U.S. National Health Examination Survey

ARTHUR J. McDOWELL

IN THE UNITED STATES, unlike in many other countries, all Federal Government statistical activities are not centralized in one statistics bureau. Rather, these activities have developed in a decentralized pattern, with different agencies responsible for the collection and analysis of statistical data in particular areas. However, a central statistical office insures that these various activities are properly integrated and prevents duplication. This is a function of the Office of Statistical Standards in the Bureau of the Budget, located, of course, in the Executive Office of the President. The Office of Statistical Standards develops and enforces some standards for the quality and comparability of data produced by Government agencies; it coordinates efforts and attempts to avoid duplication and to minimize reporting burdens. But the development and the conduct of statistical programs and activities lie almost completely within the separate agencies.

I do not intend to deal in detail with the extremely broad and varied field of Federal Government statistics. The Budget Bureau publishes a directory of statisticians in the Federal service, a useful 200-page reference book. However, a categorization of these activities will provide a setting for my description of the Health Examination Survey. The Budget Bureau has prepared a chart that groups together more than a score of agencies which have important statistical programs but which collect statistical information primarily as part of their administrative and operating responsibilities. An example of this category is the statistical program of the Internal Revenue Service, which is incidental to its work—collection of taxes.

The second category groups eight agencies having major interest in using, analyzing, and interpreting statistical data, largely, though not entirely, collected by other agencies. An example of one of these user groups is the Council of Economic Advisers.

The third category consists of general-purpose statistical agencies; those charged with collecting, compiling, analyzing, and publishing statistical data for general use. There are four such agencies, each dealing with a particular field. The Statistical Reporting Service, Department of Agriculture, deals with crop statistics, including livestock production, farmers' prices (paid and received), and farm employment. The Bureau of Labor Statistics, Department of Labor, deals with statistics on the labor force, employment earnings, productivity, and the like. The Bureau of the Census, Commerce Department, is responsible for censuses of and current statistics on population, housing, and a variety of other subjects. Finally, the newest of the general-purpose statistical agencies is the National Center for Health Statistics. This center collects, compiles, analyzes, and publishes statistics on morbidity; health care; demographic, economic, and social factors related to health; and births, deaths, marriages, and divorces.

The foregoing detail is presented, not out of bureaucratic fondness for organization charts, but because it relates to a fundamental and sig-

Mr. McDowell is chief of the Division of Health Examination Statistics, National Center for Health Statistics, Public Health Service. *This paper is an extension of his address to the joint meeting of the biostatistics section of the New York City Chapter of the American Statistical Association and the New York City Chapter of the American Public Health Association, March 25, 1965.*

The examinees shown in the photographs are models, not survey subjects.

nificant characteristic of the National Center for Health Statistics. The center, like the Census Bureau or the Bureau of Labor Statistics, is responsible for data collection and analysis and publication, but it does not carry out any substantive program to which the data are related. Persons in other Public Health Service activities or in other agencies may use the data we publish in connection with their work in disease control, disease detection, disease treatment, or health improvement. The National Center is not directly involved. Whether or not this makes it easier for the center to maintain scientific objectivity, it probably makes it easier for some users of the data to accept the fact of that objectivity.

Within the National Center for Health Statistics, the Health Examination Survey is carried out as one of the four major programs. It is one of the three aspects of the National Health Survey, authorized in 1956 by the 84th Congress as a continuing Public Health Service activity. The National Health Survey consists of three survey programs (*1*). One of these, the Health Interview Survey, collects information from the people themselves by household interviews. It is primarily concerned with the impact of illness and disability on peoples' lives and actions, and the differentials observable in different population groups. A second program, the Health Records Survey, is actually a family of record-linked surveys. It includes follow-back studies based on vital records, institutional surveys to establish sampling frames as well as to provide data, and surveys based on samples of hospital records. The third major program of the National Health Survey is the Health Examination Survey.

Health Examination Survey

The Health Examination Survey collects data by direct physical examinations and tests and measurements performed on the sample population studied. This is the best way to obtain definite diagnostic data, data on the prevalence of medically-defined illness. It is the only way to obtain information on conditions which were previously unrecognized and undiagnosed—in some cases, even nonsymptomatic. It is also the only way to obtain distributions of the population by a variety of physical, physiological, and psychological measurements.

