

VITAL and HEALTH STATISTICS
DOCUMENTS AND COMMITTEE REPORTS

Use of Vital and Health Records in Epidemiologic Research

A Report of the United States National Committee on Vital and Health Statistics

An analysis of the changing needs for data in epidemiologic research in relation to the present vital and health statistics system with recommendations for use of existing vital and health records and for developing new data for epidemiologic studies.

DHEW Publication No. (HSM) 73-1265

Washington, D. C.

March 1968

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE
John W. Gardner
Secretary

Public Health Service
William H. Stewart
Surgeon General



NATIONAL CENTER FOR HEALTH STATISTICS

THEODORE D. WOOLSEY, *Director*

PHILIP S. LAWRENCE, Sc.D., *Associate Director*

OSWALD K. SAGEN, Ph.D., *Assistant Director for Health Statistics Development*

WALT R. SIMMONS, M.A., *Assistant Director for Research and Scientific Development*

ALICE M. WATERHOUSE, M.D., *Medical Consultant*

JAMES E. KELLY, D.D.S., *Dental Advisor*

LOUIS R. STOLCIS, M.A., *Executive Officer*

DONALD GREEN, *Information Officer*

OFFICE OF HEALTH STATISTICS ANALYSIS

IWAO M. MORIYAMA, Ph.D., *Director*

DEAN E. KRUEGER, *Deputy Director*

Public Health Service Publication No. 1000-Series 4-No. 7

Library of Congress Catalog Card Number 67-62375

FOREWORD

This report, prepared under the auspices of the U.S. National Committee on Vital and Health Statistics considers ways in which vital and health statistics systems can better serve the changing need for data in epidemiologic research. These needs arise from the increasing concern with chronic noninfectious diseases and the measurement problems related to the nature of these diseases.

Recommendations are made for modifying the content of basic vital and health records, and for developing new ways of making data from these records available for epidemiologic research and other uses.

Robert L. Berg, M.D.
Chairman
U.S. National Committee on
Vital and Health Statistics

U.S. NATIONAL COMMITTEE ON VITAL AND HEALTH STATISTICS

Robert L. Berg, M.D., Chairman
Professor and Chairman, Department of
Preventive Medicine and Community Health
The University of Rochester
Medical Center
Rochester, New York

Iwao M. Moriyama, Ph.D., Executive Secretary
Director, Office of Health Statistics Analysis
National Center for Health Statistics
Public Health Service*
Washington, D.C.

Donald J. Davids
Chief, Records and Statistics Section
Colorado State Department of Public Health
Denver, Colorado

William Haenszel
Chief, Biometry Branch
National Cancer Institute
National Institutes of Health
Public Health Service*
Bethesda, Maryland

Clyde V. Kiser, Ph.D.
Senior Member, Technical Staff
Milbank Memorial Fund
40 Wall Street
New York, New York

Herbert E. Klarman, Ph.D.
Department of Public Health Administration
School of Hygiene and Public Health
The Johns Hopkins University
Baltimore, Maryland

Everett S. Lee, Ph.D.
Department of Sociology and Anthropology
University of Massachusetts
Amherst, Massachusetts

John R. Philp, M.D.
Health Officer
County of Orange Health Department
P. O. Box 355
Santa Ana, California

Donovan J. Thompson, Ph.D.
Department of Preventive Medicine
School of Medicine
University of Washington
Seattle, Washington

Theodore D. Woolsey, Ex officio
Director, National Center for Health Statistics
Public Health Service*
Washington, D.C.

SUBCOMMITTEE ON

USE OF VITAL AND HEALTH STATISTICS IN EPIDEMIOLOGIC RESEARCH

Brian MacMahon, M.D., Chairman
Professor of Epidemiology
Harvard School of Public Health
Boston, Massachusetts

Lillian Guralnick, Secretary¹
Social Science Research Analyst
Health Insurance Research Branch
Division of Health Insurance Studies
Social Security Administration*
Washington, D.C.

I. M. Moriyama, Ph.D., Ex officio
Executive Secretary
U.S. National Committee on Vital and Health
Statistics
Washington, D.C.

John Cassel, M.D., M.P.H.
Professor of Epidemiology
School of Public Health
University of North Carolina
Chapel Hill, North Carolina

Carl L. Erhardt, Sc.D., Director²
Health Intelligence Statistics
The City of New York Health Services Administration
New York, New York

Elmer A. Gardner, M.D., Director³
Community Mental Health Center
Health Sciences Center
Temple University
Philadelphia, Pennsylvania

Robert W. Miller, M.D.
Chief, Epidemiology Branch
National Cancer Institute
National Institutes of Health
Public Health Service*
Bethesda, Maryland

Donald L. Rucknagel, M.D.
Department of Human Genetics
University of Michigan Medical School
Ann Arbor, Michigan

Colin White, M.D.
Professor of Biometry
Department of Epidemiology and Public Health
School of Medicine
Yale University
New Haven, Connecticut

* Department of Health, Education, and Welfare

¹When appointed to the Subcommittee, Miss Guralnick was a Statistician, Office of Health Statistics Analysis, National Center for Health Statistics, Public Health Service, Department of Health, Education, and Welfare, Washington, D.C.

²Until July 1, 1966, Dr. Erhardt was Associate Director, Office of Research, City of New York Department of Health.

³When appointed to the Subcommittee, Dr. Gardner was Assistant Professor and Director, Division of Preventive Psychiatry, University of Rochester School of Medicine and Dentistry, Rochester, New York.

