NCHS Data and
Studies of Differential
Morbidity and Mortality

LOIS M. VERBRUGGE

University of Michigan

Public conversation, newspapers, and magazines contain fasci-
nating speculations about why some social and demographic
groups are more ill or die earlier than others. Increasingly one
finds strong statements about how changing roles and lifestyles
are affecting the health and mortality of population groups.
Researchers are hard-pressed to confirm or refute these specu-
lations. Some of the difficulty lies with data sources—for ex-
ample, lack of appropriate items or absence of tables for collected
items. Some of it lies with researchers who have not energetically
exploited available data.

This article considers three national data sources for social
and demographic differentials in morbidity and mortality.[1] My
aim is not to review known differentials, but to discuss charac-
teristics of data sources for those differentials.

The United States has numerous sources of national health
data and several for mortality data. Most of them have large
restrictions: for example, they have limited coverage of the total
U.S. population; they emphasize health facilities or health man-
power rather than health status of the population; they are
conducted infrequently or only once. I shall discuss three sources
of health and mortality data which have minimal restrictions: the

Health Interview Survey (HIS), the National Ambulatory Medical Care Survey (NAMCS), and the death registration system. All three broadly cover the national population, provide ample data on health problems or the death event, and are continuous data collection programs. The three programs are conducted by the National Center for Health Statistics.

I shall consider the *adequacy* of these data sources and *access* to them. By adequacy, I mean how well the data can measure sociodemographic differentials in morbidity and mortality. I discuss the coverage of health, mortality, and sociodemographic items now collected; their validity and interpretation; and the data's ability to explain group differentials. By access, I mean how soon and completely collected data are made available. Specifically, I discuss the ease of securing tabulations or primary data; whether available data are a subset of collected items or a complete set; how soon tabulations and tapes are ready; and how the public is informed about the data resources.

I shall point out fine features of the three programs for data adequacy and access, as well as current problems. In addition, I suggest how researchers can exploit available data for studying differential morbidity and mortality. Solution of problems and a more aggressive use of available data will help NCHS staff, health planners, and researchers provide better answers to an eager public.

Three NCHS Data Programs

Following is a brief introduction to HIS, NAMCS, and the death registration system.

1. *The Health Interview Survey (HIS)* is a population-based survey which obtains data on the health status of Americans. Conducted continuously since 1957, it is a household sample of the civilian noninstitutionalized population. At each sampled address, data for all household members are collected. Supplemental items are added to the standard HIS interview each year to secure data on particular health problems, health behaviors, or personal characteristics. For details about the survey design, see National Center for Health Statistics (1975).

2. *The National Ambulatory Medical Care Survey (NAMCS)* is a record-based survey. Started in 1973, it obtains data on visits to doctors (the most common type of health services use by the population). NAMCS is a sample of office-based physicians who provide care to ambulatory patients. For each sampled physician, special records are kept for office visits during a one-week period. For details about the survey design, see Tenney et al. (1974).

3. *The national death registration system* was initiated in 1900 and has included all states since 1933. Its aim is both legal (to provide documentation of death) and statistical (to provide information about death events and decedents). The system covers all deaths which occur in the United States. For details about the death registration system, see the Technical Appendix of *Vital Statistics of the United States*, Vol. II, Part A (National Center for Health Statistics, 1979).

Description and Explanation

Research on differential morbidity and mortality asks two main questions: How do population groups differ, and why do they differ? Descriptive data for the first question are used for reports on the nation's health, projections of future health and health care needs, and documenting trends across time. These data prompt speculation about causes, and they motivate research to test hypotheses. Analytic or explanatory, results can be used by health planners who wish to change health and health behavior. And they add to scientific knowledge about the population's health and health behavior.

To *describe* group differentials, a data source must measure characteristics of people who are ill or who have died. What criteria are used to choose the social and demographic variables for a data source? First, certain sociodemographic characteristics must be included simply to provide a picture of the nation's health. Second, it is important to identify population groups that experience especially poor health or early death so that public programs can be aimed toward them. Usually we have a good idea which groups are disadvantaged before a survey is

Intervening Variable

Sociodemographic ⟶ "Risk" ⟶ Morbidity
Status   or Mortality

Selection Variable

"Risk" ⟶ Sociodemographic ⟶ Mortality
              Status
         Morbidity

Associated Variable

Sociodemographic
Status
                              Morbidity
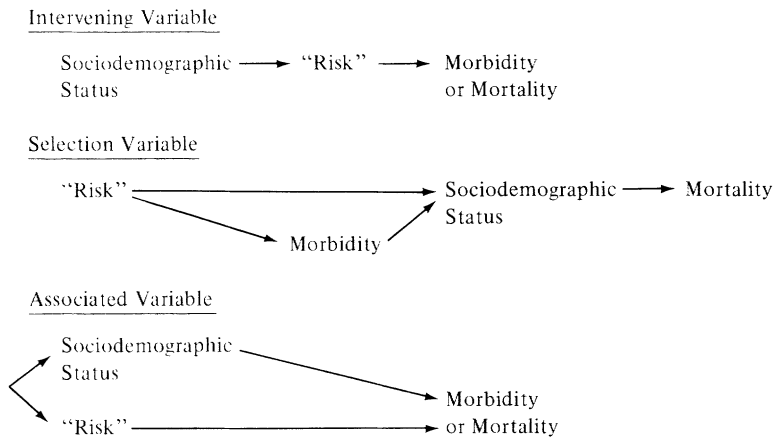"Risk" ⟶   or Mortality

Figure 1

designed, so choices are not made "in the dark." Based on these two criteria, characteristics that always rank high for inclusion are age, sex, race, marital status, socioeconomic status (education, occupation, income), and employment status. Increasingly, living arrangement is also considered a strong candidate for inclusion.

Explaining group differentials requires an additional group of variables. These are personal characteristics which are (1) correlated with sociodemographic status and also (2) causally related to morbidity or mortality.[2] The variables are of three types: intervening, selection, and associated. Intervening variables claim that being in a particular status exposes an individual to greater (or lesser) risks of poor health or death. Selection variables claim that people with certain health risks or health problems get into particular statuses. Associated variables simply show a correlation between risk and status, with no claim about their causal order. The three types are diagrammed in Figure 1.

