SELECTED COMPUTER APPLICATIONS IN PROCESSING DATA FROM A CONTINUING HEALTH HOUSEHOLD INTERVIEW SURVEY

Walt R. Simmons, Robert T. Little, and Eleanor L. Madigan
National Center for Health Statistics

Introduction

In his Introduction to the Development of Economic Doctrine, Alexander Gray comments that in no other book on the subject are so many important topics omitted.[1] That thought is an appropriate caveat for the present paper. This report presents no new statistical theory or methodology of any consequence, nor does it offer any striking patterns of programming computers. It does not even mention such features as computer memory, capacity, speed, or other characteristics of the equipment. The sole purpose is to display for those practicing statisticians who may not have had much opportunity to use computers, some illustrations in which a computer has been used to facilitate the reduction of data to more useful information. Computer processing has accomplished some of these calculations more rapidly than would otherwise have been true, and in one or two instances has made feasible a method which otherwise might not have been undertaken. Some of the illustrations may suggest more elaborate techniques which the electronic computer makes possible at tolerable cost. The applications discussed were all carried out with a Univac I, but it is believed that, with relatively minor modifications, other smaller or larger computers might have been used.

Attention is invited to one very general kind of circumstance, before considering specific examples. It has become a commonplace thought that any, or at least very nearly any, logical process which can be described in sufficiently precise terms can be programmed for computer operation. Similarly, at least in theoretical terms, and given sufficient time, any process or computation which an electronic computer can accomplish, could be done by a person with much simpler tools. Thus any statistical processing or analysis which can be described in mathematical terms can be carried out on a computer. And there is no magic which the "Giant Brains" computers use, which constructively cannot be matched by ordinary mortals.

The U. S. National Health Survey

Illustrations offered in our remarks all come from the conduct of the household interview phase of the U. S. National Health Survey. The illustrations can be viewed in better perspective if leading features of the Health Interview Survey—which will be referred to as HIS— are recalled briefly. The HIS is a continuing survey which has as its purpose determination of the incidence and prevalence of illness, impairments, and injuries; the use of medical, dental and hospital facilities, and related health phenomena for the civilian noninstitutional population of the United States. The survey is the consequence of more than 10 years of exploratory activity culminating in legislation in 1956, which directed the Surgeon General of the Public Health Service to put into operation such a system. Working under detailed specifications of the Public Health Service, the Census Bureau prepared the original survey design and collects the data. Collection is by personal interview of people in their homes, mainly through a structured questionnaire. The schedule includes some 100 questions on demographic and socioeconomic characteristics of respondents and on their health status and experiences over recent time periods.

The sample design is such that each week's interviewing constitutes a representative sample for the nation. Hence, trends can be produced for high-incidence statistics, while combinations of samples for 13 weeks, or 52, or 104 weeks, or even longer periods, can produce data for items of lower frequency in the population. Most analysis thus far has rested on interviewing for 52 weeks. Over a year the sample includes approximately 130,000 persons in 38,000 households in 6,400 land area segments in 503 PSU's (counties or groups of counties). The sample design has several times been modified in the last four years, through joint consideration and decision by the National Health Survey and the Census Bureau. But it has, throughout its life, thus far, retained its original essential features. It is a multistage, highly stratified probability sample of land segments, and of the persons resident in those areas.

A minimum amount of coding and processing of questionnaires is carried out by usual manual methods. The data are then doc-sensed, and, via punch cards, placed on magnetic tape. Most further processing is accomplished on an electronic computer.

Types of Computer Processing Discussed

The illustrations of computer processing to which we call attention in this paper may be considered as coming from four classes or functions of statistical reduction of data: (a) Editing and Coding, (b) Estimation, (c) Tabulation, and (d) Special Analyses.

Editing and Coding

It has been found convenient to describe the survey record for a sample person in terms of the punch cards which are used to transcribe that record. These are a household card; a person card; a (health) condition card for each reported illness, injury, or impairment; and a hospital card for each episode. The purpose of the Editing and Coding is to clean the data and prepare them for estimation—they are to be converted from raw data into a more efficient input of higher quality. This is a general process which is as old as the concept of tabulating statistics, but only in recent years has there been both full recognition of the fundamental importance of the editing steps, and the capacity to carry out extensive editing at tolerable costs. Computers have made the execution possible and have sharpened awareness of the value of particular editing steps.