CONTENTS

	Page
Foreword -----	iii
U.S. National Committee on Vital and Health Statistics-----	iv
Subcommittee on Use of Vital and Health Statistics in Epidemiologic Research -----	iv
Introduction -----	1
Need for Review-----	2
Procedure-----	3
Findings and Recommendations-----	3
Access to Vital and Health Records-----	3
Record Linkage-----	4
The Unique Number-----	4
Ascertainment of Death-----	5
Followup-----	6
Information on Vital and Health Records-----	6
Cause of Death-----	6
Congenital Malformations-----	7
Other Items-----	8
Development of Other Resources-----	8
Summary of Recommendations-----	8
Appendix I. Selected Bibliography on Record Linkage-----	10
Appendix II. The Underlying Cause Concept-----	12

IN THIS REPORT the changing needs for data in epidemiologic research are examined in relation to the present vital and health statistics systems.

The ascendancy of chronic noninfectious diseases as major causes of morbidity and mortality requires new types of data and changes of emphasis in existing types of data. New measurement problems are often encountered in these diseases, in which onset may be insidious, progression slow, and the interval between causes and effects long.

Each source of vital and health statistics is examined in terms of the information requested on the basic record, the present means of access to the data on the record, and the tabulations now being prepared by the National Center for Health Statistics.

Linkage of various vital and health records is seen as a means of increasing their separate value, and further exploitation of this device is recommended. Use of some universal identity number for this purpose should be explored along with the possibility of assigning such numbers at birth.

A National Death Index would be a great asset in epidemiologic research. The National Center for Health Statistics is urged to explore the technical problems in establishing this resource. Extension of the scope of pathological conditions recorded on death certificates beyond those contributing to death should be considered.

Other recommendations are made for development of vital records information on congenital malformations and for inclusion of a sample of newborn infants in the Health Examination Survey.

USE OF VITAL AND HEALTH RECORDS IN EPIDEMIOLOGIC RESEARCH

"Finding some truths, and not commonly-believed opinions, to arise from my meditations upon these neglected papers (the Bills of Mortality), I proceeded further, to consider what benefit the knowledge of the same would bring to the world; that I might not engage myself in idle, and useless speculations, but ... present the world with some real fruit from those ayrie blossoms."

John Graunt, 1662

INTRODUCTION

The practice of epidemiology, or indeed of public health, would today be inconceivable without access to vital and health records and the tabulations routinely assembled from them. The data obtained by Graunt's "antient matron" searchers have been replaced by legally required certificates of birth and death. These documents, and the breadth of their application, have changed remarkably in the last 50 years, and today they probably constitute the single most important resource of epidemiologic investigation. In addition, entirely new sources of health statistics have been developed, notable among these, in the United States, being the several programs of the National Health Survey.

With respect to epidemiologic purposes, vital and health records serve the following functions:

1. As source data for the measurement of the incidence and prevalence of disease. These measurements serve administrative as well as epidemiologic purposes. Because of the limitations of diagnostic information derived from certificates of birth and death, it is to this general purpose of providing source data that most of the newly introduced types of health statistics have been addressed.

2. For comparison of disease rates in different populations, in different parts of the same population, and in similar groups over a period of time, in order to develop hypotheses regarding the etiology of disease. Studies over time may also serve to evaluate the effectiveness of preventive or therapeutic measures.
3. For identification of groups of people at high risk of particular diseases. Such groups may be of significance in disease control or they may be particularly suitable for studies of etiology and therapy.
4. As the starting point for "follow-back" studies in which a series of cases with particular characteristics (e.g., dying from a particular disease) is identified from the primary source (in this instance, the death certificate) and supplementary information is sought from other sources (e.g., the certifying physician) relevant to etiology or other subjects of interest.
5. As the end point for studies in which subsets of the population are selected because of their unusual characteristics or environmental exposures and followed to identify diseases or other outcomes suspected of being related to the selected factors.

Erhardt¹ and Moriyama² have recently compiled some examples of the ways in which vital records are being utilized for epidemiologic research. Major projects in this country that are dependent on vital records include the extensive series of studies of veterans carried on by the Follow-Up Agency of the National Academy of Sciences, National Research Council, the large-scale studies of the effects of smoking on health by the National Heart Institute and the American Cancer Society, the followup programs of registries of cancer and other specific diseases, and a great many smaller scale studies aimed at specific problems.

It is difficult to assess the extent of epidemiologic use of the statistics published in *Vital Statistics of the United States* or in the reports from the National Health Survey. Reference to any journal of epidemiology or public health will yield examples of such uses, but extensive use is also made of them by individual investigators in searching for ideas and in preliminary testing of hypotheses—activities that largely pass unrecorded.

NEED FOR REVIEW

The emphasis of epidemiologic investigation has shifted markedly in the last two decades. A decline in interest in the infectious diseases and increase in concern with the noninfectious diseases have resulted from the change in relative importance of these categories of disease in many parts of the world, including the United States. It is also recognized that, although major tasks still remain in the improvement of control over the infectious diseases, the contribution of epidemiology to the development of control methods is largely past—new advances are being made predominantly as the result of work in the experimental laboratory and through the better application of existing knowledge. On the other hand, the identification of cigarette smoking as the major cause of this century's epidemic of lung cancer has clearly demonstrated the po-

tential contribution of epidemiologic research at the present stage of knowledge of those chronic diseases that now constitute the predominant health problems in this country.

This shift in emphasis of epidemiologic investigation has required great changes in methodology, if not in philosophy. New or newly emphasized methods are required to deal with problems such as:

1. Detection of causal associations in which decades or generations may lapse between cause and effect.
2. Measurement of existence and stage of diseases whose onset and progression are insidious and often unknown to the patient for many years.
3. Description of demographic distribution of conditions of low mortality for which the quality of medical diagnosis may vary markedly between populations and population subgroups.
4. Scientific and administrative problems of cohort studies that require large numbers and long-term followup.
5. Detection of clustering of disease in time or place at a much lower level of intensity than that observed in frank infections.

With these types of problems in mind, it has seemed useful to review existing health statistics and their sources to determine whether they should and could be modified to better serve epidemiologic needs.

The review has been further stimulated by recognition of the enormous increase that has occurred and will continue in the technical feasibility of data reduction, retrieval, and analysis on a large scale. The potential of this improved technology may be unrealized if the basic records are not appropriately modified.

