American Statistical Association at its Annual Meeting in Washington, D. C.

PUBLICATIONS OF THE 1960 CENSUSES OF POPULATION AND HOUSING

By: Howard G. Brunsman, Bureau of the Census

The general organization and arrangement of the publications of the 1960 Censuses will be quite similar to the corresponding reports for 1950. Through the use of Fosdic, the electronic computer, the highspeed printer, and the greater use of sampling, we hope to make these reports available much sooner than those for 1950.

Most of the changes in content of the reports result from changes in the content of the census schedule. For the first time, the 1960 population schedule includes a question on whether persons attending school attended a public or private school. Statistics will be shown separately for persons attending public and private kindergarten, elementary school, and high school for each census tract, urban place, and county.

Another addition to the list of questions is one on place of work of each person at work last week. These data will be coded to identify cities of 50,000 or more, and counties. They will be tabulated to show whether the residents of each urban place and each county worked in the county in which they lived. They will also be tabulated to show the central city and county of work of persons living in each census tract of a standard metropolitan statistical area. Statistics on the related subject of means of transportation to work are also included for the first time.

Language spoken in the home before coming to the United States will be reported for foreign-born persons. The schedule also contains inquiries on veteran status and on period of service. Information on number of children ever born is to be obtained for each ever-married woman in the full 25-percent sample. The item on previous residence relates to 1955, rather than to the previous year. An item on year moved into house of present residence has been added to the schedule.

As in 1950, the Bureau of the Census plans to issue a series of separate paper-covered population reports for each State. Ultimately, these separate State reports will be consolidated into a bound volume for each State.

The first of the State reports on population to be issued will be designated as series PA and will relate to the number of inhabitants in each place, county, minor civil division, urbanized area, SMSA, and State. This report will contain the only statistics for incorporated places of less than 1,000. In addition to inclusion in the bound volume for each State, the separate State parts will be assembled into a single bound volume for the convenience of users of these data.

Both the series PB and PC reports for the various States relate to the general characteristics of the population. These reports contain the greatest detail for urban places of less than 100,000, SMSA's of less than 100,000, urbanized areas, and counties. The greatest amount of detail, and substantially the same detail, is shown in these reports for each urban place of over 10,000, and for each SMSA, urbanized area, county, and State, as well as for the urban and rural parts of each State. Somewhat less detail is presented for urban places of 2,500 to 10,000, for the rural parts of counties, and, in series PC, for the rural-farm parts of counties.

Series PB contains statistics based on characteristics enumerated on a 100-percent basis and will be issued several months in advance of series PC, which is based on the 25-percent sample items. As compared with 1950, the reports will contain somewhat greater age detail, including single years of age for persons under 21. Series PB will contain tables showing characteristics for minor civil divisions, including color or race, household composition, marital status, and age. Statistics for these areas were included in the report for 1940, but not for 1950.

Series PC will contain some statistics on age of persons in the labor force and a substantial expansion in the occupation detail of employed workers. Other new tabulations include distribution of income for persons and median earnings for workers in selected occupations. Statistics on country of origin of the foreign stock in series PC will cover native of foreign or mixed parentage as well as the foreign born. The corresponding statistics in 1950 covered only the foreign born whites.

The series PD reports present statistics for States, the urban, rural-nonfarm, and rural-farm parts of States, and for SMSA's of 250,000 or more. For a few subjects, they contain statistics for SMSA's of 100,000 to 250,000 and also for cities of 250,000 or more and, very occasionally, for cities of 100,000 or more. Because of the somewhat greater detail in series PC, the series PD reports contain much less data for SMSA's of less than 250,000 and for cities than the 1950 reports.

In addition to the basic series, there will be a number of special population reports on separate subjects, such as migration, the foreign stock, income, occupation, industry, and fertility. Plans for these special reports are in process of being formulated. It is assumed that they will correspond fairly closely to their counterparts of 1950.

Turning now to housing, the new subjects include: number of bedrooms, number of bathrooms, presence of a basement, home food freezers, clothes washing machines, dryers, telephones, air conditioning, and number of automobiles, as well as presence of elevators for units in larger cities and source of water supply and type of sewage disposal for units in smaller cities and rural areas. Housing units will be classified by condition into the following three groups instead of the two groups of 1950; in good condition, deteriorating, and dilapidated.

Housing Volume I, will include a separate report for each State containing statistics based on characteristics enumerated on both a 100-percent and a sample basis. The most detailed set of tables in the report will contain statistics for each State, each SMSA and the constituent counties of each SMSA, each urban place of 50,000 or more, and each urbanized area. Other sets of tables will contain less extensive data for places of 25,000 to 50,000; 10,000 to 25,000; 2,500 to 10,000; 1,000 to 2,500, for the total of each county, and the rural and rural-farm parts of counties. Some housing subjects are enumerated in a 5-percent sample of housing units. These subjects include: number of bedrooms, fuel for heating, cooking and water heating, home food freezers, number of radios and television sets. Because of this restricted sample, statistics on these subjects are not presented for urban places of less than 25,000 or non-metropolitan counties. Only a few of these subjects are presented for urban places of 25,000 to 50,000.

In a similar manner, separate sets of tables will present less extensive data for housing units occupied by nonwhites in areas with 100 to 400, and for 400 to 2,000 housing units occupied by nonwhites than for areas with a greater number of nonwhites.

Housing Volume II, will contain analytical cross-tabulations of housing data for the United States, for each of the 9 geographic divisions, for the total of each SMSA of 100,000 or more, and for each city of 100,000 or more. The statistics for each SMSA and for the cities within the SMSA will appear in a separate report.

Volume III on City Block Characteristics will comprise about 500 reports, one for each city of 50,000 or more and one for each of about 250 places which have contracted for a block statistics program. These reports will include, for the first time, the total population, the three-way classification of condition, and the average number of rooms in owner- and renter-occupied units, in addition to the other subjects covered in 1950.

A supplementary survey relating to the components of change in the housing inventory since 1950 and to residential financing is being conducted in conjunction with the 1960 Census. The results of this survey will be presented as Volume IV on Components of Change and Volume V on Residential Financing of the 1960 Housing Census.

The characteristics of rural-farm housing will be covered in a series of analytical cross-tabulations presented for the United States as a whole and for each of about 120 economic subregions as Volume VI of the Housing reports.

The population and housing statistics for census tracts will be combined in a separate report for each tracted SMSA. Other than changes in content resulting from items on the schedule, the most dramatic change in the census tract reports will result from the tremendous increase in number of census tracts as compared with 1950. In 1960, the entire areas of 131 of the SMSA's will be tracted, as well as the areas within 240 of the cities of 50,000 or more.

XIII

SOCIAL STATISTICS: THEIR ORGANIZATION FOR POLICY USE

Chairman, Raymond T. Bowman, Office of Statistical Standards

THE USE OF STATISTICS IN THE FORMULATION AND EVALUATION OF SOCIAL PROGRAMMES

By: Octavio Cabello, Statistical Office of the United Nations*

1. Introduction

The role of statistics in guiding decisions regarding micro aspects of social programmes seems to be well established. But, the possibility of using statistics in formulating social programmes becomes uncertain as the scope of such aspects increases, and the question of whether and how statistics can be used in formulating social programmes at the national level has no obvious answer. As a result, while it may be assumed that minor decisions in the social field can be, and frequently are, based on scientific information, the major decisions regarding social policy and programmes - especially in under-developed countries - are mainly a matter of personal judgment based primarily on political expediency and traditional grounds.

The purpose of this paper is to explore the possibility of using social statistics in formulating national social programmes, to provoke wider dissemination of information on this subject in the under-developed countries and to stimulate further research on the methodology of formulating national social programmes.

The discussion here has been restricted in some cases to the field of housing, but it is hoped that similar applications may be found in other social fields.

2. Social Statistics and Social Programmes

The first problem that confronts anyone discussing social statistics is to decide what meaning to attach to the concept. In current usage the scope of social statistics ranges from a narrow coverage which includes only one aspect of social conditions, such as crime or the volume of work of welfare institutions, to the broadest possible range of subjects related to the description of the social structure, social institutions, human relations, environmental conditions, and government and private activities affecting living conditions. It seems that in spite of the attempts that have been made, social statistics cannot be identified as a unitary field by reference to common units of observation, or methods of collecting and analysing data. However, "social statistics" undefined have been the subject of continued attention in recent years, and this may be sufficient justification for preserving the expression (social statistics), defining in each case the meaning attached to it.

For the purpose of this discussion social statistics are defined as the statistics required for the formulation and evaluation of broad national social programmes. This definition focuses attention on the use or purpose of the statistics rather than on the nature of the units of observation or on the methods of collecting and analysing the data. Though non-rigorous, this definition is useful since it stresses the principle that statistics, and government statistics in particular, must always have a well defined purpose.

Little would be gained however, by defining social statistics as the statistics required for social programmes unless social programmes can be defined independently of social statistics. A social programme may be considered to be formed by the scheme of public and private activities which would contribute to maintain or improve some particular aspect of the living conditions. The aim of such a programme would be to raise the levels of living at the maximum practicable rate commensurate with long-term economic development and human resources, and in accordance with nationally and internationally accepted principles regarding human rights and responsibilities.

It may be recalled that, in 1953, a United Nations Committee of Experts on International Definition and Measurement of Standards and Levels of Living recommended the use of twelve separate components of the levels of living 1/, each component representing a well delimited aspect of the levels of living (i.e. of the actual living conditions). In view of this recommendation and of the fact that aims of social policy are usually stated with reference to particular fields, such as education, health, housing and nutrition, it will be assumed, for the purpose of this discussion, that a separate social programme will be formulated for each component of the levels of living. Such an assumption should be acceptable since the techniques and factors involved in improving living conditions in each field are different and separate authorities or government agencies are responsible for the conduct of the various programmes.

The activities involved in the various programmes are of a different nature and, therefore, the methods of formulating the programmes are also quite different. Housing programmes, for example, are concerned with maintaining an adequate supply and distribution of dwellings; education programmes with maintaining and adequate supply and distribution of teaching services and facilities; nutrition programmes with promoting adequate food consumption. Health programmes are conceptually more complex since the levels of health 2/ depend upon a great many factors including housing, education, nutrition and so forth. Some social programmes are concerned with the supply and distribution of certain essential goods or services; others with the prevalence of certain conditions such as social security or human freedoms. Statistics may be useful in connexion with every type of programme but the uses will differ according to the characteristics of each programme.

*/ The author wishes to thank Mr. W.R. Leonard and Miss Nora P. Powell for their suggestions and Miss Camille LeLong for her assistance in preparing this paper. The views expressed are the author's own and do not necessarily reflect the views of the United Nations Secretariat

Although it is expedient to consider each social programme separately, it must be recognized that they tend to be highly interrelated in the sense that the execution of one programme has direct implications for the living conditions in other social fields, e.g. malaria eradication, food supply and labour force; improvement of levels of education and housing conditions and levels of health; sanitation and nutrition, etc. Furthermore, there is an evident inter-action between social programmes and economic development. However, there is no mechanism for integrating social programmes in terms of activities and units of "well-being", such as is found in the economic area in the form of a system of national accounts, input-output tables and mathematical models, which aggregate and relate economic actions in terms of monetary units, and it is, therefore, necessary to deal with each social programme separately.

3. Uses of Statistics in Formulating a Social Programme

The task of formulating a national social programme does not consist in solving a system of equations. Mathematical models describing the effect of social measures on the levels of living, the cost of such measures and the physical inputs involved have not been established. Such models would be useful, even if based solely on non-tested hypotheses and informed estimates, since they would set forth in an explicit form the "laws" or the assumptions involved in formulating the programmes. Unfortunately they are not used in actual practice and the formulation of national programmes is largely a matter of judgement and common sense. However, carefully selected statistics can materially enlighten the bases for the decisions by providing the factual information required to ensure compatibility between social needs and resources, and the statistics recording the experience of previous years will provide perhaps the most valuable guide to planners.

