

RESEARCH USES OF VITAL RECORDS IN VITAL STATISTICS SURVEYS

MONROE G. SIRKEN

INTRODUCTION

IT is not an easy matter to keep well informed about recent technological developments involving research applications of vital records. One factor is the substantial increase in the research uses that have been made of these records in recent years. Another factor is that, customarily, matters pertaining to technology and methodology get considerably less attention than is given to the substantive findings of research projects. The net result is that the research potential of vital records, in terms of existing methodology and technology, is not being fully exploited.

The existence of a typology or index of research uses that have been made of vital records and a bibliography of published reports describing research applications would probably help to disseminate information about technical developments in this field. Preparation of these reference materials is, however, a task beyond the scope of this paper. Rather, this presentation is limited to some major types of research applica-

This article is a revision of a paper presented at the 90th Annual Meeting of the American Public Health Association, Miami, Florida, October 16, 1962.

tions of statistics that have been derived from vital statistics surveys conducted by the National Vital Statistics Division (NVSD).

First, it is necessary to briefly summarize the scope, objectives and methodology of the NVSD program for vital statistics surveys. Subsequently, research applications which have been made of statistics derived from these surveys will be discussed.

SURVEY PROGRAM OF THE NVSD

The basic objective of the NVSD survey program is to produce vital statistics which will contribute to the fund of basic data required by the National Center for Health Statistics in its program for developing a national intelligence system with respect to the character and dynamics of health problems, their trends and implications.

The program for vital statistics surveys originated about six years ago. The survey program was conceived as a stage in the evolutionary development of the registration system for producing national vital statistics. (9, 10, 11, 12) The surveys are linked to the registration system and seek to supplement the limited information reported about vital events on vital records. The basic methodology of the survey program at the present is the "followback survey linked to vital records." Three essential features of the methodology are:

1. Selection of a probability sample of vital events from files of vital records or from files of punch cards representing these records.
2. Conduct of surveys in which the mail questionnaire is the principal data collection method and persons and institutions identified on the vital records are the primary sources of information.
3. Derivation of unbiased national and regional statistics and estimation of the sampling and nonsampling errors of these statistics.

Until fairly recently, there were two parts to the NVSD vital

SIRKEN

statistics survey program. One part involved conducting the *ad hoc* surveys on a contractual basis for agencies both within and outside the Department of Health, Education, and Welfare. In the other part of the program, developmental studies were undertaken in order to improve the survey technology of the existing program and to explore and develop new survey methodology.

In 1961, the scope of the survey program was expanded to include a national mortality survey based on a sample of one out of every 330 registered deaths in the nation, or about 5,000 deaths annually. The survey program will be expanded again in 1963 to include a national natality sample survey. This survey will be based on a sample of one out of every 1,000 registered births, or about 4,300 births annually. The sample designs of both surveys will not change from year to year, and in this sense, the surveys will be continuing, but the content of the surveys will be developed on a topical basis and periodically the survey topic will be changed. Essentially, these two continuing surveys will serve as the mechanisms for conducting a series of special studies so that over the years a wide variety of national multipurpose mortality and natality statistics will be collected (8).

USES OF SURVEY DATA

The national mortality and natality surveys will enlarge the scope of vital statistics by supplementing items of information reported on the vital records with the information about the vital event which have never been obtained and most likely could not be obtained on the vital records themselves. This is probably the most important use of the survey data. The 1962 national mortality survey, for example, will produce estimates of differential utilization of hospitals and institutions in the last year of life according to socio-economic status of the decedent. Data collected in these surveys have served also as a basis for evaluating the quality of the information reported

QUARTERLY

on the vital records. This is another important use of the survey data. For example, the results of diagnostic tests reported by physicians who certified death records in the Pennsylvania Mortality Survey produced statistics on the quality of diagnostic information supporting the causes of death. (7)

In some applications, the survey data provide only part of the information needed in the research project, and the estimates derived from the vital statistics surveys are combined with estimates derived from other sources. These uses of survey data are less well known than those mentioned above. Research problems have arisen which necessitated combining the estimates derived from mortality surveys conducted by the NVSD with estimates derived from another source. The other source has been either a population survey, a health survey, or a record matching study. Each of these is briefly described below.

