Statistical Notes for Health Planners

NCHS

Number 6 February 1978

# CAUSE-OF-DEATH DATA

Theodore D. Woolsey[a]

## INTRODUCTION

In an earlier note,[1] the value and limitations of mortality data as measures of the health status of a community were discussed. The purpose of this Sixth Note is to extend that discussion and deal specifically with the usefulness to health planners of mortality statistics by cause of death. Specific precautions that the planner should keep in mind in using data of this kind are described, and some ways of analyzing cause-of-death data to improve their relevance to the health planner's concerns are suggested.

An underlying assumption in this Note is that the need for measures of the health status of a jurisdiction or a community arises principally from two responsibilities of the planner. They are:

- To be able to identify any special or emerging health problems that the residents of the community are experiencing, in order to help in determining priorities for action.
- To be able to evaluate progress in dealing with those problems and to find out whether the planning activities are being accompanied by favorable trends in the health of the people in the community.

Those two objectives are the criteria used in evaluating the usefulness of cause-of-death statistics to health planners.

Before proceeding, it is worthwhile to repeat that mortality statistics provide about the only direct measure of health covering a wide range of health problems, from acute illnesses and injuries to the major chronic diseases, that is universally available throughout the United States, in a reasonably comparable way, for every community and for every year.

## ORIGINS OF CAUSE-OF-DEATH DATA

The United States and other economically advanced countries require the filing of a death certificate whenever a person dies. This includes any infant born alive regardless of the length of life. In the United States, filing of the certificate, or "registration" of the death, is covered by State laws. These laws generally conform to a 1959 Model Act developed by the State registrars of vital records under the leadership of the National Office of Vital Statistics. A description of how the registration system operates may be found in chapter I of *Vital Statistics Rates in the United States, 1940-1960.*[2] The important fact about death registration, however, is that the local registrar, responsible for ensuring that certificates are filed for each vital event (birth, death, fetal death, marriage, and divorce or annulment), may not issue a permit to dispose of the remains unless a death certificate has been filed. Since disposing of a body without a permit is a serious violation, the registration of deaths in all but the most isolated areas is believed to be almost 100 percent complete.

Personal particulars about the deceased person are usually filled out on the certificate by the funeral director, who obtains the information from a surviving relative. The

[a] Health Statistics Consultant

funeral director also takes the certificate to the attending physician (or in the case of an unattended or violent death, to the medical examiner or coroner) for completion of what is known as the "medical certification of cause of death." This portion of the death certificate must be certified by the signature of the physician or medical-legal officer. The wording of the medical certification is virtually identical throughout the United States even though each State designs and prints its own forms. This resulted from a standardization process that occurred through the efforts of the State registrars of vital records, the National Office of Vital Statistics, the Division of Vital Statistics of the National Center for Health Statistics (NCHS), and the World Health Organization, which tries to bring about international comparability. Following extensive review of needed modifications, NCHS issues model forms of the U.S. Standard Certificates each decade. The States usually adhere closely to these in printing their own forms. The latest version of the U.S. Standard Certificate of Death with space for medical certification of cause of death is shown in figure 1.

The Division of Vital Statistics prepares and distributes instruction manuals on how to file the certification of cause of death to funeral directors, physicians, hospitals, and medical-legal officers.

In most States, the vital records office is part of the State health department. This

Figure 1. U.S. Standard Certificate of Death

(PHYSICIAN, MEDICAL EXAMINER OR CORONER)
U.S. STANDARD
CERTIFICATE OF DEATH

Form Approved
OMB No. 68R 1901

LOCAL FILE NUMBER — STATE FILE NUMBER

TYPE OR PRINT IN PERMANENT INK FOR INSTRUCTIONS SEE HANDBOOK

DECEDENT
1. DECEDENT–NAME FIRST MIDDLE LAST | 2. SEX | 3. DATE OF DEATH (Mo., Day, Yr.)
4. RACE–(e.g., White, Black, American Indian, etc.) (Specify) | 5a. AGE–Last Birthday (Yrs.) | UNDER 1 YEAR 5b. MOS. DAYS | UNDER 1 DAY 5c. HOURS MINS. | 6. DATE OF BIRTH (Mo., Day, Yr.) | 7a. COUNTY OF DEATH
7b. CITY, TOWN OR LOCATION OF DEATH | 7c. HOSPITAL OR OTHER INSTITUTION–Name (If not in either, give street and number) | 7d. IF HOSP. OR INST. Indicate DOA, OP/Emer. Rm., Inpatient (Specify)
8. STATE OF BIRTH (If not in U.S.A., name country) | 9. CITIZEN OF WHAT COUNTRY | 10. MARRIED, NEVER MARRIED, WIDOWED, DIVORCED (Specify) | 11. SURVIVING SPOUSE (If wife, give maiden name) | 12. WAS DECEDENT EVER IN U.S. ARMED FORCES? (Specify Yes or No)

IF DEATH OCCURRED IN INSTITUTION SEE HANDBOOK REGARDING COMPLETION OF RESIDENCE ITEMS

13. SOCIAL SECURITY NUMBER | 14a. USUAL OCCUPATION (Give kind of work done during most of working life, even if retired) | 14b. KIND OF BUSINESS OR INDUSTRY
15a. RESIDENCE STATE | 15b. COUNTY | 15c. CITY, TOWN OR LOCATION | 15d. STREET AND NUMBER | 15e. INSIDE CITY LIMITS (Specify Yes or No)

PARENTS
16. FATHER NAME FIRST MIDDLE LAST | 17. MOTHER MAIDEN NAME FIRST MIDDLE LAST
18a. INFORMANT–NAME (Type or Print) | 18b. MAILING ADDRESS STREET OR R.F.D. NO. CITY OR TOWN STATE ZIP

DISPOSITION
19a. BURIAL, CREMATION, REMOVAL, OTHER (Specify) | 19b. CEMETERY OR CREMATORY NAME | 19c. LOCATION CITY OR TOWN STATE
20a. FUNERAL SERVICE LICENSEE Or Person Acting As Such (Signature) ▶ | 20b. NAME OF FACILITY | 20c. ADDRESS OF FACILITY

