SOME STATISTICAL FEATURES OF THE HEALTH EXAMINATION SURVEY

Walt R. Simmons, Arthur J. McDowell, and Tavia Gordon*

1. The U. S. National Health Survey

The Health Examination Survey is one of three major vehicles being utilized by the U. S. National Health Survey in its program to provide comprehensive data on the amount and distribution of illness, injury and disability of the civilian population of the United States, on other health characteristics of that population, and on the use of medical, dental and hospital facilities. The other two vehicles are the Household Interview Survey, and the Health Records Surveys. All three activities are continuing projects. They are intended to be flexible instruments, complementing one another in providing an intelligence system on what is possibly the nation's greatest resource: the health of its population. The Bureau of the Census has participated in many aspects of the planning, sample selection, and data collection for these Surveys.

The interview survey is particularly effective in assembling data that have their intersection in the subject person himself, and concern medical and health matters of which the person has knowledge; for example, his instances of physician contact, his days of disability, his medical costs. The interview survey is not a suitable mechanism for estimating volume of undetected or undiagnosed diseases. Operating now for five years, it has been described elsewhere in some detail.[1,2,3]

The Health Records Surveys are themselves a family of undertakings characterized by two attributes: (1) The point of contact and the initial sampling unit is the facility or source which provides health care--the hospital, nursing home, personal care place, or the physician's or dentist's office; (2) the data themselves come in large part from records in these places. The records surveys of the U. S. National Health Surveys are in their early stages, and will not be further described here, except to say that they are expected to provide a wide range of information on use of health facilities and on diagnostic detail not readily available through other techniques.

*National Center for Health Statistics, Washington, D. C.

2. The Health Examination Survey

The Health Examination Survey (HES) is a process which collects health data in two essential steps: (1) a probability sample is drawn for some major sector of the national population--e.g., all noninstitutional civilian adults in the age range 18-79 years; (2) persons in this sample are given a limited physical examination, and other relevant measurements are taken, using standardized procedures. The primary objectives also are two in number: (1) to provide basic distributions of the population by a variety of physical and physiological characteristics such as height, weight, blood-pressure, body-build, and visual acuity; and (2) to estimate prevalence of specified chronic conditions. The HES is a continuing activity, which comes in parts that have been termed cycles. A given cycle concerns a particular segment of the national population and a particular set of measurements and conditions. The remainder of the present discussion will be focused on the first cycle of the HES, for which collection of data will be completed in December of this year.

3. Statistical Problems

The statistical problems of the Health Examination Survey are numerous and complicated. We can touch on only a few of them. Beyond the critical matters of general objectives, authorization, financing, capital resources, and administrative affairs, the HES exhibits substantial problems in formal survey design, in determination of unit costs, in determining initially speculated population parameters, in choice of estimating equations, in handling the nonresponse issue, in calculating precision of estimates, in training and supervision of the examining staff, in the logistics of field operation, and perhaps most notably in evaluation and control of the many measurement procedures.

The scope and content of the first cycle survey, and a number of these statistical problems have been treated in a publication of the National Health Survey.[4] Discussion in this paper is restricted to necessarily brief accounts of three problem areas in the first cycle: impact of nonresponse on the probability design, measurement problems, and basic estimating equations. In all three instances, the final story cannot yet be told since study continues on each of the topics.

4. The Impact of Nonresponse

The Health Examination Survey, like other parts of the National Health Survey, is based on legislation which specifies that the required information will be secured "on a noncompulsory basis."[4] Thus, from the outset, much attention was given to the problem of nonresponse, since a poor response rate could prevent any valid generalizations of survey findings to the population sampled. We attempted, first, to study the nature and the dimensions of the nonresponse problem, next, to take all practicable steps in the design and conduct of the survey to minimize the extent of nonresponse, and, finally, to obtain auxillary data on both respondents and nonrespondents in order to facilitate the residual imputational process.

Several surveys involving health examinations had been made in local areas in the early 1950's.[5,6,7] That experience indicated that perhaps something like one-third of the people asked to participate in a survey involving examination might fail to cooperate despite intensive persuasion efforts. In preparation for the Health Examination Survey, the National Health Survey undertook methodological studies into the motivations and attitudes involving willingness to participate in a health examination survey.[8,9]

We will make no attempt to list all of the ways in which the survey design incorporated lessons learned from the methodological studies, from earlier related surveys, or from the pilot testing of HES plans, in order to maximize response. It would not be accurate to suggest that the operating program adopted and continued every one of the characteristics which these studies implied might help the response rate. Frequently operational considerations overrode theoretical indications. Frequently, too, the practice of the art of obtaining cooperation in a health examination survey produced techniques which seemed effective. These were used without controlled experiments to determine their contribution to the desired result. Nevertheless, it is possible to identify some of the measures adopted because the methodological studies, the pilot tests, or the earlier surveys suggested they might minimize the amount of nonresponse.