A nationwide examination survey which carries out multipurpose direct examinations, tests, and measurements of a preselected representative population sample, is, I believe, an innovation in the United States and in the world. The following are some of its strengths and also, in an attempt at objectivity, some of its limitations.

Strengths. The use of a probability sample design in a nationwide health examination survey is an important strength. It makes possible generalizations concerning the nature and characteristics of the total population from which the sample is drawn, with some knowledge of how reliable those generalizations are.

A second strength of this program is the high response rates which have been achieved to date, rates which greatly decrease the danger of biased generalizations through possible differences between examined and unexamined parts of the sample.

Because the survey collects a wide range of data on every one of the sample persons examined, it is possible to investigate many different interrelationships, not just obvious differentials in disease prevalence related to demographic or socioeconomic factors (age, sex, income, education). A more unusual strength is its ability to relate one set of medical findings to another or to other kinds of data collected in the examination. For example, data on serum cholesterol levels can be related to data on blood pressure, measures of obesity, and prevalence of heart disease. Also, data on visual acuity can be related to school achievement, findings on an eye examination, scores on psychological tests, and other items. This possibility of studying known or suspected interrelationships is one of the exciting aspects of the survey.

A related and additional strength of the Health Examination Survey is that it benefits from its multidisciplinary approach to research. This program draws on and combines the talents of statisticians, physicians of various specialities, dentists, psychologists, nurses, educators, sociologists, management specialists, and others. We also have the advantages that come with the collaborative, interagency approach. The Bureau of the Census collaborates in many phases

Mobile examination center, U.S. health survey

of the survey. Other Federal agencies, such as the National Institutes of Health, the Office of Education, and the Children's Bureau, as well as nongovernment agencies, such as schools of public health, medical research centers, and survey research agencies, work with the survey.

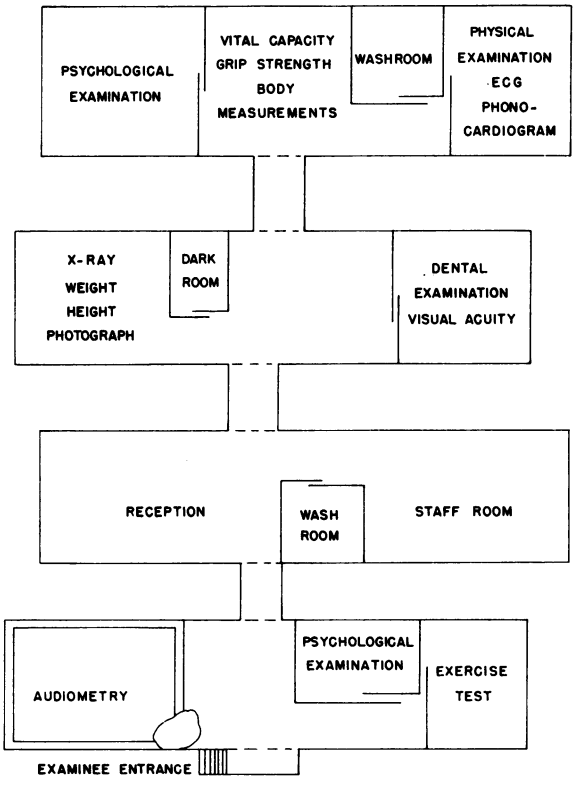
Finally, one of the survey's greatest strengths is the attention we give to the process of measurement at every stage of the examination and data-collection process, and so to the ever-present problem of nonsampling error.

Limitations. One of the limitations of the Health Examination Survey is that the populations we study are all limited in that they exclude institutionalized persons. For some studies this is a serious handicap. Second, the size of the sample included in any of the survey programs is too small to permit many of the detailed breakdowns which might be of interest. While we can get good national data, for example, we are limited to broad regions if we wish to make a geographic breakdown.

A third limitation is that of the cross-sectional—as distinct from the longitudinal—approach. We are largely limited to what we can obtain in the course of a single-visit examination, and for some studies a series of followup examinations would be necessary.

Another restriction arises from the fact that certain questions and procedures are impractical or even taboo for a survey carried out by a Government agency. Also, our sequencing precludes study of conditions with strong seasonal patterns.