In the past, the nature of health statistics and the usefulness of vital records have been limited by the fact that the records were established and continued for legal rather than health purposes. In the last 50 years, however, there has been increasing acceptance of the use of vital records for public health and scientific purposes. There is, therefore, a good prospect

¹Erhardt, C. L. Vital records are useful. *Harvard Pub. Health Alumni Bull.* 22: 7-11, 1965.

²Moriyama, I. M.: Use of vital records for epidemiological research. *J. Chronic Dis.* 17: 889-897, 1964.

that the demonstration of significant epidemiologic needs will be influential in producing changes or amplifications in vital records and statistics. While most of the information that has been added to vital records and the new sources that have developed have been based on a request for voluntary response, rather than as a legal requirement, there is no evidence that quality or completeness of information has been compromised on this account.

PROCEDURE

The subcommittee has limited its consideration to the uses of health records and statistics in investigating the etiology of disease. Although this may imply a rather narrow definition of "epidemiologic uses," it embraces the most important use of epidemiologic data. Other subcommittees of the National Committee have, of course, been concerned with medical care, demography, and other areas in which "epidemiologic" data have application.

The committee has reviewed the standard certificates of birth, death, fetal death, marriage and divorce, now in the final stages of a decennial revision, as well as the tabulations regularly developed from these sources and published in *Vital Statistics of the United States*, Programs, data forms, and reports of the three main components of the National Health Survey — The Health Interview Survey, the Health Examination Survey, and the Health Records Survey — were considered. The purposes of and current activities in special disease registries, record linkage, and genetic uses of linked vital records were reviewed.

The subcommittee did not deal with important issues of confidentiality of vital and health records and protection of the privacy of individuals. While these issues underlie any use of vital records for epidemiologic purposes and might be important in the implementation of several of its recommendations, the subcommittee did not consider itself constituted with the legal and other representation that would be required for a meaningful discussion of the issues. It should be pointed out, however, that legal and procedural safeguards against the misuse of vital record information have been part of the operation of Federal and State statistical systems for many years, and that these safeguards have been phenomenally successful. The distinction between

the use and dissemination of data on individuals as individuals and their use for describing the characteristics of groups without the identification of individuals has become widely recognized and accepted as a basis for statistical and epidemiological use of confidential records. Computers now provide new and even more effective techniques for protection against the misuse of confidential information.

The subcommittee did not take up the question of infectious disease reporting, the national component of which is now the responsibility of the National Communicable Disease Center (NCDC), Atlanta, Ga. However, it is believed that a review of this area would be useful, and the National Committee might consider the establishment of a subcommittee with special competence in infectious disease and with representation from NCDC and State Health Departments to review this field.

With respect to each of the types of records reviewed, the subcommittee considered (a) the information existent, or requested, on the basic record, (b) the tabulation now prepared from this information by the National Center for Health Statistics (NCHS), and (c) the existing means of access to the information contained on the records.

By far the most important gap between existing and potential usefulness seemed to lie in connection with item c, access to records. Since the major recommendations of the subcommittee are in this area, it will be considered first.

FINDINGS AND RECOMMENDATIONS

Access to Vital and Health Records

The problem of access to vital records has become acute with the revolutionary changes in data processing techniques already referred to. The potential for record utilization has been expanded almost beyond limit, provided the records and their storage are in such forms as permit machine access. Aggravating the problem is the fragmentation of the U.S. vital record system into more than 50 politically, geographically, and administratively independent systems. The effects of this fragmentation are especially evident in studies that require that people be followed over

long periods of time, as is usually the case in investigations of the etiology of chronic diseases.

Two particularly urgent needs are apparent: (a) methods that will improve our ability to "link" records pertaining to the same person or family, and (b) facilities to improve our ability to trace death records of persons enrolled in investigations.

Record Linkage

The potential of the information to be obtained by linkage of records is obvious. The linkage of records pertaining to an individual allows the correlation of events and circumstances of his life extending over the long periods associated with the incubation and development of chronic disease. The linkage of records pertaining to members of the same family allows studies of the genetic determinants of disease as well as the familial environment. No one seriously questions the premise that the linkage of health records of as many varieties as possible would be of great value for patient care and administrative purposes as well as for the purposes of etiologic investigations.

The practical aspects of the record linkage problem are less clear. A number of experimental record linkage systems have been set up in recent years, and a bibliography of relevant publications is given in Appendix I. Basic questions still at issue are (a) what identifying information most accurately allows linkage of records of the same individual or family, (b) what records can be linked with the greatest profit in terms of derived information, and (c) what is the most appropriate population size to be covered. Unfortunately, the answers to these questions are not generalizable to all times or all places, and although some of them have been answered for investigations of particular problems in particular places, it appears that available information is insufficient to recommend inauguration of any particular record linkage system in the United States as part of routine record procedures.

Two general recommendations do seem appropriate however. First, exploration of individual and family record linkage should continue to be pressed as a high priority research area to identify which linkages are likely to be most profitable and which methods most efficient. Second,

the basic records themselves should already be in process of modification so that (a) record linkage for the purpose of specific short-term studies, whether performed clerically or electronically, is facilitated and improved in accuracy, and (b) record systems are prepared for the incorporation of routine machine linkage when the time comes. Perhaps the most important single component of this modification would be the acceptance and wider use of the concept of a unique number to identify each individual in a population.

The Unique Number

The idea that each individual would have a unique number and that this number would identify all his health (and other) records is an old one. The advantages of such a procedure for record linkage are obvious. The widening use of the Social Security number in the United States puts the concept within the realm of feasibility for a large proportion of the population, perhaps for the first time. Several other numbering systems have been proposed, and even tried. A unique birth certificate numbering system, begun in 1947, has subsequently been dropped in many areas, presumably because of lack of use.