The unifying concept for the explanatory variables is "risk." Explaining group differentials involves identifying risks which differ notably across groups and which are also important determinants of health and death.

The best explanations demonstrate a causal ordering of the risk, sociodemographic status, and morbidity/mortality out-

come. The diagrams above for intervening and selection variables are examples. At the other pole are associations among variables without any causal order. Usually, explanations of human behavior are somewhere in between, having credible causal orderings for some of the variables. The diagram for associated variable is an example.

It is difficult to find risk variables that fulfill the two criteria stated above. We do know a good deal about how suspected risks vary among population groups, but not necessarily the causal order of the risk and the status. Relatively little is known about which personal behaviors influence chances of illness, injury, and death—in other words, which behaviors are actually risky. Without more knowledge about these causal links (between risks and dependent variables), it is too costly to include suspected risk variables on a routine basis in national data collection programs.

Adequacy of Morbidity and Mortality Data

I consider the following four aspects of adequacy.

(1) Coverage of Morbidity and Mortality

What indicators are now collected? Is anything important omitted?

(2) Validity and Interpretation of Morbidity and Mortality Items

What do the indicators measure? Do they measure it accurately? Do interpretations of indicators remain close to their actual content?

(3) Coverage, Validity, and Interpretation of Sociodemographic Items

What characteristics are queried? Are they all useful? Are any important ones missing? Are the included items valid and clearly defined?

(4) Explanatory Variables

What measures of risk are included in the data? Should any be added?

COVERAGE OF MORBIDITY AND MORTALITY

What features of illness, injury, and death are now recorded in HIS, NAMCS, and death certificates?

HIS collects data for "major" health events. These are health problems that cause some change in social role performance, physical activity, or mobility or that create the need for medical or dental care. Specifically, HIS routinely asks about illness and injury experienced in the past two weeks, restricted activity in that period, health services use in the past two weeks and past year, and long-term disability due to chronic disease or impairment. This leaves out the bulk of health problems people experience—symptoms that cause no disability or medical care. Examples of absent indicators are (a) symptoms that bother people but cause no curative actions, (b) symptoms that are self-treated and involve no contact with health services, and (c) preventive health behaviors, such as taking daily vitamins. For a copy of a recent HIS questionnaire, see Howie and Drury (1979).

NAMCS provides data on patient-physcian encounters. It contains sparse but carefully selected information about patient complaints, physician diagnoses of those complaints, services provided or ordered, and dispositions for further care.

Death certificates contain information about the death event and its setting. The principal items are cause(s) of death, autopsy status, site of death (hospital, home, etc.), and details about deaths from injury.

Adequacy of coverage should be evaluated according to each program's goal. Overall, NAMCS and death registration fulfill their aims well. HIS strives to measure Americans' health in a comprehensive way, but it ultimately measures only a small fraction of all illness and injury events and all health behaviors of the population.

VALIDITY AND INTERPRETATION OF
MORBIDITY AND MORTALITY ITEMS

Validity refers to how accurately an indicator measures a real-world event or a theoretical concept. In the past two decades, NCHS has evaluated the validity of some HIS items, comparing

interview reports with medical records. Some studies concern the validity of health services use and health expenditures (Cannell, 1965; Cannell and Fowler, 1965; National Center for Health Statistics, 1966). Others look at validity of chronic condition reports (Balamuth, 1965; Madow, 1967, 1973). In general, the validation studies conducted by NCHS and its companion agency, the National Center for Health Services Research, emphasize utilization and costs rather than health conditions.

I believe it is equally important to know about the quality of illness and injury reports. This can be illustrated for the three NCHS data sets.

1. HIS obtains extensive detail about diseases, injuries, and impairments. The validity of these reports can be studied without recourse to medical records.

First, HIS rates for specific *chronic conditions* can be compared with rates from the Health Examination Survey/Health and Nutrition Examination Survey (HES/HANES) and the Hospital Discharge Survey (HDS). These surveys use different definitions for their rates: HES/HANES medical histories provide rates for previously diagnosed conditions. ("Has a doctor ever told you that you had . . . ?" "Do you still have . . . ?") HES/HANES examinations show the clinical presence of conditions. HDS rates are for people who receive hospital care for particular conditions. HIS rates are based on conditions that cause some change in activities. These differences in definition may frustrate researchers who hold a strictly medical perspective of health. But for those with a sociomedical view, the comparisons are fascinating—comparisons of not only rates but also group differentials across the data sets. The differentials may vary if groups differ in their chances of diagnosis and their propensity to take curative actions for symptoms. Some recent HIS reports show comparisons with HANES data.

Second, some validity assessment can be done internally in HIS data. Chronic conditions are sometimes queried by a checklist. A checklist item encompasses a range of medical codes. When specific responses are coded, some will fit that range and some will not. The degree of agreement of actual codes with "legitimate" ones is a measure of validity for the checklist item.

Acute conditions are more difficult to validate through medical records or other surveys. Internal consistency checks, however, are feasible. Some imaginative thought should be given to the validation of acute conditions, since they surpass chronic ones in their frequency.

2. NAMCS offers a unique opportunity to compare self-reports and physician reports from the same record. Complaints are written in the patient's own words, and physicians later enter their diagnoses. From a medical view, one can use the diagnoses as criterion variables and then measure "errors" in symptom reports. From a sociomedical view, one can study the illness descriptions for patients with particular diagnoses and also how physicians attribute diagnoses to particular complaints—making no assumptions that diagnoses are "correct." (For an example of such analysis, see Verbrugge, 1979a.)

3. The validity of cause-of-death reports has not been investigated in a comprehensive way. (See Moriyama et al., 1966, for a fine but limited study.) NCHS recognizes the need for such research and has proposed a continuing study of the quality of medical certification for death records. Even if no diagnostic errors occurred, physician vocabulary changes over time. The same physical signs can be shifted to different medical names (and ultimately different medical codes).[3] A continuing study of medical certification would readily reveal such changes.