Among a number of factors which seemed to be related to willingness to participate was the potential examinee's knowledge about the nature of the examination and the purpose of the survey. In order to increase confidence in the survey and to allay the fears and doubts that might stem from lack of knowledge about just what was involved, several steps were taken. These included distributing in advance in the sample neighborhoods leaflets containing information about the Health Examination Survey. The sample person who made an appointment was given another leaflet which described in some detail the specific steps that would be carried out in examining him. Also, support of local groups and officials was sought and this and the medical research goals were stressed in stories in local news media.

Another factor thought to be relevant to response was the potential personal benefit to the examinee from his examination. To take advantage of this it was decided that the findings of the examination would be made available to the examinee's personal physician (or, in the case of dental findings, to his dentist) if the examinee would so instruct us.

It had been established that a possible negative motivational factor might be the inconvenience, in travel and loss of time and other ways, which the examination entailed. We attempted to minimize that by restricting the length of the examination, eliminating some possibly embarrassing procedures, careful scheduling to avoid waiting, selecting convenient locations, providing transportation, and other means.

Aside from the various motivational findings, there were other results of the presurvey studies and the pretests which influenced our survey design in the effort to maximize response. Thus, the studies had indicated that persons were less likely to agree to an examination on behalf of another member of the household than they were to agree to come themselves; consequently, we made it a practice to ask about consent to a health examination only directly of the sample person.

For another thing, it appeared important to identify the probable noncooperator early, and to handle his case in an individualized way. Motivations differ widely. Thus, to one person, the fact that our examination is free is a point in its favor; to another, perhaps concerned about government budgets, this may be a negative factor. The possibility of early detection of a disease, if it is present, is welcomed by some and dreaded by others. The plan of the survey called for the initial interviewer to make no further effort to obtain cooperation from a person once he was identified as an apparent noncooperator. Another representative of the survey, armed with all the knowledge collected by the first and with experience in handling problems of this sort, would make a later call and attempt to explain the survey more fully, correct misapprehensions and obtain cooperation.

How much the success of the Health Examination Survey owes to any or all of the many specific actions that have been taken to maximize response is not a question we can answer now, nor perhaps even later. But we have been encouraged by the survey results so far. As of the time this is written (early July 1962) the Health Examination Survey has completed operations at more than three-fourths of the separate locations throughout the United States that constitute the first cycle sample (33 of 42). The total number of sample persons identified in those areas was 6,105 and we have succeeded in examining 5,235-- or 86 percent -- of them. Thus, the rate of nonresponse to date (14 percent) is only about one-half as big as we feared it might be.

The rate referred to combines all reasons for nonresponse--refusal to cooperate, unavailability during the period of the survey for reasons other than health (e.g., away on vacation), or unavailability for reasons related to health (e.g., in short-term hospital). The rate given above relates the total number of sample persons identified in the sample households interviewed and the number of such persons examined. Because the sample is based on households, including a small number for which we are unable to obtain interviews, a further correction for presumed sample persons in noninterviewed sample households is appropriate. This correction would lower the response rate by about two percentage points, from 86 percent to 84 percent. One other measure which needs mention is the average of the percentages of response at the 33 locations. Computed this way, the uncorrected response rate is 86.6 percent.

Tables 1 and 2 show how the response rates have varied with population density group and with geographic region. As expected from the preliminary studies, the response rate varies inversely with population concentration. In rural and other urban (less than 50,000 inhabitants) areas, more than 90 percent of the sample population was examined. This decreases steadily as the concentration of population increases, until in the giant metropolitan areas only 78 percent of the sample was examined. The geographic grouping shows a fairly clear picture of lower response rates in the northeastern part of the United States as compared with the south or the west. For the total region this effect is, of course, enhanced by the relatively great number of giant metropolitan areas in the northeast but the relationship is observable in each group.