One series of runs or passes through the computer cleans the deck of intolerable data, edits fields to code range and to item consistency, and transfers household and person characteristics to appropriate person-, condition-, or hospital-record. In the same series, experiences from the condition- and hospital-records are summed and entered into the proper person-record for later use.

Among the editing processes is a programming technique, which is a type of indirect addressing, and which is used to recode each health condition and hospital episode from the detail of the International Classification of Diseases (ICD) to a broader classification of 280 groups—and then still further to groups of 38, 30, and 11 classes. Concurrently, with this recoding a total of 206 subroutines are used to edit data on diagnosis and on hospital procedure for internal consistency and plausibility.

This is accomplished by the computer examining the ICD code for validity. If found valid the number itself is reshaped by a series of mathematical manipulations into an instruction used to address memory. The generated instruction directs the computer to examine the content of the appropriate memory positions. The numerical structure of the three digit number found determines whether it is a recode of the ICD and is put into the record, or whether through additional modification becomes a transfer instruction that causes a branch to the appropriate subroutine for the prescribed edits.

To illustrate, suppose the condition ICD 611 (prostatitis) was reported. The computer would use the ICD number 611 to generate an instruction that would cause it to branch and examine the data at address 453. An inspection of the 3-digit number found at this address would determine if this number is to be used to recode ICD to one of the categories in the 280-group recode or used to cause the computer to branch to a routine that would examine the sex, age, and type of condition fields.

Since prostatitis is a disease that is chronic and can occur only in males, age 25 or over, in this case the 3-digit number or word will direct that a sequence of age, sex, and chronicity tests be made. If the case meets the tests, it will be recoded 184 as a prostate disease. For the other possibilities it will be recoded into more general categories.

At the end of each pass the computer types on the console printer the number and kind of edit changes that have been made. A supplementary listing identifies the person for which a change was made, and thus permits a human review of the original schedules and consequent corrective action, whenever the volume of editing changes of a particular kind appears to warrant such action. At the current level of editing, the volume of computer changes in data is quite modest, the annual rate being slightly over 1,000 changes for something like 13,000,000 pieces of data. Of the changes, about 250 are modified diagnoses.

Estimation

Not so many years ago estimation from probability samples was restricted largely to inflation of sample data by the reciprocal of the sampling fraction. This was followed by methods which used the simpler forms of stratification and ratio estimation. Estimation continued to be—for large-scale operations—a relatively uncomplicated procedure, partly because theory was perhaps unimaginative, but partly too, because more complex estimators

Figure 1. Condensed estimating equation, for any statistic x in the U. S. National Health Household Interview Survey.

NOTATION

Subscripts:

- g = stratum
- h = PSU
- i = Segment
- j = household
- k = person
- α = color-residence zone
- β = age-sex-color class

Statistics:

- x = a health characteristic; = 1 if present, equal 0 if not present.
- y = current population value; = 1 for each individual person.
- z = 1950 population value; = 1 for each individual person in 1950 Census.
- $\cdot$ signifies summation over missing subscripts.

Number of Units:

- m = number of sample segments
- n = designed number of sample households
- n' = responding number of sample households

(Subscripts indicate scope of inclusion where notation might otherwise be ambiguous.)

Weights:

- u = original 1st stage design weight for individual person
- v = original subsequent stages design weight for individual person
- w = design subsampling weight for out-sized segments

(Subscripts are not shown in weights, but they vary along with the statistics which they multiply.)

Summations: All summations are over the universe unless otherwise indicated.

The Estimate of x is:

$$x' = \sum_{\beta} \frac{x'_{\cdot\beta}}{y'_{\cdot\beta}} \; y_{\cdot\beta} \quad , \quad \text{in which}$$

$$x'_{\cdot\beta} = \sum_{g}^{503} \sum_{h}^{1} \sum_{i}^{m} \sum_{j}^{n'_{ghi}} \sum_{k} \sum_{\alpha} \frac{n_{ghi}}{n'_{ghi}} \; \frac{\sum\limits_{g}^{503} \sum\limits_{h}^{1} \sum\limits_{i} \sum\limits_{j} \sum\limits_{k} \sum\limits_{\beta} u \, z_{ghijk\alpha\beta}}{\sum\limits_{g} \sum\limits_{h} \sum\limits_{i} \sum\limits_{j} \sum\limits_{k} \sum\limits_{\beta} z_{ghijk\alpha\beta}} \; u \, v \, w \, x_{ghijk\alpha\beta} \quad , \quad \text{and}$$