CERTIFIER
To be Completed by CERTIFYING PHYSICIAN Only
21a. To the best of my knowledge, death occurred at the time, date and place and due to the cause(s) stated. (Signature and Title) ▶
21b. DATE SIGNED (Mo., Day, Yr.) | 21c. HOUR OF DEATH M
21d. NAME OF ATTENDING PHYSICIAN IF OTHER THAN CERTIFIER (Type or Print)

To be Completed by MEDICAL EXAMINER or CORONER Only
22a. On the basis of examination and/or investigation, in my opinion death occurred at the time, date and place and due to the cause(s) stated. (Signature and Title) ▶
22b. DATE SIGNED (Mo., Day, Yr.) | 22c. HOUR OF DEATH M
22d. PRONOUNCED DEAD (Mo., Day, Yr.) ON | 22e. PRONOUNCED DEAD (Hour) AT M

23. NAME AND ADDRESS OF CERTIFIER (PHYSICIAN, MEDICAL EXAMINER OR CORONER) (Type or Print)
24a. REGISTRAR (Signature) ▶ | 24b. DATE RECEIVED BY REGISTRAR (Mo., Day, Yr.)

CONDITIONS IF ANY WHICH GAVE RISE TO IMMEDIATE CAUSE STATING THE UNDERLYING CAUSE LAST

CAUSE OF DEATH
25. IMMEDIATE CAUSE (ENTER ONLY ONE CAUSE PER LINE FOR (a), (b), AND (c).)
PART I
(a) | Interval between onset and death
DUE TO, OR AS A CONSEQUENCE OF
(b) | Interval between onset and death
DUE TO, OR AS A CONSEQUENCE OF
(c) | Interval between onset and death
PART II OTHER SIGNIFICANT CONDITIONS–Conditions contributing to death but not related to cause given in PART I (a) | 26. AUTOPSY (Specify Yes or No) | 27. WAS CASE REFERRED TO MEDICAL EXAMINER OR CORONER (Specify Yes or No)
28a. ACC., SUICIDE, HOM., UNDET., OR PENDING INVEST. (Specify) | 28b. DATE OF INJURY (Mo., Day, Yr.) | 28c. HOUR OF INJURY M | 28d. DESCRIBE HOW INJURY OCCURRED
28e. INJURY AT WORK (Specify Yes or No) | 28f. PLACE OF INJURY–At home, farm, street, factory, office building, etc. (Specify) | 28g. LOCATION STREET OR R.F.D. No. CITY OR TOWN STATE

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE PUBLIC HEALTH SERVICE NATIONAL CENTER FOR HEALTH STATISTICS
1978 REVISION

HRA-162-1
Rev. 1/78

office queries the physician who signed the certificate if the cause-of-death statement is unclear or inconsistent.

The results of such efforts as these, in terms of reliable cause-of-death information, will be discussed later. However, it is important to give here the rationale for a cause-of-death statement as it is stated in one of the instruction manuals.[3]

"A cause of death is a disease, abnormality, injury, or poisoning that contributed directly or indirectly to death. A death often results from the combined effect of two or more diseases. These diseases may be completely unrelated, arising independently of each other; or they may be causally related to each other, that is, one disease may lead to another which in turn leads to a third condition, etc. . . . The 'immediate cause' on line (a) is the condition that directly resulted in death. . . . Line (b) is used for entering the condition that gave rise to the immediate cause. . . . Line (c) is used to record the related underlying cause, when it is not provided for in line (b). . . . In Part II should be recorded any other important disease or condition that was present at the time of death which may have contributed to death but which was not related to the immediate cause of death on line (a)."

On the Standard Certificate, spaces are also provided to record the circumstances of accidents and to specify whether autopsy results were taken into consideration in stating the cause of death.

To transform the cause-of-death data into statistics, numerical codes are assigned. For this, an internationally agreed-upon classification—*The International Classification of Diseases* (ICD)—is used. A special adaptation of the Eighth Revision ICD is used throughout the United States to classify causes of death.[4]

Until recent years the practice has been to code only the "underlying" cause of death, i.e., the cause that started the sequence of events leading to death. All commonly published mortality statistics are based on this rule. However, beginning with 1968 deaths NCHS began coding *all* causes on the certificate[b] and designed a computer program (ACME) to select automatically the underlying cause.

Although State vital statistics offices have coded the cause of death on all certificates for many years, NCHS and its predecessor offices have coded cause of death on copies of death certificates received from the States in order to produce statistics that are comparable from State to State. To perform this coding in a reproducible form requires 6-8 weeks of training with a lengthy and detailed instruction manual and quality control of the process by independent verification. Thus geographical comparisons based on statistics are risky if the coders' training has not been standardized. Centralized training courses arranged by the NCHS Applied Statistics Training Institute and the introduction by increasing numbers of States of computerized assignment of underlying causes are steadily increasing the comparability of State-coded cause-of-death statistics.

For these reasons NCHS has begun to accept the coding done by the States instead of doing the coding itself. Standardization in classifying causes of death, increasing acceptance of a computer program to select the underlying cause of death, and acceptance by NCHS of State-coded data were accomplished by State participation in the Cooperative Health Statistics System.

An added complication is introduced by decennial changes in the ICD. It has been necessary for the League of Nations and, after World War II, the World Health Organization, to revise the Classification to keep it compatible with improved medical understanding of diseases and changing terminology. This has led to discontinuities in the time series of death rates by cause of death. (The new edition of the Classification is now available and, as in the past, its adoption will require some retraining of coders.)

The pace of these decade-to-decade changes in the Classification seems not to diminish. For instance, the changes between the Sixth Revision, adopted in 1948, and the Seventh Revision, adopted in 1955, were minor, but those between the Seventh and the Eighth Revisions, adopted in 1965, were numerous and important. The Ninth edition incorporates additional major changes. However, in recent decades the classification has included more detail so that it is easier to regroup

[b]Multiple cause coding had been done in earlier years on an experimental basis. Since 1968 it has been done routinely.

categories to construct comparable time series. To bridge the gap in discontinuities, NCHS and its predecessors have usually coded a sample of death certificates using the new and the old codes at the time of the changeover. "Comparability ratios" are calculated for all major causes of death to show how the data for earlier years should be adjusted to make them comparable with the later years or vice versa.