So far in this paper we have dealt with our concern about and study of the problem of nonresponse, the operating measures taken to minimize nonresponse and the general results to date. In addition, it is, of course, necessary to make some evaluation of the extent to which the residual nonresponse group differs qualitatively from those who were examined. Considerable information bearing on this results from the fact that the design of the survey incorporates a household interview for each sample household. This interview gives information similar to that collected in the Health Interview Survey. We have this for 98 percent of the HES sample households. In addition to such demographic variables

Table 1. Average Response Rate by Population Concentration Group and by Geographic Region: First 33 Stands, Health Examination Survey

Population Concentration Group	All Regions	Northeast	West	South
All Groups----------------------------	86.6%	81.3%	87.9%	89.9%
Giant Metropolitan Areas---------------------	77.9	77.2	79.0	-
Other Very Large Metropolitan Areas----------	85.3	-	86.5	84.0
Other SMSA's---------------------------------	88.3	85.0	91.0	87.7
Other Urban----------------------------------	90.8	89.0	94	91.0
Rural--	91.7	79	92.7	95.0

NOTE: The figure shown for each cell represents the unweighted arithmetic mean of the response percentages for each of the stands included in that category: thus, the marginal total figures cannot be derived directly from the values in the particular row or column. The two cross-classifications designated by a dash (–) are ones for which no stands have been included: the two for which whole-number percentages are shown are each based on only one stand.

Table 2. Frequency Distribution of First 33 Stands by Population Concentration Group and Percentage of Sample Persons Examined (Response Rate), Health Examination Survey

Population Concentration Groups	All Stands	Percentages of Sample Persons Examined				
		94-100	87-93	80-86	73-79	66-72
All groups	33	6	12	7	6	2
Giant Metropolitan Areas	8			2	4	2
Other Very Large Metropolitan Areas	4		2	1	1	
Other SMSA's	8		6	2		
Other Urban	6	3	1	2		
Rural	7	3	3		1	

as age, sex, race, education, and income group, this provides a health history of the household member, including hospitalization, chronic disease, days absent from work due to illness, etc. Table 3 compares the response and nonresponse groups on the basis of age and sex. While the older age groups are slightly underrepresented in the first Round, there is marked agreement between the distributions for examined persons and for all sample persons.

The Health Examination Survey includes one other important means of obtaining relevant data on not-examined persons. At the time the household interview is completed, an attempt is made to obtain a signature on the medical authorization form, giving us permission to request information from the family physician's medical records. Thus, for most not-examined persons some information directly relevant to the health characteristics which the survey is attempting to measure can be obtained. This is done for all not-examined persons for whom a medical authorization was obtained and in the remaining instances the not-examined person is asked to forward to his physician a similar request. In addition, for comparison, similar inquiries are sent to the physicians of a matched sample of examined persons.

5. Measurement in the HES

Every examination has something unique about it which could affect its comparability with other examinations. For example, the Health Examination Survey uses a drink of 50 grams of glucose in its glucose tolerance test, and this could yield results incomparable with results from a challenge of 100 grams. The examination includes a venipuncture and an electrocardiogram and these procedures could affect the blood pressure of the examinee. At every point the question can be raised of the comparability of the Health Examination measurements with those obtained by other examinations using different instruments or measuring the characteristics in different contexts. Needless to say, the immense number of such points of possible incomparability requires that a choice be made of those factors which seem, on *a priori* grounds, to be most important to evaluate and most accessible. It is easy, of course, to make the wrong choices: the history of science is littered with such errors.

Such issues of comparability are critical in any study. They determine what we might call the exterior significance of the data. Internally, the problems of standardization are similar, but more easily dealt with. Essentially, they require a protocol, actions to assure conformity, and procedures which permit evaluation of residual variability. The Health Examination Survey has expended considerable efforts in this direction.

An important means of standardizing a medical examination--obvious enough but not always thought of in this connection--is the choice of examining physician. By choosing physicians of similar background and experience (almost all of the examining physicians for the Health Examination Survey are third or fourth year residents in internal medicine) it is reasonable to expect a uniformity of result that would otherwise be obtained only by prohibitively extensive training.

Having devoted reasonable efforts to developing a sound protocol, to choosing, training, retraining and supervising the examining staff, and attempting to control the quality of data collection by a variety of means, we are left with a

Table 3. Unweighted Total Sample Persons and Examined Persons by Age and Sex: First-Round of First Cycle, Health Examination Survey

Age Group	Total Sample			Examined Persons		
	Total	Male	Female	Total	Male	Female
	Percentage distribution by age groups					
Total	100.0	100.0	100.0	100.0	100.0	100.0
18-24	13.1	12.9	13.3	13.2	13.3	13.1
25-34	21.5	21.3	21.6	22.1	21.9	22.3
35-44	22.8	23.0	22.7	23.7	23.6	23.8
45-54	18.2	17.9	18.6	18.4	17.7	19.0
55-64	14.0	15.1	13.1	13.2	14.1	12.5
65-74	8.2	8.1	8.3	7.8	8.1	7.5
75-79	2.1	1.7	2.5	1.6	1.4	1.8
	Percentage distribution by sex					
Total	100.0	45.6	54.4	100.0	46.7	53.3
	Number of persons					
Total	2,614	1,191	1,423	2,205	1,030	1.175