NAMCS and death registration have no important problems for interpretation of the morbidity and mortality variables. But there are two long-standing ones for HIS.

First, HIS incidence and prevalence rates are often interpreted as measures of condition *presence*. Actually, acute incidence rates are based on conditions that cause either restricted activity or medical attention. Chronic conditions are counted if they limit activities or mobility. Thus, both acute and chronic condition rates are based on more than just condition presence. Although HIS reports state the definitions, readers still misinterpret the rates.

Second, the HIS activity limitation rates (for chronic conditions) are especially troublesome to interpret. People are asked what their usual role was in the past year, and whether

they are currently limited in doing it because of health.[4] This means that the criterion for limitation varies among individuals. Moreover, severe health problems cause some people to change their usual role. After that adjustment, they may report less limitation. To avoid this nonuniformity across people and across time, NCHS should consider measures of chronic disability based on *absolute* criteria of disability.

In summary, the validity of condition and cause-of-death reports needs more research. The topic may be approached from a purely medical view, which requires diagnostic data for a criterion. Or it can be approached from a sociomedical view, which looks at relationships among multiple measures of a condition.

At the very least, users of NCHS data must be fully informed about the measurement and coding procedures for conditions and causes of death. They must be constantly reminded about the meaning of items prone to misinterpretation. The Technical Appendices of HIS, NAMCS, and vital statistics reports are excellent devices, but problems still arise, especially for HIS.

COVERAGE, VALIDITY, AND INTERPRETATION OF SOCIODEMOGRAPHIC ITEMS

HIS offers the widest *coverage* of sociodemographic items. On a routine basis, it has age, sex, race/ethnicity, marital status, education, employment status, current occupation and industry, class of worker, family and personal income, current social role ("usual activity"), usual role in the past 12 months, living arrangement, family relationship, family size, family structure, veteran status, and national origin/ancestry.[5] The sociodemographic items in NAMCS are limited to patient's age, sex, and race.[6] Death registration has age, sex, race, marital status, and usual occupation. The variables common to all three data sets are age (date of birth), sex, and race.

Are these sociodemographic variables adequate? There is no simple answer. Inclusion of an item greatly enhances the descriptive potential of a data set, but it also boosts costs. The main function of the NCHS programs is to measure morbidity

and mortality adequately. Items about health and death events therefore take precedence over sociodemographic items. Nevertheless, one can still ask if any important variables are missing, how useful the included variables are, and how valid and interpretable they are.

Researchers have pointed to some very desirable but absent variables: NAMCS would have greater descriptive potential if marital status and a socioeconomic variable were included. Feasibility studies for NAMCS did include marital status and a general rating of socioeconomic status, but the items were dropped from the final form because missing data were too frequent (Tenney et al., 1974). Death certificates might include education instead of usual occupation, since education is a fixed characteristic for most adults. (For discussion of this issue, see Kitagawa and Hauser, 1973, and Kovar and Weed, 1977.)

Without exception, all of the included variables are interesting and important, fulfilling the two criteria stated earlier.

Relatively little attention has been given to the *validity* of sociodemographic items in the three data sets. No assessments for the HIS and NAMCS items have been made. One can assume that HIS errors are similar to those reported for census data, since HIS uses the same questions as the Census Bureau, and the Census Bureau is the collecting agency for HIS.

For death certificates, the validity or several demographic items (age, marital status, race) has been studied (Hambright, 1968, 1969). The validity of occupation reports for deaths in 1950 was assessed by matching census records and death certificates (Guralnick, 1959; Kaplan et al., 1961). Currently, occupation is seldom coded from the death certificates (only 12 states do so). Because of the great demand for mortality rates by occupation, NCHS is now examining the feasibility of coding occupation (Spirtas et al., 1979). (Note that this is a study of "codability," not of "correctness" of occupation data.)

Several problems of *interpretation* exist for sociodemographic items. First, occupation is a social status variable. Whether it can be viewed as a measure of risk (for example, exposure to job hazards and job-related stresses) is a difficult issue. It is especially pertinent for death certificates that record "usual

occupation." Usual occupation can be viewed as a measure of lifetime social status or of long-term exposure to risks attached to occupation (see Kitagawa and Hauser, 1973; Moriyama and Guralnick, 1956). There is no simple answer to this question.

Second, trends in differentials may reflect changes in risks that population groups experience. But sometimes they reflect changes in group composition. For example, the never-married group now includes more people who live together as-if-married than two decades ago. This can alter marital differentials even though the behaviors of "traditional" never-married people and "new" ones (who previously would be married) are unchanged. Researchers must be alert to the possibility of group composition effects like this.

In summary, the sociodemographic variables routinely included in the NCHS data sets should be scrutinized regularly, even though additions and deletions cannot (and should not) be made often. There is an increasing need, however, for review and change *within* included variables. This is due to the increasing need for standardization in the wording, coding, and tabulation categories of sociodemographic variables, to facilitate comparisons across federal data sets (U.S. Department of Health, Education and Welfare, 1977). The Social Science Research Council has prepared recommended wordings for items (Van Dusen and Zill, 1975). In the government, the Office of Federal Statistical Policy and Standards (1978) is working to unify practices of federal agencies. The need for standardization was recently pointed out by the Technical Consultant Panel which reviewed the Health Interview Survey (Greenberg, 1979).

EXPLANATORY VARIABLES

Are there measures of "risk" in the NCHS data sets? Is there information about stress, health habits, lifestyle behaviors, health attitudes, job hazards, and so on which might explain group differentials?

The three data sets routinely include very few risk variables. Each year, HIS asks if people receive federal assistance for medical care, disability, and income. (These can be viewed as

measures of risk, since they influence people's curative health actions.) NAMCS and death certificates have no other patient characteristics than the sociodemographic ones already mentioned.

Supplemental questions attached to the standard questionnaire are often used by HIS to obtain data on health-related characteristics. Recent examples are smoking behavior (1977), drinking behavior (1977, 1978), health habits (1977), exercise (1975), and health insurance coverage (1974, 1976, 1978). NAMCS has recently adopted this strategy also. The 1978 and 1979 surveys include supplemental questions (for one-year periods) on the Patient Record. Possible alternatives are to gather additional patient items from the physician's office records or to ask patients to fill out a short questionnaire at the time of visit.