The comparability ratios for the leading causes of death at the 1968 changeover (ratio of numbers of deaths coded to a cause-of-death rubric according to the Eighth Revision to the number coded to the closest corresponding rubric according to the Seventh Revision) ranged from .8862 for nephritis and nephrosis to 1.0440 for influenza and pneumonia.[5]

## RELIABILITY OF DATA

The general questions of the reliability of mortality statistics, and particularly death rates, were discussed in the third report in this series.[1] It was stated that although the completeness of death registration is not a problem, the undercounting of certain population groups in the census may result in overstatement of mortality for these groups. Also, residence classification differences between the numerator and the denominator tend to slightly overstate the mortality in urban places and understate that for rural areas. There is also need for caution in interpreting death rates for counties or other jurisdictions in which the population of long-term institutions is a relatively high proportion of the total. In the third report[1] attention was also drawn to the lack of stability found in death rates based upon a small number of deaths (for example, less than 100).

These precautions also pertain to death statistics classified by cause. Instability owing to small numbers may limit the usefulness of statistics by cause of death for jurisdictions with a population under 100,000—to cite an arbitrary cut-off point. A jurisdiction with 100,000 population will have, in the course of a single year, roughly the following numbers of deaths from leading causes (assuming death rates and distribution of the population by age, sex, and race are identical with the U.S. as a whole in 1975):

| | |
|---|---|
| Diseases of the heart | 336 |
| Malignant neoplasms | 172 |
| Cerebrovascular diseases | 91 |
| Accidents of all types | 48 |
| Influenza and pneumonia | 26 |
| Diabetes | 17 |
| All other causes | 199 |

Thus it is difficult to compare cause-of-death rates in two communities of this size or to compare the rates annually in any one such community. There are two remedies for this problem. One solution is to combine mortality data for several years. This does not greatly limit the usefulness of the statistics as measures of health because year-to-year changes in the mortality from most diseases is small.

A second remedy is to combine the cause-of-death data into a summary index of the community's health based upon cause-of-death classification. Certain summary indexes of mortality were discussed in *Statistical Notes for Health Planners,* No. 3.[1] Others that utilize the cause-of-death distribution are discussed in the appendix.

In addition to the problems of reliability and instability caused by population underenumeration, residence classification, and small numbers of deaths, there are five other problems peculiar to cause-of-death statistics. They are:

- Reliability in assigning codes
- Lack of continuity in trends and changes in coding rules resulting from revisions of the ICD
- Variations in medical terminology
- Thoroughness in completing the medical cause-of-death certification
- Basic reliability in determining the cause of death.

The first two of these have been discussed in the preceding section although the magnitude was not dealt with. Errors introduced by mistakes in the assignment of the numerical codes are a minor source of error compared with the others and are controllable by thoroughly training the coders, standardization of the coding rules, and computerization of the more complex decisions required. Progress is being made in all three of these areas. For example, the errors remaining after NCHS coding of causes of death probably involve no more than 1 or 2 percent of the certificates.

The assignment of a code depends upon the wording of the cause-of-death statement on the certificate. It is understood that the terminology used by physicians to describe identical clinical and/or pathological findings can vary not only over time but also from one part of the country to another at any one time. The physicians were trained in different years and at medical schools in different areas by teachers who themselves had developed individual habits in the use of terms. The quantitative effect of this difference upon death rates by cause is exceedingly difficult to measure.

Much of the variation is eliminated by coders' use of an extensive index of terms, showing where each should be classified.[4] Thus synonymous or closely related terms are coded to the same number. But lack of uniformity can still obscure health problems. A recent example which became a matter of political concern was the coding of deaths from the "sudden infant death syndrome." The medical consensus in recent years has been that a set of clinical findings define this syndrome. However, the terminology on death certificates varied greatly throughout the country, and, hence, national statistics could not be reliably compiled. Five years ago, NCHS estimates were far lower than those being publicized by private groups. Since then, terminology has become more standard; statistics for 1975 are believed to represent the problem more accurately.

The only way to reduce errors of this kind is to agree on medical terminology. The efforts of the American Medical Association and other associations representing medical records administrators in hospitals have gradually improved this situation over the years.[6]

The errors introduced through lack of thoroughness in completing the medical certification have been studied principally by comparing statements on death certificates with information from hospital records. Such studies deal with only part of the problem, however, because only about 45 percent of all deaths occur in short-term hospitals. Another difficulty is that the wording of the hospital form on which the circumstances of death are recorded is not the same as that on the death certificate.

One index of the care physicians take in certifying cause of death is the proportion of death certificates containing more than one cause. This proportion increased in the United States from 35 percent in 1917 to 58 percent in 1955.[7] NCHS statistics show that in 1968 the figure was about 75 percent, but part of this increase is due to rising proportions of persons dying at older ages since they are more likely than younger persons to have more than one condition contributing to death.

Despite these questions on how conscientious the physician is in filling out the record and how well the diagnoses reflect what has been written in hospital records, the question remains: How reliable is the evidence upon which the cause-of-death statement is based? This question has not been studied very often, but a sample of Pennsylvania death certificates for nonviolent deaths was investigated in 1956 to determine what evidence there was to form the basis of the statement. For 39 percent of the cases the quality of evidence was rated as sketchy and for 58 percent it was rated good or very good.[8]

A similar study of a national sample of cardiovascular disease deaths in 1960 resulted in the conclusion that "for cardiovascular-renal diseases as a whole, it is estimated that 70 to 75 percent of the deaths so classified may be considered as a reasonable inference or better."[9]

Most frequently cited as an overall index of the quality of the information on cause of death (though it may also reflect the physician's lack of thoroughness) is the proportion of death certificates that had to be coded to symptomatic and ill-defined causes. In 1971 this proportion was 1.3 percent—ranging from a high of 7.1 percent in one State to a low of 0.2 percent in another. In 1900-1904, when the death registration system was new, the figure for the registration area as a whole was 5.8 percent.