From a practical point of view, estimates of the economic resources, i.e. expenditure and investment by the public and private sectors, provide a suitable point of departure for the formulation of short term programmes in a particular social field. As a next step, the volume of public and private activities needed in order to maintain or improve the levels of living might be estimated and the total value calculated by using unit costs. The cost of the national programme could then be compared with the resources. If the resources appear to be insufficient to maintain or moderately raise the levels of living of a particular component, a revision of the resource allocation would need to be considered, together with the possibility of improving the efficiency and productivity of the agencies and industries involved by introducing new methods or procedures. After all possibilities are explored - and statistics would be essential for such an examination - a social programme could be outlined which would indicate the volume of activities that would be likely to be performed and the probable consequences of the programme upon the levels of living.

It is true that there are other limiting factors of social programmes, besides the economic resources, which must also be duly considered, but the system of economic statistics provides a partial linkage of great operational significance between economic development and social programmes. In effect, for practical purposes and considering only short term programmes, capitalist and socialist countries alike consider social development as a by-product of economic development, and resources for social programmes are estimated or allocated on the assumption that adequate provision is made for ensuring - as a matter of first priority - an adequate rate of economic development 3/ and a proper balance between consumption and production, in order to avoid inflationary pressures.

On the assumption that the process broadly outlined above would be followed in formulating a social programme, statistics would be required for the following major aspects of the programming operation:

a. To estimate the economic resources expected to be available for the social programme.

b. To measure living conditions by means of indicators of the levels of living.

c. To estimate the social needs, i.e. the supply and distribution of "social units", required for maintaining or improving the levels of living.

d. To test the compatibility of the estimated social needs with the expected economic and human resources.

e. To select the most convenient combination of activities, i.e. a social programme.

3.1 Uses of statistics to estimate economic resources expected to be available for the social programme

Considerable attention has been given to the application of statistics in connexion with problems of economic planning and development. Principal sources of information regarding expenditure and investment in previous years are the national accounts 4/ and the government accounts 5/ when these systems provide separate data for the several social programmes. It is not, however, an easy task to separate out the data for various categories and eventually obtain separate totals for the country as a whole. For example, the apportionment of certain items of public expenditure among related social fields involves problems of interpretation, such as the allocation of school feeding programmes to health or to education; the allocation of urban water supply projects to health, housing, or other programmes; the allocation of the cost of marsh clearance to health or to agriculture, and so forth. Some of the statistical data commonly used in evaluating the economic resources devoted to housing programmes, are the following 6/:

a. Expenditure for dwelling construction. This item, from the National Accounts includes all expenditure on new dwelling construction and major alterations to residential buildings including the value of the change in work in progress, the cost of painting and all permanent fixtures, but excluding the value of the land before improvement.

b. Rent and water charges. This item of private consumption expenditure includes all gross rent (actual and imputed gross rent on owner-occupied houses as well as actual and imputed grounds rents payable) including water charges and local rates.

These items provide information about the order of magnitude of the economic resources devoted to housing in the past. When related to the national income and other national aggregates they provide ratios which are useful in estimating future resources, on the assumption that past patterns of expenditure and allocation would prevail or change in a foreseeable direction. The following ratios may be used for this purpose:

a. Expenditure on dwelling construction as a percentage of gross domestic product.

b. Expenditure on dwelling construction as a percentage of gross fixed capital formation.

c. Rent as a percentage of private consumption expenditure.

Estimates of the gross domestic product, private consumption expenditure and fixed capital formation, over the period to be covered by the programme, will usually be obtained by the agencies forecasting economic development. On the basis of these estimates it would be possible, by applying assumed ratios based on past experience and other factors, to derive estimates of the order of magnitude of the national resources that are likely to be devoted to housing in the next few years. Such estimates, would of course need to be revised every year, in the light of new

Table 1. Expenditure for Dwelling Construction as a Percentage of Gross Domestic Product and Gross Fixed Capital Formation, and Rent as a Percentage of Private Consumption Expenditure*. Selected countries. 1956 or 1957.

Country	Year	Expenditure for dwelling construction		Rent as a
		% of GDP	% of GFCF	% of PCE
Australia	1957	...	...	8.4
Austria	1957	...	...	4.8
Belgium	1957	4.1	25.3	13.0
Canada	1957	4.5	16.7	15.4
Ceylon	1957	...	...	2.7
Denmark	1957	2.9	16.8	6.4
Dominican Republic	1957	...	...	11.1
Ecuador	1956	2.6	19.0	10.5
Finland	1957	...	...	6.4
France	1957	4.9	26.1	4.6
Ghana	1957	...	...	2.0
Greece	1957	4.8	30.5	...
Ireland	1957	2.2	16.9	4.0
Israel	1956	7.1	29.7	...
Italy	1957	5.9	27.4	3.1
Japan	1957	2.1	7.9	9.0
Korea, Rep. of	1957	0.5	4.1	7.9
Luxembourg	1957	4.0	17.2	...
Mauritius	1957	3.0	23.1	...
Morocco	1956	3.0	24.5	...
Netherlands	1957	5.5	22.0	6.9
New Zealand	1957	4.8	21.7	...
Nigeria, Fed. of	1956	2.1	20.7	...
Norway	1957	4.4	15.7	7.6
Panama	1956	...	...	16.4
Peru	1956	...	...	20.3
Philippines	1957	1.6	18.4	...
Portugal	1957	4.1	25.7	...
Puerto Rico	1957	4.2	19.1	10.5
Sweden	1957	4.8	24.2	9.2
United Kingdom	1957	2.9	18.2	8.6
United States	1957	4.2	23.7	12.6

* At current market prices.

Source: United Nations. Administrative Committee on Co-ordination. Statistical Indicators of Housing Levels of Living. Document No. ACC/WPSSP/I/4/Add.6, 6 August 1959.

information. This procedure is a very crude one but as crude as it is it may be considered too elaborate for many under-developed countries.

As an additional guide in estimating resources it may be useful to compare the situation or patterns of expenditure in various countries. Table 1 shows the values of the three ratios described above for 32 countries, in 1957 or 1956. The variation among countries is considerable and apparently there is no relation between the ratios and the state or rate of development of the countries. Also, the ratios in the countries with long historical data available show long term and cyclical variations as well as irregular changes. International comparisons, therefore, show only the range of variation of these ratios but would be rather unhelpful for the purposes of selecting suitable rates of expenditure investment in housing. But long national series would be extremely useful guides in determining the resources that are likely to become available, over a period of a few years, for housing as well as for other social programmes.

An important factor in connexion with estimating the economic resources available for social programmes is the distribution of resources between the public and private sectors. In some countries it has been found to be more efficient to finance important aspects of social programmes through government agencies, while in others the direct purchasing of essential goods and services by the individual households is considered possible and satisfactory. The general situation may be considered to be a mixed arrangement by which programmes are financed through both the public and private sectors. It is important, therefore, to have information about the extent to which the various programmes are financed by each of the two sectors. This question was discussed in a paper 7/ presented to the fifth session of the ECAFE Working Party on Economic Development and Planning, held in Bangkok, in September 1959. Table 2, which was taken from this paper, illustrates how data on resources would need to be broken down by sectors for educational programmes. A similar table could be prepared for other programmes.

A further breakdown of data on resources according to investment or current expenditure would also be of interest. The Manual for Economic and Functional Classification of Government Transactions 8/ contains a full discussion of this aspect.

Table 2. Provision and Financing of Educational Services
(in millions of national currency)

Financed by: / Provided by:		Public sector		Private sector		Total
		Central government	Provincial and local governments	Households	Enterprises	Expenditure for providing educational services
Public sector	Central government	a	---	b	---	
	Provincial and local governments	a	a	b	---	
Private sector	Non-profit institutions	b	b	a	---	
	Enterprises	---	---	b	b	
Total	Expenditure for financing educational services					

a - likely to be significant)
b - moderate significance) Notes applicable to ECAFE countries.
--- negligible)

Source: United Nations. Economic Commission for Asia and the Far East. Working Party on Economic Development and Planning. Planning the Pattern of Public Social Expenditures. Document No. E/CN.11/DPWP.5/L.7, 4 August 1959.

3.2 Use of statistics to measure living conditions by means of indicators of the levels of living

The purpose of social programmes is to raise the levels of living, and, in order to translate this purpose into positive action, it is necessary to express levels of living in statistical terms. The United Nations Committee of Experts, mentioned above, proposed in 1953, that the levels of living should be measured by reference to twelve components by means of a series of statistical indicators for each component. Recently, in 1959, an inter-agency Working Party on Statistics for Social Programmes 9/ reviewed the components and indicators in the light of information that had become available since 1953, reduced the twelve components to nine, and proposed the following priority indicators:

Component Health:

1. Expectation of life at birth.
2. Infant mortality rate.
3. Crude annual death rate.

Component Food Consumption and Nutrition:

1. National average food supplied, in terms of calories at the retail level, compared with estimated calories requirement.
2. National average food supplied, in terms of total proteins at the retail level.
3. National average food supplied, in terms of animal protein at the retail level.
4. Percent of total calories derived from cereals, roots, tubers and sugars.

Component Education:

1. Adult literacy rate.
2. Total school enrolment ratio.
3. Higher education enrolment ratio.

Component Employment and Labour Conditions:

1. Proportion of persons employed in the total labour force.
2. Ratio of male labour force in agriculture to total labour force.
3. Relative real wages in selected occupations.

Component Housing:

1. Percent of the population living in "dwellings".
2. Percent of occupied dwellings with three or more persons per room (overcrowding).
3. Percent of occupied urban dwellings with piped water.

Component Social Security:

[No indicators proposed as yet.]

Component Clothing:

[No indicators proposed as yet.]

Component Recreation:

[No indicators proposed as yet.]

Component Human Freedoms:

[No indicators proposed as yet.]

The indicators are of various classes; some reflect achievement or ultimate aims of social programmes in very broad terms, as in the case of the expectation of life at birth; others are more directly related to the activities in the respective social programme, as in the case of the total school enrolment ratio. The indicators of housing levels of living belong to the first class. They are operational indicators which define aims of social programmes and they can also be utilized for estimating the order of magnitude of the programme needed in order to improve the living conditions in respect of housing. Some of the other indicators are not so useful for formulating programmes because, though reflecting certain aspects of living conditions, their relationship to the possible actions in a particular social field are not well established. This seems to be the case with respect to the indicators of health which are not directly related to actions in the medical fields. For example, at present it appears impossible to determine what type of medical or health activities would be necessary in order to increase the expectation of life at birth by one year or reduce the infant mortality rate by five units, or the crude death rate by one unit. On the other hand, the relationship between indicators of housing or education conditions and actions in these fields are understood and can be utilized in formulating relevant programmes. There is much to be done in developing methods of measuring levels of living and particularly in developing "operational" indicators, i.e. those which can be used in formulating national or local social programmes.

3.3 Use of statistics to estimate the social needs, i.e. the supply and distribution of "social units" required for maintaining or improving the levels of living

The indicators discussed above reflect the aims of social programmes by defining in statistical terms commonly accepted aspirations in each field. The next question that may have to be answered by the policy maker or programming officer is: what is the magnitude of the job to be done, in terms of dwellings to be built, food to be supplied and distributed, medical and educational services to be made available, by the public and private sectors, in order to maintain or improve prevailing living conditions? The methodological problems involved in estimating social needs have been explored but not settled. It must be mentioned in this connexion that the Twentieth Century Fund Surveys 10/, which include estimates for the United States for 1960, contain an impressive amount of information on this subject.