From time to time, mortality risk variables have been collected through follow-back surveys. The 1964-1966 Infant Mortality Survey included such risk variables as birthweight and mother's health insurance coverage. The 1966-1968 Mortality Survey had questions on smoking behavior.[7]

Another way to obtain risk variables is to link death certificates with other data sources such as census or social security records. These contain extensive information about decedents. Some of these can be used as measures or risk. For a study based on matched census and death records, see Kitagawa and Hauser (1973). (In this study, census variables were used as measures of socioeconomic status, not risk.) An ambitious project is now under way to link social security, Internal Revenue Service, census, National Cancer Institute, and death records (Alvey et al., 1979; Kilss and DelBene, 1979).[8]

Overall, the decision to exclude explanatory variables from routine data collection in the NCHS programs is a good one. Obtaining them for selected periods provides ample opportunities for explanatory research. HIS is most active in this effort. NAMCS will soon know how willing physicians are to complete supplemental questions on the Patient Record. Another mortality follow-back survey with items on lifestyle behaviors

and exposure to chemicals and pollutants is acutely needed. Record linkage studies are arduous, but they yield excellent data about decedents.

SUMMARY

The coverage of morbidity and mortality indicators is generally fine, although HIS omits "minor" health events that are self-treated or cause no curative action. Little information is available on the validity of morbidity and mortality indicators. It is especially urgent to evaluate the validity of cause-of-death reports. The Mortality Statistics Branch of NCHS hopes to do so soon. For HIS and NAMCS, there are good opportunities to compare morbidity indicators within each data set and across health data sets. Misinterpretation of acute incidence and chronic prevalence rates is a long-standing problem for HIS and can be remedied by aggressive efforts in survey reports.

The coverage of sociodemographic variables is limited, but the choices are good. There is a critical need to standardize item wording, coding, and categories in federal data. NCHS must be willing to make changes for this purpose even though they may upset time trends in a data set. The validity of sociodemographic items has not been assessed for HIS or NAMCS, but it is reasonable to believe their validity is similar to census data. For death certificates, the quality of some sociodemographic items has been evaluated, but a study of occupation reporting is needed. There are few interpretation problems for the items.

Questionnaire supplements, follow-back surveys, and record linkage studies add information on risks (and also sociodemographic characteristics). This enhances both the descriptive and analytic potential of a data set. Supplements are an active and exciting feature of HIS. NAMCS is now experimenting with supplemental questions on the Patient Record. For mortality, a remarkable record linkage study is under way. This, however, is not a substitute for information from a follow-back survey, and another survey of decedents is sorely needed.

Access to Morbidity and Mortality Data

I consider four aspects of access:

(1) Ease of Obtaining Data

How readily available are tables and tapes for the NCHS data sets?

(2) Items Available

Are all of the sociodemographic, morbidity, and mortality items from respondent records actually used in published tables or on tapes?

(3) Timeliness

How soon do tabulations with sociodemographic differentials appear? How soon are public use tapes available?

(4) Public Information Program

Does NCHS disseminate tables and information about tapes widely?

My discussion of access is quite brief. A fuller discussion may be found in Hattwick (1979).

EASE OF OBTAINING DATA

Sociodemographic differentials from the three NCHS programs can be obtained in published tables, tabulations based on public-use tapes, and unpublished tables provided by NCHS or private vendors.

Published reports for HIS, NAMCS, and mortality are all readily available. Reports on HIS (*Vital and Health Statistics*, Series 10) are a superb resource for sociodemographic differentials. They are costly to prepare and are not sufficiently used by researchers, but they have immense importance for documenting the nation's health from year to year. The format of the HIS reports has remained quite constant—a reflection of care and foresight by NCHS staff many years ago. This constancy provides fine opportunities for the description and analysis of trends. Reports on NAMCS (*Vital and Health Statistics*, Series 13; and

issues of *Advance Data*) are developing a consistent format and are also fine resources, limited only by the scarcity of socio-demographic items available. Mortality statistics (*Vital Statistics of the United States*, Vol. II, Part A) provide extensive detail on age, sex, and race differentials.

Computer tapes for the data sets are easily obtained from the National Technical Information Service or the Scientific and Technical Information Branch (NCHS). The cost of public use tapes is low, but the data sets are sometimes too large for hard-ware facilities or data management staff to handle at nongovern-ment sites.

When outside users do not purchase tapes, they can ask for unpublished tabulations from NCHS or a private vendor. NCHS is able to provide tables that have already been prepared for staff use. But staff and time constraints usually prevent them from running special tabulations on request.Private vendors who provide these services are now appearing.

ITEMS AVAILABLE

Of all the morbidity, mortality, and sociodemographic items collected, how many are actually available to researchers in published reports or tapes?

For morbidity indicators, NAMCS's record is excellent. The number of items is limited, but all of them are used in NCHS reports and all are available on tape. HIS has many more indi-cators, but not all are available. Some questionnaire items are not coded, and some of the coded items are not placed on tapes. The amount of available data for conditions is especially slim: Details about conditions are scanned by coders and then con-densed into medical names, using a version of the International Classification of Diseases. Much of the original information remains uncoded; for example, the lay terms people use and the yes/no responses to condition checklists. This means that valuable sociomedical data on the nation's health are being gathered continuously but are not accessible for analysis.

For mortality, some information about the death event is not accessible. The absence of immediate and contributing

causes of death in published tables and public use tapes has hampered analysis of multiple causes of death (see Manton, 1979). NCHS is working to remedy this situation. Published statistics and public use tapes with these causes will soon be available for the data year 1976.

Although there are few sociodemographic variables routinely included in the data sets, they are not fully utilized. Some appear rarely in published tabulations, and many are not available on public use tapes.

Published reports on HIS typically show age, sex, and race differentials. Differentials for family income, own education, usual activity, marital status, living arrangement, occupation, employment status, family characteristics, class of worker, education of family head, industry, and personal income appear less often. (They are listed here from more frequent to less frequent.) National origin/ancestry results are forthcoming (see especially *Health, United States, 1979*). Veteran status is used for tabulations given to the Veterans Administration. HIS public use tapes contain the majority of collected variables. NAMCS reports and tapes use all of the sociodemographic variables collected.