The proportion of death certificates showing that an autopsy was done is also a general indicator of the reliability of the evidence. The most recent figure available for this is 17.6 percent in 1975.[10] This proportion has not increased in recent years. Furthermore, it is uncertain whether the autopsy results were always available at the time the death certificate was filed. Nevertheless, this method of studying the reliability of the cause-of-death statement has been successfully used in the past.[11]

Information indicating unreliability in the underlying diagnostic data and the failure of some physicians to record conscientiously what they know should not be taken as evidence of failure of cause-of-death statistics to serve adequately the purposes for which they are intended. Despite the problems cited, more about the distribution of disease and injury in the population and about the changes taking place over time has been learned by studying cause-of-death statistics than from any other source. Errors can be found in any type of disease statistics; those in the cause-of-death data are no greater than those in other types of disease statistics and are not great enough to obscure the essential information that can be used in health planning.

## CALCULATION OF CAUSE-SPECIFIC DEATH RATES

The cause-specific death rate (CSDR) is defined as follows:[1]

$$\text{CSDR} = \frac{\text{number of deaths attributed to a cause in the population at risk during a period}}{\text{population at risk at mid-point of the period}} \times 1{,}000$$

In *Statistical Notes for Health Planners*, No. 3,[1] the CSDR was discussed and only two other comments need be made here. First, though often expressed per 1,000 population, cause-of-death specific rates are usually expressed per 100,000 population to avoid the use of zeros following the decimal point. For example, the U.S. death rate for tuberculosis of the respiratory system was 1.3 per 100,000 population in 1974 or .013 per 1,000. The latter is a little less convenient to write.

Second, the term "population at risk" must be taken literally in cause-specific rates and the rates should be computed and interpreted giving thought to the question: What is the population at risk from this disease? To illustrate: The death rate for malignant neoplasms of the breast was 15.3 per 100,000 population in 1974. This rate was based upon the total population since both males and females are "at risk" of breast cancer, although the rate among females was 29.6 and that for males was 0.3. Thus the rates specific for sex convey much more information for this particular cause of death.

However, the rate for complications of pregnancy, childbirth, and the puerperium was 0.4 per 100,000 female population. The corresponding rate for both sexes, 0.2, is seldom used, since males are not "at risk." In this case, the rate per 100,000 females 15-44 years of age, which is about 1.0, is even more meaningful for comparisons, since women in that age range are usually the only ones at risk.[c]

For some diseases it is only when the death rates are shown specific for age groups that full information is conveyed because the risk varies widely at different ages. For example, while any age is at risk of dying from cerebrovascular diseases, the death rate remains very low until about 55 years of age. At ages 55-64 the rate is about 100 per 100,000. At ages 65-74 years, it triples. At ages 75-84 years it triples again, and at ages 85 years and over the death rate from cerebrovascular diseases reaches 3,000 per 100,000 population per year!

Further discussion of the types of cause-of-death mortality indexes that are particularly appropriate to health planning is in the appendix.

Table 1 shows death rate trends for some leading causes of death in the United States for the past 25 years. This table is included simply to illustrate the type of statistics that can be put together readily from existing information sources for any Health Service Area or any major jurisdiction within the area for which population denominator data are

[c] Maternal mortality is often expressed as the number of maternal deaths per 10,000 live births. This is a more precise utilization of the "at risk" concept. Even more precise would be a rate with live births and still births as a denominator.

Table 1. Death rates (per 100,000 population) for certain leading causes of death: United States, 1950, 1960, 1970 and 1975

| Causes of death | 1950 | 1960 | 1970 | 1975 |
|---|---|---|---|---|
| All causes of death | 963.8 | 954.7 | 945.3 | 888.5 |
| Diseases of heart | 356.8 | 369.0 | 362.0 | 336.2 |
| Malignant neoplasms | 139.8 | 149.2 | 162.8 | 171.7 |
| Cerebrovascular diseases | 104.0 | 108.0 | 101.9 | 91.1 |
| Accidents | 60.6 | 52.3 | 56.4 | 48.4 |
| Motor vehicle accidents | 23.1 | 21.3 | 26.9 | 21.5 |
| All other accidents | 37.5 | 31.0 | 29.5 | 26.8 |
| Influenza and pneumonia | 31.3 | 37.3 | 30.9 | 26.1 |
| Diabetes mellitus | 16.2 | 16.7 | 18.9 | 16.5 |
| Cirrhosis of liver | 9.2 | 11.3 | 15.5 | 14.8 |
| Arteriosclerosis | 20.4 | 20.3 | 15.6 | 13.6 |
| Certain causes in early infancy | 40.5 | 37.4 | 21.3 | 12.5 |
| All other causes | 185.0 | 153.2 | 160.0 | 109.3 |

available. The causes of death can also be shown in greater detail and can be subdivided by age and sex and shown separately for the white and "all other" population. There is, in summary, a tremendous amount of cause-of-death data available to the health planner.

## SOURCES OF DATA AND POPULATION DENOMINATORS FOR PLANNING JURISDICTIONS

There is a section dealing with sources of mortality data in the *Statistical Notes for Health Planners,* No. 3.[1] The same sources will, of course, be appropriate for data on cause of death. The original death certificates are filed in the State health department's vital records office. In each State there is also an office responsible for producing statistical tables from death records. These offices code causes of death and publish the results. However, some States have fallen behind in producing these statistics. NCHS, through the Cooperative Health Statistics System, now helps support a majority of States in producing more uniform and more timely vital statistics. In return, the States will supply the data to NCHS on magnetic tapes. The State governments will also provide the information to the cities, many of the larger of which are now duplicating the production of these data. However, eliminating this duplication of effort cannot occur until the States are up to date in their work because those city health departments that now produce cause-of-death statistics need them promptly.