Finally, in enumerating limitations of the



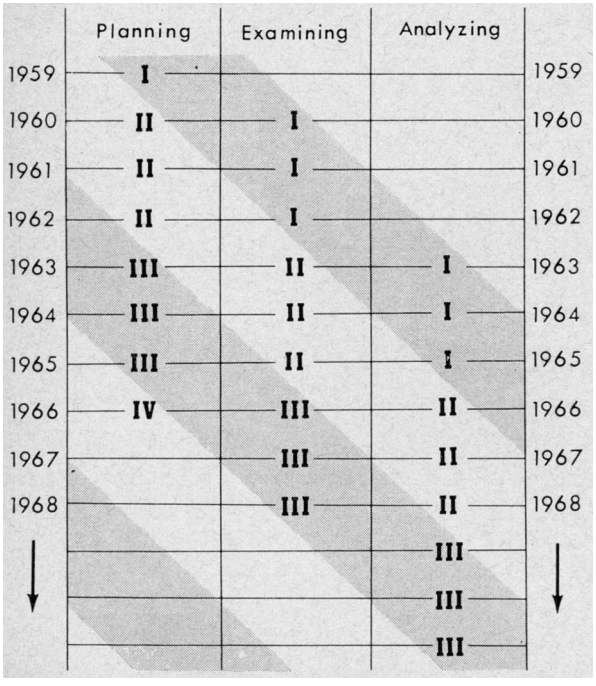
Floor plan, mobile examination center

survey, there is the "gang aft a'gley" factor. As Robert Burns implied, the execution of any survey plan, including the sampling design and the examination protocol, is likely to fall short of perfection. Our survey, of course, is not unique in having this limitation.

Survey Operation

Given the mission of collecting data on the health of the U.S. population by means of direct examinations of samples of persons, how does one proceed? The following briefly describes the plan of operation that has evolved during the past 5 years.

Cycles. To delimit this mission with respect to what can be done at a particular time, we regard the Health Examination Survey as consisting of a series of consecutive programs, each one with a specific set of goals. We refer to these successive programs as "cycles." For a particular cycle we are concerned with some segment of the total U.S. population and with some specified aspects of the health of that subpopulation. Thus, for example, the first cycle defines an adult population and attempts to obtain data on the prevalence of certain chronic diseases, notably specific cardiovascular diseases, arthritis and rheumatism, and diabetes. It also attempts to obtain distributions of the adult population with respect to a variety of measurements such as visual and auditory acuity, serum cholesterol level, blood pressure, height, weight, other body measurements, and other items (*2*). The second cycle examines a sample of a different subpopulation, children aged 6 through 11 years, and focuses more particularly on factors related to growth and development. The third cycle examines a sample of youth from 12 through 17 years of age. The content of the fourth cycle has not yet been determined.



Representation of three-level operation, U.S. Health Examination Survey, by phase of activity in each cycle

Measuring breathing capacity

Another facet of the Health Examination Survey is the concept of "three-level operation." By this I mean simply that in any given period the survey will be operating simultaneously on three different levels. Currently, for example, we are engaged in data analysis and report publication with respect to the first cycle. At the same time, we are collecting data (examining) in the second cycle, and we are also planning the third cycle. In 1966, we will begin the planning of the fourth cycle, examining will be for the third cycle, and analysis will be begun for the

second cycle. The diagram indicates how this works; it is somewhat idealized in that the lines of demarcation are not as precise as shown, and there is some shading from one activity into another.

The reasons for this three-level operation include the avoidance of complete dismantling of a field organization when data collection for a particular cycle has been completed. They also include advantages in the budget process, gains from increased staff specialization of function, and in the decidedly mixed blessing of deadlines which must be met in order to maintain the momentum of the complicated machine set in motion.

In its field organization, the survey uses mobile examination centers and traveling field-staff teams. One of these centers consists of several specially designed and constructed trailers, which are drawn by detachable tractors and set up side by side in a particular area. Covered passageways are connected to trailers, and a small clinic-on-wheels is established. This is the standard environment in which the examination is carried out.

Planning of cycles. The process of planning a cycle, or a health examination survey program, is far too complex to describe fully in this paper. I should like, however, to mention some of the necessary considerations in this planning.