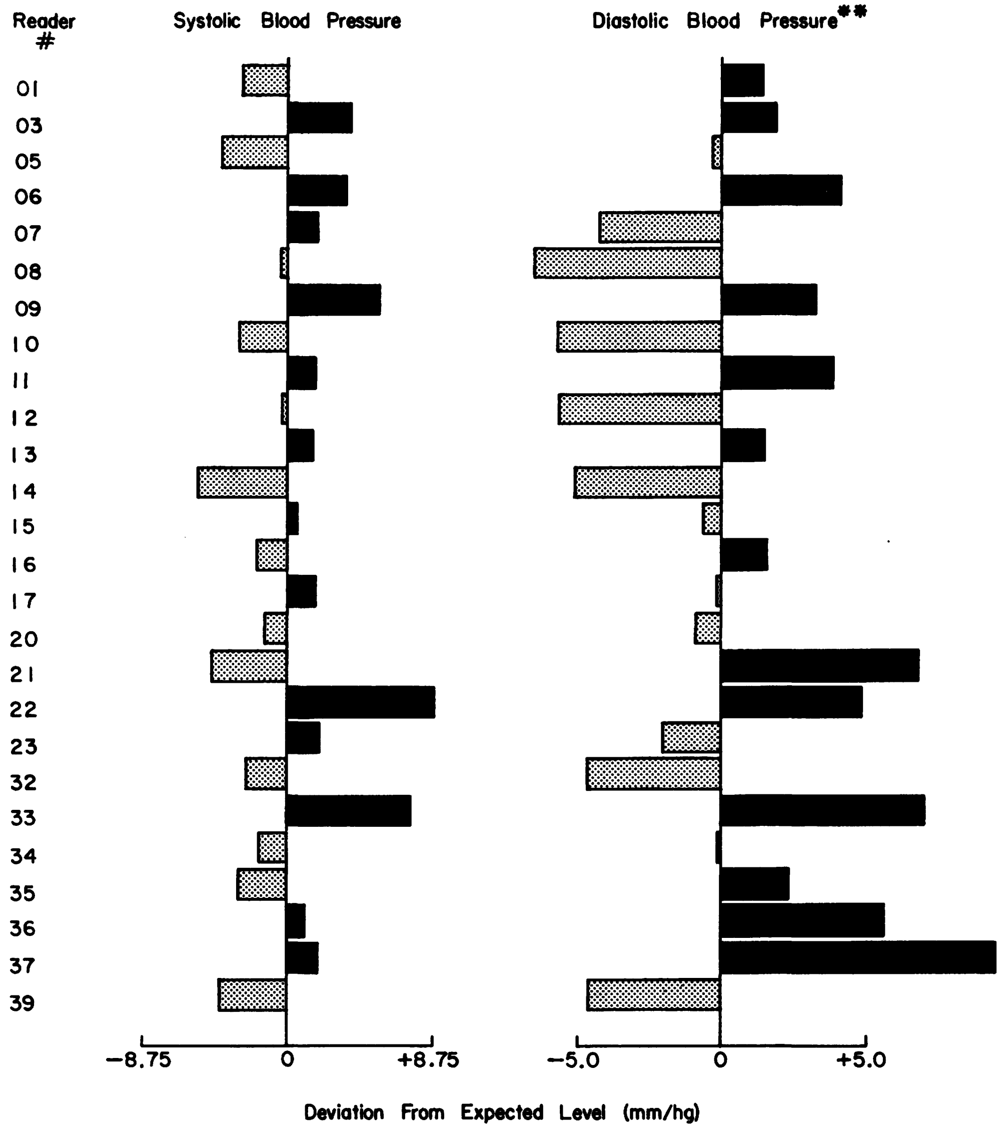
final requirement: determining the variability of measurement. This is something especially attractive to statisticians; indeed, they sometimes seem more interested in quantifying variability than in limiting it. But even limited, some variability will remain. It may be more or it may be less. It may be measurable in the context of the procedures used, or it may not. But unattended it can devour all significance.

The problems of identifying and measuring nonsampling variability arising from sources such as the examining physicians, X-ray readers, and the like, are complex. Even given that reasonable definitions and a good statistical model can be constructed to reflect the effect of nonsampling sources of error, the Health Examination Survey presents formidable administrative problems in attempting to carry out the measurement processes in such a way as to provide measures of nonsampling variability.