Over the years an efficient exchange system has been established between the States so that copies of death certificates for residents of State A for deaths occurring in State B are sent to State A. This enables each State to base its statistics on its own residents, no matter where the death occurs. This system should also work within each State so that a local health department knows about its residents' deaths occurring elsewhere. However, the success of this intrastate exchange varies from State to State.

Receiving copies of all death certificates or a magnetic tape record of each death enables NCHS to produce local statistics based on the decedent's place of residence. However, since these data are not complete until all copies of death records have been received, national statistics have not been as timely as they should be.

Another important source for health planning has recently become available. Through a contract with the Department of Community Health and Medical Practice of the University of Missouri, NCHS has financed the production of a data tape containing deaths and death rates for selected causes of death for white and Negro persons by sex and age for each county, State economic area, and State in the U.S. The statistics are for the combined years 1968-1972, and the population at risk was obtained from the 1970 enumerated population. The causes of death are grouped in 53 basic categories and 16 subtotals. These categories are not completely consistent with those published by NCHS.

These data tapes, probably the most comprehensive set of data on the health of the population currently available in equivalent geographic detail, are available through

Professor Herbert I. Sauer
University of Missouri
111 Professional Building
909 University Avenue
Columbia, MO 65201

Now NCHS is preparing an atlas of U.S. mortality rates. It will differ from the data set produced by the University of Missouri in several respects (having been compiled from deaths in 1969-1971 instead of 1968-1972, and being grouped by causes in the same manner as the annual reports, *Vital Statistics of the United States*). The atlas will also show considerable geographic detail.

The existence of appropriate population denominator data is a *sine qua non* for useful cause-of-death statistics. The numbers of deaths alone are not helpful in defining the health problems of a planning jurisdiction. However, some use has been made of indexes formed by computing the proportion of all deaths caused by particular diseases. For example, the proportion of all deaths within an age group caused by preventable diseases is an index of some interest. But it is difficult to interpret these figures in terms of the health of a community without knowing the size of the age group population. This matter will be dealt with in more detail in the appendix.

As indicated in a preceding report, detailed population breakdowns by age, race, and sex are not available for populations the size of a county or smaller in years when the census is not taken. Through 1973 the 1970 census data are probably adequate. Beyond that time estimates must be made to reflect population changes. Starting in 1985, the census is expected to be taken quinquennially instead of decennially, thereby reducing the seriousness of this problem.

The lack of needed population estimates is a difficulty pervading *all* planning, not simply health planning, and not simply the computation of death rates by cause of death. The new programs for updating local population estimates are helping to solve these problems.

## PLACE OF CAUSE-OF-DEATH DATA IN HEALTH PLANNING

In this report's introduction, an assumption was stated about certain responsibilities of the health planner that require him to examine measures of the health status of a community covered by the plans. It is in the context of those two responsibilities that the use of cause-of-death data should be examined.

To serve these two purposes, the health status data for a community should have six characteristics. They should be

- Available at regular and reasonably frequent intervals so that time comparisons can be made.
- Available on a comparable basis for other communities so that the particular community's health status can be compared with that of others.
- Available at a cost such that their value in planning substantially outweighs the cost of acquiring them.
- Sufficiently comprehensive to cover the wide range of health status problems a community in the United States is likely to face and to be representative of the entire community population.
- Compiled so that specific data can be extracted for use in focusing on particular health status problems such as diseases, injuries, or impairments.
- Compiled so that specific data can be extracted for use in focusing on particular demographic or socioeconomic groups.

Cause-of-death data evaluated by these six criteria rate well on the first three. They rate fairly well on the fourth since they are representative of the entire community population and cover a wide range of health status problems. However, there are major health status problems for which mortality is an inadequate index. These are the diseases and impairments associated with low mortality—that is, low in proportion to their importance in terms of disability, requirements for medical care, and cost to the community. These include mental diseases (of which suicide is the only significant fatal end result); arthritis; visual, hearing, speech and locomotion impairments; effects of malnutrition; and others of lesser importance.

Since coding causes of death makes it possible to disaggregate the data to show separate statistics for several specific health problems, the rating on the fifth criterion is also high. These cause-of-death rates can often be used to compare a disease problem in one community with other communities or over time even if the disease does not have a high fatality ratio. As long as the case fatality ratio does not differ widely from one area to another or has not changed drastically over 10 years, the *comparison* of death rates will adequately reflect the comparison of morbidity for the same disease.

Mortality statistics also rate well by the sixth criterion but with a proviso. Only the most basic demographic and practically no socioeconomic characteristics of the deceased person are available routinely on the death certificate. Age, race, sex, and marital status are reported, and most analyses are based on these. Such analyses help to show whether the health problem is concentrated in particular demographic groups. Occupation is reported but is not comparable with information about the exposed-to-risk population from the census.[12] Hence it is not coded routinely by NCHS.

The only way data can be analyzed by family income without elaborate matching studies is to assemble the statistics for small areas, such as census tracts, which are assumed to be homogeneous with respect to income and are characterized by use of census income figures for the tract.[13] It is also possible to add to the variables available for analysis by following back to secure further information from surviving members of the families of the decedents. Sample surveys of this type have been successfully completed, but they are expensive and time-consuming.

Despite these drawbacks, cause-of-death data rate better on the six criteria named than any other source of data. The time and space continuity and comparability that characterize these statistics are exceedingly important analytic tools in local health planning activities.

There are numerous examples of how changing trends in cause-of-death rates have called attention to emerging health problems. A familiar one is the rising rate of lung cancer mortality (at first, particularly among males) which led to programs to reduce cigarette smoking. Similarly, it was the increasing mortality from motor vehicle accidents that first led to planning strategies for dealing with this problem. While those are national examples, cause-of-death data for communities, particularly when analyzed in conjunction with statistics for neighboring or otherwise comparable communities, can also identify problems peculiar to a particular population.

Furthermore, when programs dealing with community problems are initiated, their degree of success must be evaluated. One test of their success is whether the health status problem being addressed has improved. An analysis of death rates by cause over time provides such a test. The simultaneous analysis of changes occurring in other communities can help pinpoint cause and effect relationships.