Annual mortality statistics show rates by age, sex, and race. These items are also available on public use tapes. Rates by marital status and occupation are rarely published. The main reason is insufficient staff at NCHS. Additional reasons, in the case of occupation, are concerns about validity and coding difficulty. Staff of the Mortality Statistics Branch have long recognized the need for statistics by marital status and occupation. Beginning with data year 1979, marital status rates will appear annually. Whether occupation rates are produced, and how frequently, will depend on results of validity and coding feasibility studies.

In survey research, an important precept is "If you do not intend to use a variable, do not collect it." The three NCHS data sets should abide by this precept. More morbidity information should be coded from HIS interviews. If this cannot be done on a routine basis, special studies should be undertaken. All sociodemographic items now collected for HIS and death certificates should be used regularly. NCHS should evaluate reasons for non-

coding and infrequent use of items, test the validity of items which have potentially large errors, delete items of low validity, and try to use retained variables fully in published reports. All good-quality variables should be available on public use tapes.

TIMELINESS

The speed of producing reports and public use tapes has concerned NCHS staff for a number of years. Producing them in the midst of nonstop data collection activities is difficult for all three programs. Nevertheless, the lag time for HIS and mortality was reduced substantially in the 1970s. (NAMCS was a newcomer in the decade, started in 1973.) Annual summaries for HIS and NAMCS and final mortality rates are available within 18 months of the end of a data year. Public use tapes are available at about the same time. Publication of detailed mortality rates (in *Vital Statistics of the United States*) takes several years more. The improvements in the past decade have been welcomed by researchers and health planners. Increased staff in all three programs would help reduce lag times further.

PUBLIC INFORMATION PROGRAM

NCHS provides good general information about its publications and tapes. Descriptive booklets are well-organized and widely distributed. The public information office (Scientific and Technical Information Branch) responds rapidly and appropriately to outside requests for publications.

The 1979 Data Tape Users Conference sponsored by NCHS is an important aspect of the public information program. It allowed users to convey both satisfactions and problems about access to NCHS data. Conferences like this should be held on a regular basis.

There is, however, a broader problem of public information for health and mortality data. Many federal agencies collect morbidity and mortality data for the U.S. population or special population groups. It is possible to be familiar with NCHS data sets but know virtually nothing about others. Government re-

ports such as the annual Health Statistics Report (Department of Health, Education and Welfare, 1978) provide an inventory and brief description of data programs, but they are not aimed at data users. Ideally, there should be a clearinghouse which provides information and access to health and mortality data generated anywhere in the federal government.

SUMMARY

Data for HIS, NAMCS, and mortality are easily obtained in published reports and from public use tapes. Procedures for securing unpublished tabulations are changing, shifting to purchased services from private vendors rather than special computations done by NCHS staff. Some sociodemographic, morbidity, and mortality items for HIS and death certificates are not available. NCHS should evaluate and report the reasons for low use and seek to remedy the situation. Given the limited number of sociodemographic items and their great importance for health planning and research, public use tapes should include all of the collected variables. Efforts to reduce lag time between data collection and publication should continue, particularly for staff reports on special topics and bound volumes of mortality statistics. Public information about NCHS data is generally good, but the lack of a central information office for federal morbidity and mortality data is frustrating and leads to serious underutilization of data.

Research Questions

There are many unanswered questions about how and why social and demographic groups differ in their health and mortality.

DESCRIPTIVE QUESTIONS

Some important differentials are unknown because of adequacy and access problems. For others, the data are adequate and accessible but have not been closely examined by researchers.

Differentials at One Time Point

Questions: How do occupation groups differ in health, length of life, and cause or death? Do men and women in the same occupation differ in their illness and death rates? How do education and marital groups differ in mortality?

Absence of some sociodemographic differentials is due mainly to inadequate or inaccessible data. HIS routinely collects information on occupation and industry, but published tabulations are scarce. NAMCS has no items on patients' socioeconomic status. The death certificate has no items on education or income; they can be obtained only by follow-back survey or record linkage (see Kitagawa and Hauser, 1973). Occupation is on the certificate, but mortality rates by occupation are seldom published because of concerns about data quality. (The most recent rates are for 1950.)

There are fewer problems for demographic differentials. The most striking one is the lack of recent mortality differentials by marital status. They were published for 1940, 1949-1951, and 1959-1961, but not for 1969-1971. NCHS hopes to provide them soon on an annual basis.

Trends in Differentials Over Time

Questions: Do differentials widen or narrow over time? Have males and females become more similar in their chronic condition rates or in their use of health services? Have race differences in suicide widened or narrowed in the past decade?

HIS and death registration offer fine opportunities for descriptive and analytic work on trends in differentials. HIS now has over 20 years of annual rates. Mortality statistics extend farther. For both data sets, trends in rates are discussed more often than trends in differentials. Mortality data have been used extensively for trend analysis, but HIS is underutilized. (For examples of trend analyses, see Klebba, 1971; Klebba et al., 1973, 1974; Kleinman et al., 1979; Land and McMillen, 1978, 1979; Omran, 1977; and Verbrugge, 1976a, 1979b.) Published data await further work by researchers.

Comparisons Across Data Sets

Question: How do hypertension rates differ in HIS and HES/HANES? Do sex differentials for hypertension look the same in interview and examination data? Do occupation differentials look similar for morbidity and mortality?

These questions all require comparisons across two or more data sets. Many comparisons can be done now with published data, and researchers are beginning to do so. They may encounter problems when tabulation categories differ across data sets. This may prevent some detailed comparisons, but it should not halt the comparative effort entirely. (For comparisons of HIS and HANES rates for specific conditions, see Moss and Scott, 1978; Thom, 1979; Wilder, 1974; and Wilson et al., 1977. For comparisons of sex differentials in morbidity and mortality, see Verbrugge, 1976b, 1980.)