Ideally, what would be desired would be replicate measurements of the same characteristics on randomized samples of persons. For characteristics delineated entirely in the course of the clinic examination, this has thus far proved too difficult to arrange. In order to try to approximate this type of evaluation it was arranged, however, to have two examining physicians at most locations, and examinees at a location were assigned alternatively to each physician. Analyses of differences in physical examination findings between physicians indicate, as might be expected, that there is often a greater variability between physicians than chance assignment of examinees would be expected to yield. A good example is blood pressure (Figure A). Data for each physician, on which the chart is based, have been adjusted or standardized for age-sex composition. The exhibited deviations contain components of chance, geography, and perhaps other factors as well as inter-physician differentials, but observed differences between physicians are significant for both systolic and diastolic data. As would be anticipated on an a priori basis, between-physician relative variability is markedly greater for diastolic than for systolic pressures.

Another method of gauging the variability of measurement is by using data collected by replicate measurements of nonsample persons, either by our observers or by other observers. To a limited extent we have done this as part of our regular training. This was also done in a special methodological study which involved (among other things) complete cardiovascular examinations to be replicated on a series of persons.[10] Since

Figure A. Blood Pressures Obtained by Round I Physicians: Deviations From Expected Levels*



*The expected level for a physician is obtained by using the age-sex distribution of his examinees and the average age-sex specific blood pressure levels. Comparisons based on first blood pressures taken.

**Fourth phase.

NOTE: Scales for systolic and diastolic deviations are proportional to their respective standard deviations. The number of persons examined by a physician ranged from 56 to 248, averaging about 120.

neither the Health Examination physicians nor examinees were involved in the methodological study, the results provide only general guidance. They did, however, lead us to discount palpation of the peripheral arteries as a diagnostic finding.

For characteristics which are measured outside the clinic, replicate measurements are feasible and in some cases have been undertaken. Blood glucose and serum cholesterol concentrations are determined at a central laboratory, and for a sample of cases, aliquots are shipped to another laboratory for determination. Some blind replicates have also been introduced into the regular series of laboratory determinations. And, of course, the laboratories also have their own quality controls.

Also evaluated outside the clinic are the electrocardiograms, the chest X-rays and X-rays of the hands and feet. Each of these are evaluated in replicate determinations. The X-rays of the hands and feet may be used to illustrate the problem.

As part of the examination for arthritis and rheumatism, X-rays are taken of the hands and feet of each examinee. These are sent to three physicians especially qualified to evaluate such films. The pair of films for each person is examined independently by each of these three physicians, who note on a standard form any abnormalities observed. In particular, evidence of osteoarthritis is graded from 0 (absent) to 4 (present and severe). Should any two observers assign grades to a film which differ by more than one step, the films for that person are reread in a review session, first independently and then jointly, and a final grade assignment is agreed to. A grade of 2 or more is considered definite osteoarthritis. This is the only evidence used in the Health Examination Survey for a diagnosis of osteoarthritis.

Bypassing problems of etiology, associated symptoms, and medical sequelae (as we do), there remains the grading of the X-ray findings. This is accomplished by first providing a "normal standard" which is X-rayed along with the hands and feet. The standard is a metacarpal bone from an apparently normal person in his 30's. It is encased in lucite and X-rayed along with the hands and feet of each examinee.

Abnormalities once recognized must be graded by comparing the extent of abnormality with a series of graded pictures. For the hands, these pictures are actual X-ray photographs. For the feet, the standard pictures are in the mind of the beholder.

We show data from some of the early readings of Health Examination Survey films (table 4). It will be noted that, despite selection of readers of recognized competency, similar background, and having identical instructions, there is still a distinct difference between the three observers. Not only is the distribution of findings different among readers, but there is a distinct variation between first and second readings by the same

Table 4. Grade Assignments by Different Readers in Replicate Reading of 154 X-rays of the Hands and Feet (Stand 04), Health Examination Survey

Osteoarthritis Grade	First Reading			Second Reading		
	Reader A	Reader B	Reader C	Reader A	Reader B	Reader C
	Number of films graded for hands					
0----------------------	13	28	34	16	33	42
1----------------------	58	78	67	63	81	48
2----------------------	76	41	46	67	33	52
3----------------------	5	5	5	6	5	10
4----------------------	2	2	2	2	2	2
Average Grade-----------	1.51	1.19	1.18	1.45	1.10	1.23
	Number of films graded for feet					
0----------------------	45	55	71	40	48	82
1----------------------	50	67	52	49	69	41
2----------------------	57	29	29	61	35	29
3----------------------	2	3	2	4	2	2
4----------------------	-	-	-	-	-	-
Average Grade-----------	1.10	0.87	0.75	1.19	0.94	0.68

NOTE: Readers differ from each other (1% level), but agree with themselves (5% level) in reading pattern.

reader. The within-reader variation is not statistically significant, however, under a hypothesis that distributions of first and second readings are samples from the same universe.