Social needs must be related either to a point of time (e.g. a certain date), as in the case of estimates of the housing shortage, or to a certain time period, as in the case of food supplies. The volume of work to be done, i.e. the social programmes, however, can only be related to a period of time which may be one year, three, five or other number of years. A housing programme, for example, may be designed to absorb the existing housing shortage during a 5, 10, or 20 year period and, consequently, the magnitude of the annual job to be done will change in accordance with the length of the programme. The following discussion is concerned with estimates of needs in terms of goods or services to be provided by the programme over a period of several years.

The possibility of making objective estimates of social needs may be illustrated with reference to the field of housing. In this case the question is: how many dwellings and housing units of other classes 11/ would need to be built in a country or major geographic division, during a given period, in order to ensure that the percentage of people living in dwellings remains stable, or increases, and that both the density of occupation (persons per room) and the percentage of dwellings with 3 or more persons per room remain stable or decrease? Questions of this type have been answered for many countries at various periods; among others for Australia, Belgium, Chile, Colombia, Denmark, France, Japan, Netherlands, Peru, Sweden, United Kingdom, United States and Venezuela, and a review of the assumptions and methods used is being prepared in the United Nations Statistical Office. In making such estimates, statistics have been used in connexion with the following aspects of the problem:

a. To estimate the number of dwellings that would be needed to house the population currently living in housing units considered unacceptable from a structural point of view, e.g. huts, caves, "viviendas callampas", etc., or who are without shelter of any kind.

b. To estimate the number of dwellings required to reduce the density of occupation.

c. To estimate the number of dwellings required to eliminate undesirable and involuntary doubling-up of families and households.

d. To estimate the number of regular dwellings that need to be repaired or replaced because of their state of maintenance.

e. To estimate the number of dwellings that will need to be replaced because of obsolescence, disaster or conversion to non-residential use, etc.

f. To estimate the number of dwellings required to house the future population increase, and the potential new households, over a period.

Estimates of total needs have been obtained by combining the partial estimates indicated above or by other means.

In the case of housing programmes the essential "goods" is the dwelling, and in this discussion only one type of dwelling has been considered. Other programmes are concerned not with one type of goods or services but with the supply and distribution of several or many goods and services, separately or in alternative combinations. In order to estimate social needs it would be necessary to define which are the essential units in each case. For primary education programmes the essential units in each case. For primary education programmes the essential unit may be the teacher, for health programmes there would be several units and the doctor might be one of them. The determination of which are the basic social units to be supplied or made available to the population in the case of each programme is an essential question which must be settled in order to estimate the volume or number of units needed and whether such needs could be satisfied with the available resources.

3.4 Use of statistics to test the compatibility of the estimated social needs with the expected economic and human resources

It would be of little use to estimate the volume of activities necessary to maintain or raise the levels of living without verifying whether such a programme is feasible from the point of view of resources available. Such verification would need to be conducted in respect of financial resources; human resources (such as doctors, engineers or labourers); expected availability of material elements (such as drugs, vaccines, books, foods or building materials); and the capacity of the agencies and industries concerned to undertake their share of the programme. The compatibility of social needs and resources is tested by comparing the estimated needs with the corresponding resources. Information on the following three aspects would be necessary, with respect to each programme, in order to conduct such tests:

a. What are the goods and services, i.e. the "social units" which need to be supplied to the population in order to maintain or raise the levels of living.

b. How many "social units" are needed to produce a certain effect on the statistical indicators of the levels of living.

c. What is the average cost of the "social units" involved in each programme.

Information on these three aspects can be obtained in the case of housing programmes but it may be more difficult to obtain in the case of other programmes. The "social units" would change from country to country according to circumstances, social organization and technical knowledge prevailing at the time the programme is formulated. The problem involved in obtaining the information described in b is complicated by the fact that the same changes on certain indicators of the levels of living (e.g. crude death rate) may be obtained by the effect of different "social units", or combinations of them, and that in the social field it is not normally possible to isolate the results of specific actions. However, this complication is essentially peculiar to health programmes whose objectives are vaguely defined by the indicators of the levels of health (see part 3.2). The solution might be facilitated by narrowing down the field of health to aspects directly and primarily related to medical or sanitary actions, or by conceptually sub-dividing the health component of the levels of living into "sub-components".

The testing of the compatibility of social needs and physical resources implies a knowledge of the physical inputs of various kinds that constitute a certain "social unit". In the field of housing for example, it would be necessary to know how much of each class of basic materials, how many hours or months of work of non-skilled, skilled and professional manpower, etc., go into the construction of an average dwelling of a certain type. These inputs when multiplied by the number of dwellings needed to obtain the desired effect on the housing levels of living would yield estimates of physical needs for housing programmes which could then be compared with the available resources.

The capacity to produce or import certain essential elements (for example building materials in the case of housing programmes, vaccines and drugs in the case of health programmes, books in the case of education programmes, basic foods, etc.), as well as the availability of professional and skilled personnel, are essential limiting factors for social programmes, just as essential as the financial resources. The case of countries which could finance an adequate social programme but do not have enough teachers, doctors, engineers, nurses, or statisticians is not unusual. A much more frequent case is, of course, that in which both financial and physical resources are scarce.

The availability of professional personnel has especial importance for social programmes, as can be illustrated by the following example taken from "The Approach of Operational Research to Planning in India" by P. C. Mahalanobis: "There are, at present, about 65,000 fully qualified (six-year trained) doctors in India ... About two thousand six-year trained doctors are turned out in India every year, and the cost of training each doctor is about forty or fifty thousand rupees. Under existing conditions it may take 60 or 70 years to provide one doctor for every two thousand persons in the rural areas on an average ... I have thought it desirable, therefore, to include in the Draft Planframe a proposal to bring some health service to every home in the country within a reasonable time, possibly in 10 or 15 years ..., by establishing two new cadres of 2-year and 4-year trained health assistants as a first step to a national health service throughout the country ... One 6-year trained physician would be in charge of a group of 5 or 6 health assistants; ..." 12/

The testing of needs and economic resources in monetary terms can be made with respect to the total resources of the country or with respect to the households' income. The use of statistics for testing the consistency of national resources with the magnitude of social programmes is of great importance, since social programmes must be realistic but sufficiently adequate to ensure a gradual improvement of living conditions. The total economic resources needed for a social programme would be determined by relating to the number of "social units" needed their unitary cost. The aggregate cost of a national programme may then be compared with the resources estimated to be available for that programme (see part 3.1).

The following example taken from the field of housing will serve as an illustration: In a report on housing in Peru 13/ it is estimated that it would be necessary to construct 2,339,700 dwellings

Table 3. Cost of 30-year Housing Programme, Beginning 31 December 1956.

Areas	New dwellings to be constructed			Existing dwellings to be repaired		
	Number of dwellings	Cost per dwelling (Soles 1956)	Total investment (Soles 1956)	Number of dwellings	Cost per dwelling (Soles 1956)	Total cost of repairs (Soles 1956)
Metropolitan	657,200	40,000	26,288	68,100	10,000	681.0
Urban	669,400	30,000	20,082	255,800	7,500	1,918.5
Rural	1,013,100	10,000	10,131	687,600	2,000	1,375.2
Total	2,339,700		56,501	1,011,500		3,947.7

Source: Perú. Comisión para la Reforma Agraria y la Vivienda. Informe sobre la vivienda en el Perú. Imprenta Casa Nacional de Moneda, Lima, 1958.

and to repair some 1,011,500 over a 30-year period, beginning on 31 December 1956, in order to improve the housing conditions to certain levels now considered acceptable in that country. This would represent an investment of approximately 60,000 million soles at 1956 prices, as indicated in Table 3.

The total cost of the programme is related to the national income as follows: 14/ "The cost of the housing programme, excluding communal services, has been estimated to be 60,000 million soles, to be invested in 30 years. Since the national income and, parallel with it, the capacity for dwelling construction will increase during this period, the most convenient form of distributing such investment over the 30-year period is by equating it to a fixed proportion of the national income.

"On this basis and on the assumption that both the population and the national income will increase following a geometric progression, with annual rates of 2.1 and 5 percent, respectively, the annual investment can be estimated.

"The investment for the first year would be 900 million soles, and this amount would be increased by 5% per year, that is, in the same proportion that the national income increases in constant prices. In this form 60,000 million soles would be invested over the 30-year period.

"Considering that the national income in 1957 was of the order of 25,000 million soles ... and that the available proportion for net investment was 16 percent (average for the last five years), the resources for investment would be 4,000 million soles. The 900 million investment needed for the first year of the housing programme represent approximately 3.6 percent of the national income and 22.5 percent of the net investment, which are evidently very high percentages. ..."

The use of statistics to determine the extent to which households are able to finance their essential social needs out of their regular income is quite common. The question to be answered may be proposed in the following form: Given a certain average level of income, a distribution of households by income groups and corresponding patterns of expenditure, what proportion of the households could, or could not, afford to purchase the "social units" considered in social programmes (housing, food, education, medical services, etc.) at current or expected prices? For example, the question of whether households can afford a type of minimum satisfactory dwelling is normally explored in the formulation of national housing programmes.

In a fairly recent study on the economic significance of housing in Central America and Panama, it is estimated that approximately 46 percent of the urban families in El Salvador could not pay more than 70-88 percent of a monthly rent equivalent to 1/100 of the average cost of a dwelling, such rent being considered adequate to finance the dwelling construction, ("economic rent") 15/. In another study dealing with the housing situation in Colombia, about 20 percent of urban families were found unable to pay a rent equivalent to the amount needed to finance the construction of a minimum acceptable dwelling in forty years 16/.

The form in which the information on household income and expenditure may be used in the formulation of social programmes deserves separate consideration, as indeed, do all the other major uses of statistics briefly examined here. It is not possible to go any further into this matter in this discussion.

3.5 Use of statistics to select the most convenient combination of activities, i.e. a social programme

After social needs are estimated and matched with the estimates of available resources, the next logical step would be to attempt to bring needs and resources into balance by adopting feasible targets or goals for social programmes. Such targets could not be entirely objective since they would be based on assumptions regarding future population growth 17/, national income and relationships between social measures and levels of living, and would be the result of compromise between desire for improvement of one aspect of the levels of living (e.g. housing), the need to maintain an even rate of improvement in other components and the need to maintain a sustained economic growth. The targets, furthermore, should be revised periodically as new information becomes available. The statistical bases on which the targets would be established would be known and could be examined by anybody interested in the subject, a fact that should have considerable political significance, especially in under-developed countries.

The resources may be found to be smaller or larger than the needs for a certain social programme. Normally the balance of needs and resources for any programme will vary depending on the type of resources. There may be sufficient labour and insufficient building materials, sufficient financial resources and insufficient professional personnel, and so forth. Whatever the situation may be, the use of statistics in exploring the possibilities of equating needs and resources would be essential.

If after exploring all the possibilities, the resources are insufficient even to maintain the existing levels of living, statistics can also be used to evaluate the consequences of such a situation. In the case of housing, to mention a very real example, statistics could be used to estimate the extent to which the housing shortage would increase in the event that an insufficient number of dwellings were built over a period of years. Statistics, however, would not be very helpful in judging the political and social consequences of allowing housing conditions to deteriorate consistently over a period of several years.

4. Use of Statistics for the Evaluation of Social Programmes

While the formulation of a social programme is an *a priori* operation, oriented to the future, the evaluation must be made *a posteriori*, and should, therefore, be based on past experience.

In this discussion a social programme has been considered to be composed of public and private activities, whether or not consciously oriented. When the programme has been established by a systematic procedure, such as the one broadly outlined in part 3 of this paper, it could be evaluated in the following ways:

a. By measuring the ultimate results of the programme by the observed changes in the indicators of the levels of living.

b. By testing the assumptions made in formulating the programme as regards the expected behaviour of the households, future population growth, national income and expected results of social measures in the light of observed facts.

c. By assessing the administrative efficiency of the agencies concerned.