The wider the use of a particular number, the greater is the amount of information that can be linked. At the same time, the more frequently an individual is asked for that number the more likely he is to remember it correctly or have it available, which again encourages its use for other purposes. Thus, there is a circular effect operating in favor of the more widespread numbering system. This consideration argues overwhelmingly for the adoption of the Social Security number as the unique identifying number for health record purposes. If there was doubt about this previously, it was dispelled by the implementation of the Medicare program, in which the Social Security number is the basic identifying device.

The Social Security number is not ideal. A proportion of the population still has no number; a person may, by request, have more than one number; numbers are not usually assigned until wage earning begins; the lack of any apparent pattern to the assigned number (i.e., a person

cannot directly relate his number to his birth date or any other readily remembered set of figures) makes for frequent errors in recording and transcribing. Some of these problems will decrease as time passes and use of the numbers spreads. Others may be mitigated by deliberate efforts.

In the latter context, for example, it is noted that NCHS and representatives of State Health Departments are planning to confer with the Social Security Administration (SSA) to explore the possibility that a number assigned to an individual at birth by the State Health Department (from a list previously supplied by SSA) could later become the Social Security number when the individual makes application for assignment of a number.

Hospitals should be asked to incorporate Social Security numbers as part of their basic identifying information. (Presumably, this will in any event be necessary for Medicare patients.) They should be asked to explore the feasibility of the use of the Social Security number in place of their own record numbering system. This would greatly improve access to records of a particular patient identified by means of his Social Security number in some other context.

The U.S. Standard Certificate of Death now contains a place for the Social Security number. The States are urged to retain this item in their own certificates and to press for its completion in all possible instances. Whether or not it subsequently becomes possible to assign a "Social Security number" at birth, it would be of great value to incorporate Social Security numbers of both parents on the Certificate of Live Birth. Family record linkages would be enormously facilitated by this procedure.

Ascertainment of Death

U.S. death certificates contain a considerable amount of information regarding the decedent, and the presumed cause of his death. The quality of the information is high and the completeness of registration is excellent. These certificates constitute a most valuable source of information for epidemiologic studies.

The major problem arises in locating the death certificate of a specific individual unless

the date and place of his death are known with accuracy. The registration system is centralized only with respect to certain advisory functions and the preparation of national statistical tabulations. After the national statistical tabulations have been prepared, the centralized copies of the basic records are discarded. To locate the death certificate of a person whose place of death is not known might involve separate searches in more than 50 different States and cities. There is considerable variation across the country in the ease with which searches of these archives can be conducted. In addition, there is an understandable reluctance on the part of some Registrars to undertake searches for certificates which may not exist in their archives. It should be noted that, in some jurisdictions, while the cost of the search is borne by the Registrar's office, the reimbursement goes directly into the State treasury and does not compensate the Registrar's office for the additional labor involved.

As only one example of the type of problem that this fragmentation can provoke, it may be pointed out that in a current study of the mortality among steelworkers undertaken by the National Cancer Institute, more than 30 clerical steps are involved in tracing the subjects from employment to death. In addition to the size of the task involved in tracing deaths, there is also a problem with completeness, since it is, in practice, never practical to search for untraced cases in all 50 jurisdictions. Searches are usually restricted to two or three States with the highest rates of immigration from the area of the study.

The idea of a *National Death Index* located in the National Center for Health Statistics has been discussed, more or less casually, from time to time in the past. The index would permit identification of the fact, place and date of death, and, probably, State death certificate number. Access to information on the death certificates themselves would, as now, be through individual State registries.

It is believed that the mounting volume of studies of the chronic diseases alone makes the establishment of such an index an urgent necessity. There are, in addition, needs for such an index in connection with Medicare, and other medical care and commercial purposes. The cost of the index could be justified entirely by its

epidemiologic uses, but the additional medical care and commercial uses appear to eliminate cost as a serious deterrent.

In epidemiology, the index would be used primarily for studies of prognosis or cause of death among persons having particular diseases or undergoing particular exposure, for example:

1. Selected occupations, e.g., steelworkers, the rubber industry, hard-rock and uranium miners, smelters, etc.
2. Professional societies whose members have unusual laboratory exposure, e.g., American Chemical Society, American Society of Immunologists.
3. Groups undergoing health status evaluations (as predictors of longevity or causes of death) e.g., participants in the National Health Survey, Kaiser-Permanente multiphasic examinations, studies of infant development and health status, executive health (preventive maintenance) programs, and the American Cancer Society's study of more than 1 million persons.
4. Other suspected special risk groups, e.g., children who received SV 40 or other contaminated vaccines or drugs, persons exposed to ionizing radiation, blood donors, cases of accidental poisoning, etc.
5. Members of special disease registries--cancer, mental illness, congenital malformation.

It is therefore recommended, that a *National Death Index* be established. Problems that require detailed technical study include the amount and nature of information that will be required for accurate identification, cost of the index, and the extent to which this cost can be met by commercial uses.

Followup

In some studies it is sufficient to identify the dead members of the study group--members not so identified being assumed to be alive. However, when death ascertainment is known to be

incomplete, or when pathologic end points short of death are being studied, it is usually necessary to trace individuals over periods of years or decades.

In this country, in the presence of considerable migration and the absence of continuous population registration, a national health service or other national identification system, such a followup can be extremely difficult. Nevertheless, there are numerous resources and methods that can be utilized. Many workers appear to be unaware of the range and potential of these resources. In connection with its own research program the Division of Radiological Health (DRH), USPHS, prepared in 1962 a roster of followup facilities.³ More than 30 sources are described, together with the legal basis on which they may be utilized and the usual procedures involved. It is believed that it would be most useful to make such a document readily available through publication, and it is recommended that NCHS arrange for the preparation and publication of a revised version of the DRH manuscript or a newly prepared document with similar intent. Costs of various procedures should be included.

