Quality Control in a National Health Examination Survey

DHEW Publication No. (HRA) 74-1023

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service
Health Resources Administration
National Center for Health Statistics
Rockville, Maryland

Vital and Health Statistics—Series 2, No. 44
Reprinted as DHEW Publication No. (HRA) 74-1023
August 1973

Library of Congress Catolog Card Number 70-608550

CONTENTS

SYMBOLS

Data not available	---
Category not applicable	...
Quantity zero	-
Quantity more than 0 but less than 0.05	0.0
Figure does not meet standards of reliability or precision	*

QUALITY CONTROL IN A NATIONAL HEALTH EXAMINATION SURVEY

Wesley L. Schaible, *Division of Health Examination Statistics*

INTRODUCTION

The Health Examination Survey

The Health Examination Survey is one of the major survey programs conducted by the National Center for Health Statistics under the authority of the National Health Survey Act of 1956. This act provided "(1) for a continuing survey and special studies to secure on a non-compulsory basis accurate and current statistical information on the amount, distribution, and effects of illness and disability in the United States and the services received for or because of such conditions; and (2) for studying methods and survey techniques for securing such statistical information, with a view toward their continuing improvement."

To obtain information about the health of the U.S. population the National Health Survey is divided into three survey programs. [1] One of these, the Health Interview Survey, is primarily concerned with the impact and social dimensions of morbidity. It collects data by continuously sampling and interviewing the civilian, noninstitutional population of the United States. A second, the Health Resources Program, provides statistics on the health of the institutional population, the utilization of medical facilities, and health manpower. The third program of the National Health Survey is the Health Examination Survey (HES). [2] In the Health Examination Survey primary emphasis is on the collection of data by direct examinations and tests on a probability sample from the civilian, noninstitutional population of the United States. Such examinations and tests can yield standardized information about diagnosed conditions, including those which persons may fail to report or may be incapable of reporting in a survey based upon individual interviews. They can also reveal previously undiagnosed, unattended, and nonmanifested conditions.

The overall plan of the Health Examination Survey is to conduct successive, separate surveys, each with specific objectives. Each of these surveys has been referred to as a "cycle." Thus far in HES the objective of each cycle has been to obtain data on specific health characteristics of a certain age segment of the U.S. population.

Collection of sample data in Cycle I began in October 1959, and was completed in December 1962.[3] The target population consisted of all civilian, noninstitutionalized adults in the United States aged 18-79 years. The probability sample, which was based on households, identified 7,710 sample persons of whom 6,672 were examined. The 2-hour examination focused particularly on certain cardiovascular diseases, arthritis and rheumatism, and diabetes. Various other examination data were collected, including measurements of visual and auditory acuity and blood pressures; electrocardiograms; findings of medical and dental examinations; blood analyses; hand, foot, and chest X-rays; and numerous body measurements. Additional health history data were gathered by interviewers and self-administered questionnaires.

The data collection phase of Cycle II began in July 1963 and was completed in December 1965.[4] This cycle was concerned with noninstitutionalized children 6-11 years age; and from a total of 7,417 sample children, 7,119 were examined. The 3-hour examination was designed to obtain measures related to healthy growth and development. The types of data collected included visual and auditory acuity, blood pressures, electrocardiograms and spirograms, medical and dental examinations, hand-wrist and chest X-rays, psychological tests, and numerous body measurements. Additional data were obtained by means of several questionnaires including one on household composition, two concerned with health history and habits, and one, completed by school personnel, on the child's academic and social achievement. Also, a copy of the birth certificate of each consenting sample child was requested from the appropriate State registrar to verify date of birth and provide supplemental information.

Data collection in the Cycle III sample began in March 1966 and was completed in March 1970.[5] Of 7,514 sample youths identified 6,768 were examined. The segment of concern in this cycle was the noninstitutionalized youths aged 12 through 17 years. As in Cycle II, a 3-hour examination focused on factors and conditions related to healthy growth and development. A physician performed a medical examination designed with special attention given to items relating to adolescent growth and development. A psychologist recorded the youth's performance on various tests to study the growth of certain aspects of thinking, socialization, and motor coordination. Other parts of the examination included a dental examination; tests of visual and auditory acuity, color vision, respiratory function, and grip strength; exercise tolerance on a treadmill, an electrocardiogram; an X-ray of the chest and one of the hand and wrist; and weight, height, skinfold thickness, and other body measurements. A blood sample was taken for hemoglobin and hematocrit, serological tests, and extensive blood typing. The serum chemistries determined were total cholesterol, uric acid, and protein-bound iodine. In addition to data collected in the various examinations, further information was obtained from seven questionnaires: a household interview questionnaire which provided household composition and demographic information, a marital history of the parents, a medical history of the youth by the parent, a health habits and history of the youth completed in the home by the youth, a health behavior questionnaire completed by the youth at the examination center, a questionnaire administered by the nurse to obtain menstrual information from female examinees, and a questionnaire completed by the school which the examinee attended. Also, a copy of each consenting sample youth's birth certificate was requested from the appropriate State registrar. As in Cycle II the certificate was used to verify date of birth and obtain supplemental information.

Field Operations

The examinations were performed in mobile examination centers which were transported throughout the country to the sample areas. In Cycle I, two such centers were operated simultaneously in a "leap-frog" pattern along the scheduled route of 42 sample locations. Two centers were again employed in Cycle II, visiting 40 sample locations. In Cycle III, one center was used to return to the same 40 locations as in Cycle II. By using the same locations, and thus the same households, in addition to the primary objective of obtaining the desired data for youths aged 12 through 17, longitudinal data were obtained for approximately 35 percent of those youths who were examined in Cycle II.

In Cycle I, examinations were scheduled individually at half-hour intervals with all examinees going through the same sequence of tests. In Cycles II and III it was felt that the youths would be more at ease if a number of them were present in the center at the same time. After considering alternative scheduling arrangements, it was decided to bring six youths to the center at the beginning of each morning and afternoon. The examination was performed according to a schedule designed to eliminate situations in which one test might have an undesirable effect on a subsequent test. Efficient utilization of facilities and staff time, a second consideration in scheduling, led to differential sequencing of examinees.

For discussion purposes the HES field staff may be divided into three interrelated but distinct groups. The first group consists of Bureau of Census interviewers and their supervisors, who are not permanent members of the HES staff but are assigned to HES from the Census District in which the examination center is located. These interviewers make the initial contact with the households and obtain basic information pertinent to HES operations.

The second group is composed of the Field Operations Managers, Administrative Assistants, and Health Examination Representatives, all of whom work in or out of an office located near the site of the examination center. These members of the HES field staff are responsible for additional interviewing in households which have eligible youths, checking for the presence and consistency of questionnaire information, and arranging for transportation and various other services. They are also responsible for the many other administrative functions necessary for an effective operation. The Health Examination Representatives serve a particularly important and delicate role in obtaining the cooperation so necessary in surveys of this type.

The third group is the examination center staff, which thus far has consisted of physicians, dentists, psychologists, nurses, technicians, and clerical assistants.

The sample design, which was similar throughout the three cycles, is that of a multistage, stratified probability sample of loose clusters of persons in land-based segments. The successive elements dealt with in the process of sampling are primary sampling unit (county or group of contiguous counties), census enumeration district, segment (a cluster of households), household, eligible person, and finally, sample person. The basic design is essentially self-weighting, although operational efficiencies require some modification of sampling rates. However, in all situations the probability of inclusion of every sample person is known. For Cycle I, the variables of stratification were geographic location and population density. For Cycles II and III the 1960 Census data were available, and rate of population change between 1950 and 1960 was added as a third variable of stratification. This extremely abbreviated treatment of the subject may be supplemented by the more complete discussions given in publications more concerned with this aspect of the survey.[3-6]

QUALITY CONTROL IN DATA COLLECTION

Two sources of error may enter into data collection activities—— sampling error, that error which occurs because data were gathered from a sample rather than from the entire population of concern; and nonsampling error, a somewhat loosely defined collection of "other sources of error." If a sample is chosen in an appropriate manner, the sampling error can be estimated, a feature which is of concern in the design of all sample surveys.

