

Current Methods of Collecting Statistics Of Health and Health Problems

By O. K. SAGEN, Ph.D.

The methods of collecting statistics relating to health and health problems are as multifarious as the subject of health itself. The extent and variety of health statistics are limited only by the resources available and the imaginations of those in this work. Although it is impossible to describe all statistical collection methods relating to health, an insight can be gained into what is going on and what may be done by classifying and outlining the principal sources of health statistics.

First, agreement is needed on what is meant by health and health statistics. The word "health" according to the dictionary means "state of being hale or sound in body, mind, or soul; especially freedom from physical disease or pain." The aim of public health efforts is to bring this salubrious state to the general population. The efficacy of the public health effort is evaluated by progress toward that goal. The role of public health statistics is to describe the current state of health, measure its improvement or deterioration, point out salient problems, and provide analytical material for evaluating and guiding the public health effort. Although this definition may seem excessively broad, it is necessary to adopt such a concept to encompass the field of the present-day public health effort and its associated statistical collection methods.

Dr. Sagen is chief statistician at the Illinois Department of Public Health, Springfield, Ill. This paper was presented before the Second Conference on Public Health Statistics of the School of Public Health of the University of Michigan at Ann Arbor, June 16.

Since the health of the individual is affected by his social and economic environment, public health statistics must include a considerable amount of information on the social and economic conditions of the people, as well as the biological aspects of their sickness and health. For this reason, the public health statistician needs to keep in close touch with, and have available, statistics on population, housing, employment, welfare, and related social conditions. These data are not collected and compiled by the public health statistician except as portions of social statistics draw upon data collected in health work, such as the vital statistics of birth, marriage, and death. The public health statistician has the responsibility of utilizing the statistics in these related fields by interpreting and applying them to health problems.

Methods of Collecting Data

There are three distinct ways in which data for health statistics originate: (a) statistics compiled from returns and reports required by law, principally those of birth, marriage, death, and reportable diseases; (b) statistics derived from program activities and operation of health facilities; and (c) survey statistics from which data are gathered purely for the sake of the resulting statistical information. Also, some collections of data may represent a loose combination of types, for example, mobile X-ray screening in a tuberculosis program wherein a survey for statistical data is combined with a case-finding device in the operation of an antituberculosis program.

Reporting by Law

Collection of statistics required by law has the advantage of legal authority to compel reporting, but it also has certain disadvantages for improving the administrative mechanism and changing the content of reporting. For example, the fact that the registration of births and deaths in many local areas is not in the hands of the health department is an administrative disadvantage. Even one State health department does not have birth and death registration responsibility. The disadvantage of legal specifications for statistical reporting is illustrated by the difficulty encountered with fetal death statistics. In order to extend the reporting to include all products of conception, it becomes necessary to amend the laws in almost every State.

In many instances, statistics originate in official licensure and regulation of various enterprises and occupations. The licensure of medical practitioners, nurses, and other health service occupations affords some statistical information on manpower resources in the field of health. However, more useful information arises from the licensure and regulatory activities over hospitals, nursing homes, diagnostic laboratories, and other institutions from which indications of illness can be obtained through required reporting. In the field of environmental health the licensure and regulatory function is the key to most of the information on the problems of environment. Notable examples are regulation of industrial health hazards; waste disposal; water, milk and food supplies; and housing sanitation. These services are often combined with a health service program, with the service aspect strengthened by the entree afforded by police power, while the law enforcement is facilitated by the persuasive aspects of useful service gratefully received.

Program Data

The proliferation of health services, facilities, and programs in recent years has given emphasis to the development of health statistics and to routine reporting of the statistical data. The administration of the services requires systematic record keeping which in turn generates statistical data. Compilation and analysis of

these statistics become essential in administering, evaluating, and operating the projects, and here the official and nonofficial agencies share common problems. An example is the tuberculosis X-ray screening program in which the voluntary tuberculosis association conducts surveys side by side with the official health department. In this situation the statistics of the activity assume added importance because they provide the basis for coordinating twin efforts which cannot be allowed to clash or to duplicate and overlap each other. In other cases, the voluntary agency will operate the screening venture and the official agency will do the follow-up. Here again, the statistics provide an indispensable link in keeping the project in alignment and in balance.

The operation of health services—to people, to lower animals, or to environmental elements—brings out two competing demands for statistics. On one side is the need for statistics in the administrative work of planning and operating the projects, while on the other the so-called clinical or research statistics are needed to evaluate methodology and reveal new knowledge on the subject of the service. Since competent statistical work is expensive, the resources often are not sufficient to provide both kinds of statistics for the project, and a large amount of data arising out of health service operations is collected but not used, or is improperly handled. Another factor is the shortage of biometricians for work in the health sciences. Despite these limitations important statistical work is being done in the various health services, and utilization of the statistics is on the increase.

The phrase "health services" is used here to embrace the whole area of preventing, mitigating, and remedying disease, disability, and injury. Even this is not sufficient when we think of sources of health statistics obtainable from various insurance plans designed to mitigate economic loss to individuals due to sickness, disability, or death. Life insurance statistics for a long time furnished about the only source of consistent information covering illness and disability for a large part of the population. The experience of the hospitalization insurance and medical care plans has now been added to this source. In these activities, the gathering

of health statistics comes as an essential by-product of the operating records for the enterprise.

As stated earlier, the salient characteristic of the statistics from health services is that the statistical data are derived by a systematic handling and processing of the records used in operating the service. Good records are essential to efficient operation, as well as to reliable statistics. This becomes apparent the larger the service grows and the longer it operates. Health departments have encountered this fact in their experience with case registries for tuberculosis and cancer. However, we must keep in mind that there is a tendency toward over-elaborate statistics from clinical data. A mass of unmanageable data results from trying to assemble every scrap of information. Clinical data are best utilized for specific research studies, with limited objectives carefully planned in advance of collecting the information.