Indeed it may be inquired whether any two interpreters of any medical document ever, in the strictest sense, have the same standards. Thus, if one is to construct a statistical model designed to estimate the prevalence of osteoarthritis, he must first ask what possible meaning he can attribute to this parameter. Surely he can never purport to mimic reality unless he assumes that the parameter itself constitutes a variable. But it is conceivable that if our ambitions are somewhat less we can, in some fashion, define the parameter as an intersection of the idealizations of various observers. We could, for instance, accept as prevalence the expected value of a specified measurement procedure.

Another thing to note about the replicate determinations is that the probability of "correctly" identifying a film varies from film to film. The same series of films was re-evaluated by accident. On the average, the larger the number of observers calling a film abnormal the first time, the larger the number calling it abnormal the second time. (Table 5) If a statistical model is used which presumes a uniform probability of correctly identifying a film, the facts will introduce a correlation between the determinations of the independent observers and spite the model.

Without attempting to suggest at this time a solution to these specific problems, a more general point may be worth making. Statisticians are carefully trained to determine if the data actually satisfy the assumptions of any model they apply to it. The fact is that in a large variety of cases the assumptions are not met, despite the fact that within the specific data this failure may not be demonstrable. If we attempt to mimic the actual complexity of the facts we generally arrive at a point where our model breaks down. A good model should allow us to arrive at two estimates--the first, an estimate (preferably unbiased) of the number of "abnormal" persons; the second, an estimate of the probability that a given person is "truly" abnormal. Such a model should work in a world where different observers

Table 5. Distribution of 154 X-rays of Hands and Feet (Stand 04), by Number of Readers Declaring Film Positive, for Initial Readings of Three Readers, and for Second Readings of Same Readers

Number of Readers Declaring Film Abnormal* in First Reading	Number of X-rays				
	Total	Number of Readers Declaring Film Abnormal* in Second Reading			
		0	1	2	3
<u>Hands</u>					
All cases---------	154	66	29	27	32
0----------------------	64	57	5	2	-
1----------------------	34	8	16	9	1
2----------------------	18	1	6	7	4
3----------------------	38	-	2	9	27
<u>Feet</u>					
All cases---------	154	84	30	17	23
0----------------------	90	72	16	2	-
1----------------------	30	12	14	4	-
2----------------------	10	-	-	5	5
3----------------------	24	-	-	6	18

*Evidence of Osteoarthritis of grade 2 or more.

have different definitions of "abnormal" and where the probability of "correctly" characterizing a person varies from person to person. We are still looking for such a model.

6. Estimation

We turn now to the conversion of measurements into estimates of population parameters. The paper does not present a final estimation technique for the HES. It describes a problem, identifies what we consider leading features of the estimation process, and outlines a pattern of thinking and a method of approach to solution of the problem.

The sample design is described in considerable detail in reference 4. It is a multistage stratified probability sample of loose clusters of persons in rather small land segments. Overall it will include about 6,600 persons in approximately the same number of households in some 2,100 land segments in 42 primary sampling units (PSU's) in continental United States. A PSU is a Standard Metropolitan Statistical Area, a county, or a group of 2 or 3 contiguous counties. The master design is self-weighting with respect to persons, but for a number of reasons appropriate inflation factors are not precisely constant for all examinees.

Estimation and calculation of variances must, of course, take into consideration the complex sample design. But for most of the present discussion it will be useful to treat that basic design as though it were a simple random sample.

Consider a simple random sample of n from N individuals in a population for which u_i is the unknown true conceptual value of a characteristic of the ith person (e.g., $u_i = 1$ if the person has hypertension and = 0 otherwise). The population mean value is the parameter $\bar{u}$. An operational approximation to this concept is defined so that if it could be attained it would reflect a value v_i for the ith individual (say, $v_i = 1$ if the individual has prevailing systolic blood pressure ≥ 140 and = 0 otherwise), and $\bar{v}$ for the mean of the population. A survey conducted under a specified set of procedures yields a value x_i ($x_i = 1$ if the measurement obtained for systolic pressure ≥ 140; = 0 otherwise) for the ith unit, and thus is a second order approximation to the conceptual u_i.

The statistician undertakes to process the x-data in such a fashion that they yield estimates of $\bar{v}$ and consequently shed light on judgments concerning the parameter $\bar{u}$. The usefulness of survey data will depend heavily upon the degree of relevance of the v-values to the u-values. In the HES, attention has been given this matter in the selection of u-values which are measurable and of v-values which medical experts agree are indicators of the corresponding u-values, and which can in practice be standardized. Preparation for decisions in this area included contract pilot studies of methods and processes, extensive medical consultations, and two full-dress field trial rehearsals of tentatively chosen content of examination.