Nonsampling errors, generally neglected in the statistical literature until recently, are now commonly considered in the planning, conduct, and evaluation of surveys. Increasing the sample size, a method frequently used to reduce sampling errors, is not effective in the reduction of nonsampling errors, which instead of diminishing as the sample size is increased may remain constant or perhaps become larger. Therefore, in large samples such as those used in HES, nonsampling errors are of primary concern and it is on the identification, evaluation, and control of nonsampling error that quality control in data collection is centered. Quality control not only implies a concern with keeping output (usually repetitive) within certain levels of quality, but also the diverse types of activity which in general promote the quality of the product. In HES, the products are national estimates of various U.S. population characteristics such as those briefly described in the introductory section, and the prescribed level of quality is that of the highest attainable accuracy and precision within the usual limitations dictated by acceptable procedures and reasonable costs.

Any attempt at producing a relatively complete list of types or sources of nonsampling errors would be a lengthy task. However, many types of nonsampling errors which arise in data collection are included in the following broad and occasionally overlapping categories:

Conceptual errors—e.g., errors due to survey design, definitions, or specifications.

Nonresponse errors—e.g., errors due to noncoverage, lack of respondent cooperation or recall, collecting agent omissions, illegible entries, lost records.

Measurement process errors—e.g., errors due to lack of environmental control, poorly calibrated measuring instruments, badly worded questionnaires, improper examiner or subject influence in the performance of measurements.

Processing and analysis errors—e.g., errors in coding, punching, computation, presentation, interpretation.

The purpose of this report is to present general procedures which are important in HES efforts to identify, control, and evaluate nonresponse and measurement process errors including related conceptual errors; a brief section on HES processing procedures is also presented.

Reduction of nonresponse, including missing data, is a primary concern of the quality control program. Various publicity practices and interviewing techniques are employed in the effort to reduce nonresponse. A second major concern of the total quality control effort is those errors associated with the measurement procedures and instruments. Detailed procedures relating to the data collection process are followed to help insure that HES data are of high quality. Aside from initial procedures designed to reduce errors, constant surveillance and evaluation is necessary, especially in a lengthy survey such as HES. Closely related to this aspect is the necessity of providing, when possible, data of such a nature that objective and quantitative statements of residual errors and uncertainties associated with reported values can be included in reports of survey findings. The final major responsibility of the quality control program is in the processing of data where various verifications, edits, and consistency checks aid in the discovery and reduction of errors.

Aside from classification of types of nonsampling errors, it is particularly convenient and useful to divide nonsampling errors into variable and systematic errors. Errors of a variable nature are usually due to a combination of known and unknown factors which singly are usually insignificant. Their combined effect on obtained measurements and especially on estimates such as means are lessened by the fact that they are, by their nature, as likely to overstate as to understate the true measurement, and so there is a tendency for errors to "cancel out." Systematic errors or biases, on the other hand, are deviations from the true measurement which are the result of circumstances which, if known and understood, would be expected to produce some predictable direction and magnitude of difference between the true measurement or estimate and its obtained value. For example, if the illumination for a visual acuity test is less than that required for good results, a bias will result in the data, since the examinees' responses will tend to be consistently different (in the direction of poorer visual acuity) from that which would have been obtained with proper illumination.

Investigations of nonsampling error often employ models containing bias and variable error components. More complex models may also consider biases which operate for only a portion of the data collection activities and also bias and variable errors identified by major or similar sources. Although models separating variable and systematic errors are often intuitively appealing as well as theoretically necessary, their actual application is not lacking in problems, both theoretical and practical. For example, there can be many different solutions to the seemingly basic problem of deciding when an error which is basically constant but not strictly so throughout data collection, should be treated as a bias or as a variable error. Further discussion along these lines is presented with the evaluation of residual measurement errors where a simple model is utilized.

Much time and effort in HES is devoted to the reduction of nonsampling error and to the many problems involved in the evaluation of residual errors. All HES personnel involved with the collection of data are also actively involved in quality control efforts, and the initiation and execution of quality control procedures are a result of a concerted effort involving many members of the HES staff as well as numerous other consultants. Many general procedures presented here are common to all cycles but modifications of specific techniques are often necessary due to the population segment being sampled during a particular cycle. More specific procedures relating to particular data areas are included in reports of findings.

MEANS OF REDUCING NONRESPONSE

One type of nonsampling error that occurs in perhaps all surveys, especially where subject participation is voluntary, is that due to nonresponse. The amount of bias created by nonresponse generally, but not necessarily, varies with the amount of nonresponse. The usual course of action to reduce nonresponse bias, therefore, is to obtain as great a response as possible. From the outset HES has recognized the problem of nonresponse and has experienced success in its control and minimization. The procedures presently used are a result of experience and knowledge obtained from pretests, previous cycles,[7] methodological studies,[8,9] and other surveys.

Advance Publicity

If those persons chosen as a part of the sample fully understand the Health Examination Survey and its purpose, they are much more likely to give their complete cooperation. In HES, the informational effort is directed toward persons who might be expected to fall in the sample and toward organizations and agencies whose support for the survey is not only desirable in itself but might also influence the response of sample persons.

In Cycle I, HES medical officers made personal visits to State and local health departments, medical and dental organizations and societies, and in Cycle II, to State education officials as well. During these visits the medical officers explained the purposes of the survey, the specialized nature of the examination, and the confidentiality of the data collected. They also pointed out that no results of the examination would be disclosed to the examine, but that he would be encouraged to sign an authorization permitting a report of the findings of his examination to be sent to his physician or dentist. In Cycle III, which returned to the same sampling locations as Cycle II, an explanatory letter was sent to these officials in lieu of a personal visit.

Approximately 6 weeks before examinations began in each area, a member of the HES field staff made personal contact with health, school, and other concerned officials. In Cycles II and III the cooperation of school officials was particularly valuable, since youths participating often had to miss a half-day of school on the day of examination. Also schools were requested to complete a questionnaire concerning social and academic achievement of the youths.

Two or 3 weeks before examinations began a second letter was sent to these officials, as a reminder and to give the exact dates and location of the operations. At this time the Census Bureau mailed to each household to be contacted a postcard stating that the Census Bureau would be visiting within a few days in connection with a survey being conducted for the U.S. Public Health Service. Shortly before the commencement of interviewing by the Bureau of the Census personnel, the local newspapers were provided with a news release. Although this publicity reached a much larger segment of the population than HES attempted to examine, a local newspaper article was often a means of communicating with the desired segment of the population. Newspaper articles were valuable also to the Health Examination Representatives during the second household interview, providing them with a printed document from a local source which could be used to impart information and establish the authenticity of the HES program during a rather extensive explanation of the program.

A previous publication[7] evaluating the effect of the publicity on sample persons in Cycle I gives examination rates of those reached through various types of publicity. These rates indicate that in general publicity had a positive effect on response. Publicity was also associated with success in obtaining and scheduling an appointment at the time of initial interview with the sample person.

Interviewing Techniques

The personal interview, a technique used to collect data, was also an effective means of disseminating information to sample persons about the Health Examination Survey. The first interview conducted by Bureau of the Census interviewers, had as the primary purpose the obtaining of household composition and demographic data. During this preliminary interview, the Census interviewers also recorded any pertinent information they felt would be useful to the Health Examination Representatives (HER's) who make followup interviews. In Cycle I, the Census interviewer also told the sample person about the HES program and concluded the interview by offering the opportunity to have an examination. This offer was made only to the sample persons and not to a proxy respondent since a methodological study[9] had shown that proxy respondents were less willing to commit others to an examination than themselves. If the offer was not accepted, all attempts to persuade the respondent were delayed for the HER's interview.

In Cycles II and III, the Census interviewer was not a real source of information concerning the survey, deferring most respondent questions for the HER's followup. During this second interview, the HER's presented a rather detailed explanation of the HES program and extended the offer of an examination. Even though the content of the interview was carefully planned, the HER's were allowed some latitude in the presentation to aid in establishing proper rapport. The HER's were instructed that each case was to be treated individually and that with few exceptions, cooperation could be achieved if there was sufficient insight into the real matters of influence. The general approach was that each case should be conscientiously pursued in a professional manner, both directly by the HER's and indirectly through other potential influences until an examination was achieved or until there was, without a doubt, no chance of achieving an examination. Exceptions to these instructions were allowed in the case of persons who were manifestly unable to come to the examination center or where further pursuit would create problems in public or professional relations. Within this approach, the HER's were given considerable latitude and independence in dealing with each case so long as the efforts and approaches were straightforward and factual. In some uncooperative cases, where a change in interviewer personality might have been beneficial, a different interviewer was assigned. In other cases, cooperation was sought through another member of the household or some other influential intermediary.