Information on Vital and Health Records

Having in mind the undesirability of burdening routinely prepared documents with information required only for ad hoc and limited purposes, there are few recommendations to be made regarding the content of existing vital and health records. However, some areas of concern were identified.

Cause of Death

The present standard certificate of death restricts information on pathologic conditions existing in the decedent to those which the certifying physician considered as contributing to the death. In fact, however, with many chronic condi-

³Staff of the Cooperative Thyrotoxicosis Therapy Follow-Up Study: *Resources for Locating Patients*. Division of Radiological Health, U.S. Public Health Service, 1962. (mimeographed)

tions it may be impossible to determine in individual cases whether or not an existent condition contributed to the death; such a determination may be possible only in terms of statistical probability based on comparison of mortality rates in groups of persons with and without the condition. In many instances, therefore, selection of pathologic conditions to appear on the certificate is arbitrary and inconsistent. A memorandum on this matter is reproduced in Appendix II as a basis for discussion.

The matter, while clearly a theoretical problem, did not appear of great practical moment so long as statistical tabulations, and, generally speaking, access to certificates, were on the basis of a single pathologic condition selected as the "underlying" cause. This single selection made quite obvious the arbitrary nature of the information presented and the necessity to restrict the use of diagnostic information to pathologic conditions likely to be selected as the underlying cause if they were present.

However, NCHS is now committed to early initiation of a program of coding and tabulating multiple causes of death listed on individual certificates. It then becomes much more crucial which conditions are included on and which omitted from the certificate. From two points of view--(a) for investigating in statistical terms which conditions do in fact contribute to death and in what degree, and (b) improving the death certificate as a general research tool--it seems important to shift the emphasis in certification from conditions supposedly contributing to death to conditions present at the time of death, or of which residua are present at the time of death.

How such additional information might best be collected has not been evaluated. Thus, in designing a possible Part III of the cause of death question on the death certificate, seeking information on other significant conditions present, it is not clear whether more complete certification would be obtained by means of an open-ended question or by a checklist of conditions such as diabetes, hypertension, congenital malformation, and so on. Such questions may be approachable through follow-back studies starting from current certificates, and it is suggested that such studies be undertaken as a basis for

specific proposals for the next revision of the standard certificate.

Congenital Malformations

As other causes decline, the congenital malformations become responsible for an increasing proportion of mortality and morbidity in infancy and childhood. Because certain forms of malformation have been clearly shown to be preventable, and because of the relatively short interval between cause and effect, this category of diseases is a prime target for investigative effort. Birth certificates can provide a major resource for this effort by providing access to (1) very large numbers of cases of common malformations, (2) large numbers of cases of common malformations with characteristics of particular value, e.g., being a twin. (3) series of cases of malformations of such rarity that series cannot be assembled from clinical sources, and (4) population-based series that can be used to investigate geographic and temporal clustering.

Birth certificates can never be expected to provide a complete ascertainment of malformations--partly because of the impossibility of defining a malformation. However, descriptive features make it quite evident that different etiologic factors are operative in different categories of malformation, and, in some instances, within a single diagnostic category of malformation. Birth certificates will be a useful resource even if their information is limited to those major, externally evident, anomalies which could be expected to be fully reported in a reasonably competent system.

It is to be hoped, therefore, that all States will retain the malformation item proposed on the standard live birth certificate, and make efforts to improve reporting of this item. With respect to the maintenance of surveillance of malformation reporting, in order to detect geographic or temporal clustering at an early stage, it is recommended that a comparison be made between reporting on birth certificates and systems, such as have been set up in New Jersey and British Columbia, that are based on reporting independent of the birth certificate.

It is further recommended that consideration be given to the possibility of examination of a

Development of Other Resources

nationally selected sample of newborn infants (including late fetal deaths) as one of the cycles of the Health Examination Survey. With respect to congenital malformation specifically, such a survey would provide estimates of the prevalence of congenital malformations detectable at birth—heretofore unavailable except with respect to the patients of a few highly specialized hospitals—and a standard against which to evaluate other means of estimating malformation rates. Because of the special problems of conducting examinations immediately after birth, such a survey would require different sampling procedures from those that have been used so far in the Health Examination Survey. However, sampling of maternity facilities and of blocks of time within such facilities (rather than direct sampling of individuals born) would seem to provide a relatively simple method for assembling a representative national sample of births.

Other Items

Level of education of parents has replaced occupation on the standard certificates of live birth and fetal death. The replacement results from the difficulty of interpreting and coding the occupation item, and belief that, as an index of social class (which is the main purpose of the occupation item), education is equally valid. In view of the very limited use of occupation as recorded on death certificates, an attempt might also be made to collect education data on death certificates in selected areas. The greater difficulty of obtaining education data on decedents, as contrasted with parents, is recognized, and in the evaluation phase the item should probably be sought as a supplement to, rather than replacement for, information on occupation.

In many studies of the effects of social and psychologic stress it would be advantageous to have the names (rather than simply the number) of the children as an item on the certificate of divorce. It is understood, however, that there is a reluctance on the part of Registrars to press for this item.

It seems worthwhile to identify two resources that do not appear to have been utilized to their fullest potential for epidemiologic investigation, although there are no specific recommendations as to how their utilization might be increased.

The first consists of the records of deaths coming under the purview of the medical examiner. The medical examiners' investigations of such deaths are detailed and the records are public. Medical examiners in general appear to be favorably disposed to the utilization of the records for scientific investigation, and, in some instances, have initiated epidemiologic investigations based on their material. The records offer extensive possibilities for analysis alone, or as the starting point for investigations in which additional information is sought.

The second resource consists of the records of health insurance schemes. The value of such records is demonstrated by the series of reports based on the experience of the Health Insurance Plan (HIP) of Greater New York. While it is true that HIP incorporates medical care as well as an insurance program, essentially similar records are usually required as the basis for payment even when the medical service is not performed by the insuring company. Remarkably little use has been made, for example, of the voluminous material that must exist in Blue Cross-Blue Shield plans relevant to studies of the familial clustering of disease, the association of diseases in individuals, and the natural history of chronic diseases.