POPULATION SURVEYS

Estimation of specific death rates for the general population presents a difficult problem when the mortality data are not available from the death records and the population data are not available from the census. Recently, a sample survey methodology was developed to cope with this problem. The methodology, which produces unbiased estimates of specific mortality rates, involves conducting two sample surveys. One of these, a survey linked to the death record, provides unbiased estimates of the numerators of the rates. The other, a survey of the corresponding population at risk, provides unbiased estimates of the denominators of the rates.

This "dual sampling" methodology was applied in a national lung cancer epidemiological study (4) sponsored by the National Cancer Institute. The objective of the study was to obtain national estimates of lung cancer death rates specific for age, smoking habits and residence history. The mortality estimates, required for the numerators of these rates, were produced by a national lung cancer mortality survey in which

information pertaining to cigarette smoking habits and residence histories were collected from family informants for a national sample of lung cancer deaths that occurred during 1958. To obtain the estimates required for the denominators of the lung cancer mortality rates, arrangements were made with the Bureau of the Census to obtain comparable information on smoking habits and lifetime residences for a sample of the national population included in the Current Population Survey. (1)

HEALTH SURVEYS

Currently, national estimates of the annual volume of hospital utilization are based on data obtained in household interviews conducted weekly by the Health Interview Survey (HIS) of the National Health Survey Division. (13) The interviews obtain information on hospital utilization by each resident member of the household during a reference period of specified duration preceding the week of the interview. The household interviews do not, however, obtain information on hospital utilization by the persons who would have been residents of the household except for the fact that they died during the reference period. Thus, the estimates derived from the health survey relate only to that portion of the population at risk that survived the reference period and were alive at the time of the interview. To obtain estimates of the volume of hospital utilization by the entire population, the estimates derived from the HIS have to be supplemented by estimates of hospital utilization by persons who died during the reference period.

In 1958, the National Health Survey Program arranged for the NVSD to develop and test a method of estimation and data collection to fill the gap in hospitalization data. The procedure developed makes use of information about hospital utilization in the last year of life. (14) A method for collecting these data by means of surveys linked to the death record was demonstrated in a pilot study conducted in the Middle Atlantic

States and, subsequently, the methodology was applied in the 1961 National Mortality Survey.

RECORD MATCHING STUDIES

By matching death certificates for persons who died after the Decennial Census to their census enumeration schedules, the information about the decedent reported on the death record is supplemented by the demographic and socioeconomic information about the decedent and other persons in his household collected in the census. The major technical difficulty with this methodology for expanding mortality statistics has been the limited success of the matching operation. Past efforts (3, 5) at matching death and census records have been uniformly unsuccessful in locating the census schedule for about one-fifth of the decedents. Lack of census information for a proportion of decedents represented by a nonmatch rate of this size would produce potentially biased results that could scarcely be tolerated in most studies.

In a recent study of social and economic differentials in mortality (6) that involved matching 340,000 death certificates for persons who died in the 4-month period May to August, 1960 with the 1960 census schedules, the matching problem was resolved by conducting a mortality survey in which the census information was collected for a subsample of about 10,000 decedents. The mortality survey was conducted in advance of the operation for matching death certificates with census schedules in order to avoid the risks of delaying data collection operations. Consequently, of the 10,000 decedents in the survey, it is expected that for about one-fifth, or 2,000 decedents, the death certificates will not be matched with census schedules; and for about four-fifths, the match will be successful. The survey information collected for the former group will provide estimates of the census characteristics of decedents whose death certificates will not be matched with census schedules. The survey information collected for the latter group will be compared with the respective information

reported on the census schedules, thus providing a statistical basis for assessing the validity of the estimating procedure making use of survey data in lieu of census data for the unmatched cases.