The measurement of levels of living by means of statistical indicators has been discussed in part 3.2 and the evaluation of assumptions, "laws" and theories or of the administrative efficiency of agencies cannot be approached in a general way. Such evaluations can only be made with reference to the particular assumptions, laws etc. of individual programmes. In the final analysis, the only comprehensive yardsticks of the success or failure of social programmes are the statistical indicators of the levels of living.

5. Abstract

Social statistics are defined as the statistics required for social programmes, and a social programme as the scheme of public and private activities which have a direct bearing on a particular aspect of living conditions. It is assumed that there is a separate programme for each component of the levels of living, e.g. health, education, housing.

The uses of statistics for the following aspects of social programming are examined and illustrated with examples from the field of housing: estimating of economic resources for the programme; measurement of living conditions by statistical indicators; estimating of social needs, e.g. "dwellings", to maintain or improve the levels of living of growing populations; testing of the consistency of social needs and economic, human and material resources; and adoption of goals for social programmes consistent with expected resources.

Several forms of evaluation are briefly discussed, and the indicators of levels of living are considered to be the only comprehensive yardsticks of the success of failure of social programmes.

The purpose of the paper is to provoke dissemination of information on this subject among under-developed countries and to stimulate research on the uses of statistics for social programming.

Footnotes

1/ United Nations. Report on International Definition and Measurement of Standards and Levels of Living. Document No. E/CN.3/179-E/CN.5299. (Sales No. 1954.IV.5.)

2/ Health is defined in the constitution of the World Health Organization as follows: "Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity". See: United Nations. Final Acts of the International Health Conference, held in New York from 19 June to 22 July 1946. Document No. E/155, October 1946.

3/ In capitalist countries by means of an adequate rate of investment in the means of production. In socialist economies by securing a rate of increase of the output of means of production greater than the rate of increase of the output of consumers' goods.

4/ United Nations. A System of National Accounts and Supporting Tables. Document No. ST/STAT/-Series F./No.2/Rev.1. (Sales No. 1959.XVII.-11.)

5/ United Nations. A Manual for Economic and Functional Classification of Government Transactions. Document No. ST/TAA/M/12 - ST/ECA/49. (Sales No. 58.XVI.2.)

6/ The evaluation of the economic significance of housing programmes is discussed in: United Nations. Administrative Committee on Co-ordination. Statistical Indicators of Housing Levels of Living, methodological study of a preliminary nature. Document No. ACC/WPSSP/I/4/Add.6, 6 August 1959.

7/ United Nations. Economic Commission for Asia and the Far East. Working Party on Economic Development and Planning. Planning the Pattern of Public Social Expenditures. Document No. E/CN.11/DPWP.5/L.7, 4 August 1959.

8/ Op. cit.

9/ United Nations. Administrative Committee on Co-ordination. First Technical Report of the Working Party on Statistics for Social Programmes. (Geneva, 16-22 September 1959). Document No. ACC/WPSSP/I/9/Rev.1, 22 October 1959.

10/ Dewhurst, J. Frederic and Associates. America's Needs and Resources, A New Survey. The Twentieth Century Fund, New York, 1955, 1,148 pages.

11/ The concepts of "dwellings" and "housing unit" used in this discussion are those defined in: United Nations. General Principles for a Housing Census. Document No. ST/STAT/SER.M/-28. (Sales No. 58.XVII.8.)

12/ Reprinted from Sankhyā, The Indian Journal of Statistics, Vol. 16, Parts 1 and 2, 1955. "Supplementary Consideration", chapter 5, para. 7.

13/ Perú. Comisión para la Reforma Agraria y la Vivienda. Informe sobre la vivienda en el Perú. Imprenta Casa Nacional de Moneda, Lima, 1958.

14/ Ibid., p. 57-58. (Unofficial translation.)

15/ United Nations. Economic Commission for Latin America. Significación económica de los programas de vivienda en Centro América y Panamá. Document No. E/CN.12/CCE/AC.6/5.

16/ Colombia. Corporación Nacional de Servicios Públicos. Departamento de Vivienda. Déficit y demanda de vivienda en Colombia. Series Estudios Socio-Económicos No. 3. Bogota, 1956.

17/ Implied in projections of national income as well as in the estimates of future needs.

HEALTH, EDUCATION, AND WELFARE INDICATORS AND TRENDS

By: Luther W. Stringham, U.S. Department of Health, Education, and Welfare

Much effort has been devoted in the Department of Health, Education, and Welfare to bringing together statistical information on a wide variety of subjects that relate to the human resources of the Nation and to presenting this information in a form that will be useful to those who are concerned with the health, education, and welfare of the people.

One objective of this activity has been to assemble, organize, interpret, and present statistics for executive use. In this connection there has been intensive examination of data of many classifications--economic, social, cultural, etc.--to find the best measures of national developments and appropriate indicators of current and emerging problems.

A part of the effort to make statistical data conveniently available to policy officials has involved the development of the monthly Health, Education, and Welfare Indicators and its annual supplement, the Health, Education, and Welfare Trends. It is the purpose of this paper to describe the evolution of these documents and to report on the plans that are now underway to make them available to the public.

A Reexamination of Statistical Reporting to Departmental Officials

In January 1957 the Program Analysis Officer of the Department initiated a reexamination of the manner in which statistical data were being received and used in the Office of the Secretary. He was concerned especially with whether trend information on appropriate subjects was reaching officials of the Department promptly and in a form convenient for their use. 1/

While substantial progress had been made during 1955 and 1956 in improving the upward flow of information on program developments, the Program Analysis Officer continued to be dissatisfied with the amount, variety, and timeliness of the information--especially statistical information--regularly presented to Departmental officials. Although statistical materials were being supplied in a variety of publications, memoranda, reports, charts, and special presentations, there were still gaps and delays in the channeling of needed information.

There was a need for better methods of obtaining, organizing, and interpreting program data and other statistics compiled by the operating agencies of the Department and for securing and using pertinent data from outside sources. Better mechanisms were needed to improve the flow of data from those whose main responsibility is to produce statistics to those whose job it is to use them.

Executive Needs for Statistical Data

The first step in the reexamination of statistical reporting was to establish the needs of Departmental officials for statistical information. In this connection several basic questions were raised. What are the special requirements of a Cabinet officer for measures of national trends? Under what circumstances will he use such information? How do his needs compare with those of other officials, such as the President and Congressional leaders? How do the Secretary's needs differ from those of other officials within his own Department, such as the Surgeon General, the several commissioners, and other administrators? The views of the Secretary and of his staff assistants on these matters were solicited. A review also was made of actual requests for data that had been made over a period of time by top officials of the Department.

Although the needs of a Cabinet officer obviously would vary from person to person and from time to time, it was agreed that any Secretary should have channeled to him a fund of background knowledge that would provide correct impressions of developments and highlight current and emerging problems. This information would help him formulate judgments, encourage him to raise questions that would lead to further analysis of particular problems, and provide facts upon which he could rely in conversations with the President, in Cabinet discussions, in meetings with national organizations, and at press conferences and before Congressional committies.

It was agreed that the Secretary of Health, Education, and Welfare, with his wide range of interests and responsibilities relating to the human resources of the Nation, should be provided with up-to-date statistics on many subjects and in considerable depth. These subjects should include (1) major social and economic developments; (2) national trends in health, education, and welfare; and (3) progress in the many programs administered by the Department. Some of the specific topics under each of these categories can be cited by way of illustration.

Background information on social developments would include the dynamics of population growth, including its present and prospective size, composition, and distribution; the process of urbanization; mobility of the population; and changes in community life. Births, deaths, life expectancy, family formation, juvenile delinquency, illegitimacy, desertion, and dependency would be other subjects on which background information ought to be provided.

In the economic field basic facts would deal with business fluctuations, long-term rates of growth, increases in productivity, distribution of wealth and income, changes in price levels and in specific prices, growth of the labor force, and trends in the extent and duration of unemployment. Economic and social data would include a breakdown of national expenditure, revealing the share of public and private spending on activities and programs affecting the health, education, and welfare of the people. Included would be significant facts on government finances at all levels.

In the field of health pertinent data would include the health status of the population, including trends in morbidity and disability; the demand for and supply of health manpower; health facilities and their utilization; hospital construction; medical care prices and expenditures; health insurance; medical research expenditures; nutrition; sales of drug preparations; and environmental health factors, including radiological contamination, water pollution, and air pollution.

In education they would include the size of the school age population, retention rates, levels of achievement, prospective costs, school staffing and expected needs for teachers, construction of educational facilities, financing of school operations and construction, and national needs for specialized talent and for vocational education.

In the welfare field necessary data would include public and private social welfare expenditures; coverage under social security programs; public assistance recipients and payments; relationships between economic developments and needs for public assistance; and effects of price changes on recipients of social insurance and public assistance. In the area of vocational rehabilitation would be included the numbers of persons disabled; the number who might be restored to productive occupations; and their potential economic contribution.

A Canvass of Available Information

The second step in the review involved a survey of available statistics that would meet the requirements described above. As a part of this survey numerous Departmental and other governmental publications were review, including the *Social Security Bulletin*, *Public Health Reports*, *Monthly Vital Statistics Report*, *Construction Review*, *Survey of Current Business*, *Treasury Bulletin*, and many reports of the Departments of Commerce, Agriculture, and Labor. Extensive conversations were held with representatives of the operating agencies of the Department and with persons in other government agencies. Internal reports were analyzed and office files containing unpublished data were examined. Nongovernmental sources, such as the American Hospital Association, also were considered. In short, an effort was made to secure a comprehensive view of statistical resources, regardless of their origin or classification, that could contribute to an understanding of developments in health, education, and welfare.

This canvass revealed a substantial number of more or less pertinent time series relating to (1) population; (2) labor force; (3) unemployment; (4) national income and output; (5) prices; (6) governmental finances; (7) construction; (8) research and development; (9) credit; and (10) food consumption. Specifically related to health, education, and welfare were the vital statistics, morbidity, and other health data regularly compiled by the Public Health Service; the Office of Education's biennial and other educational surveys covering facilities, enrollments, staffing, and financing; the Social Security Administration's program data on social insurance, public assistance, credit unions, and child health and welfare; and other compilations pertaining to social welfare developments in the Nation as a whole. Other program statistics were available on vocational rehabilitation, food and drug law enforcement, St. Elizabeths Hospital, and on other program operations of the Department. In this survey we also noted various subjects on which information is gathered by private agencies and the availability of projections, such as of population and school enrollments, that are of particular interest to policy officials engaged in program planning.

The Concept of "Indicators" of Health, Education, and Welfare

The review of these resources revealed many time series that were compiled on a weekly, monthly, or quarterly basis, and an even larger number that were compiled annually or at some less frequent interval. Although all of them were "available" for executive use, either in a published or other form, only a small fraction in fact "trickled up" to policy officials.

With this background, we approached the third step in the reexamination of the reporting system; namely to a consideration of a suitable vehicle for the presentation of pertinent data. At hand, of course, was the *Economic Indicators*, prepared by the Council of Economic Advisers for the Joint Economic Committee. The *Economic Indicators*--with each page consisting of a brief text, graph, and statistical table--offered an attractive format that had stood the test of time. We also had studied the United Nations "Report on International Definition and Measurement of Standards and Levels of Living," which describes various types of "indicators" that reflect changes in human well-being.

In view of the established executive needs, the body of thought relating to "indicators" of human well-being, the available statistical resources, and a tested format, we decided to try what seemed like a novel idea--that is, to develop a monthly document to be modeled after the *Economic Indicators*, and to be called the *Health, Education, and Welfare Indicators*. In line with the Secretary's preference for really current data, even though on a limited number of subjects, we decided first to concentrate on those for which weekly, monthly, or quarterly data were produced.