Various "selling points" might also be factors in the high response rate achieved by HES. In some cases the HER might emphasize the fact that the sample person had the chance to receive an expensive medical examination absolutely free.

Also, all sample persons were advised that findings are treated in a confidential manner. The Census postcard apprising the household of the survey carried a statement of confidentiality. During the initial interview, confidentiality was again assured. Also, the forms used for data collection carried statements of confidentiality.

Examination Policies

Several policies related to the examination itself were important in the reduction of nonresponse. Every effort was made to arrange a suitable time for the examination, including night or weekend appointments for those persons who found the normal weekday appointments inconvenient.

In addition, HES provided transportation to and from the examination center. This was not only a service for those who did not have readily available transportation but it was also a valuable technique in obtaining examinations from persons who might have failed to keep their appointments without some further stimulus. In Cycle II, the HER who had conducted the interview and obtained the consent for examination was responsible for transporting the children. In Cycle III, with an older age group and fewer HER's, local transportation companies provided this service, under contract with HES. An adult escort employed by HES accompanied the youths. Where HES transportation was refused, a mileage allowance was made available to those who provided their own transportation.

The examinations were designed to include a minimal number of tests that might be distasteful and no tests which were potentially harmful to the examinees. If, despite this general practice, a particular test was completely unacceptable to a specific examinee, it could be omitted if this action assured his cooperation in the remainder of the examination. Also of importance was the effort made by the examining staff to answer questions and explain procedures and equipment to the examinees. The survey was in an area long enough for this atmosphere and news about the examination to be relayed by examinees to their friends who might also be part of the HES sample. Therefore, a sample person who initially refused the examination might reconsider and consent after listening to a peer who has found the experience informative and enjoyable. Finally, there was an immediate followup and rescheduling of those who cancelled or failed to keep an appointment. In some cases those who failed to keep an appointment due to fear or various apprehensions could be contacted and repersuaded in time to be examined only a few minutes later than originally scheduled.

Evaluation of Nonresponse

For the purpose of discussion nonrespondents can be divided into two groups: those persons who should have been, but were not, identified as sample persons and those identified as sample persons but who were not examined. In HES the first group was composed almost entirely of those who were absent from the sample household during the entire examination period and for whom no information could be obtained from family, neighbors, or other knowledgeable sources. The information defining whether a person residing in one of these sample households was indeed a sample person was therefore unavailable. But if usual rates of occurrence of sample persons per sample household were applied to these households, the resultant expected loss of sample persons would be small. In Cycle I this group comprised 1.6 percent of the total sample and in Cycles II and III, approximately 0.1 percent. The greater percentage in Cycle I is largely explained by the age segment being sampled. Practically every household will contain an adult but not necessarily a member of the younger and more restricted age segments sampled in Cycles II and III. In these two cycles, information was obtained from neighbors or others familiar with the age composition of these sample households, further reducing this classification of nonrespondents,

The second group of nonrespondents, those identified but not examined, is larger and worthy of more discussion. This group was composed of those who died or moved away between the first Census contact and the examination, those manifestly unable to be examined due to illness or severe physical disability, and those who refused to participate. The latter made up the great majority of the nonresponse experienced by HES.

In Cycle I, of 7,710 sample persons 6,672 were examined giving a response rate of 87 percent. In Cycle II, of 7,417 sample children 7,119 were examined, a response rate of 96 percent. In Cycle III 90 percent of the 7,514 sample youths were examined. Although these rates are good for a survey of this type, as long as there is nonresponse, the final estimates are subject to a potential nonresponse bias. If the nonrespondents differ from the respondents with respect to a particular characteristic, then usual estimating procedures will produce a biased estimate for that characteristic. However, a small nonresponse rate is usually considered acceptable; first, it is generally unavoidable and second, the characteristics of the nonrespondents would have to be substantially different to produce even a small bias in the estimates of the characteristics being measured. In practice, the biasing effect of nonresponse is never fully known although insight into potential effects is often gained by the consideration of known characteristics associated with the nonrespondents. This is especially true if there is an association between the known characteristics of nonrespondents and the information of interest to the survey. Basic data on nonrespondents, such as age, race, sex, and urban-rural status, are often known to survey personnel. Thus, in Cycles II and III, for example, since height and weight (unknown for nonrespondents) were generally associated with age (known) more could usually be inferred concerning the biasing effect of the nonrespondents on the estimate of height and weight than if no association were present. In HES rather detailed basic information was obtained on practically all sample persons by means of questionnaires administered before any mention of the examination. Comparisons can therefore be made on a number of variables between those persons in the sample who consented and were examined and those who were not examined. This approach has been used in analyzing factors related to response in the Cycle I sample.[7]

Another approach used to evaluate nonresponse is to investigate examination rates by various age, sex, and other demographic variables. In Cycle I, the examination rate was highest for the youngest persons and diminished with increasing age. In the age group 18-24 years, 90.2 percent of the sample was examined, whereas in the age group 75-79 years, the examination rate was 74.3 percent. In Cycle II, the examination rates by single years of age ranged only from 95.7 to 96.2 percent. By sex, the examination rate in Cycle I was 88.3 percent for men and 85.0 percent for women. In Cycle II, the rate was 96.5 percent for males and 95.5 percent for females. Looking at population density, in Cycle I the examination rate of five population density classes ranged for 77.7 to 92.0 percent, and in Cycle II eight classes ranged from 93.1 to 98.6 percent. In both cycles there was a distinct relation between response and population density, with rural areas having the highest response and highly urban areas the lowest. These examples show that although overall response rates may be good, individual cells in various groupings may have a somewhat low response. The fact that the nonresponse bias may be larger in these lower response cells should be kept in mind when interpreting the data. More complete accounts considering these and other variables have been published previously.[2,3]

Although the above methods of analysis do give insight into the problem, a quantitative evaluation of nonresponse bias necessitates an estimate of the characteristic of interest for the nonrespondents. In practice, the value of the characteristic of interest for the nonrespondents is never known, and therefore, the nonresponse bias cannot be accurately computed. But many ingenious techniques have been devised to help deal with both general and specific aspects of the nonresponse problem. In surveys where efforts to win cooperation are not particularly intense, a technique sometimes employed is to subsample the nonrespondents and through increased persuasion obtain enough response in the subsample to make estimates for all nonrespondents. With these estimates adjustment for nonresponse bias can be made. Several problems are inherent in this approach, the most obvious being that the subsample of nonrespondents most probably will contain those who will not respond even to the increased efforts, so that the method used to adjust for the nonresponse bias would itself be subject to a nonresponse bias. In surveys such as HES, where intensive persuasion efforts are made on all subjects, those who are not cooperative after these efforts are "hard core" nonrespondents; and any attempt to achieve even limited cooperation, much less to obtain a satisfactory response from a subsample, would be futile.

In Cycle I information about nonrespondents obtained from existing records proved valuable in evaluating the impact of nonresponse. During the household interview each sample person was asked to give the name and address of his personal physician and to indicate how long it had been since he had last seen him. In each household the respondent was asked to sign a form authorizing his physician to release medical information to the National Health Survey. If a nonexamined person had signed such a medical release and given the name of a personal physician whom he had seen in the preceding 2 years, an inquiry was sent to his physician. If the person had not signed a release, the inquiry form would be sent to him with a request that he forward it to his physician for completion. A similar inquiry form was sent to an examined person from the same place who was of the same sex and, as nearly as possible, the same age. Although there were some problems in obtaining usable medical information for the nonexamined, the data collected were comparable to that for examinees. The study concluded that it was improbable that the nonresponse introduced a serious bias in the findings of the survey.[3]

Whether or not point estimates of characteristics of interest are made for nonrespondents, the evaluation of possible limits is often valuable in subject areas where reasonable limits can often be assumed with the aid of knowledge and experience in the subject matter area. If the sample mean is the estimate of concern, a quantity equal to that which would have been obtained with 100 percent response is $P_r \bar{Y}_r + P_{nr} \bar{Y}_{nr}$, where P represents the proportion of response (P_r) or nonresponse (P_{nr}) and $\bar{Y}$ the estimate of the sample mean for respondents $(\bar{Y}_r)$ and nonrespondents $(\bar{Y}_{nr})$. With the above notation the nonresponse bias can be expressed as $P_{nr} (\bar{Y}_r - \bar{Y}_{nr})$. When the response is extremely high as in Cycle II and P_{nr} is on the order of .04 or even if the category of interest is of relatively high nonresponse and P_{nr} is on the order of .06 (the highest nonresponse in all age-sex categories) the difference between the response and nonresponse values $(\bar{Y}_r - \bar{Y}_{nr})$ would need to be quite sizable to produce more than a negligible bias. However, in Cycle I in the 75-79 years of age group, where the nonresponse was .26, such a strong statement could not be made. For estimation purposes the method chosen to deal with that nonresponse which did occur was imputing to nonrespondents the characteristics of "similar" respondents. This was accomplished by multiplying, within classes of sample persons, the weights of respondents by the reciprocal of the proportion responding. In all cycles classes were defined by age and sex within each geographic location. Since basic weights of sample persons within a location were generally the same this adjustment had essentially the same effect on estimates of totals and means as assuming that $(\bar{Y}_r - \bar{Y}_{nr})$ was negligible in each class, i.e., that the nonresponse bias was negligible in each class.