Differentials that Are More Comprehensive

Questions: How do population groups differ in their self-treatment of symptoms, in the number of curative actions they take for symptoms, in their use of nonmedical health professionals? How do they describe their ailments? How do groups differ in their immediate cause of death and in their multiple causes of death?

Many important and fascinating questions cannot be answered because morbidity and mortality indicators are not collected or because collected ones are inaccessible. HIS has restricted coverage of morbidity indicators, and condition data are incompletely coded. Mortality data for immediate and contributing causes have been inaccessible for several decades (but will soon be available).

ANALYTIC QUESTIONS

The main purpose of the three NCHS data sets is to provide high-quality descriptive data on morbidity and mortality. But some analytic questions might be handled quite easily, requiring

small increments in data collection and promising high analytic yield. I offer the following examples.

Risk Factors

Previously, I suggested criteria for including risk variables that will explain sociodemographic differentials: They should vary notably among population groups, and they should be well-documented, important causes of disease or injury. Smoking behavior fits these criteria. Whenever possible, it should be included in NCHS health and mortality data. (This could be through supplementary or routine questions in HIS, supplementary questions for NAMCS, and follow-back surveys for mortality.)

The Issue of "Selection"

Sometimes health risks and resulting poor health influence an individual's social status. This is a little-studied issue. Ideally, one needs data on the timing of both health events and social status events over a long period; for example, a dated history of chronic illness and marital status. (Occupation and employment status are also important social statuses, which may be influenced by poor health.) Longitudinal studies can gather such data prospectively. But retrospective interviews can query health and social histories, or just the onset time of current conditions and social status. HIS now asks about onset time of conditions but not of social statuses.[9] NCHS might consider a one-year supplement with questions about onset dates for chronic conditions and current marital, occupation, and employment status. (Respondents can also be asked directly if a health problem caused them to change their activities permanently.)

Current vs. Usual Status

Risks of illness and death are often cumulative in nature. In other words, it is repeated exposure rather than immediate ex-

posure that matters. More consideration must be given to cumulative factors associated with social status. For example, HIS might include measures of usual occupation (over one's lifetime) as well as current or recent (past year) occupation. If usual status produces especially large differentials, this might suggest risk variables for further research.

More generally, the relationship between current and usual social status is not well known. This is an especially important question for occupation, since HIS asks for current status but death registration asks for usual status. Comparisons of occupation-specific rates across the two data sets depend on knowing how similar current and usual occupation are.

Morbidity and Mortality of the Elderly Population

I have focused on problems in measuring and understanding sociodemographic differentials. Improvements in the adequacy of NCHS data and access to them will result in better knowledge of those differentials and, obviously, better knowledge of the rates for each population group.

The elderly population is of special importance since it generally has the highest prevalence rates for chronic conditions and chronic disability, highest health services use, and highest mortality rates. As an increasing fraction of the U.S. population becomes elderly, the demand increases for good data about their health and mortality.

In the long run, NCHS's best strategy is to provide adequate and accessible data for all population groups, so that a great variety of demands can be met.

Conclusion

Description of differentials necessarily precedes their explanation. NCHS is responsible for collecting and reporting data that show morbidity and mortality differentials for population

groups. I have discussed three NCHS data programs, asking how adequate and accessible their data are for description of socio-demographic differentials. For adequacy, I considered (1) the coverage, validity, and interpretation of morbidity and mortality indicators; (2) the same features of sociodemographic items; and (3) the inclusion of explanatory (risk) variables. For access, I discussed (1) the ease of obtaining data, (2) items available in published reports or on tapes, (3) timeliness of data release, and (4) public information about reports and tapes.

Here I summarize satisfactory and troublesome aspects of each data program.

HEALTH INTERVIEW SURVEY

The Health Interview Survey (HIS) is designed to measure Americans' health status. But HIS is selective in its coverage of morbidity indicators, omitting many conditions and health actions experienced by individuals.

The validity of reported health services use and expenditures has been studied extensively. Less attention has been given to validity of condition reports, although there are many opportunities to compare indicators within HIS and also with other national health surveys. Any unpublished studies of condition validity by NCHS should be reported.

Incidence and prevalence rates in HIS reflect both the presence of conditions and also disability or medical care taken for them. This restriction is not sufficiently clear to data users.

Coverage of sociodemographic variables is good, but their validity has not been discussed.

Wording for some items differs from Bureau of the Census practice, and there is a need for standardization of sociodemographic variables in federal data systems.

HIS's use of annual supplements is praiseworthy. They augment information on morbidity, risks, and sociodemographic characteristics.

Access to HIS data is generally good, but the interview's richness is not fully exploited. Emphasis on medical coding of

conditions ignores the *sociomedical* aspects of respondent reports.

Although all sociodemographic items are coded, some are rarely used in HIS reports and some are not on public use tapes. HIS staff should make every effort to exploit the interview data by more complete coding, tabulation, and public access to items.

NATIONAL AMBULATORY MEDICAL CARE SURVEY

The National Ambulatory Medical Care Survey (NAMCS) is a carefully condensed record of visits to doctors' offices. Its aims are narrow, and I believe it achieves them well. There are few problems of adequacy or access.

Coverage of patient characteristics is slim, but the items that exist are well-used and accessible.

Other sociodemographic or risk variables should be gathered on an occasaional basis by using office records, by asking patients to fill out a short questionnaire at the time of visit, or by adding items to the Patient Record.

DEATH REGISTRATION

The statistical purpose of death registration is to measure length of life, causes of death, and other aspects of the death event for the U.S. population.

The coverage of mortality items is adequate.

The validity of cause-of-death reports is largely unknown and needs careful study.

There are no difficulties in interpretation of mortality items.

Coverage of sociodemographic items is small. There is debate about the usefulness of usual occupation, and researchers have suggested that education is a preferable measure of socio-economic status.

The validity of occupation reports is not well known and needs research. Whether to view occupation as a measure of socioeconomic status or of risk is also an issue.

Follow-back surveys have collected risk variables from time to time. These surveys are relatively easy to conduct, and they provide immensely valuable data on individual behaviors which may be related to death (its timing and cause). Record linkage studies are difficult to conduct, but they yield detailed data on socioeconomic characteristics of decedents.