Comparisons with other jurisdictions are particularly useful to the planner since the performance of other communities can serve as a standard for identifying special problems and for a "control group" when seeking to determine whether a particular program initiative was responsible for the change that has occurred. However, precautions must be taken in geographic comparisons. Allowances must be made for differences in the characteristics of the population[1] and for other nonrelevant factors affecting the comparison. This includes the existence of a large resident institution in a community, which inflates its death rates for particular causes.

In addition to identifying special health problems and evaluating the effectiveness of countermeasures, cause-of-death statistics can be used as a basis for resource allocation; for cost-benefit analysis in deciding among program alternatives; and in budget formulation. There is not space in this note to go into the methods that have been used for turning cause-of-death statistics to these kinds of planning uses. However, the third report in this series[1] dealt with the construction of indexes of "years of life lost" based on mortality statistics. The same methods can be used for particular causes of death, and such indexes are particularly relevant in cost-benefit analyses. The discussion in that report of the relationship of mortality to incidence and case fatality should be read again in relation to cause-of-death statistics. The example given in that discussion of two strategies for dealing with motor vehicle accident mortality is an excellent one for illustrating both the usefulness of and precautions that must be taken in using cause-of-death data.

A general point that should be stressed in the use of statistics on causes of death for health indexes is that they must be used imaginatively and with a clear knowledge of the kind of health index needed for the purpose at hand. There are ways of disaggregating the data, or of using them selectively, to make them far more sensitive as indexes of particular health problems. They can even be used, as illustrated in the appendix, to measure the quality of health care in the community.

## REFERENCES


[1]National Center for Health Statistics: Mortality, by J. C. Kleinman. *Statistical Notes for Health Planners,* No. 3. Health Resources Administration. Rockville, Md., Feb. 1977.

[2]National Center for Health Statistics: *Vital Statistics Rates in the United States, 1940-1960,* by R. D. Grove and A. Hetzel. PHS Pub. No. 1677. Public Health Service. Washington. U.S. Government Printing Office, 1968.

[3]National Center for Health Statistics: *Physician's Handbook on Medical Certification: Death, Fetal Death, Birth.* PHS Pub. No. 593-B. Public Health Service. Washington. U.S. Government Printing Office, 1967.

[4]National Center for Health Statistics: *Eighth Revision International Classification of Diseases, Adapted for Use in the United States.* Vol. I, Tabular list, 1967; Vol. II, Alphabetical index, 1968. PHS Pub. No. 1693. Public Health Service. Washington. U.S. Government Printing Office.

[5]National Center for Health Statistics: Comparability of mortality statistics for the Seventh and Eighth Revisions, International Classification of Diseases, by J. A. Klebba and A. B. Dolman. *Vital and Health Statistics.* Series 2-No. 66, DHEW Pub. No. (HRA) 76-1340. Health Resources Administration. Washington. U.S. Government Printing Office, Oct. 1975. table I.

[6]Gordon, B. L., ed.: *Current Medical Information and Terminology,* 4th ed. Chicago, American Medical Association, 1971.

[7]Spiegelman, J.: *Introduction to Demography,* Rev. ed. Cambridge, Mass., Harvard University Press, 1968, ch. 3.

[8]Moriyama, I. M., et al.: Inquiry into diagnostic evidence supporting medical certifications of death. *Am. J. Public Health* 48: 1376-87, Oct. 1958.

[9]Moriyama, I. M., et al., Evaluation of Diagnostic Information Supporting Medical Certification of Cardiovascular Disease Deaths. Unpublished paper presented at American Public Health Association meeting, Kansas City, Mo., Nov. 13, 1963.

[10]National Center for Health Statistics: Final Mortality Statistics, 1975. *Monthly Vital Statistics Report—Advance Report,* Vol. 25, No. 11, Supp. DHEW Pub. No. (HRA) 77-1120. Health Resources Administration. Rockville, Md., Feb. 11, 1977.

[11]James, G., et al.: Accuracy of cause-of-death statements on death certificates. *Public Health Rep.* 70(1): 39-51, Jan. 1955.

[12]Kitagawa, E. M., and Hauser, P. M.: *Differential Mortality in the United States: A Study in Socioeconomic Epidemiology.* Cambridge, Mass. Harvard University Press, 1973.

[13]National Center for Health Statistics: Selected vital and health statistics in poverty and non-poverty areas of 19 large cities, by S. J. Ventura et al. *Vital and Health Statistics.* Series 21-No. 26. DHEW Pub. No. (HRA) 76-1904. Health Resources Administration. Washington. U.S. Government Printing Office, Nov. 1975.

[14]Guralnick, L. and Jackson, A.: An index of unnecessary deaths. *Public Health Rep.* 82(2): 39-51, Feb. 1967.

[15]Rutstein, D. D.: *Blueprint for Medical Care.* Cambridge, Mass. The Massachusetts Institute of Technology Press, 1974. ch. 13.

[16]Rutstein, D. D., et al.: Measuring the quality of medical care—a clinical method. *N. Engl. J. Med.* 294:582-588, Mar. 1960.

# APPENDIX

# SUMMARY INDEXES OF HEALTH MAKING USE OF CAUSE-OF-DEATH DATA

## INTRODUCTION

In suggesting indexes of the overall health of a community's residents, based upon cause-of-death statistics, this report will again take up where *Statistical Notes for Health Planners*, No. 3,[1] left off. In that report the following passage appears:

"The importance of general mortality data to the health planner is primarily as a clue to other health problems. A large part of mortality is caused by illness for which there is little effective medical care intervention. Factors such as socioeconomic status which affect mortality may be beyond the ability of the health planners to modify, at least in the short run."

Thus, indexes of general health based upon mortality data should be sensitive to the mortality that can be influenced by high quality medical care available to all. The indexes should not be affected by factors such as age-sex structure of the population for which the health care system can not be held responsible.