Certain general guidelines are established by a long-range budget projection and an estimate of likely personnel and dollar resources available. Other guides are provided by the known availability of personnel and equipment and by the National Center's time-factor policies with respect to permissible lag between instituting a program and publishing the results.

At an early stage the broad target of a particular cycle, which includes the population group to be studied and the general nature of the desired objectives, is determined. Thus, our third-cycle program concerns youth aged 12 through 17 years and focuses on growth and development and on factors related to the adolescent stage. In making these determinations, we consult a great deal with various agencies and individuals concerned with health problems. A formally constituted Advisory Committee to the Surgeon General on the National Health Survey provides us with a broadly based group of expert advisers. Advice is also sought from scores of other persons.

When this broad-target determination has been made, we are little more than started on the long road of planning the cycle. Then the consultation effort is greatly expanded, and an attempt is made to identify the kinds of information that health workers feel should be collected in such a survey. We meet with various specialty groups to determine what kind of examinations are feasible in our setting. We develop criteria for inclusion, such as minimum expected levels of prevalence of a disease which our sample will be adequate to detect, time and cost factors, acceptability of examination procedures, appropriateness of the Health Examination Survey mechanism to obtain the desired data, expected reproducibility of data we can collect, and so on.

Frequently the results of such early consultation indicate that it will be necessary to carry out a methodological study to develop the required procedure or to calibrate our modified procedures. Examples include developmental studies, such as that on the single-visit cardiovascular examination carried out for us by Stamler (*3*), in preparing for the first cycle. An example of calibration studies is that which compared the use of the American Optical Company's Sight-Screener machine with

Psychological test

the Snellen type measurement (*4*). Another is the study which compared the modified glucose tolerance test, used in the first cycle, with the standard test (*5*).

Planning a cycle may also require feasibility studies of proposed procedures. It will certainly include pilot testing of individual portions of the examination and then of the total operation. These, of course, are followed by revisions, further developmental work, and more pretesting. Finally, the 2-plus years of lead time allotted to plan the program race by, and there is still some last-minute rush.

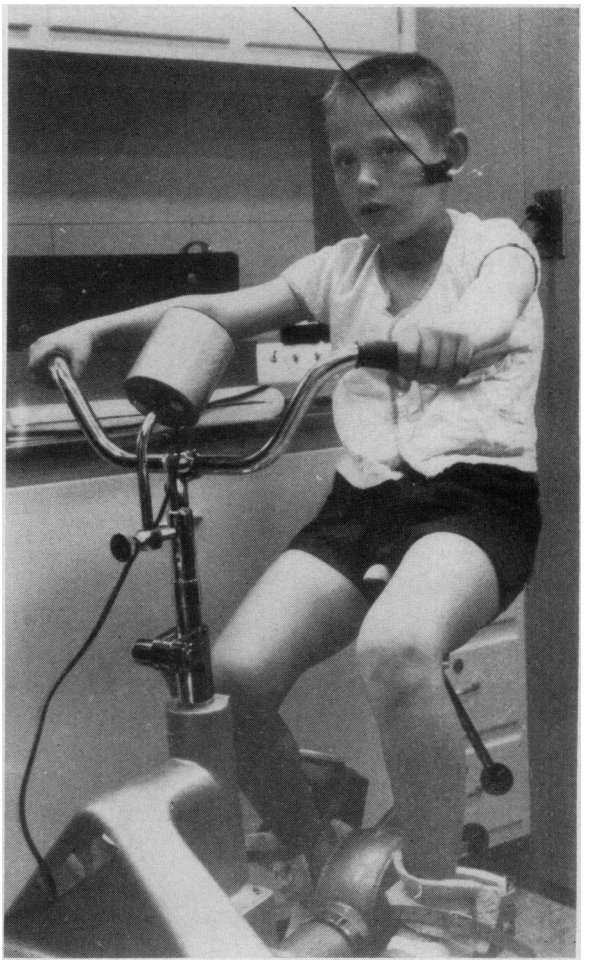
Sample design. The sample design for the Health Examination Survey has been essentially the same in each of the first three cycles, although there have been minor changes. The design is similar to that used in the Bureau of Census Current Population Survey, and this work is carried out jointly with the Census Bureau. The sample is a multistage, stratified, probability sample of loose clusters of persons in land-based population segments. The successive elements in the process of selection are primary sampling unit (PSU); census enumeration district; segment (a cluster of households); household; potential (eligible) sample persons; and finally, sample person. In the first stage of the sampling, the 1,900 areas into which the entire United States is divided are combined into about 40 strata. In this stage, there is stratification by broad geographic region, by population density groupings, and—except for the first cycle—by the percentage of change between the 1950 and 1960 censuses. Then one PSU is selected from each stratum, with a probability of selection proportional to size. A modified Goodman-Kish technique of controlled selection is used (*2*).