Control of Missing Data

In all operations where data are collected, missing data in the records of respondents are a potential problem. Missing data often create problems during data processing, editing, and analysis; but generally the responsible act occurs during the data collection process, and therefore missing data are considered here as a type of nonresponse rather than a problem related to data processing or analysis. As in any nonresponse problem, the most reliable means of reducing the bias introduced by missing data is the prevention of the nonresponse. There are some unavoidable losses of data: omission of a test when it might be detrimental to the examinee, refusal of an examinee to participate in a particular test, or inability to obtain an acceptable performance of a procedure. The examining staff were expected to use discretion regarding these unavoidable losses. They also were responsible for preventing any avoidable loss of data. Much data in HES were recorded by hand on standardized forms, and illegible or insufficient entries were essentially missing data. Emphasis was therefore put on standardized, accurate, complete, and legible entries in the recording of all data. Each examiner was responsible for reviewing each record as soon as he had finished making entries in it to be sure there were no avoidable omissions or errors. If an omission was unavoidable, he entered an explanatory note on the form. Additional personnel were used as recorders in certain parts of the examination to facilitate recording and to check for errors, omissions, and inconsistencies.

In addition immediate review of records in the examination center while the examinees were readily available was another valuable technique in the reduction of missing data. After the test and the initial review, the examiner entered the time in and out, his initials and comments on a control record and returned the folder to the clerical assistant. The clerical assistant again reviewed all records and checked to see that all tests had been performed and properly recorded before the examinees left the examination center. In some cases extra effort had to be exerted to produce a record before the examinees left the examination center. For example, all X-ray films were developed before the examinees left the center and inadequate films were replaced. To prevent loss of data from equipment-malfunction back-up equipment was kept in the examination center for certain tests.

Review of questionnaires, especially those which the respondents were responsible for completing, was productive in the discovery and reduction of missing data. As in the collection of data by examination, it was advantageous to review questionnaires in the field where missing or inconsistent data could be readily corrected by personal contact with the respondent or a responsible member of the sample household. All HES questionnaires were reviewed in the field, some quite extensively, some less so. The only exception was those Cycle II and III questionnaires which were left with schools during the summer and were not returned until the survey had left the area. The majority of school questionnaires, however, were returned and reviewed while the survey was in the area and missing information could readily be obtained. Edits were performed during processing on these as well as all HES data, but the advantage of the field review in reducing missing data was lost in this particular case. Breakage of blood vials during shipment to the laboratories or as the result of accidents in the laboratories themselves was another example of a source of missing data where the time of occurrence prohibited remedial action by the field staff.

Preventive procedures greatly reduced the problem of missing data in HES. The usual procedure for treating that which did exist was to allow the analyst who was familiar with the subject area to determine an appropriate method of imputation. On those parts of the examination where the estimates to be made were considered sufficiently critical, imputations of missing data were made. Where some but not all of the elements of a particular part of the examination were completed for an individual, the missing elements were usually imputed by matching those results which had been obtained on this part of the examination with those for other examinees of the same age-sex-race group whose examinations were complete and then randomly selecting within this frame the values for missing elements. When an entire data area of the examination was missing imputation was made by using values of an examinee with complete results from the same age-sex-race group. Whenever the extent and probable impact of missing data was serious enough to warrant concern or when imputation techniques were of special interest, the extent and methods of treatment were included in reports of findings.

MEANS OF REDUCING MEASUREMENT PROCESS ERROR

A second type of nonsampling error which deserves consideration equal to if not greater than that devoted to nonresponse is measurement process error. Measurement process error is used here to denote that error which occurs in the determination and performance of the measurement procedures and includes errors involving questions of validity and of reliability. It is not used to encompass recording and transcribing errors which are somewhat different and will be considered in another section of this report. Hereafter measurement process error will be referred to simply as measurement error.

It is simple to conceptualize a measurement composed of some "true" value plus some measurement error. But in the collection and analysis of field data, the exact values of these components for a particular measurement cannot be obtained. Therefore the traditional estimates of central tendency and sampling error will be subject to some degree of bias depending upon the nature of the measurement errors. The lack of specific values of measurement errors associated with a particular measurement does not preclude reasonable estimates of the overall measurement error, although knowledge of the values could be useful. With these estimates in mind, judgments can be made about the desirability of one process or procedure over another in the effort to reduce measurement errors. In HES the most direct attack on the problem of measurement error was to adopt all feasible precautions and procedures to minimize measurement errors in the collection process. These efforts may be divided into three major areas: planning, standardization of the testing environment, and standardization of the testing processes.

Planning

The first and certainly one of the most important steps in the conduct of a survey is defining precisely the information to be collected and deciding the best way to collect this information. Ideally, how the data are collected should be taken into account in deciding what data are to be collected. If methods for obtaining certain information are unreliable and it is not feasible to develop a satisfactory method, then the resources and effort which would be used to collect the unreliable data might be better used to obtain other information which may be collected by more reliable methods. Advisors, both from within HES and from other sources, are actively involved in the determination of what data are to be collected and the methods which should be used to collect these data.

On the basis of advisor recommendations, methodological studies, general knowledge of the problems involved, and experience gained in previous cycles, tentative plans for a cycle are developed. The development of the final design from these tentative plans entails an intensive series of pretests, evaluations, and resultant modifications. For example, in the preparation for Cycle III the proposed design, which was similar to that used in Cycle II, was pretested three times. The first pretest was in Brooklyn, New York, on 93 youths representative of the age spread and roughly equal by sex. This very preliminary pretest was conducted in conjunction with the nineteenth location of Cycle II. The week prior to the pretest was spent in orientation and on-the-site training for all members of the field staff. A flexible examination schedule was maintained during the pretest to allow evaluation of various scheduling patterns, such as sex separations and varied numbers of examinees at one time. Exit interviews were conducted with all examinees to obtain their reactions and suggestions for improvement.

After much evaluation, additional planning, and modification involving facilities, procedures, equipment, sequencing, and allotment of times for various parts of the examination, a second pretest was held in Detroit, Michigan. Like the New York pretest it was performed in conjunction with a Cycle II location. Again, the examination schedule was purposely light to allow time for training and continuing evaluation. After summary evaluations and a few new modifications, the largest and final pretest was conducted in Wilmington, Delaware. Of 188 pretest youths 163 were examined. By this time, most of the data collection procedures were in the last stages of refinement and a number of consultants contracted by HES visited the examination center to help with additional training of examining personnel and to plan final details concerning procedures and equipment. The results of this pretest were carefully considered and, after a few minor modifications and revisions, preparations were made to begin examination of the Cycle III sample. At the first sample location, the initial week of examinations was devoted to examining nonsample youths. These "dry runs," as they are termed, provided a final training period and on-the-spot testing of equipment and procedures.

In all HES cycles, after the test procedures have been pretested and are ready for the start of a cycle, detailed written instructions are compiled. Written instructions are essential to the examining staff for training, reference, and review. They are especially valuable in lengthy surveys to prevent minor changes and drifting in examiner techniques and procedures. In addition, written instructions serve as an essential reference for those who analyze the data.

Standardization of Test Environment

Many measurements will vary depending on the environmental influences present at the time of the measurement. Temperature, humidity, noise level, light intensity, and visual distractions are a few of the many environmental factors which could directly influence the results of various measurements taken in the Health Examination Survey. The problem of standardization of environment in HES is greatly reduced by performing the examination in mobile trailers which are transported from location to location.