The relevance of v-values to u-values is necessary, but only pays off when the x-values from field measurement are reasonably faithful representations of the defined v-values. We have just referred to some of our efforts to deal with this measurement problem.

Sampling and Stratification. In what may be termed the classical approach to estimation, it is assumed that membership of the universe is completely known; sampling is without flaw; response is perfect; the x-measurements are unique, and are taken, recorded and processed without error. The population mean $\bar{X}$ is estimated by $\bar{x}'$, a linear combination of x-observations.

Algebraically this process is summarized with the equation

$$\bar{x}' = \frac{x'}{N} = \frac{1}{N}\left[\frac{N}{n}\sum_{i=1}^{n} x_i\right] \tag{1}$$

in which the sample observations have been weighted or adjusted by the reciprocal of the sampling fraction. Its sampling variance is well known and, under a broad range of conditions, $\bar{x}'$ is normally distributed about $\bar{X}$.

Many alternatives to this "classical" estimator are known. No attempt is made here to encompass the total field of reasonable possibilities. Note is taken that a worthy objective is to seek an estimator which has not more than a modest bias and which has a relatively low mean square error. Toward this objective, the effect is being explored of application to HES sample observations of other adjustment factors in addition to the sampling weight. Three "types" of adjustment are considered, although all three have much in common.

Ratio Estimation. First, for an item such as hypertension, there clearly are differences in prevalence among different age-sex groups. Since there are available from the Census good independent estimates of population by age and sex, there is a gain in applying a second adjustment for age and sex control through ratio estimation. This might yield the estimator $\bar{x}''$, where

$$\bar{x}'' = \frac{x''}{N} = \frac{\sum_a \sum_i \frac{N_a}{\sum_i \frac{N}{n} y_{ai}} \frac{N}{n} x_{ai}}{N}, \qquad (2)$$

in which: x_{ai} is observation for ith person in ath age-sex class.
y_{ai} is unity for all persons in ath age-sex class, and equal to zero otherwise, and
N_a is control number of persons in ath age-sex class.

More generally, if $x'_{a\lambda}$ is the basic inflation estimate for any $\lambda^{\underline{th}}$ subclass of the $a^{\underline{th}}$ group, and y'_a the sample inflation estimate of population in the ath class, then the ratio estimate of rate for the total $\lambda^{\underline{th}}$ subclass becomes:

$$\bar{x}''_{.\lambda} = \frac{x''_{.\lambda}}{y''_{.\lambda}} = \frac{\sum_a \frac{N_a}{y'_a} x'_{a\lambda}}{\sum_a \frac{N_a}{y'_a} y'_{a\lambda}}. \qquad (3)$$

The estimate of the aggregate $x''_{.\lambda}$ has lower variance than would the corresponding inflation estimate given sufficient correlation between $x'_{a\lambda}$ and y'_a. Note that this ratio estimation has a type of effect similar to that which would arise from stratification by age and sex, with proportional allocation. For substantially sized sample groups, the ratio estimate is effectively unbiased.

The form in equation (3) has the operational advantage that consistent estimates of $x''_{.\lambda}$ can be secured for any subclass λ simply by adding sample data $x_{a\lambda i}$ which have been weighted by the fixed multiplier $\left(\frac{N_a}{y'_a} \; \frac{N}{n}\right)$.

Other Auxiliary Data. Another type of possible adjustment of sample observations is suggested by the fact that several additional items of demographic and health data collected for each of the 6,600 examinees through household interview, are also collected in another NHS survey for a much larger sample of 390,000 persons. This circumstance makes it possible to utilize the larger survey for weighting the smaller in essentially the same way that double-sampling or poststratification is sometimes employed, but at a nearly zero additional cost. This process is called another type of adjustment, although in one frame of reference it is not different from the age-sex ratio adjustment just considered. The order of relative reduction in variance through introduction of such a process is given by the statistic:

$$R = \frac{(H-1)B}{nW + (H-1)B} \qquad (4)$$

where

H is the number of pseudostrata,

W is average within-stratum variance, and

$B = \frac{1}{H-1} \Sigma n_h (\bar{x}_h - \bar{x})^2 - W$ is between-stratum variance on a unit basis, in which $\bar{x}_h$ is the mean of the $h^{\underline{th}}$ of H pseudostrata, n_h is number of persons in the $h^{\underline{th}}$ stratum, and all calculations are from sample data.

Thus the procedure will be helpful to the degree that $(H-1)B$ is large as compared with nW.

The use of the 390,000 sample to weight the 6,600 sample could introduce a bias into the process if the two surveys used different procedures. Since the two surveys were carried out under practically identical instructions, with the same auspices, and in large part with the same interviewers, we believe the risk of bias from this source is trivial.