$$y'_{\cdot\beta} = \sum_{g}^{503} \sum_{h}^{1} \sum_{i}^{m} \sum_{j}^{n'_{ghi}} \sum_{k} \sum_{\alpha} \frac{n_{ghi}}{n'_{ghi}} \; \frac{\sum\limits_{g}^{503} \sum\limits_{h}^{1} \sum\limits_{i} \sum\limits_{j} \sum\limits_{k} \sum\limits_{\beta} u \, z_{ghijk\alpha\beta}}{\sum\limits_{g} \sum\limits_{h} \sum\limits_{i} \sum\limits_{j} \sum\limits_{k} \sum\limits_{\beta} z_{ghijk\alpha\beta}} \; u \, v \, w \, y_{ghijk\alpha\beta} \quad , \quad \text{and}$$

$y_{\cdot\beta}$ is the official independent control estimate for the β th age-sex-color group for the total U. S. civilian non-institutional population.

NOTE: The adjustment $\frac{y_{\cdot\beta}}{y'_{\cdot\beta}}$ is applied separately to each sample datum in the β th age-sex-color class would require calculation which was too costly in terms of dollars or time. Again the computer has made recent theoretical advances operationally feasible. At the risk of being guilty of huckstering, but in the effort to illustrate the point, we display a somewhat condensed version of the basic estimating equation for the Health Interview Survey.

Without belaboring the point, but with the reminder that this displayed estimating equation does not reflect all the detail of actual estimation, and noting that the complex estimating process is applied to each of the 130,000 persons interviewed each year, it is obvious that the computer permits an elaborateness in estimation which would rarely, if ever, be attempted by less powerful computational devices.

An important characteristic of computer estimation is that it permits differential weighting of original observations at very small cost. Thus, it invites the use of efficient survey designs which otherwise might be foreclosed.

An illustration of the kind of estimating steps taken is given here by a description in outline form of the last phase of the process, which is an adjustment of all data by ratio factors to bring them into agreement with independent official population estimates prepared by the Bureau of the Census. The algebra of this step is to multiply every datum in each of 60 age-sex-color cells by the ratio of a population control to the sample population estimate for that cell before adjustment. Two computer runs are required to accomplish this adjustment, controlled by machine instructions designed to (1) read all data records into the computer, isolate, classify, and tally each weighted person-record into the proper one of 60 age-sex-color classes and code the record with that class indicator; (2) introduce into computer memory the independent estimates for the 60 specified classes; (3) carry out the necessary arithmetic to calculate ratio factors and store these factors in a set pattern in memory for use in the next pass of the data records; (4) again read the data records, use the class code to select the proper adjustment factor stored in memory, and multiply the previous weight by that factor to produce a new weight; and (5) put this adjusted weight into all records associated with that person.

This phase of estimation is not a complex computer operation compared with several others in the HIS survey. However, the same adjustment on conventional punched-card equipment would develop into a complicated series of processing steps and likely require some supporting manual operations. Time and cost would be prohibitive whereas only four hours of computer running time for one year's data are required to complete this process.

Tabulation

An element of the estimation process, which is not explicitly evident from the equations, is that the computer is programmed to convert each observed person, condition, or hospital episode to an estimate of that part of the total universe which it represents. Hence, in principle, simple addition of the converted data for any class produces a population estimate for that class. This conceptually simple process is slightly more involved in operation for two reasons. The first of these may be thought of in general terms as a matter of sorting, classification, and rearrangement of records. Ordinarily tabulation cannot be accomplished by a simple automatic retrieval of data stored in the computer. It is necessary through one of several search and assembly procedures to collect cases which belong in the same cell, and then count them—or add their adjusted weights—to produce the desired table. The other special action which must be taken is adjustment of the data to the desired time scale. Basic weights in the estimation procedure are set at such a level that direct tabulation of the sample over 13 weeks of interviewing produces an average value of the relevant statistic for a 13-week period. When 52 weeks of interviews are tabulated, a multiplier of 1/4 must be introduced into the tabulation to preserve the feature of average-over-the year. For some statistics, such as number of physician visits, the reference period for the direct statistic is a 2-week interval. If total physician visits in a year are desired, and 52 weeks of interviews are tabulated, the correct multiplier is 6.5, because the basic weighting applied to the sum of 2-week experience for each of four quarters produces an estimate for eight weeks, which must then be multiplied by 6.5 to produce a 52-week aggregate. For other combinations of number of weeks of interview, length of reference period, and type of statistic other multipliers are appropriate. In all cases, the scaling is a simple matter, and great flexibility in tabulation is possible with a minimum of new programming.