Upon arrival at a sample location the trailers are parked parallel to one another and are connected by enclosed passageways to form the examination center. Within the center there is a sound-proof room in which the hearing test is conducted. The temperature is regulated and in the portion of the examination center where the exercise tolerance test is performed, the humidity is also kept within certain limits. Light intensity for visual testing is controlled. In addition, the arrangement of rooms within the center is carefully planned to take account of the possible effect of a particular area on an adjacent area, another factor in the standardization of the test environment.

Little environmental control is possible in the data gathered by home administered questionnaires other than that resulting from their administration in the home and at a time when the respondent is not distracted by work, visitors, or unusual situations. But this is of less concern, since most HES data are gathered in the controlled environment of the mobile examination center, and the type of data gathered by questionnaire are subject to little environmental influence compared with those gathered in the examination center.

Standardization of Testing Processes

Standardization of testing (measurement) processes is, for the most part, achieved by the use of mechanical devices, appropriate operational procedures, care in selection and training of examiners, and procedures designed to reduce the impact of subject errors. The magnitude of systematic error or bias associated with a testing process is generally difficult to ascertain although in some cases the direction of the bias is known and in certain instances the magnitude can be estimated. Aside from the bias associated with testing processes, the variable errors are also of concern. These variable errors are generally more easily quantified in the resultant data than biases for the primary reason that it is difficult to discover biases. In the examinations given by HES, the process variation associated with the test procedures was usually potentially greater and therefore of more concern than that associated with environmental factors. Many of the procedures employed to identify and reduce process variation were also important in the identification and reduction of certain biases.

Use of mechanical devices

A well constructed measuring device is less subject to large variable and systematic errors than a human examiner; therefore, one method of reducing errors in a testing process is to use mechanical devices. Devices which produce "hard documents," in such forms as printouts, tracings, photographs, and magnetic tapes, are particularly valuable since additional procedures which normally would allow human errors are performed mechanically.

Although the use of mechanical devices helps reduce errors in a measurement process, it must be recognized that these devices themselves are subject to variation and must be calibrated regularly. Calibrations may be performed only once at each location or as frequently as before and after each subject. The HES Field Staff Procedure Manual contains instructions for calibrations to be performed by the examining staff as well as testing procedures and other pertinent information. In some cases the resources for calibration of the more sophisticated instruments were not available at the examination center, and either a technician experienced with the instrument had to come to the examination center or the machine had to be sent away for calibration. Various arrangements and systems of back-up equipment were used when instruments were away from the examination center.

Since mechanical devices are under direct examiner control or supervision, human error may readily enter into the measurement process depending upon the degree of examiner involvement. The degree of involvement varies from situations in which there is active involvement, as when a physician or nurse takes a blood pressure, to those in which the examiner is rather passively involved, as in the turning of a switch on an instrument. In HES the attempt to reduce examiner errors was approached in several ways. Whenever possible the measurement procedures (aside from the use of mechanical devices) were highly standardized and the amount of subjective judgment necessary was small. In measurements where the degree of human involvement and judgment was large, a system of checks, when practicable, was included in the measurement procedure by using a second person as a recorder-observer. For example, the recorder for the dental examination served a necessary function by relieving the dentist of recording findings throughout the examination, but in addition, she was responsible for double-checking the forms for incomplete and inconsistent findings. In taking body measurements the more active participation of the recorder-observer demanded that he have training in body measurement procedures equal to that of the examiner, and the technicians responsible for taking body measurements interchange in the roles of examiner and recorder-observer. Repetition of measurements was another method used to reduce the impact of testing process errors, including examiner errors. For instance, in obtaining blood hematocrit, a technician routinely performed two determinations on each subject, and in addition a second technician repeated readings on a sample basis.

Selection and training of examiners

Those examiners who were conscientious and had personalities suitable for the type of work were less likely to introduce as great a degree of measurement error as those who were not so dispositioned. Obviously, proper basic training was another prerequisite to becoming a member of the examining staff, but aside from previous experience and knowledge an initial training period and frequent retraining in the specific techniques of HES were necessary for each examiner to become properly skilled. The permanent HES advisory staff supervised this training. When necessary, consultants augmented this training both in the field and in their own facilities.

The length of time necessary to complete extensive surveys such as those conducted by HES creates problems of drift in techniques that are not as likely to occur when data are gathered in a shorter period of time. The practice of providing all members of the examining staff with detailed written instructions covering test procedures helps to achieve consistency in measurement techniques throughout a cycle. The forms that the field staff used were quite structured, and most data were recorded as numbers. Retraining is also important for the achievement of consistency in measurement techniques in HES. Time spent retraining might range from a few minutes for reviewing a single item with a fellow examiner to several days' retraining in an entire area. An invaluable policy in the standardization of procedures and in the training and retraining of personnel is the use of time before and during the survey for the examination of nonsample persons. At the beginning of each location, the day prior to the start of regular examinations was set aside for the examination of a few volunteer nonsample persons. The light schedule allowed ample time for reviewing procedures, clarifying any questions regarding drift in technique, and assuring that all equipment was properly prepared for the examination of sample persons on the following day.

Reduction of subject errors

Subject error is one of the most difficult types of measurement error to evaluate or control. Subject cooperation and concentration are extremely important when a mental response is being elicited. In HES this type of cooperation was perhaps most important in the psychological testing. Subject cooperation and concentration were also often important when active physical involvement was required as in the spirometry where data on maximum expiratory flow rates were gathered. If a maximum response was to be measured, the subject had to give a maximum effort. To this end, the technicians routinely gave vocal encouragement before and during the subject's exhalation.

Procedures designed to reduce subject error may be quite similar to or even coincide with those designed to reduce examiner error. Subject errors as well as examiner errors deserve corrective procedures in the taking of body measurements. That the subject is not standing or sitting in the proper position will tend to introduce errors in many body measurements. This is more of a problem than immediately apparent, since the technician in making some of the measurements, is not able to observe whether the subject has deviated from the correct position. In HES this particular source of subject error was controlled by using a second technician as a recorder who was also responsible for seeing that the subject was in the proper position for all measurements.

Use of repeated measurements is effective in reducing subject error as well as other testing process errors. An example is found in the determination of threshold hearing levels, where the threshold recorded was the lowest decibel reading at which responses were obtained in at least 50 percent of the trials using a minimum of three trials.

PROCESSING ERRORS

In HES the specific methods by which data were recorded vary: some data were recorded directly on magnetic tape for immediate computer use; other data were recorded in the form of hard documents such as X-rays, height photographs, or weight printouts; but the majority of HES data thus far have been recorded by checking boxes and making written entries in words or numbers in appropriate spaces of standard forms. Of these three methods the first is obviously superior for purposes of reducing human errors in recording, coding, and punching. But there were many areas of data collection in HES where present technology and high costs prohibited the use of this method. In the ongoing work in HES, efforts are being made to reduce the proportion of data recorded on forms and increase the proportion recorded on magnetic tape or as hard documents.

The processing of data is an activity in which errors which have occurred previously but have not been discovered and corrected can be detected. But it is also an activity in which new errors can easily occur. The objective of the HES data processing program is to detect as many previous errors as reasonably possible while providing for tight control on process errors. Although extensive clerical edits to detect missing and in some cases inconsistent data are performed in the field, the actual processing of data can be considered to start after the data arrives at headquarters from the field at the end of each location. At this point practically all examination and questionnaire data are subjected to clerical edits of some nature. Heights and weights recorded on the forms are all verified by comparison with the original photographs and printed records. Approximately a 5-percent sample of all examination forms are reviewed on a routine basis as a part of the posting and evaluation of replicate data. The Census household questionnaires are reviewed again for erroneous exclusion or inclusion of those who should or should not be sample subjects as well as for other errors. Edits are performed on samples of other questionnaires. Except for a few special cases all forms and questionnaires are microfilmed before being sent for coding and punching; if records are later lost in shipment or processing, the data are preserved. Microfilm also provides a convenient, compact record of individual forms for immediate reference. After microfilming, the records are sent directly for coding, punching, and editing, according to detailed specifications and instructions prepared in HES. In the design of forms consideration was given to the reduction of clerical coding so that extensive coding is not required. To reduce coding and punching errors *all* coding and punching work is verified. Edits are then performed and printouts returned to HES where discrepancies and disallowed values are checked against the microfilmed records and rectified. The data are then forwarded to the Division of Data Processing, NCHS, for transfer to tape and more extensive edits and consistency checks. Except for the problem of imputation for missing values referred to earlier, the data are then considered ready for analysis.

