

III. STATISTICS FOR THESIS TOPICS FROM THE NATIONAL CENTER FOR HEALTH STATISTICS

THEODORE D. WOOLSEY

I have been asked to report on what is available in both published and unpublished official statistics from the National Center for Health Statistics for use in doctoral dissertations and masters' theses. This assignment covers a lot of ground, and perhaps the best way to start is by giving a thumbnail sketch of the various data collection programs of the Center.

First, and best known, is the program of national vital statistics. The records of births, deaths, fetal deaths, marriages, and divorces are, and will be, social and health indicators of enormous importance. They signal the milestones of life and are the main determinants of population growth.

By arrangement with the state health department or other responsible agency the Center purchases from each jurisdiction copies, usually on microfilm, of the vital records, and codes and punches the information, and from the punch cards produces magnetic tape for input to a computer.

For economy and speed a 50 per cent sample of the birth records is selected within the National Center for Health Statistics before the processing starts, but all the death certificates are used. Not all states have central collections of marriage and divorce records. For

this and other reasons the national collection of these records is not complete. Marriage records are obtained from 35 states and divorce records from 21. This is a situation our Division of Vital Statistics is trying hard to correct, but there are many obstacles.

In addition to the collection of data from the registration of vital records, the Center has a program of health surveys, authorized by the National Health Survey Act of 1956. These surveys are of four different types:

1. A continuing national survey to collect facts about morbidity, disability, and use of health services and facilities, by means of interviews in samples of households. This survey emphasizes the social and economic dimensions of health.
2. A series of surveys to collect physiological and clinical information from national samples of the general population by examining samples of people. This survey stresses the medical and dental aspects of health.
3. Surveys to collect facts from establishments providing hospital, medical, nursing, or personal care, about the characteristics of their patients or the establishments themselves.
4. Surveys based on the vital records as sampling frames. These include at the present time a survey to obtain more information about decedents by querying the relatives and the institutions in which medical care was received prior to death, and a methodologically similar survey of infants.

What kinds of data can be derived from these data collection activities? While this may be apparent from the nature of the systems, it might be well to enlarge on the output aspects.

The vital records analysis produces statistics on fertility, mortality (including, of course, causes of death), the rate of natural increase, life tables, and rates and characteristics of marriages and divorces.

From the surveys we obtain data on incidence and prevalence of diseases, injuries, and impairments, a number of measures of disability, utilization of medical and dental services, hospitalization, health insurance coverage, personal health expenditures; also, physiological characteristics, anthropometric characteristics, distributions

of clinical variables; number and types of institutions, medical and nursing characteristics of the institutional population, and detailed information about patients discharged from hospitals.

The vital records surveys produce details of the births and deaths beyond what is found on the certificates themselves. For example, the amount of hospital care used in the last year of life is determined for deaths in the sample, and future birth expectations will be sought for mothers in a sample of births.

All of the types of information outlined above might be thought of as dependent variable data. There are also many independent variables that are dealt with—the various types of classifying characteristics.

On the vital records these variables and attributes are limited, but two important ones are available which are not, or are not yet, available from the surveys. These are: geographic detail and trends over time. Furthermore, as I have pointed out, the vital records surveys permit great extension of the information for classifying the births and deaths. In these surveys and in the establishment surveys, however, the corresponding classification of the exposed-to-risk population must be obtained from available census estimates or independent population surveys.

The possibilities for classifying variables in the interview and examination surveys are limited only by the size of the sample and the ability of the respondents to provide reliable information. In the current forms of these surveys the variables are of three types: demographic, socio-economic, and environmental. I shall not list them in detail.

So much for the kinds of data we produce. Now, what about their availability to students?

A great deal of the information is circulated. We try to publish as much of the data as we can and as much as we think is worth using. However, as in every statistical agency, there are also unpublished data in our files.

In the case of vital statistics, there tends to be a large amount of unpublished tabulated data, particularly for mortality. On the other hand, in the surveys there is less, and what is available is in

less detail because of sample size limitations. We also have magnetic tape and, for a limited time, the basic records.

Of particular concern in this panel discussion is the matter of access to unpublished data by students. Let me state some of our policies concerning this:

In the first place, the basic records are available only to employees of the Center or persons working for the Center under contract. This is primarily because of the confidential nature of the records. The only exception to this is that vital records are sometimes made available with state approval to official agencies or established private research groups.

As for punch cards or magnetic tape, these are only infrequently turned over to nonemployees in the case of vital statistics. Exceptions are made for persons in the United States Public Health Service or in the Department of Health, Education, and Welfare, by special arrangement. In the case of the survey data the policy is even more restrictive because of the added complexities introduced by the sampling process.

Unpublished tabulations as they come from the computer are available by special arrangement for approved projects but usually only for persons working at the Center. Unpublished tables derived from the tabulations, on the other hand, are frequently made available to outside workers. Of course, the amount of time that can be spent in preparing such tables is limited by the availability of staff. The same holds true for specially prepared computer runs. Special tabulations of vital statistics are occasionally made, with reimbursement for cost.

We have found that there is a strong relationship between the stage of preparation of our reports and the amount of staff time required to help other people in their use of the data; that is, when data are turned over to others at an early stage in the process, the time required for our staff to explain them is likely to be quite considerable. For example, giving outside workers magnetic tape from the surveys involves so much programmer time that we practically never release it.

On the other hand, statistical results ready for publication have had explanatory notes, definitions, sampling errors, etc., prepared in writing; hence little additional staff time is required.

We feel that we should not divert a great deal of staff time for *individual* users when our highest priority job is to produce the data in published form as rapidly as possible for the use of *all* consumers.

This does not mean that we do not provide unpublished data for people interested in research or special estimates. It does mean that we try not to let this interfere with our major responsibility—the production of statistical reports. It also means that we have not been able to make any special arrangements to help students in search of thesis materials, though we are always happy to see the data used in this way.

Perhaps it seems that we are not doing all we should to assist students. I would be most interested to have comments.