Attention is called to a standard procedure used in NHS to facilitate programming for tabulation, although the process is not dependent on tabulation by computer. We refer to the regular practice in planning tabulations of preparing annotated dummy tables. These tables are prepared as early as possible in the entire collection process, preliminary versions usually being made before the questionnaire has been

Figure 2.

Fiscal 1961 Quarterly

PREVALENCE OF TOTAL CHRONIC CONDITIONS CLASSIFIED BY VARIOUS MEASURES OF SEVERITY AND DAYS OF RESTRICTED ACTIVITY, BED DISABILITY AND WORK OR SCHOOL LOSS ASSOCIATED WITH THESE CHRONIC CONDITIONS ACCORDING TO SEX AND AGE

Sex and Age	Number of Chronic Conditions						Number of Days		
	Total	Total Medically Attended	With bed days in the year	With Restricted Activity Days in 2 Weeks	With Bed Disability Days in 2 Weeks	With Days Lost From Work or School in 2 Weeks	Of Restricted Activity	Of Bed Disability	Lost from Work or School
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Both Sexes									
Under 15									
15-24									
15-16									
17-24									
25-44									
45-54									
55-64									
65-74									
75+									
Male									
Ditto Age									
Female									
Ditto Age									

1. Include all chronic condition cards with "1" in column (aa)
2. Medically attended - column C code 1
3. Restricted Activity Col. G
   Bed Days Col. H
   Work & School Loss Col. I & J
4. Cols. 1-6 x basic weight
   Cols. 7-9 x 6.5 weight

$$S_2 \times A_7 \left[(T + SP_1 + DBY + DRW + DBW + DWSW)_c + (DRW + DBW + DWSW)_d \right]$$

designed. A typical illustration of such a table is shown here as figure 2. Those symbols which are not self-explanatory in the table are code terms which are a shorthand description of NHS tables.

In 1961 approximately 775,000 separate cells of tabulated data were assembled for this project with a total running time of about 700 hours.

Special Analyses

Clearly there is no limit to the possible special analyses which might be undertaken when one has many thousands of items of data on tape and the power of a computer with which to relate some of those data to others. The most important point we could make here probably is just that: the variety of possible analysis is restricted only by one's imagination, and that extensive actual analysis becomes feasible because the computer can make millions of comparisons and computations on individual items in a matter of minutes.

We shall restrict our discussion to a single illustration. This concerns a matter on which at NHS we have expended a good bit of energy, but feel we have just scratched the surface of the work that needs to be done.

Among the questions in the interview is a series which asks whether the person had a hospital episode within the past year, and, indirectly through admission date and duration of stay, whether and when a hospital discharge occurred. Regular processing of these data produces an estimate of hospital discharges experienced over the year previous to interview by the population living at the time of interview. Except for discharges within that period for persons no longer alive at the end of the period, this estimate ought to be equivalent to total number of discharges from hospitals annually. For appropriate definitions of hospitals and of discharges, no very precise count of this latter figure is known from any source outside the NHS. Informed speculation did lead to a preliminary guess, however, that the NHS tabulated data on number of discharges showed a total which likely was 15-20 percent smaller than the unknown true count would be. A pilot study of hospital experience of decedents in their last year of life[2] suggested that decedents might account for something of the order of one half of these 15-20 percentage points. This left an unexplained deficiency possibly in the neighborhood of 7-12 percent. Several plausible hypotheses might "explain" the difference. One of these postulates a recall-loss or memory decay which increases with distance between the occurrence of an event and the time of interview. This notion has a certain intuitive plausibility, and has received attention from a considerable number of people, but published findings are quite limited. (See e.g., references 3, 4, and 5.) A more precise statement of this hypothesis would recognize recall-loss as the resultant not of pure memory failure but of a complex of factors including such elements as differential evaluations over time, financial impacts, and a variety of motivations.