SURVEILLANCE AND EVALUATION OF RESIDUAL MEASUREMENT PROCESS ERROR

Monitoring Systems

Despite efforts to reduce measurement process errors, residual errors of a magnitude large enough to warrant concern occur with some regularity. There is, therefore, a real and urgent need to have a system whereby these residual errors can be monitored. As stated previously, the concept of quality control is based on the desire to obtain end products of a certain quality. Therefore, one of the main purposes of a monitoring system would be to indicate whether or not the measurements produced by a certain measurement process attained the desired quality. A second major purpose would be to make possible quantitative summary descriptions of residual measurement errors to aid in the interpretation of survey data.

In the Health Examination Survey several types of monitoring systems were used. One of the more systematic of these was the review of forms and questionnaires. Records review is a common practice in efforts to monitor and control errors although the effectiveness of this method is generally limited to certain types of errors, specifically, missing, inconsistent, and impossible values. In addition, this method can be used to detect errors in interpretation of records, such as X-rays, electrocardiograms, and spirograms where readings of hard documents can be checked for correctness of interpretation against the document itself. Various errors can, of course, have an influence on the process of obtaining the hard document; and, therefore, procedures designed to control errors are as important in the creation of a hard document as in any other area of collection. In HES, as previously described, extensive review is performed in the examination center before the examinees depart and during data processing.

Perhaps the most direct monitoring system used in the Health Examination Survey was the observation of the measurement process as it was being applied to an examinee. Medical, dental, and psychological advisors from HES and other advisors and consultants regularly visited the examination center to observe examination procedures and retrain examiners if necessary. A good example of how routine observation was used as a monitoring system can be found in the taking of body measurements. The one examiner, in addition to acting as a recorder and aiding in the positioning of the examinee, was also responsible for observing and correcting any errors in measurement technique.

The most extensive system of monitoring used in the Health Examination Survey in Cycle III was the collection and evaluation of replicate data. Replicate measurements are useful for a variety of reasons which include use as a means of increasing precision of estimates of individual measurements, as a training technique, and as a monitoring system which includes the objective of final evaluation of measurement errors. These objectives are not incompatible, and replicate data collected primarily for one of these objectives often indirectly, if not directly, accomplish one or both of the remaining two. For this reason replicate data are most often collected with a combination of these objectives in mind. The single most important source of replicate data in Cycle III was the replicate examinations where approximately 5 percent of the regular examinees were returned to the examination center for a second complete examination except for drawing blood and taking X-rays. Other sources of replicate data are discussed later in this report.

Biases and Controls in Replicate Measurements

A major source of uncertainty in estimates derived from replicate measurements is failure to make the replicate measurement under the same conditions and in the same manner as the original measurement. This uncertainty is difficult to evaluate and most evaluations are restricted to subjective statements concerning the direction and/or size of the bias and the need for concern in the analysis of data. Several policies regarding Cycle III replicate examinations were valuable in the attempt to obtain replicate measurements taken under the same conditions and in the same manner.

Replicate examinations were not conducted during a specific time set aside exclusively for them, but whenever possible were interspersed among the regular examinations. An original examination was given priority over a replicate examination in that no replicate examination was scheduled if it occupied time needed for a regular examination. In practice there was often space to interject replicate examinations in the schedule without interfering with regular examinations. However, this priority plus the fact that replicates were drawn from those examined had the effect of increasing the likelihood that a replicate examination would be scheduled toward the end of the examination period. Nevertheless, the attempt to space the replicate examinations in the schedule was a valuable policy in that the interspacing of replicate and original examinations created an atmosphere more conducive to the replicate examination's being conducted in essentially the same manner as the original.

The examiners had been informed of the purpose and importance of the replicate examination program. It was emphasized that they should not vary their procedures on a replicate examination or in any way try to collect "better" data than they normally would. Thereafter, the conduct of a replicate examination was not given any greater emphasis than any other instruction since overemphasizing "sameness" might have created more bias than it would have eliminated.

During the original examination neither the examiner nor the examinee knew whether or not the examinee would be returned for a replicate examination. During the replicate examination examiners were not specifically informed that an examinee was a replicate although no attempt was made to conceal this fact since in an examination as lengthy as that given in HES the examinee would undoubtedly be remembered by several if not all examiners. Even though an examinee might be remembered it was extremely unlikely that an examiner would remember a specific measurement after a time lapse of 2 or 3 weeks. Some bias might be introduced by the examiner's knowledge of the replicate status of an examinee, but it would seem that generally this bias might be quite small when compared to the measurement error and in some cases to the biases associated with the knowledge and familiarity gained by the examinee during the original examination. Examinee bias can be important especially in measurements where a response is elicited or when due to the time lapse, the true value of the measurement has changed. The effect of learning is certainly a confounding factor in areas such as psychological testing and to a lesser extent in measurements such as determination of hearing levels, where familiarity with the testing devices, procedures, and personnel may well influence the results. Since the time lapse was usually 2 or 3 weeks, some appreciable changes might occur in certain measurements such as weight. However, for most of the data collected the actual change can only be very small and this effect may usually be neglected.

In Cycle III replicate data were obtained on approximately 70 percent of those selected for replicate examinations. One explanation for this low rate is that the persuasion and follow-up efforts were not as intensive as for regular examinees. This is a partial result of giving priority to regular examinees if interviewer or examination time was limited. There also seems to be an increased objection to returning for a second examination, as demonstrated in the most frequent reasons for refusal: "One time is enough" and "I can't miss school again."

Selection of Replicate Examinees

The selection of Cycle III examinees for replicate examinations was random within certain restrictions imposed by practical considerations. One of the restrictions was that replicates were selected only from those examined during the first week and a half of the approximately 3½ weeks of examinations. This time period was chosen to facilitate the interspersing of replicate examinations with originals in the examining schedule without interfering with the time allotted for original examinations and without scheduling additional time to accommodate replicates. In a voluntary survey it is obviously impractical to follow a scientific random process in scheduling subjects, so those scheduled during the first week and a half are not, in the strict sense, a random sample of all those scheduled. But evidence that replicates might be considered "representative" is found in the fact that youths of certain ages, locations, incomes, etc., are not routinely more likely to be scheduled during any particular segment of the examination schedule. However the availability and desires of the subjects do influence the composition of the replicate sample. For instance, an examinee whose participation in an original examination was achieved only after repeated contacts by survey personnel might be excluded from a replicate examination since it is unlikely that he would have received an original examination during the first week and a half. The schedule of locations considering time of year, sequencing of examinations, relation to other events which might make subjects more or less available, and other related aspects give no obvious discriminatory factor. After examining these and other relatively minor considerations there appears to be no reason to believe that the subjects scheduled and examined during the first part of a stand differ from those scheduled and examined during the latter portion of a stand with respect to the data gathered.

Another restriction on complete randomness in the selection of examinees for replicate examinations was the exclusion of those examinees who were "geographically inconvenient" to the examination center. "Geographically inconvenient" was arbitrarily defined as a distance of 30 miles or greater; although if conditions dictated, exceptions were sometimes allowed. A primary consideration in choosing a site for the examination center was the centrality of the location in relation to the sample segments (a segment is a cluster of households). Since segments were drawn with probability proportional to population, most segments were in relatively populated areas; and so the examination center was also in or adjacent to a relatively populated area. Therefore, the subjects deleted by this 30-mile restriction usually resided in relatively less populated areas; so this restriction may create a bias in the replicate data if, in fact, characteristics and errors of concern differed by population density. Even if differences did exist, the total effect of this restraint was not great since it excluded only approximately 10 percent of the eligible examinees. There were other minor restrictions of medical and operational nature imposed on the complete randomness of the replicate sample, but they were not readily associated with large differences. Also they deleted at most only 1-2 percent of the eligible examinees and for these reasons are of small consequence.

Since the purpose of replicate examinations is to give information about errors, the matter of concern between those excluded and those eligible for selection is not the possible differences in the values of measurements but the possible differences in the errors associated with the measurements as shown by the discrepancy between two measurements on the same subject. For example, measurements may vary markedly by some demographic classification, but this is not so relevant as the question of whether or not the errors vary by this classification. It should also be noted that although subjects did influence measurement errors, the environment, procedures, and examiners were also highly influential. The consideration of these additional influences causes a completely random selection of subjects to be of somewhat less concern.