After some 40 PSU's are selected in the first stage, we select in each PSU a number of census enumeration districts and then, successively, segments and households. Random numbers are used in making these selections; at the same time a control is imposed to maximize distribution throughout the PSU of segments selected.

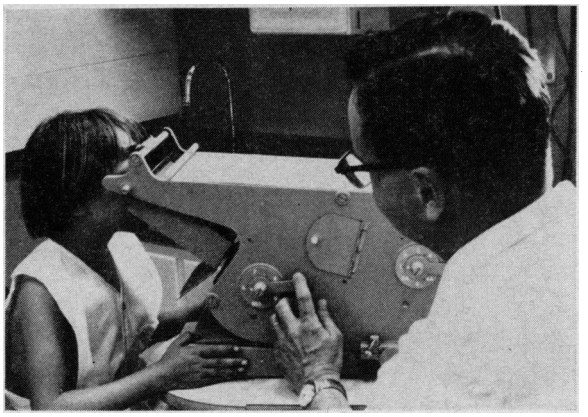
The next stage of sample selection is performed in the field through actual interview visit to every one of the identified households. During this visit certain information is collected to enable us to perform the final stage of sample selection. In the final stage all potential sample persons are listed, and then the list is systematically reduced to bring the total numbers within manageable limits.

The size of the sample is determined separately for each cycle, but for each of the first two cycles it is approximately 8,000 persons. In the sample of children aged 6 through 11, therefore, we have roughly 1,000 children in each single year of age. The sampling fraction for that cycle is about 1 in 3,000.

Response rates. At the outset of the survey one of the major concerns was the probable response rate. This program is, of course, voluntary, but does not use volunteers. If too high a proportion of the carefully selected sample was unwilling to cooperate in a health examination survey, the findings from the examinations of the remainder might be seriously biased.



Measuring heart rate after exercise



Measuring visual acuity

The experience of other examination surveys was not reassuring. Studies of the Commission on Chronic Diseases in Baltimore (*6*) and in Hunterdon County, N.J. (*7*) had succeeded in examining only about two-thirds of their samples. The experience of Chen and Cobb in Pittsburgh (*8*) had been similar. It was feared that the U.S. National Health Examination Survey might experience about the same one-third level of nonresponse, and this might seriously affect the representativeness of the findings.

Currently, however, with a completed first-cycle field operation and with second-cycle examinations nearly three-fourths complete we feel more secure in this particular area. In the first cycle we were successful in examining about 86 percent of the adults who were identified in the sample. In the second cycle the picture is even brighter. We have examined more than 96 percent of the sample children at the locations visited so far, which indicates that a parent would rather have his child examined than himself or, more frequently, herself.

The figures I quote for response rates are means for all sample children at all locations. There is considerable variation from one PSU to another. For the adult program, the range is from about 66 to 98 percent, while for the children aged 6 through 11 years the range is much narrower—from 90 to 100 percent. We have actually examined 100 percent of the sample at three different locations in the children's program.

There is an interesting differential in the first-cycle response rates by PSU, with response related inversely to population density. The response rates in rural areas were mainly in the 90–95 percent range, while most of those in the great metropolitan areas were in the 70's. We also have some information concerning the nonexamined sample persons. The operation provided that we get considerable questionnaire information from the households on the initial visit, and we have this on nearly everyone. Then too, we sent inquiries to the physicians of large numbers of the nonexamined sample persons and likewise to matched samples of examined persons. The analysis of the available data does not reveal differences between examined and nonexamined which would be expected to bias the findings.

Measurement error. If in the Health Examination Survey we have grown less concerned about bias from nonresponse, it does not follow that we have no other concerns. Among other things, we have grown steadily more concerned about measurement error.