The summer and fall of 1957 were devoted to the preparation of an experimental issuance. By November 1957 the first milestone was reached, with the reproduction of a "trial issue,"2/ that contained graphs, tables, and text on some 30 subjects. The "trial issue" was submitted for comment to the Secretary and other officials, to the Department's Regional Directors, to the heads of the operating agencies of the Department, and to selected research and program analysis personnel. Outside the Department the Assistant Director for Statistical Standards of the Budget Bureau and the members of the Council of Economic Advisers were asked to evaluate the proposed report.

Answers were solicited to six questions: (1) Are the subjects those on which officials of the Department feel a need for current information and will this type of information serve their purposes? (2) Is the organization of material appropriate for presenting trend information on topics of concern to the Department? (3) Is the format suitable for easy reference? (4) What subjects are not adequately covered and what additional statistical series should be provided for? (5) How frequently should it be issued? (6) What distribution should it have?

On the whole the reaction to the "trial issue" was very favorable. The replies were most enthusiastic from officials with broad policy responsibilities and from individuals with wide program interests. There was a favorable reaction to the proposed organization, although the question was raised as to whether the format of the Economic Indicators should be followed quite so rigidly.

A Companion Document: Health, Education, and Welfare Trends

The appraisal of the "trial issue," though favorable indeed, focused attention on the limited number of current monthly series on many health, education, and welfare subjects and the consequent desirability of making effective use of annual data in order to achieve a sufficiently rounded subject-matter presentation. It was realized that a monthly report alone would not serve as the best medium for the presentation of annual statistics and for the interpretation of series that changed slowly over time.

From these considerations emerged the idea of an annual supplement to the Health, Education, and Welfare Indicators that would present long-term annual trend data. We decided that such a companion document should be developed and that it might appropriately be called the Health, Education, and Welfare Trends.

In February 1958, the Secretary directed that the Office of Program Analysis should proceed with (1) the further development and issuance of the monthly Health, Education, and Welfare Indicators and (2) the planning of an annual volume of Health, Education, and Welfare Trends. It was agreed that both of these documents would, pending their further development, be treated as internal administrative reports but that they would be made available, in limited quantities, to persons outside the Department who would like to participate in their evaluation and to offer suggestions with respect to their further development.

In accordance with the Secretary's decision, the first regular issue was prepared in April 1958 and regular monthly issuance has occurred since that time. The reception of the Health, Education, and Welfare Indicators within the government agencies was cordial from the beginning. Some 300 copies were ordered, at their own expense, by the operating agencies of the Department. From the beginning copies were sent to the members of the Council of Economic Advisers, the Bureau of the Budget, and the White House Staff. With successive issues additional agencies were added, usually at their own request, such as the Legislative Reference Service of the Library of Congress, the U.S. Information Agency, the U.S. Delegation to the United Nations, and Congressional Committees. In July the President and the Cabinet were made aware of the progress that had been made.

During 1958 attention was divided between the development of new charts and tables of current monthly data for the Health, Education, and Welfare Indicators and of trend materials for the annual supplement. 3/ By the end of 1958 over 100 presentations had been prepared for the two documents. Seventy-one of the charts and tables were then brought together in the 1959 Edition of the Health, Education, and Welfare Trends. In this volume were presented data on (1) population, vital statistics, economic growth and stability, and government finances; (2) health manpower, medical facilities, mental health, and incidence of disease; (3) school enrollments, graduates, teachers, educational facilities, and education finances; (4) social insurance, public assistance, child health and welfare, and credit unions; and (5) vocational rehabilitation. The topics covered in the 1959 Edition of Health, Education, and Welfare Trends did not by any means exhaust the subjects that might have been included. There were no tabulations, for example, on medical care expenditures, health insurance, water pollution, classrooms, family formation, and metropolitan growth; and only a few projections were incorporated.

Appraisals by Persons Outside Government

From the inception of this effort care was taken to secure appraisals from interested persons, both inside and outside the government. Initially these evaluations were to insure that significant types of data were not being overlooked and that appropriate use was being made of the statistics presented. This practice made us aware, at quite an early stage, that, although the initial focus was on Departmental needs, these compilations would be of interest to a much wider audience. We began to realize that persons concerned with social trends generally would be interested in the same types of information that are required by government officials.

Sensing this interest several copies of each monthly issue of the Health, Education, and Welfare Indicators and later of the 1959 Edition of the Health, Education, and Welfare Trends were sent to persons in nongovernmental organizations for their comment, including the National Social Welfare Assembly, Social Science Research Council, the Population Council, National Bureau of Economic Research, American Federation of Labor and Congress of Industrial Organizations, American Red Cross, American Medical Association, United Community Funds and Councils of America, and the American Public Welfare Association. Copies also were sent to members of the American Statistical Association who, at the 1958 annual meeting, had indicated a desire to participate in the evaluation of the project. In addition, the Social Legislation Information Service was helpful in making copies available for comment to interested organizations among its constituents.

Public response to these documents has been cordial and indicative of the breadth of the potential audience. Great interest has been evidenced by those with a business orientation as well as those working in such fields as labor, health, and welfare administration. Almost without exception we were urged to make these documents available for sale to the public.

As a result both of the public and private interest and an appreciation of the potentialities of these documents, funds were included in the Department's appropriation for the fiscal year 1960 to make possible their further development for use by the public as well as by government agencies. It is anticipated that arrangements for public sale will be accomplished during 1960.4/

Prospects for the Future

Although the Health, Education, and Welfare Indicators and the Health, Education, and Welfare Trends are still in their infancy, it may already be anticipated that they can perform an important role beyond that of serving as convenient statistical collections. They will reveal gaps in information on many subjects, and they can serve as the catalyst for the development of new measures and for the collection of new data. They should facilitate study of the whole field of social trends. Finally, a body of data of this type may suggest improvements in the organization of interrelated series and facilitate the analysis of social problems.

1/ At that time Charles B. Lawrence, Jr., was the Program Analysis Officer for the Department. As his Assistant, I was assigned responsibility for carrying forward the study of trend reporting described in this paper and for developing the HEW Indicators and HEW Trends. When used in this article the pronoun "we" refers specifically to Mr. Lawrence and the author.

2/ It is of interest to note that coincidental with this development a similar proposal was being made to the Joint Economic Committee. In a paper submitted to the Subcommittee on Fiscal Policy and published on November 5, 1957, Katherine Ellickson, AFL-CIO, stated:

> "The Joint Economic Committee could usefully initiate a set of measures of human well-being to provide objective data on progress made and wastes still requiring attention. The Committee's publication, Economic Indicators, has been most useful in providing monthly information on the Nation's economic stability and growth. Why not add a supplementary section devoted to trends in the people's welfare?"

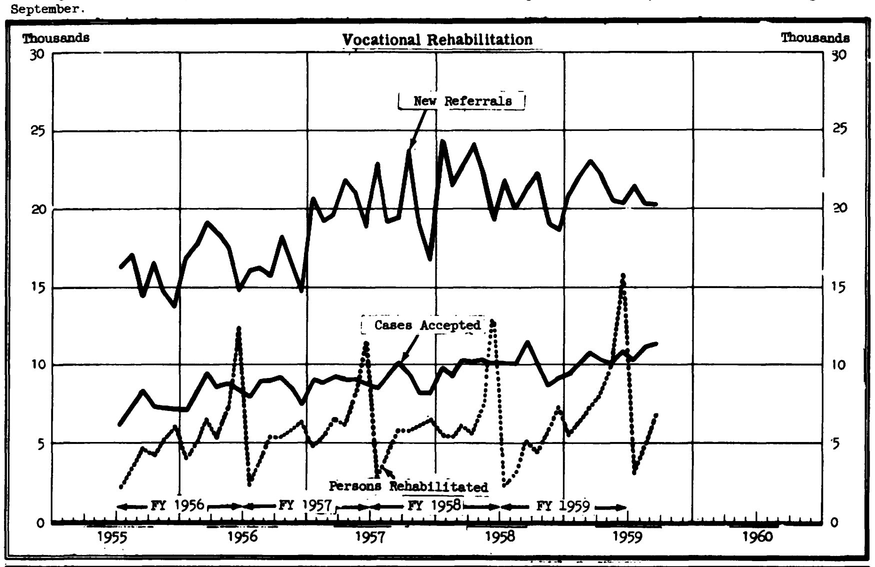
She then suggested various "indicators" illustrative of those that might be considered for inclusion: death rates at various ages, birthrates and population growth, families, housing adequacy, national health, accident rates, levels of educational attainment, social insurance coverage, etc. (Federal Expenditure Policy for Economic Growth and Stability, Joint Economic Committee, 85th Congress, 1st Session, November 5, 1957, p. 979.)

3/ At the end of this paper is a sample page from the HEW Indicators and a sample page from the HEW Trends.

4/ Editors Note: The 1960 Edition of Health, Education, and Welfare Trends was published in April 1960 and is now available for public purchase through the U.S. Government Printing Office, Superintendent of Documents, at 50 cents a copy.

VOCATIONAL REHABILITATION

At the end of the first quarter of FY 1960, State agencies had completed 14,500 rehabilitations--40 percent above the level one year earlier. In new plans initiated, the 3-month total of 30,000 is 8 percent ahead of the first quarter of FY 1959. Active cases on hand at the close of the period reach 183,400--the all-time high for September.



Period[1]	Number during period: New referrals	Cases accepted	Plans completed[2]	Cases served	Cases closed: Rehabilitated	Other reasons[3]	Active cases at close of period: Total number	Ready for employment
1940	-	18,374	-	65,624	11,890	6,560	47,174	6,570
1949	158,364	99,202	-	216,997	58,020	25,262	133,715	10,983
1950	154,819	92,009	-	225,724	59,597	25,186	140,941	13,375
1951	153,295	90,603	-	231,544	66,193	25,783	139,568	12,948
1952	154,470	83,922	-	228,481	63,632	27,397	137,452	12,242
1953	153,969	84,397	-	221,848	61,308	27,366	133,174	11,858
1954	145,289	78,045	-	211,222	55,825	28,627	126,770	-
1955	176,067	82,269	-	209,030	57,981	23,476	127,573	-
1956	197,040	93,555	77,630	221,156	65,640	21,059	134,457	13,337
1957	218,638	104,125	84,622	238,592	70,940	23,063	144,589	15,561
1958	255,632	113,855	94,541	258,439	74,317	25,281	158,841	18,584
1959	252,252	121,559	100,273	280,384	80,740	28,533	171,111	19,686
1958								
July	21,722	10,105	6,843	-	2,057	1,300	165,573	19,318
August	20,227	10,143	7,485	-	3,074	1,459	171,183	19,616
September .	21,470	11,519	13,370	-	5,179	1,989	175,534	19,931
October ...	22,375	10,196	9,197	-	4,468	1,826	179,436	20,282
November ..	19,047	8,798	7,275	-	5,875	1,730	180,629	20,411
December ..	18,766	9,130	7,377	-	7,213	2,278	180,268	20,617
1959								
January ...	21,051	9,779	8,237	-	5,545	2,244	182,258	21,185
February...	21,861	9,754	7,707	-	6,344	2,305	183,363	21,034
March ...	23,054	10,887	8,382	-	7,351	2,772	184,127	21,411
April	22,069	10,312	8,447	-	7,900	3,053	183,486	20,547
May	20,541	10,104	7,694	-	10,011	2,820	180,759	19,780
June	20,069	10,832	8,259	-	15,723	4,757	171,111	19,686
July	21,684	10,304	7,484	-	3,155	1,773	176,487	20,588
August	20,270	10,995	8,386	-	4,538	1,877	181,067	20,800
September .	20,280	11,213	13,995	-	6,768	2,085	183,427	20,504
October ...								
November ..								
December ..								