For NHS data, it is possible to calculate for each reported discharge the interval between discharge and date of interview. A full description of this process requires several pages, and is not necessary to the present account. It is sufficient to call attention to several key features: (1) All dates involved, some of which were originally expressed in days, some in weeks, and some in months, must be translated to a common scale to permit direct arithmetic operations; (2) persons which had reported hospitalization, none of which upon edit was in the year previous to interview, should be excluded; (3) episodes of persons who were still in the hospital on the Sunday night before interview should be excluded, since they did not constitute discharges; (4) provision must be made for treating instances of "information not available" for each item of original entry; (5) several classes of case can be distinguished on the basis of relationships among admission dates, length of stay, and interview date; since no single method of computation is best for all classes, a routine was introduced which chose the exact form of calculation for discharge date according to the class of the case. For those interested, an outline of the procedure is given in figure 3.

This matter of interval between discharge and interview was given attention because of the following hypothesis. It ought to be true that, within sampling error, approximately equal numbers of hospital discharges occur in the first month prior to the interview date, in the second month before interview date, the third month, or during the k^{th} month before interview for k, say, not greater than 20. If the NHS data could be tabulated to show the number of hospital discharges reported by length of interval between discharge and interview—called henceforth, the J-interval—the relationship between the a priori hypothesis of level experience and reported data could be analyzed.

The topic is an intriguing one. This paper might easily have devoted its full allotment of space, and more, to discussion of associated theory. But that is not the function of the present forum. Accordingly, we shall treat the subject in terms of a very simple model. Our purpose is not to argue necessarily that this is a best choice of model, but to illustrate how the computer aids in solving the problem of this model.

Assume that for a specified population there are $\bar{D}$ hospital discharges in a unit time period, say a lunar month or 4-week period. $\bar{D}$ is assumed to be constant. Let q_x be the probability that a discharged person dies x lunar months after the discharge—for this analysis, the lunar month is considered an indivisible time unit; q_o is the probability that the dischargee is dead at the time of discharge. The quantity r_J is the rate of reporting discharges by living persons for a specific J-interval—i.e., J lunar months between discharge and interview. The ratio $r_j = 1$ represents perfect reporting; r_j less than unity represents net underreporting; r_j greater than unity represents net overreporting. With this formulation, the reported number of hospital discharges for a given J interval is:

$$D_J = \bar{D}\,(1 - \sum_{x=o}^{J} q_x)\; r_J .$$

The quantity in parentheses is the probability that a dischargee is still alive J months after discharge. This quantity is necessarily equal to or less than unity and is monotonically decreasing for increasing values of J. If the recall-loss hypothesis holds in its simplest form, r_J is also monotonically decreasing, and at $J = 0$ is equal to unity. Under these assumptions, D_o is equal to the number of discharges of live persons in a month, and D_J is a monotonically decreasing function of J. The 2nd order power series approximation to the D_J function might be written

$$D'_J = a_o + a_1 J + a_2 J^2$$

and the estimated level monthly volume of discharges as $\bar{D}' = \dfrac{D'_o}{1 - q_o}$.

Using the methods sketched earlier, the computer calculated from observed data, the quantities D'_J for each value of J, for J = 1 to 13. (Actually the first computer calculations were made for J in terms of days; they were assembled into lunar month terms manually, but it would have been a simple matter to have converted to lunar months on the computer.) Calculations were made separately for each of

Figure 3. Outline of Calculation for Interval between Discharge and Date of Interview

Part A. Initial Data and Notation

X: = 1 if case has 1 + days in hospital in previous 12 months

= 0 otherwise

Y: = 1 if still in hospital on previous Sunday night before Census date

= 0 otherwise, including D.K.

A': = month and year of admission

A: = A' converted to a coded day, with June 1956 coded zero; July 1956 coded 1, etc. In essence, the code assumes that all admissions occur on the 15th or 16th of the month. Thus an admission of August 1957 is coded A = 365 + 31 + 15 = 411 [Coding done by table-look-up process.]

C': Week of interview

C: C' translated into day code. Table assumes interviewing takes place on Tuesday, which is median day. [Also by table-look-up process.]

D: "12 months before interview week" is calculated in computer as C minus 367

E: "Number of hospital days in previous 12 months," is reported on questionnaire

F: Derived Date of Discharge (See below)

J: Derived interval between discharge and interview date: J = C - F

Part B. Computer Operations

(a) Select all persons for which X = 1, and base subsequent work on hospital cards for these persons

(b) Translate date of admission A' into coded date A

(c) Translate date of interview C' into coded date C

(d) Test Y: If Y = 1, set F = C - 500, J = 500

Y = 0, skip to (e)

(e) Calculate D = C - 367

(f) Test A:D: If A ≤ D, D + E = F

If A > D, A + E = F,

[But if "still in hospital" is unknown, or if A' or E is unknown, then F = C - 600, and J = 600]

(g) J = C - F

(h) Put J and C into record.