The majority of health statistics arise from cross-sectional data on conditions as they exist at a particular time. Cross-section statistics fail to reveal what goes on in the never-ceasing process of change in people and in their conditions. Studies based on longitudinal data fill that gap by following the course of events in the same group of individuals through repeated observations over a period of time. Fruitful sources of longitudinal data are the health services given for chronic diseases such as cancer, for chronic or recurrent infections such as tuberculosis, venereal disease, and rheumatic fever, and for impairments such as those due to poliomyelitis, premature birth, and dental caries. Also, of equal if not of greater importance are longitudinal studies on persons who are apparently well at the outset. Compared to cross-sectional studies, longitudinal studies are costly if they are made only to complete follow-up. Indiscriminate enthusiasm for longitudinal studies has stimulated too many duplications of effort. For example, it should not be necessary to conduct identical follow-up studies in all cancer programs over the country, yet at one time this procedure was seriously advocated for cancer registries.

Emphasis is now given to obtaining statistical summaries of services provided by health departments, particularly in local health de-

partments which want to know what each of their programs is doing. Many local health departments are keeping count of nurses' visits, of the number of children served by well-child conferences and in school health programs, the number of immunizations given, sanitary inspections and services rendered, and other activities of the local health department. Administrative data of this kind have appeal but are not as useful as they should be since the attempt has been to cover too much territory. The problem of how to use such data effectively in evaluating local health programs is undergoing investigation.

The mass of statistics that can arise from conducting health services and programs makes it important that health agencies exercise discrimination in planning statistical work. The primary necessity is to plan how much statistical work will be needed to operate the program effectively and to furnish a simple evaluation of its effectiveness. A decision is needed on what and how much research is to accompany the program and on what the potentialities are for learning something that is both significant and new. This is the joint responsibility of the program director and the health statistician.

Survey Statistics

The survey method of obtaining statistics is the oldest of the three approaches discussed here. In the United States the earliest comprehensive survey was the Federal census of population in 1790. The object of the survey method is to obtain cross-sectional information on the distribution of certain carefully defined characteristics in a population, which may be animate or inanimate, human, or any other species of fauna or flora. For a long time the survey method was restricted to a complete enumeration of the population under study. As a result of the development of statistical theory, we now have techniques for obtaining valid information characterizing an entire population by enumerating only a part of it. These statistical sampling methods are advantageous because of the relatively low cost at which reliable facts can be learned about a large population. The technique also furnishes information on the degree of precision which

can be safely attached to the statistical results. Extending the use of sampling surveys will open a veritable gold mine of knowledge on health conditions. So far, the most notable health study of this kind is the National Health Survey of 1935-36.

The various mass-screening techniques for tuberculosis, heart disease, diabetes, and syphilis; for dental, hearing, and eyesight defects; and for other diseases conducted individually or in combinations as "multiphasic" screening, provide challenging methods for making observations of large groups of the population. The principal object of screening has been case finding and health education so that early diagnosis can lead to early remedial treatment for as many as possible. Statistically, the value of findings in such surveys has not been as useful as desired because the screening applications have been only incidentally designed to yield statistics. If screening projects were conducted so that statistical sampling techniques could be applied, the data would yield information on distribution of the defects surveyed in the general population. Nevertheless, valuable data have been assembled through these screening processes.

It would be futile in anything short of a textbook even to list all applications of survey methods in health statistics. The Public Health Service has perhaps done more of this work than all other health agencies combined. Perusal of a file of *Public Health Reports* for recent years will provide a fairly comprehensive bibliography of the various situations to which survey techniques are being applied.

Quantity and Quality

The sources of statistics sketched are sufficiently comprehensive to indicate that there is no dearth of data on health as far as quantity is concerned. Current methods of collecting health statistics tend to emphasize quantity above quality. Part of this emphasis is due to the conditions which health statistics are called upon to describe and the large populations among which such conditions exist. An unfortunate circumstance is that statistics too often connote mere aggregates of numbers in

tables in which one can search for and find an interesting numerical "fact." Throughout the country health statistics as a whole still suffer from much poor handling of data. Insufficient precautions in assembling data, improper processing, and unskilled analysis produce unusable and misleading "statistics." Year by year we are seeing improvement as the users of health statistics become better informed and as the statistical personnel in health work become more proficient.

State-Federal cooperation is doing much to improve the quality of State and national vital statistics. Current vital data are of a higher and more uniform quality throughout the entire country than ever before in history. The striking increase in the completeness of birth registration for the last decade bears witness to some of this improvement (1).

Health statisticians generally are getting sharper tools and learning how to use them. Great strides have been made in obtaining more uniform and precise methods of classifying original data, which is the first and most critical step in compiling statistics. The general adoption of good classification standards means that our health statistics all speak the same language. The greater use of mechanical equipment and the development of new machines expressly for processing statistical data have shortened the processing time, brought about greater numerical accuracy, and permitted statistical analyses which hitherto have involved too much labor.

The mechanical aids and improved compilation methods do not reduce the need for solid thinking in producing statistics. The object of the entire statistical process is to produce useful conclusions on health problems, conclusions that are valid and simply expressed. This represents an intellectual exercise that cannot be avoided. Current statistical practice applied to health data needs to give greater recognition to this principle if it is to utilize fully and properly the material already available in such abundance.

REFERENCE

- (1) Shapiro, Sam, and Schachter, Joseph: Birth registration completeness, United States, 1950. *Pub. Health Rep.* 67: 513-524 (1952).