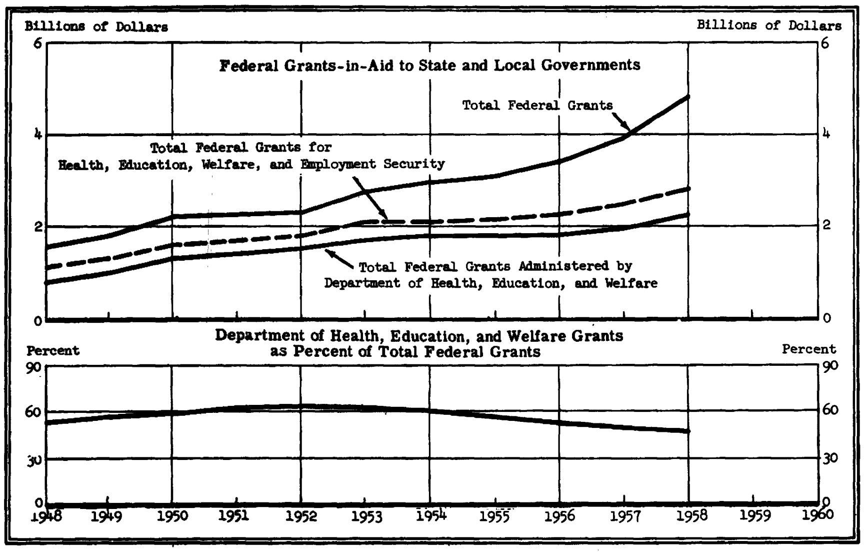
1/ Annual data are on a fiscal year basis; monthly data are for the calendar year designated. 2/ New individual rehabilitation plans initiated for clients. Prior to July 1958, data cover plans completed during the period. These two series are roughly comparable. 3/ Closed before or after a rehabilitation plan was initiated, because services were declined or not needed, or because of migration, illness, aggravated disability, personal factors, etc. Prior to July 1958, data were reported on a quarterly basis only.

Source: Office of Vocational Rehabilitation

SAMPLE PAGE FROM HEALTH, EDUCATION, AND WELFARE INDICATORS

FEDERAL GRANTS-IN-AID TO STATE AND LOCAL GOVERNMENTS

Federal grants-in-aid to State and local governments totalled $4.8 billion in the fiscal year 1958. Of this total $2.8 billion, or nearly 59 percent, were for grants for health, education, welfare, and employment security programs. The percent of all Federal grants accounted for by those administered by the Department of Health, Education, and Welfare has decreased from 64.7 percent of the total in 1952 to 47.2 percent of the total in 1958.



(In millions of dollars)

Fiscal year	Total Federal grants-in-aid 1/	Grants for health, education, welfare, and employment security 2/: Total	Percent of total Federal grants	Health	Education	Welfare	Employment security 3/	All other grants 4/	Grants administered by Department of Health, Education, and Welfare 5/: Total	Percent of total
1948	1,575.4	1,136.6	72.1	55.3	113.3	810.3	157.7	438.8	834.4	53.0
1949	1,835.5	1,323.5	72.1	66.6	68.9	1,026.7	161.1	512.0	1,053.8	57.4
1950	2,208.0	1,644.8	74.5	123.8	69.9	1,236.6	214.5	563.2	1,314.3	59.5
1951	2,250.1	1,718.6	76.4	174.3	80.3	1,288.3	175.6	531.6	1,430.6	63.6
1952	2,327.0	1,806.5	77.6	187.4	143.5	1,292.5	183.2	520.5	1,505.6	64.7
1953	2,756.8	2,061.0	74.8	172.8	246.7	1,444.0	197.5	695.8	1,746.3	63.3
1954	2,956.2	2,128.4	72.0	140.2	235.2	1,552.8	200.1	827.8	1,810.9	61.3
1955	3,093.9	2,154.0	69.6	119.2	277.8	1,568.0	188.9	940.0	1,817.3	58.7
1956	3,438.2	2,278.1	66.3	133.2	252.1	1,632.5	260.3	1,160.1	1,837.3	53.4
1957	3,933.0	2,496.2	63.5	163.2	253.6	1,759.9	319.5	1,436.8	1,966.4	50.0
1958	4,800.6	2,824.5	58.8	193.5	283.7	2,023.1	324.1	1,976.1	2,266.6	47.2
1959										

1/ This tabulation of funds made available to State and local governments under Federal grant-in-aid programs was prepared in the Social Security Administration from data compiled by the Treasury Department on a "checks issued" basis from reports of the administering agencies. Additional detail is regularly published in the Annual Reports of the Secretary of the Treasury and in the June issue of the Social Security Bulletin. The term "grants-in-aid" as used in this tabulation is confined to Federal funds made available for cooperative Federal-State or Federal-local programs administered at the State and/or local level and for those programs in which the bulk of the funds is channeled through agencies of State and local governments. Emergency grants and the value of grants-in-kind are included when they conform to this definition. Federal aid granted directly to individuals and private institutions and reimbursements to State and local governments for expenses incurred by them as agents of the Federal government in administering programs primarily national in character have been excluded. Federal payments in lieu of taxes and shared revenues also have been excluded. Detail may not add to totals due to rounding. 2/ For further breakdowns of the funds made available to State and local governments under health, education, and welfare grant programs see the following additional charts and tables: Federal Grants-in-Aid for Medical Services and Health Facilities Construction; Federal Grants-in-Aid for Education; and Federal Grants-in-Aid for Public Assistance and Child Welfare Services. 3/ Includes employment service and unemployment insurance administration, 1948 to date; administration of veterans unemployment self-employment allowances, 1948-53; and beginning 1956, distribution to State accounts in Federal unemployment trust fund of certain tax collections. 4/ Consists principally of highway construction, housing and urban redevelopment, airport construction, natural resources, and agricultural commodity programs. 5/ The Department of Health, Education, and Welfare was created in 1953. Grant programs designated as administered by the Department of Health, Education, and Welfare for years prior to 1953 are those (a) presently administered by the Department that were in operation prior to its establishment; and (b) other grant programs that were administered by agencies now located in the Department.

Source: Treasury Department and
Social Security Administration

DISCUSSION

By: Katherine Pollak Ellickson, AFL-CIO

Mr. Cabello's paper opens up stimulating vistas of what might be accomplished in this country if we followed the pattern of social programming that he outlines. But his approach is not one that has often been put into effect in the United States on a national scale. Instances do exist, especially at the local level, as for example when communities plan for new school buildings to meet expanding population. But in the equally important field of housing, local authorities are not proving too successful in attempting to meet housing needs for all the people. And very little of what Mr. Cabello calls social programming is now put into practice at the national level. Decisions are made to enlarge or cut federal appropriations by a series of political steps that would be difficult to explain even in many hours to our friends in other countries who follow a different approach.

Numerous factors contribute to the absence of such social programming at the national level. Among them are the size and diversity of our country, the division of responsibility between federal, state and local governments, the dominant faith in private enterprise, and the objections of important groups to government action, particularly if it may compete with their own operations or limit their freedom.

Perhaps long-range planning will come into better repute here as a result of the challenge presented by other countries. In the meantime, the statistics published in Health and Welfare Trends and the Indicators reflect the U. S. pattern of recording what has happened and what is happening rather than what remains to be done if basic human needs are to be met.

Both these HEW publications contain much useful material, especially on the Department's programs. It is gratifying that their value is to be recognized by making them generally available. Ready accessibility of this type of information will make it possible to include up-to-date facts in many discussions of public policy, thus hopefully resulting in sounder substance. Even though many types of social data are theoretically available to busy persons, it is too easy to neglect exerting the effort to locate them when a speech is to be made or a report drafted.

Much valuable material is prepared by government agencies, or is available in their files, which does not readily come to public attention. One reason has been the reluctance of the Congress to appropriate sufficient funds for more effective informational work by the government.

Vast business expenditures in the field of public relations are accepted as essential for successful private enterprise. Since such expenditures are tax deductible, surely the American people should be furnished in attractive and usable form the important social statistics which have been gathered by public agencies.

In addition to letting us all have these two new publications, the Department of Health, Education and Welfare should also improve the format and expand the contents of its other publications. The Social Security Bulletin, for example, deals with programs affecting every man, woman and child, but it is strictly limited in size and uses small type that discourages rather than attracts many readers.

Rather than spending more time praising the new publications -- which really speak for themselves --, I shall suggest certain additions or modifications which will assist to "highlight current and emerging problems," which is one of the purposes avowed by Mr. Stringham. It is of course one with which I am in sympathy. Civic organizations, like cabinet officers and lawmakers, need to know about current and emerging problems as well as about operating statistics and past accomplishments.

In more instances the data should be shown in relation to changes in the value of the dollar or on a per capita basis. Since population is growing rapidly and the dollar depreciates over the years, failure to adjust for such changes is apt to lead to a false optimism. An upward curve may turn downward when properly adjusted.

Through such changes and the addition of new material the contents might make a greater contribution to the formulation and evaluation of social programs.

Mr. Cabello very properly emphasizes the importance of comparing social needs with resources. In the United States, contrary to the situation in most countries, resources exist or can be created through rapid economic growth to meet the most pressing social needs within the foreseeable future. The great difficulty in this fortunate country is to focus attention on the unmet needs which too often are glossed over because they are involved with existing social, economic and political patterns which some people do not want to see changed.

Some of the social needs affect even the relatively well-to-do members of the population, including many statisticians. All of us are endangered by continuing shortages of medical personnel and facilities. We may welcome the data on hospital construction on page 33 of Trends, and be comforted by the fact that the chart now shows an upward movement especially as compared with the period from 1933-47. But it would be socially constructive to include also official Public Health Service estimates as to the backlog of hospital beds as compared to needs. The apparent progress has been offset by the growth of the population and obsolescence, so that the shortage today is still over 800,000 beds.

The shortage of nursing home facilities of good quality is even more acute but is not mentioned. According to U. S. Public Health Service standards, more than 400,000 nursing home beds are required in the United States. Only 270,000 exist, and of these 116,000 are not acceptable on the basis of fire and health hazards.

The statistics presented on the numbers of doctors show a stable supply but do not highlight the decline in relation to population.

In the field of education, the data might well be arranged in a manner that would futher illustrate need for social programming. Basic data are presented on separate pages on the number of children of school age and the number of teachers. But the reader interested in the relationship between these two sets of data would have to go through his own calculations to see how serious the situation has become for all of us.

The special social needs of the low-income groups likewise are of general interest. After all, the health and welfare of human beings directly influence the size of our national output and the stability of our society.

Objective measures of the health and well-being of the various segments of the population are not easy to obtain. Certainly the editors of statistical compilations are not personally responsible for the absence of more basic research measures of this kind. But there is a basic challenge to appropriate agencies to develop essential information and to make it available with appropriate analysis.

By way of additional illustration, I will use the experience of the Advisory Council on Public Assistance charged by the Social Security Amendments of 1958 with examining the national program and reporting to Congress. We could get dollar figures on outlays and individual payments such as are carried in Trends, but it was much more difficult to obtain sufficient information on how well the program is fulfilling its function of maintaining the health and welfare of the people who turn to public assistance as a last resort. The lack of up-to-date information on actual budgetary needs of families of various types was one of the handicaps. At our request the staff developed some estimates of unmet needs which are to be published for the first time in the forthcoming report of the Advisory Council. Our recommendations include specific proposals that should lead to more basic information, including the development of national budget guides to help the states prepare their own family budgets.

When we examined present deficiencies in medical care, we were aware of tremendous gaps in the program and much poor quality service. In that area too we are urging the collection and publication of more information.

Even though the press should not note these points until the official release date, which may be January 4, I hope Dr. Bowman is lending a friendly ear.

In my statement to the Joint Economic Committee to which Mr. Stringham refers in his paper, I suggested certain measures of human well-being to provide objective data on progress made and problems still requiring attention. Some are in the Trends but others are not. I still think an effort should be made to include measures of housing adequacy and slum conditions, the number of families at various income levels by significant family types, the percent of children under 18 in broken homes, and the number of children attending school on part-time shifts.

The objective should be to reveal areas of social strain and personal maladjustment as a continuing challenge to develop more effective social programming.