## THE SMR AND CMF FOR CAUSES OF DEATH

The building blocks of the SMR and the CMF are the age-sex-specific death rates because these rates are not affected by the age-sex distribution of the population. Hence, they are more suitable for planning purposes than the crude death rate of the community.[1] However, in attempting to summarize the health of a community using age-sex-specific death rates, the planner encounters the problems of (1) the lack of stability resulting from small numbers of deaths in the numerators of some of the rates and (2) the inherent difficulty of forming judgments based on examining a large array of data.

Furthermore, as pointed out earlier, much of the mortality is caused by illness for which medical care currently has little to offer.

Using an array of age-sex- and cause-of-death-specific rates helps in pinpointing disease and injury problems for which the health care system is responsible. However, the drawbacks of small numbers and large arrays of data are exacerbated.

The previous statistical note[1] showed how the crude death rate and the death rates adjusted by the indirect and direct methods can be expressed as weighted averages or ratios of weighted sums. These indirect and direct methods include their respective companion ratios, the standard mortality ratio (SMR) and the comparative mortality figure (CMF). The elements of all these ratios are the community's age-sex-specific death rates and weights consisting of the proportion of the population or that of the standard community.

$$\text{crude death rate} = \frac{d}{p} \times 1{,}000 = \sum m_a \frac{p_a}{p}$$

$$\text{SMR} = \frac{\sum m_a \frac{p_a}{p}}{\sum M_a \frac{p_a}{p}}$$

which is proportional to the age-adjusted death rate, indirect method, and

$$\text{CMF} = \frac{\sum m_a \frac{P_a}{P}}{\sum M_a \frac{P_a}{P}}$$

which is proportional to the age-adjusted death rate, direct method.

In these equations, $d$ stands for numbers of deaths and $p$ represents the community's population. The subscript $a$ is for an age group, while $d$ and $p$ without the subscript $a$ stand for the total for all age groups. The $m_a$ is the age-specific death rate in the community; $M_a$ is a standard set of age-specific death rates, often taken to be those for the United States. Lower case $p$ refers to the population of the community while $P$ is the population in the standard community, again, usually the United States as a whole. The summation $\Sigma$ is over all age groups.

The SMR and the CMF are summary indexes of mortality that have been commonly used to hold constant the effect of differing age distributions when comparing mortality between two communities or the same community at two different points in time. These indexes have also been applied to time or space comparisons for a particular cause of death. While they do remove the effect of differing age distributions, they are *not* sensitive to that part of the mortality that can be influenced by high quality medical care available to all.

The reasons for this disadvantage of the SMR and the CMF are explained in the appendix to Statistical Note No. 3.[1] It is stated there that "for use in planning, the emphasis on the elderly is unfortunate, since *mortality* in this group is probably least amenable to planning intervention." The reason for the emphasis on mortality among the elderly in these indexes is that the numerator of the index, which is the part of the index measuring the health of the subject community, contains the sum over all age groups of either the actual number of deaths in the subject community or the number of deaths that would be expected in the standard community population if the subject community death rates by age were experienced.

The insensitivity of the SMR and the CMF becomes even more evident when particular health problems using cause-of-death data are examined. Table I presents an extreme case of deaths from intracranial vascular lesions in Santa Cruz County, California, and DeKalb County, Georgia (the same counties used in Statistical Note No. 3[1] to illustrate the effect of different indexes). Note the large proportion of deaths in the highest age groups.

In the United States as a whole, there were nearly 100,000 deaths from intracranial vascular lesions among persons 80 years and over in 1974. This is nearly one-half of all deaths from this group of causes and is about 5 percent of all deaths at any age from all causes. Yet what can medical science, public health, and the best possible health delivery system do to reduce deaths from intracranial vascular lesions at age 80 years and over?

## THE YEARS-OF-LIFE-LOST INDEX FOR CAUSES OF DEATH

Health indexes suitable for health planning purposes must therefore avoid forms that depend on the summing up of deaths for individual causes of death. One alternative, suggested in the earlier report, is an index based on "years of life lost" by all persons dying before the age of 70 years. Each death before age 70 is assigned years of life lost equal to the difference between the actual age at death and age 70; deaths at age 65 years and over are omitted. Thus the younger the age at which death occurs, the more heavily that death influences the index.

The effect of this emphasis on deaths at younger ages is shown in Table I. Years of life lost (YLL) from intracranial vascular lesions and expected years of life lost are shown in columns 4 and 6 for Santa Cruz County and in columns 10 and 12 for DeKalb County. While the SMR for DeKalb County exceeds that for Santa Cruz, the YLL indexes for the two are almost identical. This results from the effect upon the SMR of the higher DeKalb County death rates from stroke at ages 65 years and over. These rates have no effect on the YLL index.

The years-of-life-lost index is computed as the ratio of the actual years of life lost in the subject community to the expected years of life lost in that community if age-specific death rates from a standard community were being experienced. The index is equally appropriate for comparing total mortality or mortality from a single cause of death over time or between communities.

## INDEXES BASED ON "PREVENTABLE" OR "UNNECESSARY" DEATHS

In considering how mortality data can best be used to measure a community's degree of health, several authors have proposed a direct approach to the question: which deaths are preventable? If a community's health could be measured in terms of its preventable deaths, such an index would be optimally sensitive to those aspects of the mortality that *can* be affected by the intervention of plan-

Table I. Deaths and average annual death rates (per 100,000 population) from intracranial vascular lesions (ICD nos. 430-38) among white males in Santa Cruz County, California, and DeKalb County, Georgia: 1968-72

**Santa Cruz, California**

| Age | Population (1970) | Number of deaths[1] (1968-72) | Average death rates[1] | Years of lost life[2] (1968-72) | Expected deaths[3] (1970) | Expected years of life lost (1970) |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Under 1 year | 847 | 0 | 0.0 | 0 | .04 | 2.8 |
| 1-4 years | 3,274 | 0 | 0.0 | 0 | .03 | 2.0 |
| 5-14 years | 10,279 | 0 | 0.0 | 0 | .06 | 3.6 |
| 15-24 years | 10,201 | 1 | 2.2 | 50 | .15 | 7.5 |
| 25-34 years | 6,413 | 2 | 6.9 | 80 | .22 | 8.8 |
| 35-44 years | 5,661 | 3 | 11.8 | 90 | .65 | 19.5 |
| 45-54 years | 6,051 | 8 | 29.4 | 160 | 2.18 | 43.6 |
| 55-64 years | 5,402 | 28 | 115.2 | 280 | 6.44 | 64.4 |
| 65-74 years | 4,920 | 96 | 433.6 | ... | 21.06 | ... |
| 75-84 years | 2,781 | 153 | 1,222.6 | ... | 38.45 | ... |
| 85 years and over | 649 | 95 | 3,252.9 | ... | 21.84 | ... |
| All ages | 56,478 | 386 | 151.9 | 660 | 91.12 | 152.2 |