Lastly, in connection with the major programs of medical care now being initiated—notably Medicare, the Title 19 provisions of Medicare as they relate to children, and the heart-cancer-stroke and community mental health centers—it is hoped that adequate statistical and epidemiologic consultation will be obtained at this early stage to ensure the ultimate usefulness of the medical records.

SUMMARY OF RECOMMENDATIONS

1. Linkage of various vital and health records should continue to be explored as a high priority research area, to identify which linkages are likely to be most profitable and which methods most efficient.
2. The Social Security number should be accepted as the most practical numerical identification of individuals, and incorporated into all vital and health records. Where possible, the Social Security number should become the actual identification number for the specific record of the individual. Hospitals should be asked to consider this last possibility.
3. NCHS, representatives of State Health Departments, and the Social Security Administration should explore the possibility of assigning a Social Security number to an individual at birth. This recommendation supposes that the Social Security Administration is itself not already considering the use of a birth numbering system for its own purposes.
4. Social Security numbers of parents should be added to the certificates of live birth and fetal death.
5. Because of the great epidemiologic usefulness of a *National Death Index*, NCHS should explore with some urgency the technical problems involved in the establishment of such a resource. Apart from estimates of cost and the extent to which this would be mitigated by commercial use, there is need for information on the amount and nature of the information required to identify decedents, and the computer technology most appropriate to such an extensive operation. It should be stressed that an index is meant, not a repository for the death certificates themselves.
6. A document should be prepared and published for the assistance of investigators in followup studies, along the lines of that prepared for their own staff by the Division of Radiological Health. Sources of followup, legal bases, usual procedures, and costs should be included.
7. The basis for inclusion of pathologic conditions on the death certificate may need to be extended beyond the present restriction to conditions presumed to have contributed to the death. Information might be sought on all significant conditions present at the time of death, or of which significant residua are present at the time of death. The best way of obtaining this information should be sought in follow-back studies based on current certificates in preparation for the next revision of the standard certificate.
8. Efforts should be made to extend, improve, and utilize information on congenital malformations reported on vital records, particularly the birth certificate.
9. Consideration should be given to the possibility of examination of a national sample of newborn infants (including fetal deaths) as one of the cycles of the Health Examination Survey.



APPENDIX I

SELECTED BIBLIOGRAPHY ON RECORD LINKAGE

- Acheson, E. D.: *The Oxford Record Linkage Study, Report on the Second Year's Operations*. Oxford Regional Hospital Board. Oxford, England 1963.
- Acheson, E. D.: The Oxford Record Linkage Study, a review of the method with some preliminary results. *Proc. Roy. Soc. Med.* 57:11, 1964.
- Acheson, E. D.: The Oxford Record Linkage Study, a central file of morbidity. *Brit. J. Prev. & Social Med.* 18:8, Jan. 1964.
- Acheson, E. D., and Evans, J. G.: The Oxford Record Linkage Study. *Biometrics* 2:367, 1963.
- Acheson, E. D., Truelove, S. C., and Witts, L. J.: National epidemiology. *Brit. M. J.* 1:668, 1961.
- Bahn, A. K.: Methodological study of population of outpatient psychiatric clinics, Maryland 1958-59. *Public Health Monograph* No. 65. PHS Pub. No. 821. Washington, D.C., 1961.
- Bahn, A. K.: The development of an effective statistical system in mental illness. *Am. J. Psychiat.* 116:798, 1960.
- Bahn, A. K.: Person matching by electronic methods. *Communications of the Association for Computing Machinery* 5:404, 1962.
- Bahn, A. K., and Bahn, R.: Considerations in using social security numbers on birth certificates for research purposes. *Pub. Health Rep.* 79:937, 1964.
- National Office of Vital Statistics: Some aspects of vital statistics registration in Canada, by A. E. Bailey. *Vital Statistics—Special Reports* Vol. 23, No. 22. Public Health Service. Washington, D.C., 1947. pp. 252-61.
- National Center for Health Statistics: Summary report on Tennessee's "Record of Events," by M. R. Baird, and E. H. Halpin. *The Public Health Conference on Records and Statistics*, PHS Doc. No. 473. Public Health Service. Washington, D.C., Dec. 1959
- Central Bureau of Statistics: Population changes 1961. *Official Statistics of Sweden* (Population and Vital Statistics) Stockholm, 1963.
- Davidson, L.: Retrieval of misspelled names in an airline passenger records system. *Communications of the Association for Computing Machinery.* 5:169, 1962.
- Dunn, H. L.: Record linkage. *Am. J. Pub. Health.* 36:1412, 1946.
- Dunn, H. L.: A national identity registration system to synthesize social statistics. *Estadistica, Journal of the Inter American Statistical Institute* 11:605, Sept. 1953.
- Harris, R.: Letter to Dr. Councell dated Mar. 8, 1965.
- Kennedy, J. M.: Linkage of birth and marriage records using a digital computer. *Atomic Energy of Canada Ltd.*, Report No. 1258. Chalk River, Ontario, 1961. p. 18.
- Kennedy, J. M.: The use of a digital computer for record linkage. *Proceeding of the Seminar on the Use of Vital and Health Statistics for Genetic and Radiation Studies*, 1962. pp. 155-159.
- Kennedy, J. M., Newcombe, H. B., Okazaki, E. A., and Smith, M. E.: List processing methods for organizing files of linked records. *Atomic Energy of Canada Ltd.*, Report No. 2078, 1964.
- Kjelsberg, M., and Metzner, H.: Record system used in the Tecumseh Community Health Study. Personal communication.
- Marshall, J. T.: Canada's national vital statistics index. *Population Studies* 1(2):204, Sept. 1947. (Printed in Great Britain.)
- Masi, A. T., Sartwell, P. E., and Shulman, L. E.: The use of record linkage to determine familial occurrence of disease from hospital records (Hashimoto's disease). *Am. J. Pub. Health* 54(11): 1887-1894, Nov. 1964.
- McKeown, T. and others: Session VII, panel discussion, session on epidemiological studies. *Second International Conference on Congenital Malformations*. The International Medical Congress, Ltd., New York, 1964. pp. 341-351.
- Minet, P. L.: Fertilité précoce d'une cohorte de mariages dans une province canadienne. Biology Branch. *Atomic Energy of Canada Ltd.* Basel 14:186-196. Chalk River, Ontario, 1964.
- National Center for Health Statistics: Study group on record linkage. *The Public Health Conference on Records and Statistics*, PHS Doc. No. 603. Public Health Service. Washington. U.S. Government Printing Office, Nov. 1964.
- National Center for Health Statistics: Planning session for study group on record linkage. *The Public Health Conference on Records and Statistics*, PHS Doc. No. 603.1. Public Health Service. Washington. U.S. Government Printing Office, Jan. 1965.