Nonresponse. Earlier in this paper we have discussed the importance of nonresponse in the HES survey. At this point, we stress a hard fact: there is no way in which danger from nonresponse can be entirely eliminated. What one does is to adopt a course which he judges is a reasonable compromise among bias, variance, cost, and operating feasibility. In the HES, extensive steps were taken to minimize nonresponse, and substitute measurements for nonrespondents have been explored as possible techniques. But in essence, the estimation procedure will impute to nonrespondents the data for respondents. The question then resolves into one of what subclasses shall be recognized in the imputation procedure. We suggest that three guidelines are useful in resolving this question: (1) The classes should be ones in which variation in key statistics between classes is large compared with variation within classes; (2) the response rates between classes should be different; (3) the number of respondents in each class should be large enough to avoid letting any respondent represent too

Table 6. Listing of Headings and Stubs for Exploratory Tabulations From Round I Data of the Health Examination Survey[a]

Tabulation Number	r - class	c - class	Number of rc - cells	Avg. value of n_{rc}
1---------------	Age-sex (14)	Health Class[b] (4)	56	40
2---------------	Age-sex (8)	PSU (14)	112	20
3---------------	Age-sex (14)	Income Class (3)	42	50
4---------------	Age-sex (14)	Population Density (5)	70	30
5---------------	PSU (14)	Income Class (3)	42	50
6---------------	Population Density (5)	Geographic Region (3)	11[c]	200
7---------------	Super-stratum (11)[c d]	Race (2)	22	100
8---------------	PSU (14)	Persuasion Utilized[e] (2)	28	80

[a]Numbers in parentheses indicate number of categories in the classification.

[b]Presence of cardiovascular and arthritic conditions according to interview.

[c]Some cells are vacant.

[d]Super-stratum is a combination of population density and region.

[e]Degree of effort invoked to induce response.

many nonrespondents. Fortunately these considerations are similar to those governing the choice of pseudostratification, and the nature of the appropriate adjustment is similar.

Round I - Experimentation. The 42 HES stands consist of 3 Rounds of 14 stands. Each round is a probability sample of the U. S. Data from Round I are being tabulated separately, partly to produce a few preliminary survey results, but especially to study variations in data for each of eight key statistics by a variety of cross classifications. The cross-tabulations are those indicated in Table 6. For each row, column, and cell, response rates and prevalence rates of each of the eight statistics are being calculated and will be analyzed for relevance to the problems which we have just discussed.

7. Closing Comment.

Our remarks necessarily have concentrated on but a few of the statistical features of the HES. The account, however, has identified a considerable range of problems. We hope to have conveyed the impression that in the National Health Survey we are giving real attention to some of them. We hope also to stimulate others to seek solutions.

REFERENCES

1. U. S. National Health Survey. *Origin and Program of the U. S. National Health Survey.* Health Statistics, Series A-1. PHS Publication No. 584-A1. Public Health Service. Washington, D. C., May 1958.

2. U. S. National Health Survey. *The Statistical Design of the Health Household-Interview Survey.* Health Statistics, Series A-2. PHS Publication No. 584-A2. Public Health Service. Washington, D. C., July 1958.

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

4. U. S. National Health Survey. *Plan and Initial Program of the Health Examination Survey.* Health Statistics, Series A-4. PHS Publication No. 584-A4. Public Health Service. Washington, D. C., May 1962.

5. Commission on Chronic Illness: *Chronic Illness in a Large City: The Baltimore Study.* (Chronic Illness in the United States Vol. IV) Harvard University Press 1957.

6. Commission on Chronic Illness: *Chronic Illness in a Rural Area: The Hunterdon Study.* (Chronic Illness in the United States Vol. III) Harvard University Press 1959.

7. Chen, E. and Cobb, S.: *Further Study of the Non-Participation Problems in a Morbidity Survey Involving Clinical Examination.* Journal Chronic Diseases 7:321-331. April 1958.

8. U. S. National Health Survey. *Attitudes Toward Cooperation in a Health Examination Survey.* Health Statistics, Series D-6. PHS Publication No. 584-D6. Public Health Service, Washington, D. C., July 1961.

9. U. S. National Health Survey. *Cooperation in Health Examination Surveys.* Health Statistics, Series D-2. PHS Publication No. 584-D2. Public Health Service, Washington, D. C., June 1960.

10. U. S. National Health Survey. *Evaluation of a Single-visit Cardiovascular Examination.* Health Statistics, Series D-7. PHS Publication No. 584-D7. Public Health Service, Washington, D.C., December 1961.