**DeKalb, Georgia**

| Age | Population (1970) | Number of deaths[1] (1968-72) | Average death rates[1] | Years of lost life[2] (1968-72) | Expected deaths[3] (1970) | Expected years of life lost (1970) |
|---|---|---|---|---|---|---|
| | (7) | (8) | (9) | (10) | (11) | (12) |
| Under 1 year | 3,325 | 2 | 13.4 | 139 | .14 | 9.7 |
| 1-4 years | 12,669 | 2 | 3.5 | 134 | .11 | 7.4 |
| 5-14 years | 37,883 | 2 | 1.2 | 120 | .23 | 13.8 |
| 15-24 years | 29,136 | 1 | .8 | 50 | .44 | 22.0 |
| 25-34 years | 27,383 | 3 | 2.4 | 120 | .93 | 37.2 |
| 35-44 years | 24,223 | 10 | 9.2 | 300 | 2.79 | 83.7 |
| 45-54 years | 20,201 | 34 | 37.4 | 680 | 7.27 | 145.4 |
| 55-64 years | 11,128 | 43 | 85.9 | 430 | 13.28 | 132.8 |
| 65-74 years | 4,901 | 96 | 435.3 | ... | 20.98 | ... |
| 75-84 years | 1,867 | 158 | 1,880.6 | ... | 25.82 | ... |
| 85 years and over | 335 | 60 | 3,980.1 | ... | 11.27 | ... |
| All ages | 173,051 | 411 | 52.8 | 1,973 | 83.25 | 452.0 |

[1]NCHS processed only a 50-percent sample of death certificates in 1972. The deaths shown are the sum of all deaths in 1968-71 plus a half of the 1972 deaths. To compensate for this, the average number of deaths per year was computed by dividing this number by 4.5 before computing the death rate. The population shown is the enumerated population in April 1970.

[2]Years of life lost is computed from the deaths in column (2) as explained in the text.

[3]Expected number of deaths and years of life lost are computed by applying U.S. values to the county population in 1970.

Santa Cruz: $SMR_S = \frac{386 \div 4.5}{91.12} = .94$ DeKalb: $SMR_D = \frac{411 \div 4.5}{83.25} = 1.10$

$YLL_S = \frac{660 \div 4.5}{152.2} = .96$ $YLL_D = \frac{1,973 \div 4.5}{452.0} = .97$

ning and an improved medical care delivery system.

There have been two general lines of attack on the problem of determining the preventable deaths.[d] The first and oldest is to examine age-specific or age-cause-specific death rates in a wide range of communities and to determine the best (i.e., lowest) death rates achieved *somewhere*. The assumption is that within particular age groups or age-sex groups of the population the lowest rates that *have been* achieved *could be* achieved anywhere. Hence deaths in the subject community in excess of those that would occur if the optimum death rates were experienced represent preventable deaths. In one form or another this idea dates back to William Farr, the Registrar-General of England, in the middle of the last century.

Guralnick and Jackson, in their paper "An Index of Unnecessary Deaths," present an exposition of one form of this method.[14] Its benefits are credibility, simplicity, and availability (since such indexes can be easily computed from currently available data).

The other line of attack is newer and more sophisticated, with respect to the age-cause-specific death rates used as criteria. This method uses the judgments of experts in medical care and biomedical research to determine the lowest death rates within each age and cause-of-death group that could be achieved by (1) the best scientific knowledge available or likely to be available by a reasonable projection of present biomedical research efforts and (2) the application of this knowledge in an optimum system of health care delivery to all people. The use of data on "unnecessary untimely deaths" as one basis for a health program guidance system has been strongly argued by Rutstein.[15]

Presumably, the age and cause-of-death rates formed by such expert judgment would be at least as low as the values of the cause-specific rates used in the Guralnick-Johnson index. These judgments might rely heavily on time trends of the death rates projected to some future date as well as on the knowledge of the experts. Once the age and cause-of-death rates had been determined, however, they could probably be used as a standard until further research rendered them obsolete.[16]

While there is no agreement yet on a *complete* list of entirely preventable causes of death or on the extent to which other causes might be reduced by the full application of present knowledge, there is a broad consensus that deaths from certain causes should not occur. A definite improvement in the usefulness of cause-of-death statistics could be accomplished quite simply and quickly by applying that consensus to the analysis of the statistics now available for local areas to define their particular health problems.

[d] The terminology has varied among authors. Some refer to "preventable" deaths; others to "unnecessary" deaths; and others to "unnecessary, untimely" deaths. However, the underlying concept seems the same.

## SYMBOLS

| | |
|---|---|
| Data not available | - - - |
| Category not applicable | . . . |
| Quantity zero | - |
| Quantity more than 0 but less than 0.05 | 0.0 |
| Figure does not meet standards of reliability or precision | * |

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service
National Center for Health Statistics
3700 East-West Highway
Hyattsville, Maryland 20782

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

DHEW Publication No. (PHS) 78-1237

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF HEW
HEW 390

FIRST CLASS MAIL

U.S. MAIL

Statistical Notes for Health Planners is a cooperative activity of the National Center for Health Statistics, Office of the Assistant Secretary for Health, and the Bureau of Health Planning and Resources Development, Health Resources Administration.

Information, questions, and contributions should be directed to Mary Grace Kovar, Division of Analysis, NCHS, Room 2-27, Center Building, 3700 East-West Highway, Hyattsville, Md. 20782.