Access is infrequent and irregular for occupation-specific rates, marital-specific rates, and multiple causes of death. The codability of occupation reports is now being studied, and publications with marital and multiple cause data are forthcoming.

Ideally, the Mortality Statistics Branch of NCHS should augment its work on item validity; increase the regular publication of rates by occupation, marital status, and multiple cause; and conduct another follow-back survey. These can be accomplished only with substantial increases in staff.

What about the task of explanation, of answering "why"? This responsibility belongs mainly to researchers outside NCHS. Their first job is to describe differentials in an interesting and thoughtful way. Then hypotheses can be stated and data can be collected or located to test them. Often the NCHS health surveys or death certificates will not be adequate for this explanatory work. When they are augmented by supplementary items, follow-back survey, or linked records, they become good resources for explanatory research.

This sequence of activities by NCHS and researchers is a reasonable one, but it is not especially common. Researchers often develop hypotheses and design analytic studies based on a sociological or epidemiological theory. But explanatory research can also originate in thoughtful reviews of real-world descriptive data. The best situation is to have both a theory and high-quality descriptive data. HIS, NAMCS, and mortality data are currently underutilized by researchers. I hope that more active partnership between NCHS staff and outside researchers, between description and explanation, between "who" and "why" can be created. This collaboration will improve the collection

and presentation of descriptive data by NCHS, and also the explanation of morbidity and mortality differentials by researchers.

NOTES

1. I use the term "morbidity" in a comprehensive way to include illness and injury, short-term and long-term disability, health services use, medicine use, and other curative and preventive behaviors. Some researchers prefer to use the word in a more restricted sense, covering only measures of illness and injury. "Health" is used as a synonym for morbidity.

2. In this article, I consider only nonecological characteristics. I exclude geographic and environmental variables such as place of residence, region of the United States, and pollution level of residential area.

3. For example, in the past decade many bronchitis/emphysema/asthma deaths have been shifted to the title "chronic obstructive lung disease." Rates for bronchitis/emphysema/asthma have dropped sharply as a result.

4. The question on usual role is: "What was _____ doing most of the past 12 months: keeping house, working or doing something else?" (The words "keeping house" are used only for adult females.)

5. NAMCS data also have some characteristics of the physician, the physician's office, and date of visit, but these are not our concern here.

6. Usual activity is a person's main current role (preschool, school, work, keeping house, retired, other). Living arrangement indicates if someone lives alone, with relatives, or with nonrelatives. Family relationship means relationship to head of household. Family structure indicates the presence of one or both parents and other adult relatives. National origin/ancestry is a recent addition (1976) to HIS.

7. There is an earlier National Mortality Survey of 1962-1963 deaths, but it contains no risk variables, as defined here. (See Mathis, 1969.)

8. National longitudinal surveys can ultimately be primary sources of mortality data. Survival and death can be related to information about individuals' social roles and life-styles. If desired, survey deaths can be linked with death certificates to augment information about the death event. Current national longitudinal surveys must anticipate this research and begin to view death as an event of substantive importance, not just a cause of panel attrition.

9. The onset question is: "When did _____ first notice his [condition name]?" Answers are precoded by interviewers as follows: last week, week before, past two weeks-DK which, 2 weeks to 3 months, over 3 months up to 12 months, more than 12 months ago. For analysis of selection, finer detail in the categories would be necessary.

REFERENCES

Alvey, W., F. Aziz, and T. Mason
    1979   "Quality of mortality reporting in SSA linked data: some preliminary descriptive results." Presented at the Population Association of America meetings, Philadelphia, April.

Balamuth, E.
1965 Health Interview Responses Compared With Medical Records. Vital and
Health Statistics, Series 2, No. 7. Rockville, MD: National Center for Health
Statistics.
Cannell, C. F.
1965 Reporting of Hospitalization in the Health Interview Survey. Vital and Health
Statistics, Series 2, No. 6. Rockville, MD: National Ceter for Health Statistics.
Cannell, C. F. and F. Fowler
1965 Comparison of Hospitalization Reporting in Three Survey Procedures. Vital
and Health Statistics, Series 2, No. 8. Rockville, MD: National Center for
Health Statistics.
Greenberg, B. G.
1979 "The National Health Interview Survey—recommendations by a technical
consultant panel." Presented at the Third Biennial Conference on Health
Survey Research Methods, Reston, Virginia, May.
Guralnick, L.
1959 "Census-NOVS study of death certificates matched to census records." Mil-
bank Memorial Fund Q. 36: 144-153.
Hambright, T.
1969 Comparability of Marital Status, Race, Nativity, and Country of Origin on
the Death Certificate and Matching Census Record, United States, May-
August 1960. Vital and Health Statistics, Series 2, No. 34. Rockville, MD:
National Center for Health Statistics.
1962 Comparability of Age on the Death Certificate and Matching Census Record,
United States, May-August 1960. Vital and Health Statistics, Series 2, No. 29.
Rockville, MD: National Center for Health Statistics.
Hattwick, M.
1979 "Using health data collected by CDC and NCHS." Presented at the Conference
on Demographic and Health Information for Aging Research: Resources
and Needs, Bethesda, Maryland, June.
Howie, L. J. and T. F. Drury
1978 Current Estimates from the Health Interview Survey, United States, 1977.
Vital and Health Statistics, Series 10, No. 126. Hyattsville, MD: National
Center for Health Statistics.
Kaplan, D. L., E. Parkhurst, and P. K. Whelpton
1961 The Comparability of Reports on Occupation from Vital Records and the
1950 Census. Vital Statistics-Special Reports, 53(1). Washington, DC: Na-
tional Office of Vital Statistics.
Kilss, B. and L. DelBene
1979 "Further developments on possible collaborative interagency epidemiological
research." LASS Working Notes No. 6, Office of Research and Statistics,
Social Security Administration.
Kitagawa, E. M. and P. M. Hauser
1973 Differential Mortality in the United States. Cambridge: Harvard Univ. Press.
Klebba, A. J.
1971 Leading Components of Upturn in Mortality for Men, United States, 1952-67.
Vital and Health Statistics, Series 20, No. 11. Rockville, MD: National Center
for Health Statistics.
Klebba, A. J., J. D. Maurer, and E. J. Glass
1974 Mortality Trends for Leading Causes of Death, United States, 1950-69. Vital
and Health Statistics, Series 20, No. 16. Rockville, MD: National Center for
Health Statistics.