Newcombe, H. B.: Detection of genetic trends in public health. *Effect of Radiation on Human Heredity*. World Health Organization, Geneva, 1957. pp. 157-168.

Newcombe, H. B.: Record linkage as a means of studying the somatic effects of low doses of radiation. *Science* 130: 954-959, 1959.

Newcombe, H. B.: Population genetics, population records. *Methodology in Human Genetics*. Holden-Day, Inc., 1962. pp. 92-113.

Newcombe H. B.: Untapped knowledge of human populations. *Transactions of the Royal Society of Canada*. Vol. LVI, Series III, Section 111, June 1962. pp. 173-180.

Newcombe, H. B.: Risk of fetal death to mothers of different ABO and RH blood types. *Am.J.Human Genet.* 15(4): 449-464, Dec. 1963.

Newcombe, H. B.: Pedigrees for population studies, a progress report. *Cold Spring Harbor Symposia on Quantitative Biology*, 29:21, 1964.

Newcombe, H. B.: Panel discussion, session on epidemiological studies. *Second International Conference on Congenital Malformations*. The International Medical Congress, Ltd., New York, 1964. pp. 345-349.

Newcombe, H. B.: Environmental versus genetic interpretations of birth-order effects. *Eugenics Quarterly*, Nov. 1964. p. 36.

Newcombe, H. B.: Screening for effects of maternal age and birth order, in a register of handicapped children. *Ann. Human Genet.* 27:367-382. London, 1964.

Newcombe, H. B.: Use of vital statistics. United Nations World Population Conference, Belgrade, Yugoslavia, Aug. 30 to Sept. 10, 1965.

Newcombe, H. B., Axford, S. J., and James, A. P.: A plan for the study of fertility of relatives of children suffering from hereditary and other defects. *Atomic Energy of Canada Ltd.*, Report No. 511, Nov. 1957. p. 50.

Newcombe, H. B., James, A. P., and Axford, S. J.: Family linkage of vital and health records. *Atomic Energy of Canada Ltd.*, Report No. 470. Chalk River, Ontario, July 1957. (Reprinted Aug. 1961 and May 1964.)

Newcombe, H. B., and Kennedy, J. M.: Record linkage, making maximum use of the discriminating power of identifying information. *Communications of the Association for Computing Machinery* 5(11):563-566, 1962.

Newcombe, H. B., and Kennedy, J. M.: Demographic analysis and computer programs. United Nations World Population Conference, Belgrade, Yugoslavia, Aug. 30 to Sept. 10, 1965.

Newcombe, H. B., Kennedy, J. M., Axford, S. J., and James A. P.: Automatic linkage of vital records. *Science* 130(3381): 954-959, Oct. 1959.

Newcombe, H. B., and Rhynas, P. O. W.: The cost of individual follow-up studies of large populations. Proceedings of the International Population Conference-Paper No. 16. (Reprinted as Atomic Energy of Canada Ltd., Report No. 1255, Sept. 1961. p. 8.)

Newcombe, H. B., and Rhynas, P. O. W.: Family linkage of population records. *On the Use of Vital and Health Statistics for Genetic and Radiation Studies*. United Nations Pub., Sales No. 61. XVII(8):135-154, 1962.

Newcombe, H. B., and Rhynas, P. O. W.: Child spacing following stillbirth and infant death. *Eugenics Quarterly* 9(1): 25-35, Mar. 1962.

Newcombe, H. B., and Tavendale, O. G.: Effects of father's age on the risk of child handicap or death. *Am.J.Human Genet.* 17(2):163-178, 1965.

National Office of Vital Statistics: A coordinated system of vital records and statistics, the need for coordination, by H. L. Dunn. Public Health Service. Washington, D.C., July 1953.

Phillips, W., Jr., and Bahn, A. K.: Experience with computer matching by names. *Proceedings of the Social Statistics Sections, American Statistical Association*, 1963.

Phillips, W., Jr., Bahn, A. K., and Miyasaki, M.: Person-matching by electronic methods. *Communications of the Association for Computing Machinery* 5:404, 1962.

Phillips, W., Jr., and Gorwitz, K., and Bahn, A. K.: Electronic maintenance of case registers. *Pub.Health Rep.* 77: 503, 1962.

Schwartz, E. E.: Some observations on the Canadian family allowances program. *The Social Service Review* Vol. XX, No. 4. Dec. 1946.

Statistical Commission: Twelfth session, Item 9 of the provisional agenda E/CN/3/293. *United Nations Economic and Social Council*. Feb. 7, 1962. p. 38.

Vinge, Margit, Head, Documentation Unit, Central Bureau of Statistics, Stockholm. Letter dated Mar. 9, 1965, to Louis R. Stolcis with attachments.