CONCLUSIONS

The effective dissemination of today's flood of technical information presents a serious problem in virtually all branches of science and technology. (12) There is a need, I believe, for more effective dissemination of information concerning recent technological and methodological developments involving research applications of vital records. In this connection, several types of research applications of statistics derived from sample surveys linked to death records were described in this report.

Undoubtedly, more effective dissemination of technical information on research uses of vital records would further stimulate methodological developments along these lines, and I believe that these innovations are long overdue. Recent developments, particularly in data collection methods, in other fields of health statistics appear to have surpassed those in vital statistics. This, I believe, is one of the reasons why vital statistics have lagged behind other types of health statistics in recent years. Thus, whereas a few years ago, mortality statistics were virtually the only regularly available national morbidity statistics, at the present time more is known about the morbidity and medical care experience of the general population than is known about the terminal experience of the population prior to death.

The initial objective in developing the methodology referred to in this report as "surveys linked to the death record" was to expand national mortality statistics to include the terminal morbidity and medical care experience and other related health and behavioral experiences that are not reported on the death certificate. The other types of research applications of data derived from these surveys, presented in this report, represent unexpected dividends on the original investment in methodological research.

QUARTERLY
REFERENCES

1. Bureau of the Census: CONCEPTS AND METHODS USED IN THE CURRENT EMPLOYMENT AND UNEMPLOYMENT STATISTICS PREPARED BY THE BUREAU OF THE CENSUS. [Series P-23, No. 5], May, 1958.
2. Gray, Dwight E.: Information and Research—Blood Relatives or In-laws? *Science*, July 1962, 137: 263–266.
3. Guralnick, Lillian, and Nam, Charles B.: Census-NOVS Study of Death Certificates Matched to Census Records. The Milbank Memorial Fund *Quarterly*, April, 1959, xxxvii, (2): 144–153.
4. Haenszel, William, Loveland, Donald B. and Sirken, Monroe G.: Lung-Cancer Mortality as Related to Residence and Smoking Histories. I. White Males. *Journal of the National Cancer Institute*, April 1962, 28(4): 947–1001.
5. Kaplan, David L., Parkhurst, Elizabeth and Whelpton, Pascal K.: The Comparability of Reports on Occupation From Vital Records and the 1950 Census. *Vital Statistics—Special Reports*, June 1961, 53(1).
6. Kitagawa, Evelyn M., and Hauser, Philip M.: Methods Used in a Current Study of Social and Economic Differentials in Mortality in EMERGING TECHNIQUES IN POPULATION RESEARCH. New York, Milbank Memorial Fund, 1963, pp.
7. Moriyama, Iwao M., Baum, William S., Haenszel, William M., and Mattison, Berwyn F.: Inquiry into Diagnostic Evidence Supporting Medical Certifications of Death. *American Journal of Public Health*, October, 1958, 48(10): 1376–1387.
8. National Vital Statistics Division: Sample Survey Program of the National Vital Statistics Division. *Proceedings of the Ninth Conference on Records and Statistics*, June 1962. Washington, D. C.
9. Sirken Monroe G., and Brown, Morton L.: Quality of Data Elicited by Successive Mailings in Mail Surveys. *Proceedings of the Social Statistics Section*, 1962, American Statistical Association. [To be published.]
10. Sirken, Monroe G. and Dunn, Halbert L.: Expanding and Improving Vital Statistics, *Public Health Reports*, June 1958, 73(6): 537–540.
11. Sirken, Monroe G., Pifer, James W. and Brown, Morton L.: Design of Surveys Linked to Death Records. National Vital Statistics Division, Washington, D. C., September 1962.
12. —————: Survey Procedures for Supplementing Mortality Statistics, *American Journal of Public Health*, November 1960, 50(11): 1753–1764.
13. U.S. National Health Survey: Hospital Discharges and Length of Stay: Short-stay Hospitals, United States, 1958–1960. *Health Statistics*, [Series B-No. 32], April, 1962.
14. U.S. National Health Survey: Hospital Utilization in the Last Year of Life. *Health Statistics*, [Series D-No. 3], January, 1961.