2 years' data. Derived values of D'_J were obtained by fitting the 2nd order curves to observed data by usual least squares methods. In these calculations, the value $q_o = 0.035$ was used, which is consistent with other evidence on proportion of dischargees who are dead on discharge.[2, 6]

Of only incidental relevance to the present account, but of possible interest to readers are the following results. In both years the 2nd order curve gave very good fit. In 1958 variance about the regression curve was but 9 percent as large as variance between months, and in 1959 variance about the regression curve was only 7 percent of between-months variance. Using 13 $\overline{D}'$ as the "adjusted" estimate of total annual discharges, the process implied that the total deficiency in reported discharges—including the omission of discharges of persons not living at time of interview as a "deficiency"—is 15 percent for both 1958 and 1959 data. (A preliminary calculation for 1960 data yields a figure of 17 percent.) Thus these results are not inconsistent with the speculations regarding totals using external data, and may well be more precise.

The value of the computer calculation as contrasted with other possible modes of calculation is emphasized when one looks beyond these first global results. Both a priori judgment, and other exploration[7] suggest that the shape of the curve $D'_J = f(J)$ might vary for differing age-sex classes, or length-of-stay groups, or other categories, The computer programming makes it a simple matter to tabulate D_J values for any age-sex, length-of-stay, or other subclass which is identifiable in the case record. This is not the place for discussion of substantive results, but we do note that values have been computed for a considerable number of such subclasses, and are being studied. They apparently contain a good bit of information; for example, the rate of decline or dropoff in reported discharges with increasing J-interval is sharper for hospital stays of very short duration than for stays of longer duration—a tendency that might have been expected.

Other avenues are immediately suggested by these results, including multivariate analysis of the initial figures, and a considerable variety of schemes for adjustment of reported data in the estimation process. One step into these areas has been taken by the NHS in its official publications. Study of the shape of the D'_J curves has led to the view that data on discharges reported to have occurred within 6 months of the date of interview are considerably superior to those reported for the 6-month period which lies in the range of 6 to 12 months prior to interview—although it does not necessarily follow that a 6-month reference period in the questionnaire would be an improvement over a 12-month reference period. In line with this view, hospital estimates for 1959 and 1960 are being based only on discharges reported for the most recent 6 months of the 12-month reference period in the questionnaire. Computer processing of basic data has made this change a low-cost by-product of the J-analysis operation. The use of the 6-month period produces an estimate of total number of discharges which from 1959 is 8 percent larger than the 1-year reference period would yield, and for 1960 is 10 percent larger. Quantitatively this procedure would appear to have removed something over one half of the total deficiency. In conjunction with data on discharges for decedents, the process might remove quantitatively almost the entire deficiency. Whether the resultant figure is completely unbiased may not be certain.

We shall say no more here about this particular problem. It was introduced only to illustrate the flexibility and adaptability of computer processing to nonroutine types of analysis. We do wish to reemphasize our belief that the long-range benefits of computer applications in social statistics will be found not so much in reduced costs or in increased speed, but rather in the twin characteristics of methodological strength and flexibility which make feasible a more intensive reduction of data and more perceptive analysis.

REFERENCES

1. Alexander Gray: *Introduction to the Development of Economic Doctrine.* Longmans, Green & Co., London, 1934.

2. U.S. National Health Survey. *Hospital Utilization in the Last Year of Life.* Health Statistics, Series D-3. PHS Publication No. 584-D3. Public Health Service. Washington, D. C., January 1961.

3. Percy G. Gray: The Memory Factor in Social Surveys. *Journal of American Statistical Association.* Vol. 50, pp. 344-363, Washington, D. C., 1955.

4. Robert Ferber: On the Reliability of Responses Secured in Sample Surveys. *Journal of American Statistical Association.* Vol. 50, pp. 788-810, Washington, D. C., 1955.

5. R. K. Som: *On Recall Lapse in Demographic Studies.* International Population Conference (Proceedings, p. 50). Vienna, 1959.

6. Guide Issue. *Journal of the American Hospital Association.* Chicago, 1957.

7. U.S. National Health Survey. *Reporting of Hospitalization in the Health Interview Survey.* Health Statistics, Series D-4. PHS Publication No. 584-D4. Public Health Service. Washington, D. C., May 1961.