1973   Mortality Trends: Age, Color, and Sex, United States, 1950-69. Vital and Health Statistics, Series 20, No. 15. Rockville, MD: National Center for Health Statistics.
Kleinman, J. C., J. J. Feldman, and M. A. Monk
1979   "The effects of changes in smoking habits on coronary heart disease mortality." Amer. J. of Public Health 69, 8: 795-802.
Kovar, M. G. and J. A. Weed
1977   "Considerations in using individual socioeconomic characteristics in the analysis of mortality." Proceedings of the American Statistical Association (Social Statistics Section), 1977.
Land, K. C. and M. M. McMillen
1979   "A macrodynamic analysis of changes in mortality indexes in the United States, 1946-75: some preliminary results." Social Indicators Research.
1978   "Demographic and socioeconomic determinants of morbidity and disability trends in the United States, 1958-77." Working Papers in Applied Social Statistics, No. WP7811. Urbana: Department of Sociology, University of Illinois.
Madow, W. G.
1973   Net Differences in Interview Data on Chronic Conditions and Information Derived from Medical Records. Vital and Health Statistics, Series 2, No. 57. Rockville, MD: National Center for Health Statistics.
1967   Interview Data on Chronic Conditions Compared with Information Derived from Medical Records. Vital and Health Statistics, Series 2, No. 23. Rockville, MD: National Center for Health Statistics.
Manton, K.
1979   "U.S. mortality data: utility and shortcomings." Presented at the Conference on Demographic and Health Information for Aging Research: Resources and Needs, Bethesda, Maryland, June.
Mathis, E. S.
1969   Socioeconomic Characteristics of Deceased Persons, United States, 1962-1963 Deaths. Vital and Health Statistics, Series 22, No. 9, Rockville, MD: National Center for Health Statistics.
Moriyama, I. M. and L. Guralnick
1956   "Occupational and social class differences in mortality," pp. 61-73 in Milbank Memorial Fund (ed.) Trends and Differentials in Mortality. New York: Milbank Memorial Fund.
Moriyama, I. M., T. R. Dawber, and W. B. Kannel
1966   "Evaluation of diagnostic information supporting medical certification of death from cardiovascular disease," pp. 405-419 in W. Haenszel (ed.) Epidemiological Approaches to the Study of Cancer and Other Chronic Diseases. National Cancer Institute Monograph No. 19. Bethesda, MD: National Cancer Institute.
Moss, A. J. and G. Scott
1978   Characteristics of Persons with Hypertension, United States, 1974. Vital and Health Statistics, Series 10, No. 121. Hyattsville, MD: National Center for Health Statistics.

National Center for Health Statistics
  1979 Vital Statistics of the United States 1975. Vol. II, Part A (Mortality). Hyatts-
       ville, MD.
  1975 Health Interview Survey Procedure 1957-1974. Vital and Health Statistics,
       Series 1, No. 11. Rockville, MD.
  1966 Interview Response on Health Insurance Compared with Insurance Records,
       United States, 1960. Vital and Health Statistics, Series 2, No. 18. Rockville,
       MD.
Office of Federal Statistical Policy and Standards. Department of Commerce
  1978 A Framework for Planning U.S. Federal Statistics for the 1980's. Issued
       July 1978.
Omran, A. R.
  1977 "Epidemiologic transition in the U.S.: the health factor in population change."
       Population Bulletin 32, 2.
Spirtas, R., H. M. Rosenberg, D. Burnham, and J. Priebe
  1979 "Occupation and industry information from the death certificate: preliminary
       assessment on the quality of reporting." Presented at the American Statistical
       Association meetings, Washington, D.C., August.
Tenney, J. B., K. L. White, and J. W. Williamson
  1974 National Ambulatory Medical Care Survey: Background and Methodology.
       Vital and Health Statistics, Series 2, No. 61. Rockville, MD: National Center
       for Health Statistics.
Thom, T.
  1979 "Sources of statistics on hypertension in the elderly." Presented at the Con-
       ference on Demographic and Health Information for Aging Research: Re-
       sources and Needs, Bethesda, Maryland, June.
U.S. Department of Health, Education and Welfare. Public Health Service
  1978 Health Statistics Report, Fiscal Year 1978. Issued May 1978.
  1977 Health Statistics Plan, Fiscal Years 1978-1982. Issued February 1977.
Van Dusen, R. A. and N. Zill (eds.)
  1975 Basic Background Items for U.S. Household Surveys. Washington, DC:
       Center for Coordination of Research on Social Indicators, Social Science
       Research Council.
Verbrugge, L. M.
  1980 "Sex differentials in morbidity and mortality," in A. H. Stromberg (ed.)
       Women, Health, and Medicine. Palo Alto, CA: Mayfield Publishing.
  1979a "Complaints and diagnoses: sex differences in patient descriptions and physi-
       cian attributions of illness." Revised version of a paper presented at the
       American Public Health Association meetings, Los Angeles, November 1978.
  1979b "Recent trends in sex mortality differentials in the United States." Presented
       at the Population Association of America meetings, Philadelphia, April.
  1976a "Females and illness: recent trends in sex differences in the United States." J. of
       Health and Social Behavior 17 (December): 387-403.
  1976b "Sex differetials in morbidity and mortality in the United States." Social
       Biology 23: 275-296.

Wilder, C. S.
    1974   Prevalence of Chronic Circulatory Conditions, United States, 1972. Vital and Health Statistics, Series 10, No. 94. Rockville, MD: National Center for Health Statistics.

Wilson, R. W., K. M. Danchik, and M. A. Monk
    1977   "Hypertension," pp. 27-41 in Health, United States, 1976-1977. DHEW Publication No. HRA 77-1232. Rockville, MD: National Center for Health Statistics and National Center for Health Services Research.