Additional Replicate Data

Although the full scale replicate examinations were the single most important source of replicate data in Cycle III, other replicate measurements were performed for the purposes of monitoring, training, and obtaining more precise estimates of individual characteristics. As mentioned previously, blood was not drawn and X-rays were not taken during the full scale replicate examination. However, replicate data were obtained for all blood work and X-rays. Each day for 15 consecutive days at each location an additional tube of blood was drawn from two subjects. The pairs of tubes, which were sent to a laboratory in the same shipment but under different identifying numbers, were systematically allotted to the laboratories performing Cycle III work for HES. In addition, for certain tests the laboratories routinely performed their own replicate determinations on every blood sample. Whenever differences larger than predetermined tolerances occurred the analysis was repeated. Replicate X-ray interpretations were performed by evaluators on a sample basis.

The body measurements were performed by four examiners and therefore, this replicate data was then divided into two types, inter- and intraexaminer. Interexaminer replicates were collected at the beginning of each location where the four technicians each performed a set of measurements on one of the two dry run examinees selected for replicates. The results were compared with predetermined tolerances and if indicated, measurements were repeated. However, the original measurements were preserved for analytical purposes. Technician pairing was rotated on a systematic basis at each location. In these replicate data, the replication was performed the same morning or afternoon as the original, whereas with full scale replicates the period between examinations was usually a week to 10 days. During the course of each location, every technician repeated his own measurements on one regular examinee. The duplicate measurements were not performed immediately after the original measurements, but usually at the end of the half-day period. Again results were compared to a set of tolerances and differences were resolved by the supervisory technician. Replicate body measurement data were also gathered during selected retraining periods.

In the dental area, the general plan concerning the collection of replicate data was that two dental advisors alternated periodic replicate data gathering sessions with the examining dentists. In the course of a typical session, replicate measurements were gathered on approximately 35 sample youths during a period of 4 or 5 days. The duplicate examination immediately followed the original and discrepancies were discussed and resolved while the subject was in the chair. In addition to the replicates between advisors and examining dentists, the two advisors also performed periodic replicates on one another under field conditions.

In the medical area, there also was a certain amount of replicate data collection, primarily for training purposes. During the first several days of each location, the dry run, and several examination days, both the examining physician and medical advisor evaluated each examinee, comparing and discussing findings of the examination.

Evaluation of Residual Measurement Error

Measurement error, as stated previously, includes those nonsampling errors associated with the determination and performance of the measurement procedures. Since a reported value whose measurement error is entirely unknown is of questionable value, an evaluation of measurement error in the reports of findings of sample surveys such as HES is highly desirable.

Many problems still exist in evaluation of measurement errors in sample surveys. Research is underway, but at present both rigorous theory and detailed operational protocol are in short supply. In HES specific methods are in experimental stages. This section contains a very brief introduction to the effects of measurement errors and problems of evaluation, and to some of the methods by which HES evaluates residual measurement error.

Evaluation is facilitated by the fact that measurement errors may be classified as variable or systematic (bias). The following simple model incorporates this classification. It separates the k^{th} obtained measurement on the j^{th} individual (x'_{jk}) into three components: 1) the "true" measurement (x_j), 2) a systematic measurement bias (b), and 3) a variable measurement error (e_{jk}); that is $x'_{jk} = x_j + b + e_{jk}$. The "true" measurement (which is unobservable) would be the value obtained by the application of the perfect measurement process. This process would measure precisely the characteristic intended and contain no biases, conceptual or other. In addition, repeated measurements on a single subject would not vary unless the characteristic being measured actually did change during the period of data collection. The bias component represents those errors which are systematic and remain constant throughout the entire collection process. The variable measurement error component represents those errors which may vary from measurement to measurement and from day to day and are influenced by uncontrollable factors. It also includes errors which can be controlled but whose magnitude does not justify the effort or expense. If, as is often possible, these variable errors can be considered random, with one composite variable error associated with each obtained value, the task of evaluating the impact of bias and variable errors is made considerably easier.

Assuming the above model, statements can be made about the effect of measurement errors on estimates computed from sample data, specifically estimates of standard error and of central tendency such as the mean, median, and mode. Biases as represented in this model alter traditional estimates of central tendency by a quantity equal to the net bias. Because of this influence, the standard error which is not affected by a bias, alone is an unsatisfactory measure of the accuracy of estimates such as sample means when the bias is large compared to the standard error. In such cases, a measure of accuracy which contains a bias component is the mean square error which is simply the sum of the squared bias and the variance of the estimate. But estimation of this measure necessitates a point estimate of the bias and since the bias component is squared does not indicate the direction of the bias. In practice, point estimates of bias can rarely be made with reasonable certainty and the direction of the bias is important in the interpretation of results. Therefore the general approach in HES is to present estimates of standard error and to discuss suspected biases giving, where possible, an evaluation of the magnitude and direction.

Because of the very large sample, the effect of variable errors on HES estimates of central tendency is of less concern than it might be in other surveys. Variable errors can conveniently and appropriately be assumed to have mean zero in many measurement processes. It is likely, however, that the mean variable error obtained in samples will not be zero, in which case an estimate of the population mean would be in error. With a large sample, the mean variable error is less likely to deviate from zero by more than a small quantity. To evaluate this deviation, HES will use replicate data to obtain estimates of the error variance.

Variable measurement errors can also influence estimates of standard error. To estimate standard error, HES uses a technique called balanced half-sample pseudoreplication.[10,11] The half-sample estimate of standard error is calculated in a manner similar to that of the traditional estimate. The half-sample method, however, requires deviations of a sample (selected according to a particular scheme) of half-sample statistics from the total sample statistic rather than deviations of individual values from the total sample statistic. Each half-sample statistic is obtained from the individual values in one of two primary sampling units. This estimate may be written as:

$$\left[\sum_{i=1}^{L} \frac{(\bar{y}_i - \bar{y})^2}{L} \right]^{1/2}$$

where $\bar{y}$ is the total sample estimate of parameter $\bar{Y}$

$\dot{\bar{y}}_i$ is an estimate of $\bar{Y}$ utilizing data from 1 of 2 PSU's

L is the number of half-sample replicates.

The most common approach in problems of estimation is to assume that the obtained values are nearly correct and to use them in estimating procedures as if they were "true" values. Also only one measurement is usually made on each unit; i.e., $k = 1$, so that for simplicity the k subscript will be dropped. In this case, using the notation of the above model, the half-sample replication estimate of standard error used in reports of findings would be

$$\left[\sum_{i=1}^{L} \frac{(\bar{x}'_{i\cdot} - \bar{x}')^2}{L} \right]^{1/2}$$

But if it is allowed that obtained measurements are not correct but in fact contain possible biases and variable errors, then substituting for x' the following expression is obtained:

$$\left[\sum_{i=1}^{L} \frac{(\bar{x}_{i\cdot} - \bar{x})^2}{L} + \sum_{i=1}^{L} \frac{(\bar{e}_{i\cdot} - \bar{e})^2}{L} + 2 \sum_{i=1}^{L} \frac{(\bar{x}_{i\cdot} - \bar{x})(\bar{e}_{i\cdot} - \bar{e})}{L} \right]^{1/2} \qquad (1)$$

It can be seen that, as in usual estimates of standard error, a bias which operates as the bias component in the above model will have no influence on half-sample estimates of standard error. In practice, biases are more complex and different biases may be operating in different situations and time periods, in which case they would affect estimates of standard error. Considering the relative magnitude of biases of this type which are known to exist in HES it seems that generally biases should have only a small effect on HES estimates of standard error.