While Mr. Cabello does not deal with this point, it would also be helpful in social programming to consider the cost of not meeting current needs in terms of unnecessary deaths, burdens on mental institutions and jails, continuing dependency, and the loss of valuable manpower.

XIV

QUANTIFICATION AND MEASUREMENT IN BIOLOGICAL AND SOCIAL SCIENCES

Chairman, William G. Cochran, Harvard University

Efficient Conversion of Non-Metric Information into Metric Information—Robert P. Abelson, Yale University, and John W. Tukey, Princeton University

EFFICIENT CONVERSION OF NON-METRIC INFORMATION INTO METRIC INFORMATION

By: Robert P. Abelson and John W. Tukey
Yale University and Princeton University

The title of this paper may prove misleading. "Conversion of non-metric information into metric information" may sound like getting something for nothing. In fact, we are concerned with getting a convenient modest something for an inconvenient modest something. Further, our methods are thus far limited to a particular class of situations. We do not have anything like a universal recipe for converting the qualitative into the quantitative, though a good deal can be done within the confines of the situations with which we are here concerned.

Consider n points to which we wish to assign numerical values, $X_1, X_2, \ldots X_n$. Suppose we have insufficient information to provide a natural or "correct" assignment of numerical values; our knowledge is limited to a set of constraints on the values, i.e. a set of inequalities on the X's. (For example, $X_1 \leq X_2 \leq \ldots \leq X_n$).

Consider first the purposes that an assignment of numerical values can serve in such a situation. It has become somewhat fashionable, particularly in certain areas of the behavioral sciences, to frown upon "arbitrary" numerical assignment (scaling) procedures. Some take the position that in the absence of a compelling rationale for numerical assignment, no numerical assignment whatever should be attempted. (Stevens, 1951). Thus if a set of points are known only up to a rank order, one is limited to the declaration of an "ordinal scale". Further manipulations using the scale are limited, so the dictum goes, to techniques appropriate to ordinal scales -- in particular, to those non-parametric statistical techniques designed for the analysis of rankings. The net effect of this dictum is to restrict the flexibility of statistical analysis severely and unnecessarily.

Reliance in such circumstances upon non-parametric procedures seems to us to be unwise, not because such procedures always lack power (90% power is no cause for disdain), but because they are poorly adapted to the variety of uses one requires for good insight into bodies of data. Often when adaptation to new uses is attempted, it is only at considerable sacrifice of power (as in the situations discussed here). Furthermore, <u>the typical state of knowledge short of metric information is not rank-order information</u>; ordinarily, one possesses something more than rank-order information. For example, one may know that X_1, X_2, and X_3 are ordered and in addition that X_2 is closer to X_3 than it is to X_1. Non-parametric techniques which take full advantage of such types of situation are generally unavailable. We would like to probe more deeply here, to gain some idea of what lies between rank-order scales and metric scales.

Consider now the kind of problem for which a numerical assignment procedure is useful. Suppose that the n points represent levels of an independent variable and that we wish to carry out the regression of a dependent variable (about which we have metric information) upon this independent variable (about which we have only non-metric information). To be even more specific, the independent and dependent variables might be imbedded in an analysis of variance design where we were interested in forming a single degree of freedom contrast among the levels of the independent variable. The appropriate coefficients to use in forming such a contrast would be a direct outcome of an assignment of numerical values.

To sum up thus far: we seek a procedure for assigning numerical values to a set of n entities, given a set of inequalities which the assigned values must obey. The problem is of interest because a) it sheds light upon the nature of knowledge intermediate between rank-order knowledge and metric knowledge, and b) the solution makes powerful regression techniques, particularly the formation of contrasts, applicable to many situations when the entities represent levels or versions of an independent variable.

<u>The criterion for good numerical assignment</u>.

The sequence of n numerical values to be assigned must obey certain inequalities. Likewise, the "ideal" values, the values one would assign if one had full scale knowledge, must obey the same inequalities. That is, both the sequence we choose and the sequence we ought to have chosen lie in the convex set of sequences permitted by the inequalities. Denote the chosen sequence by $[X_1, X_2, \ldots X_n]$ and the ideal sequence by $[Y_1, Y_2 \ldots Y_n]$. A convenient and reasonable criterion of the success of our choice is the <u>square of the formal product-moment correlation between [X] and [Y]</u>:

$$r^2 = \frac{\left[\sum_{i=1}^{n} (X_i-\bar{X})\,(Y_i-\bar{Y})\right]^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2 \quad \sum_{i=1}^{n}(Y_i-\bar{Y})^2}$$

To avoid confusion, one must note that this correlation coefficient is purely "formal" and is not to be thought of in terms of a bivariate distribution from which points are sampled. This r^2 plays a key role in the specific application discussed earlier. In testing the significance of a contrast, the power of the test increases directly with r^2. The same r^2 is almost ubiquitous in other aspects of regression analysis.

There is an obvious difficulty with r^2 as a criterion: one does not know the ideal sequence, [Y]. The sequence [X] is of our choice; but, in our ignorance, [Y] might be any sequence within a certain range of possibilities. A further choice must be made in order to provide a usable criterion. On the one hand, one might make some kind of distributional assumption about the possible [Y]'s, and average r^2 over this <u>a priori distribution</u>. It is difficult to do this in any reasonable and meaningful way. (Indeed the resulting mathematical problem is rather difficult to attack.) On the other hand, one might make the conservative, fixed assumption that <u>the Y - sequence may well be such as to minimize r^2 for the chosen X-sequence</u>. This minimum r^2, for [Y] satisfying the inequalities and [X] fixed, is the criterion we have chosen to assess any fixed [X]. The mathematical problem then becomes a maximin problem: how should one choose [X] such that the minimum r^2 is maximized?

In other words, we play a "game against Nature" in which we fear the worst. For any choice of numerical assignments for [X], assume that Nature chooses a set of "true" values [Y] which obey the inequalities but yield r^2 min., the lowest possible squared product-moment coefficient with [X]. We play the game by choosing [X] such that r^2 min. is maximized. (We refer to this choice as the "maximin sequence", denoted by [C]). This results in a guarantee that r^2 cannot be less than a certain value, denoted as r^2 maximin, so long as Nature obeys the inequalities. The strategy amounts to optimizing the conservative guarantee, rather than maximizing some kind of average value.

Mathematical properties of the problem.

Since the criterion is correlational, the units and origins of the sequences [X] and [Y] are immaterial to the maximin problem. In what follows, only the relative spacing of the n numerical values is of consequence, while the units and origins are chosen for convenience.

With a given set of inequalities there is associated a special set of sequences which we call "corners", such that any admissible sequence [Y] can be generated as a positive linear combination of the corner sequences. (The name corner arises from the geometrical conception of the permissible sequences as a convex set of vectors). As an example, consider the rank order case: $Y_1 \leq Y_2 \leq Y_3 \leq Y_4$. Fixing $Y_1 = 0$, a simple set of corners is the triplet:

(0, 0, 0, 1)
(0, 0, 1, 1)
(0, 1, 1, 1)

Any [Y] satisfying rank order (with $Y_1 = 0$) can be expressed as a positive linear combination of these corners.

The corners provide the key to the maximin solution, via the following two theorems proved by Tukey. Proofs are given in our more extended article. (Abelson and Tukey, 195[illegible]).

Theorem I. For any fixed [X], minimum r^2 is reached for [Y] equal to one of the corner sequences. In other words, whatever choice we make for [X], Nature plays her most damaging game at one of the corners. Consider the rank order case again, and suppose we "play" the equal-interval sequence -3,-1,1,3. Nature achieves r^2 min. = .600 by playing 0,0,0,1 or 0,1,1,1. No worse than this can be done to us when we play -3,-1,1,3. (However, we have a better play in the maximin sense).

Theorem II. (Oversimplified) The maximum r^2min is achieved by the sequence which correlates equally with all corner sequences. In the rank order case with n = 4, we need simply find the sequence (C_1, C_2, C_3, C_4) which correlates equally with (0,0,0,1), (0,0,1,1), and (0,1,1,1). This is a matter of simultaneous linear equations in the unknown C's which are readily solved.

Theorem 2 as it has been given here is not correct for all sets of inequalities. In particular, the theorem fails when the sequence which correlates equally with all corners does not itself satisfy the appropriate inequalities. Further complications arise when there are more than (n-1) corners in a given case. The fuller paper goes into the subtle details involved. The correct but more involved theorem will simply be stated here in passing: For any system of inequalities with its associated corner sequences, there exists one and only one sequence which a) is a positive linear combination of a set of the corner sequences such that b) it correlates equally highly with all these corners and c) more highly with corners not in the set, if any.

Results for the rank order case

First we present in some detail the results for the rank order case. Then, more briefly, the results for other cases. Throughout we use for the values of the maximin sequence the convenient normalization

$$\sum_{i=1}^{n} C_i = 0 \qquad \sum_{i=1}^{n} C_i^2 = 1/r^2 \text{ maximin}$$

The maximin sequence in the rank order case for n=4 is: -.866, -.134, .134, .866. For n=8, the values are: -.935, -.289, -.144, -.045, .045, .144, .289, .935. For indefinitely large n, the limiting values for the extreme points are: -1.000, -.414, -.318, -.268..........,.268, .318, .414, 1.000.

The solution is markedly non-linear. The values at the two ends have very large relative separation from the next values inwards. This comes about because the solution is guarding against the possibility that Nature will play 0, 0, ... 0,0,0,1 or 0,1,1,1, ... 1,1. A linear sequence can fail rather badly against these possibilities, especially for large n. However, in practice one is often unwilling to acknowledge sequences as pathological as a,a,... a,a,a,b as reasonable possibilities for the "true" sequence. If so, then one may attempt to rule out such unusual sequences from Nature's repertoire. This means reformulating the inequalities so that these pathological corners do not occur. This is possible in a number of ways, all of which require that something more stringent than mere rank order be assumed. When this is done, one finds that the end values of the maximin sequence are not forced to lie so far from the body of the sequence as in the rank order case, and a linear sequence does not fare as poorly (in the maximin sense) as a basis for numerical assignment. Of this, more later.

The following brief display gives r^2 maximin in the rank order case for various values of n; by way of comparison, the values of r^2 min. against a linear assignment are shown.

n	5	10	20	50
r^2 maximin	.596	.478	.406	.339
r^2 min.(linear)	.500	.273	.143	.059

100	200	500	1000	n
.303	.274	.244	.225	2/2+log.(m-1)
.030	.015	.006	.003	3/(m+1)

Asymptotically, r^2 maximin approaches zero, but very slowly, whilst r^2 min. for a linear sequence approaches zero rather rapidly.

The form of the maximin sequence may be roughly approximated with a simple pattern of integers by the following device: write down a linear sequence with mean zero, quadruple the extreme values and double the next-to-end values. At n=8, for example, this quick approximation to the maximin sequence would be (-28,-10,-3,1,1,3,10,28). For n less than 50, the r^2 min. for this approximation is at least 90% as high as r^2 maximin. This scheme, which we dub as the "linear-2-4" sequence, is easily remembered.

If Nature is really playing a near-linear sequence, then of course we would be better off by playing a linear sequence than by guarding against wild behavior of Nature by playing the maximin solution or its surrogate, the linear-2-4. If we would like to achieve higher r^2 in case Nature's behavior is near-linear without risking too great a drop in r^2 below the maximin value in case Nature's behavior is wild, a good hedge for small n is to choose a "linear-2" sequence; that is, a linear sequence with the end values doubled. In passing, it might be mentioned that "rankits", like linear coefficients, fare poorly when Nature is behaving wildly.