APPENDIX II

THE UNDERLYING CAUSE CONCEPT

At the November 1961 Geneva meeting of the Subcommittee on Classification of Diseases of the Expert Committee on Health Statistics, it was suggested that the "concept of underlying cause might be reexamined to see if a more satisfactory conceptual basis for classifying causes of death could not be developed." The concept can be reviewed in the light of uses made of information obtained on death certificates, and from the viewpoint of the data available for decedents. No matter how the medical certification is designed, its source is the existing observations on the decedent, and its statistical function is to serve public health and medical needs.

Cause of death information is now used: (1) to describe the most important current health problems; (2) to record frequency of certain rare, fatal diseases; (3) to serve as a starting point of epidemiologic studies of specific populations characterized by the disease to which the death was assigned; and (4) to assess the frequency of conditions associated with death, but which are not the direct cause of death.

Does the underlying cause of death supply data that are appropriate to these uses? If not, what questions on the medical certification would elicit a more valid response for these applications? Are there other uses not now served that could be met with a change in the kind of information requested?

The first use, to describe the most important current health problems, has changed, at least in the United States, in the last decade. The National Health Survey can now provide data on the major illnesses found among the living population.

While it was never sound to use mortality data as an estimate of morbidity, this was commonly done in the past. Such a use is no longer defensible. The function of the death certificate is still that of providing information on those diseases that cause death. These data augment those obtained on morbidity by supplying data for diseases that cause death without previous signs or symptoms; by evaluating the severity of disease through the fact that it has caused "premature" death; by supplying counts for diseases, that, owing to their rare occurrence, or occurrence in a limited population (say, under 1 month of age) are not easily reached through survey methods. In all of these cases, the concept of an underlying cause of death can provide valid information. But if disease incidence were viewed as a continuum of information beginning with the occurrence of signs and symptoms in the living population, followed by disabling illness and finally, death, then the information collected on the death certificate should be restructured to parallel the "prevalence" concept used in the collection of morbidity data. The underlying cause concept cannot provide prevalence data, or the counts needed for case fatality data.

The second use, to record the frequency of rare diseases with high fatality rates is clearly met by the underlying cause concept. There should not be any difficulty in collecting such information under other definitions of the cause of death.

The third use, as a starting point of epidemiologic studies, is now practically limited to studies of diseases selected as the underlying cause of death. Many investigators would pre-

fer to follow every occurrence of a disease in the population of decedents rather than the population of persons whose deaths were assigned to this disease. The larger population can be approached by coding all the diseases reported on the death certificate. The mention of these diseases will still be limited to the physician's interpretation of the response required by the present certificate. A thoughtful certifier may select only those diseases related to the death, while another physician may list all serious diseases present at death. If the epidemiologist wishes to know for how many persons a particular disease contributed to death, it may be possible to obtain the information from the current certificate when all conditions reported on the record are coded. If the epidemiologist needs information on how many persons died with a disease, this count cannot be obtained with the present form of the medical certification.

The fourth use, to assess frequency of conditions associated with death but not the direct cause of death, has been cited as the chief reason for coding all the conditions reported on the death certificate, rather than the underlying cause alone. As pointed out in the previous paragraph, the wording on the present form does not elicit a report of all conditions present at death, not even all serious conditions. Response to specific queries concerning diseases present at death has demonstrated that the entries on the certificate represent the physician's judgment or understanding of the nature of the report expected from him. The form must be redesigned if conditions present at death are needed.

In summary, the underlying cause concept cannot generally provide prevalence or incidence data. It can provide counts for diseases causing death, or contributing to the event of death. Where there is more than one disease contributing to the death, the present wording of the certificate does not encourage a complete response. Where there is a serious disease present that is unrelated to the sequence of events resulting in death, the certificate specifically discourages such an entry.

Once a decision is reached on what data are needed about death for medical and public health

purposes, the requirements must be phrased to fit into the way death certificates are completed, and what is known about decedents. The medical and social information at hand for a decedent may vary from that assembled for the person under systematic lifelong medical care to the one-time posthumous observation made by the medical examiner of an unidentified person. For the former individual there may be a record of his physical condition at regular intervals, each illness episode, his last illness, and perhaps an autopsy protocol. On the basis of this information, what should be recorded on the death certificate? How can the questions or instructions in the medical certification be phrased to elicit the data needed? Can these questions be so phrased to serve also for the death in which little information is known?

What approach should be taken to permit separation of data by quality—that based on complete histories; on attending physician records for a terminal illness only, on autopsy only, etc? Some exploration of quality of diagnostic information and its effect on medical certification has been reported in special studies, such as one by I. M. Moriyama and others.⁴ Thus far, there has not been any consideration of establishing a routine collection of information on the quality of medical certification. Is it possible or desirable to make such measures part of the collection system for mortality statistics?

If it can be decided what data should be collected through the medical certifications on death records, it may become easily evident as to whether these data can be obtained through the underlying cause approach, or through a new concept. A new concept, and the questions or instructions needed to produce the desired answers would then need to be tested. An experiment would need to be devised to learn whether or not the new form is producing the expected data.

⁴Moriyama, I. M., Dawber, T. R., and Kannel, W. B.: Evaluation of Diagnostic Information Supporting Medical Certification of Deaths from Cardiovascular Diseases. *National Cancer Institute Monograph* No. 19, Jan. 1966. pp. 405-419.

OUTLINE OF REPORT SERIES FOR VITAL AND HEALTH STATISTICS

Public Health Service Publication No. 1000

- 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 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 on 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 20. Data on mortality.*—Various statistics on mortality other than as included in 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 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, medical experience in the last year of life, characteristics of pregnancy, etc.

For a list of titles of reports published in these series, write to: Office of Information
National Center for Health Statistics
U.S. Public Health Service
Washington, D.C. 20201

DHEW Publication No. (HSM) 73-1265
Series 4 -No. 7

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

OFFICIAL BUSINESS
Penalty for Private Use \$300

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF HEW

HEW 396



THIRD CLASS
BLK. RT.