Our attack on nonsampling variability is fourfold. First we try to minimize it. We would avoid it completely if we could, and we sometimes do, but only when we decide not to perform a particular procedure or collect a specific datum because the measurement error in it would vitiate its value. Usually, however, we try first of all to minimize the measurement error. Second, we try to monitor it, to be aware of it during the process—not only long afterwards. This is done in part to further minimize the measurement error but also in part to try to measure it—the third attack. Finally, we try to account for the measurement error in the analysis and publication of findings of the examination.

The Health Examination Survey staff is concerned with measurement error in each phase of operation: in planning a specific examination or procedure, in conducting it, and in analyzing its results. We develop a highly standardized examining procedure and, of course, provide detailed written operating instructions and carry out training of staff. Even the decision to have many or few examiners is based, in part, on consideration of measurement error. We stress the process of recording data and emphasize the absolute necessity for uniformity, accuracy, legibility, and completeness. We try

to mechanize the recording process whenever possible, and make use of such devices as tape recorders, automatic printing of results, and photographic recording of scale reading to minimize recording error.

Another important advantage to the use of such instrumentation is in measuring the measurement error. Statisticians would sometimes like to be able to answer the question of how much measurement error there is in the survey by complete replications of the entire measurement process. To do this completely is obviously out of the question. However, when we can produce a hard document, such as a tape recording of heart sound, an X-ray film, an electrocardiographic tracing, or the like, we can then obtain replicate readings and interpretations and not only reduce the error that would otherwise be present, but also have some measure of reliability of the readings.

We do a good bit of this replicate interpretation of hard documents. In addition we are able, to some extent, to make use of randomization of assignments to examiners in order to get information on interexaminer differences. In addition to this, we do a limited amount of actual replication of the examination procedure, both in specific parts of the examination and even occasionally with the entire examination.

I will only mention the problem of estimation and the related problems of measuring the variability of the data being analyzed. Our complex survey design results in some special difficulties in this area. We are engaged in some methodological research and are receiving assistance from a number of experts. At present, we are using a half-sample replication technique in the computer tabulations of data, and we obtain estimates of variance in the same runs that give us the findings. The steps in the estimation process also include use of ratio estimation and poststratification techniques.

Relationships with other workers. The many relationships which the Health Examination Survey has with the statistical and health scientific communities have been suggested in the description of the program planning. We also have many close working relationships through methodological research contracts, and we have contracted with a number of scientists to collaborate with us in analysis and publication of specific segments of data from the survey. Hopefully, many of the survey findings from our examinations and methodological studies, all of which will be published and distributed, will prove useful and will repay, to some extent, the scientific workers for their generous assistance.

REFERENCES:

(*1*) U.S. National Health Survey: Origin, program, and operation of the U.S. National Health Survey. PHS Publication No. 1000, Series 1, No. 1. U.S. Government Printing Office, Washington, D.C., 1963.

(*2*) U.S. National Health Survey: Plan and initial program of the Health Examination Survey. PHS Publication No. 584–A4. U.S. Government Printing Office, Washington, D.C., 1962.

(*3*) U.S. National Health Survey: Evaluation of a single-visit cardiovascular examination. PHS Publication No. 584–D7. U.S. Government Printing Office, Washington, D.C., 1961.

(*4*) U.S. National Health Survey: Comparison of two-vision testing devices. PHS Publication No. 1000. Series 2, No. 1. U.S. Government Printing Office, Washington, D.C., 1963.

(*5*) U.S. National Health Survey: The one-hour oral glucose tolerance test. PHS Publication No. 1000, Series 2, No. 3. U.S. Government Printing Office, Washington, D.C., 1963.

(*6*) Commission on Chronic Illness: Chronic illness in a large city: The Baltimore study. *In* Chronic illness in the United States. Vol. 4. Harvard University Press, Cambridge, Mass., 1957.

(*7*) Commission on Chronic Illness: Chronic illness in a rural area: The Hunterdon study. *In* Chronic illness in the United States. Vol. 3. Harvard University Press, Cambridge, Mass., 1959.

(*8*) Chen, E., and Cobb, S.: Further study of the nonparticipation problems in a morbidity survey involving clinical examination. J Chronic Dis 7: 321–331 (1958).