Other orderly cases

Definitions

I. Symmetric Rank order

$$X_1 \leq X_2 \leq X_3 \leq \ldots \leq X_{n-1} \leq X_n$$
$$D_1 = D_{n-1},\ D_2 = D_{n-2},\ D_3 = D_{n-3} \ldots$$
where $D_1 = (X_2 - X_1),\ D_2 = (X_3 - X_2) \ldots$
$$D_{n-1} = (X_n - X_{n-1})$$

II. Symmetric, Extremes Bunched

$$X_1 \leq X_2 \leq X_3 \leq \ldots \leq X_{n-1} \leq X_n$$
$$(D_1 = D_{n-1}) \leq (D_2 = D_{n-2}) \leq (D_3 = D_{n-3}) \cdots$$

III. Non-symmetric, Extremes Bunched

$$X_1 \leq X_2 \leq X_3 \leq \ldots \leq X_{n-1} \leq X_n$$
$$D_1 \leq D_2 \leq D_3 \ldots;\ D_{n-1} \leq D_{n-2} \leq D_{n-3} \cdots$$

IV. Symmetric, Extremes Spread

$$X_1 \leq X_2 \leq X_3 \leq \ldots \leq X_{n-1} \leq X_n$$
$$(D_1 = D_{n-1}) \geq (D_2 = D_{n-2}) \geq (D_3 = D_{n-3}) \ldots$$

V. Non-symmetric, Extremes Spread

$$X_1 \leq X_2 \leq X_3 \leq \ldots \leq X_{n-1} \leq X_n$$
$$D_1 \geq D_2 \geq D_3 \geq \ldots;\ D_{n-1} \geq D_{n-2} \geq D_{n-3} \cdots$$

VI. "Diminishing Returns"

$$X_1 \leq X_2 \leq X_3 \leq \ldots \leq X_{n-1} \leq X_n$$
$$D_1 \geq D_2 \geq D_3 \geq \ldots \geq D_{n-1}$$

(or mirror image)

Maximin r^2

n	I	II	III	IV	V	VI
3	1.000	1.000	.750	1.000	.750	.933
4	.853	.947	.909	.974	.667	.887
5	.853	.974	.778	.947	.625	.834
6	.786	.940	.874	.922	.599	.827
7	.786	.964	.787	.901	.578	.806
8	.744	.936	.856	.882	.561	.789
9	.744	.958	.791	.865	.548	.774
10	.714	.935	.845	.850	.536	.761
11	.714	.953	.793	.837	.526	.750
12	.692	.935	.838	.826	.517	.740
13	.692	.950	.794	.815	.507	.731
14	.674	.935	.832	.805	.501	.724
15	.674	.949	.795	.796	.492	.716
16	.659	.934	.827	.788	.487	.710
17	.659	.948	.795	.780	.479	.704
18	.646	.934	.825	.773	.475	.699
19	.646	.947	.796	.767	.467	.694
20	.636	.934	.823	.762	.464	.689

Maximin weights exemplified: n = 8

	I	II	III	IV	V	VI
C_1	-.707	-.477	-.548	-.707	-.935	-.935
C_2	-.293	-.395	-.418	-.219	-.110	-.160
C_3	-.225	-.312	-.289	-.131	-.055	-.062
C_4	-.189	-.230	-.159	-.044	0	.036
C_5	.189	.230	.159	.044	0	.134
C_6	.225	.312	.289	.131	.055	.231
C_7	.293	.395	.418	.219	.110	.329
C_8	.707	.477	.548	.707	.935	.427

Further cases for small n

In the literature on psychological scaling, the case in which the first differences of a ranked sequence are ranked is called an "ordered metric scale" (Coombs,1950). The cases treated above are a limited coverage of this variety of scale. Next we consider all possible ordered metric scales with n=3,4,or 5. Also, we consider for n=4 all possible "higher-ordered metric scales" (Siegel, 1956). These are cases in which all differences of a ranked sequence are ranked. In addition, the rank order case for n=3 and 4 is considered with a numerical constraint on the relative size of the biggest or the smallest interval.

Ordered metric scales with n=3

There is only one case here. We have $X_1 \leq X_2 \leq X_3$ and $X_2 - X_1 \geq X_3 - X_2$. (The other possibility is simply a mirror image of this one). The maximin sequence can be approximated with the simple integer sequence (-7,2,5) with r^2 min. = .923.

n=4

Maximin coefficients and maximin r^2, for all cases of n=4 involving only simple inequalities among differences.

1. The three differences $(X_2 - X_1)$, $(X_3 - X_2)$ and $(X_4 - X_3)$ are represented by the digits 1,2, and 3.
2. When the inequalities specify that a particular difference is greater than another, the larger difference is written first (e.g. the system of inequalities $(X_3 - X_2) \geq (X_2 - X_1) \geq (X_4 - X_3) \geq 0$ is written 213.)
3. When the relative size of two differences is not specified, they are enclosed in parentheses. (e.g. the system of inequalities $(X_2 - X_1) \geq (X_3 - X_2) \geq 0$; $(X_2 - X_1) \geq (X_4 - X_3) \geq 0$ is written 1 (23).

System	C_1	C_2	C_3	C_4	r^2
(13)2	-87	00	00	87	.667
(12)3	-87	-13	50	50	.789
1(23)	-87	05	27	55	.887
132	-87	05	27	55	.887
123	-87	04	29	54	.887
2(13)	-66	-34	34	66	.909
213	-66	-34	42	58	.941

n=5

System	C_1	C_2	C_3	C_4	C_5	r^2
(124)3	-89	-20	00	20	89	.595
(123)4	-89	-20	00	55	55	.694
(14)(23),(14)23	-89	00	00	00	89	.625
(13)(24),(13)24, (13)42	-89	-10	-10	55	55	.704
(12)(34),(12)34	-89	-20	33	36	40	.801
(23)(14),(23)14	-55	-55	00	55	55	.833
1(234)	-89	00	08	25	57	.840
1(34)2,1342, 14(23),1432	-89	01	04	29	55	.843
1(24)3, 1423	-89	-04	12	29	52	.853
1(23)4,12(34),(12)43, 1324,1234,1243	-89	-05	13	32	50	.854
2(134)	-69	-40	16	35	58	.892
2(14)3	-69	-40	26	26	58	.901
24(13)	-61	-49	16	36	58	.914
23(14)	-56	-54	14	40	56	.918
2(34)1,2341,2431	-56	-54	16	36	57	.921
2(13)4	-69	-40	18	44	47	.923
2413	-61	-49	21	30	58	.927
21(34),2134,2143	-69	-40	21	38	50	.931
2314	-63	-47	19	37	53	.935

Higher-ordered metric scales for n=4

These cases can be specified as elaborations of ordered metric scales. The notation is as in the previous displays, with the addition of constraints upon the sums of differences. For example (in the abbreviated notation), $1+2\geq 3$ signifies $D_1+D_2\geq D_3$.

System	C_1	C_2	C_3	C_4	r^2
123; $1\geq 2+3$	-87	07	36	43	.933
123; $1\leq 2+3$	-73	-19	35	58	.972
132; $1\geq 2+3$	-87	16	16	55	.908
132; $1\leq 2+3$	-72	-10	13	69	.974
213; $2\geq 1+3$	-66	-34	50	50	.952
213; $2\leq 1+3$	-73	-19	35	58	.972

A further case with n=3

Rank order is known, and the ratio of the larger D to the smaller D does not exceed K. (We do not know which D is larger, however).

K	r^2
9	.824
4	.893
2	.964
1.5	.987

The maximin sequence for any K can be represented most simply by (-1,0,1).

n=4

Rank order is known, and the largest D (whichever it is) does not exceed the fraction p of the range.

p	C_1	C_2	C_3	C_4	r^2
.40	-67	-24	24	67	.981
.50	-71	-24	24	71	.893
.60	-71	-26	26	71	.865
.70	-75	-22	22	75	.816
.80	-79	-16	16	79	.765
.90	-83	-14	14	83	.711

Rank order is known, and the smallest D, whichever it is, is not less than the fraction q of the range.

q	C_1	C_2	C_3	C_4	r^2
.05	-82	-15	15	82	.723
.10	-78	-16	16	78	.796
.15	-74	-18	18	74	.865
.20	-71	-19	19	71	.924
.25	-69	-20	20	69	.971
.30	-68	-22	22	68	.988

Discussion

Results for a large number of cases have been presented. Many of the maximin r^2's are seen to be in the .80's or .90's. This is quite good for most analytic purposes. Thus a little non-metric information will go a long way when it is converted to metric information. A comparison of cases makes it clear that maximin r^2 is most readily boosted above the rank order value when the inequalities put bounds upon the external intervals of the sequence. "Extremes bunched" is a more favorable case than "extremes spread"; for n=4 and 5, r^2 is higher when an internal interval is known to be biggest than when an external interval is known to be biggest. Restrictions on the fraction of the total range allotted the biggest interval result in powerful increases in r^2; this comes about because huge external intervals are thereby prohibited. One way to summarize this class of results is to say that scales with big gaps in the middle are more "robust" than scales with big gaps in the tails.

Certain other general conclusions are apparent from the results: symmetry is a fairly powerful condition; higher-ordered metric scales can be very close indeed to numerical scales; and so on.

Nevertheless, many of you are no doubt wondering about the proof of this pudding. How often can maximin sequences actually be put to good use?

The answer is not cut-and-dried. Consider the rank order case. Here the values of maximin r^2 are only fair; moreover, the maximin sequence has an unfamiliar flavor. The end values are moved far out to guard against wild plays of Nature. Are we seriously recommending that for a rank order case with, say n=6, the contrast (-20,-6,-1,1,6,20) be used to capture the single degree of freedom associated with the rank order?

The investigator might reject the appropriateness of this contrast. He might say, "It is too bizarre. Give me straight-forward linear weights, or perhaps rankits. I do not foresee that Nature will play tricks on me". Our reply would be, "If you say your non-metric information is rank order and nothing more, then you implicitly acknowledge the possibility of a "true" sequence of the form (a,a,a,a,a,b). A conservative man would protect himself against such a possibility. If you say that this possibility is inconceivable, then you really have more non-metric information than mere rank order. If you could define this extra knowledge precisely, it would lead to another maximin sequence, one that might strike you as intuitively more reasonable .

Here lies the heart of the situation. Quite commonly, when we say we only know rank order, we actually know more than this, but don't know how to express what else it is that we know. Typically, our excess "knowledge" is to the effect that the scale is no worse than mildly curvilinear, that Nature behaves smoothly in some sense. This is a more vague and general conception than any of the highly specific cases considered in this paper. The maximin method needs extension to this general case, the problem being to specify the inequalities and corners in some reasonable way. The same problem, seen from a different standpoint, has been apperceived by Mosteller (1958). The problem is clear; the solution is not.

The murkiness of the general "rank-order-plus-smoothness" case should not obscure the fact that in a good many practical situations the maximin approach can straightforwardly be used to good effect. Perhaps the leading candidate for a clear-cut case is the ordered metric scale for n=3. Excellent use can be made of the contrast (-7,2,5) in the situation where X_2 is known to lie between X_1 and X_3 but nearer to X_3 than to X_1. One instance of the use of this contrast is already in the literature. (Sarnoff, 1960). It is hoped that other instances of practical application will make their appearance in the near future.

REFERENCES

Abelson, R.P. & Tukey, J.W. (1958) Efficient utilization of non-numerical information in quantitative analysis: I. General theory, the case of simple rank order, and some other systems of simple irregularities. Princeton University (dittoed).

Coombs, C.H. (1950) Psychological scaling without a unit of measurement. Psychol. Rev., 57, 145-158.

Mosteller, F. (1958) The mystery of the missing corpus. Psychometrika, 23, 279-289.

Sarnoff, I. (1960) Reaction formation and cynicism. J. Personal., 28, No. 1

Siegel, S. (1956) A method for obtaining an ordered metric scale. Psychometrika, 21, 207-216.

Stevens, S.S. (1951) Mathematics, measurement, and psychophysics. In S.S. Stevens (Ed.) Handbook of experimental psychology. New York: Wiley.