Variable errors can effect estimates of standard error, especially when data are presented, as HES' usually are, by classifications such as age, sex, and race. If the estimates of standard error within each classification are not to be affected by variable errors, then the half-sample as well as the total sample must have mean error zero within each classification. Due to the relatively small numbers in some categories this is unlikely in HES data and the error variance will generally contribute to the estimate of standard error. If no assumptions are made about the sample mean values of variable errors, any change in the half-sample estimate of standard error due to variable errors depends on the error variance term and the covariance term in formula (1) above. As is seen in this expression if there is zero or a positive covariance between mean half-sample errors and mean half-sample "true" values, then the estimate will be increased. If there is a negative covariance, the estimate of standard error is increased if the variance of mean half-sample errors is greater than twice the covariance, unchanged if the two quantities are equal, and decreased if this variance is less than twice the covariance. In general variable errors might be expected to slightly increase half-sample estimates of standard errors in HES, due partially to the error variance and perhaps partially to the covariance. It would seem that the extent of the inflation due to the error variance is small since in most of HES data the variable errors are believed to be small in relation to the magnitude of the "true" measurements. Also preliminary indications from replicate data imply that at least in those measurements investigated this component is not of great concern. More definite general statements about the effect of measurement errors are difficult to make since in the wide range of data collected by HES the size, variability, and nature of errors are somewhat diversified. In reports of findings, attempts are made to apprise the reader of the impact of errors on the findings.

Many ingenious methods of evaluating biases have been developed by researchers to meet the particular demands of the situations at hand. Some of the more traditional methods used in HES will be briefly mentioned here.

Systematic errors can produce considerable bias in estimates of certain parameters and unfortunately are in general difficult to detect. When one and only one examiner is involved in the measurement process, the question of systematic errors can perhaps be addressed only by comparing the results with similar studies. In many cases this is not an entirely satisfactory solution since often it is difficult to find studies that are comparable in both the populations measured and in the techniques and environments used. Replicate data, aside from providing a means of evaluating variable errors, can also be used to evaluate certain types of systematic errors. For example, if in addition to the single regular examiner an "expert" takes measurements on a subsample using the same measurement process, then the systematic error associated with the examiner as compared to the "expert" can be evaluated. In HES a scheme of this type was most extensively used in the dental area. If two regular examiners apply a measurement process with different results, comparison, although still useful, is less revealing since it is usually difficult to determine which examiner was in error without referring back to an "expert." If a larger number of examiners is used, comparisons between examiners become more meaningful since, in general, with approximately equal training and expertise, an individual examiner who disagrees with the group is in error.

In Cycles I and II the practice of comparing data collected by different examiners was extended to the comparing of data collected by two examination teams in two different examination centers. The use of two teams and centers greatly reduced the time needed to collect the survey data and also allowed comparison of results which could give indications of the presence or absence of biases due to an examining team or facility. In addition, at a few locations in Cycle II full scale replicate data were collected, enabling comparisons of data gathered on the same subjects by the two teams in the two different examination centers.

As mentioned earlier, there are many questions to be answered in the evaluation of residual measurement errors and it is hoped that more substantial material can be presented at a later date. Also more specific information can be found in reports of findings.

REFERENCES

[1]National Center for Health Statistics: Origin, program and operations of the U.S. National Health Survey. *Vital and Health Statistics*. PHS Pub. No. 1000-Series 1-No. 1. Public Health Service. Washington. U.S. Government Printing Office, Aug. 1963.

[2]National Center for Health Statistics: Plan and initial program of the Health Examination Survey. *Vital and Health Statistics*. PHS Pub. No. 1000-Series 1-No. 4. Public Health Service. Washington. U.S. Government Printing Office, July 1965.

[3]National Center for Health Statistics: Cycle I of the Health Examination Survey, sample and response. *Vital and Health Statistics*. PHS Pub. No. 1000-Series 11-No. 1. Public Health Service. Washington. U.S. Government Printing Office, Apr. 1964.

[4]National Center for Health Statistics: Plan, operation, and response results of a program of children's examinations. *Vital and Health Statistics*. PHS Pub. No. 1000-Series 1-No. 5. Public Health Service. Washington. U.S. Government Printing Office, Oct. 1967.

[5]National Center for Health Statistics: Plan and operation of a Health Examination Survey of U.S. youths 12-17 years of age. *Vital and Health Statistics*. PHS Pub. No. 1000-Series 1-No. 8. Public Health Service. Washington. U.S. Government Printing Office, Sept. 1969.

[6]National Center for Health Statistics: Sample design and estimation procedures for a National Health Examination Survey of Children. *Vital and Health Statistics*. Series 2-No. 43. DHEW Pub. No. (HSM) 72-1005. Washington. U.S. Government Printing Office, Aug. 1971.

[7]National Center for Health Statistics: Factors related to response in a Health Examination Survey. *Vital and Health Statistics.* PHS Pub. No. 1000-Series 2-No. 36. Public Health Service. Washington. U.S. Government Printing Office, Mar. 1968.

[8]National Center for Health Statistics: Cooperation in Health Examination Surveys. *Vital and Health Statistics.* PHS Pub. No. 1000-Series 2-No. 9. Public Health Service. Washington. U.S. Government Printing Office, July 1965.

[9]U.S. National Health Survey: Attitudes toward cooperation in a Health Examination Survey. *Health Statistics.* PHS Pub. No. 584.-D6. Public Health Service. Washington. U.S. Government Printing Office, July 1961. (Out of print)

[10]National Center for Health Statistics: Replication, an approach to the analysis of data from complex surveys. *Vital and Health Statistics.* PHS Pub. No. 1000-Series 2-No. 14. Public Health Service. Washington. U.S. Government Printing Office, Apr. 1966.

[11]Simmons, W. R. and Baird, J. T., Jr.: Pseudo-Replication in the NCHS Health Examination Survey. Paper presented at the Social Statistics Section proceedings of the American Statistical Association, 1968.

———— ooo ————

☆ U. S. GOVERNMENT PRINTING OFFICE : 1973: 543-879/21

VITAL AND HEALTH STATISTICS PUBLICATION SERIES

Series 1. Programs and collection procedures.—Reports which describe the general programs of the National Center for Health Statistics and its offices and divisions, data collection methods used, definitions, and other material necessary for understanding the data.

Series 2. Data evaluation and methods research.—Studies of new statistical methodology including: experimental tests of new survey methods, studies of vital statistics collection methods, new analytical techniques, objective evaluations of reliability of collected data, contributions to statistical theory.

Series 3. Analytical studies.—Reports presenting analytical or interpretive studies based on vital and health statistics, carrying the analysis further than the expository types of reports in the other series.

Series 4. Documents and committee reports.—Final reports of major committees concerned with vital and health statistics, and documents such as recommended model vital registration laws and revised birth and death certificates.

Series 10. Data from the Health Interview Survey.—Statistics on illness, accidental injuries, disability, use of hospital, medical, dental, and other services, and other health-related topics, based on data collected in a continuing national household interview survey.

Series 11. Data from the Health Examination Survey.—Data from direct examination, testing, and measurement of national samples of the civilian, noninstitutional population provide the basis for two types of reports: (1) estimates of the medically defined prevalence of specific diseases in the United States and the distributions of the population with respect to physical, physiological, and psychological characteristics; and (2) analysis of relationships among the various measurements without reference to an explicit finite universe of persons.

Series 12. Data from the Institutional Population Surveys—Statistics relating to the health characteristics of persons in institutions, and their medical, nursing, and personal care received, based on national samples of establishments providing these services and samples of the residents or patients.

Series 13. Data from the Hospital Discharge Survey.—Statistics relating to discharged patients in short-stay hospitals, based on a sample of patient records in a national sample of hospitals.

Series 14. Data on health resources: manpower and facilities.—Statistics on the numbers, geographic distribution, and characteristics of health resources including physicians, dentists, nurses, other health occupations, hospitals, nursing homes, and outpatient facilities.

Series 20. Data on mortality.—Various statistics on mortality other than as included in regular annual or monthly reports—special analyses by cause of death, age, and other demographic variables, also geographic and time series analyses.

Series 21. Data on natality, marriage, and divorce.—Various statistics on natality, marriage, and divorce other than as included in regular annual or monthly reports—special analyses by demographic variables, also geographic and time series analyses, studies of fertility.

Series 22. Data from the National Natality and Mortality Surveys.—Statistics on characteristics of births and deaths not available from the vital records, based on sample surveys stemming from these records, including such topics as mortality by socioeconomic class, hospital experience in the last year of life, medical care during pregnancy, health insurance coverage, etc.

DHEW Publication No. (HRA) 74-1023
Series 2-No. 44

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
PUBLIC HEALTH SERVICE
Health Services and Mental Health Administration
Health Resources Administration
5600 Fishers Lane
Rockville, Maryland 20852

OFFICIAL BUSINESS
Penalty for Private Use $300

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF HEW
HEW 390

U.S. MAIL

THIRD CLASS
BLK. RT.