Editorial: The Use of Existing Databases in Morbidity and Mortality Studies

The measurement of morbidity and mortality from chronic or long incubation-period diseases in the community is important for several reasons. The resulting data provide a basis for generating hypotheses related to the etiology of chronic diseases; for determining the effectiveness of both preventive and treat ment efforts; and for evaluating needs for specific health services, both clinical and preventive. The traditional epidemiological approaches evaluate the distribution of disease in relationship to time, place, and geography, and the characteristics of individuals or persons.

Most of the population measures of chronic diseases have included mortality statistics, registries of well-defined and relatively easily diagnosed diseases, such as cancer[1] and insulin-dependent diabetes mellitus,[2] and data collected in special cross-sectional surveys such as the National Health Interview and Examination Survey,[3] as well as from longitudinal studies using carefully defined independent and dependent measurements over time, such as the Framingham Study.[4]

The cost of specific longitudinal studies to measure risk factors and morbidity, incidence, and prevalence is substantial. By necessity, these studies have to be restricted to relatively small samples of individuals and a limited geographic distribution. Cross-sectional surveys such as the National Health Examination Survey[3] and registries for cancer or other disorders provide a broader coverage of the population, but at the expense of limited measurements of independent variables, risk factors, and especially longitudinal measures of independent variables. The information about disease and other clinical outcomes is often limited.

Morbidity surveys of chronic diseases have a long tradition in the United States and other countries. The Hagerstown studies provided one of the first attempts to evaluate morbidity in the population by using questionnaires related to health and disability.[5] The Commission on Chronic Illness in Baltimore[6] and New Jersey[7] in the 1950s used the combination of health interviews, screening, and clinical physical examinations to estimate the prevalence of disease in a population. These studies demonstrated an extremely important fact: the likelihood of ascertainment of a disease in a population is a function of the amount and type of available health services and is independent of the extent of the disease in the population. Thus, the more health services available in the community, and the higher the technology deployed in those services, the greater the likelihood of identifying specific disease. A second key observation was the inverse relationship between measures of socioeconomic class, such as income, occupation, or education and the extent of morbidity and mortality in the population.

Recently there has been growing interest in using existing databases collected for reasons other than epidemiological studies to estimate morbidity and relationships to risk factors within populations. Large databases can be found in health maintenance organizations,[9] membership lists of professional organizations,[10] and the routinely collected records of health care providers.[11–13] They are increasingly being viewed as alternatives to traditional but more costly longitudinal and cross-sectional epidemiological studies.

Existing databases are obviously of value, as long as the investigator recognizes the limitations of the data, limitations in the quality of the original data collection, and the biases that arise from associated variations in ascertainment of disease and selection for treatment. Existing databases collected in the past also provide opportunities for low-cost follow-up studies, similar to traditional historical occupational cohorts. In interpreting studies, problems both of the quality of the original data and, most important, of differential follow-up and measurements of outcome must be carefully considered.[14]

The papers presented in this issue of the Journal provide good examples of the use of existing data sets and selected samples. In their American Cancer Society study, Thun and colleagues used volunteers to collect information about risk factors for cancer from other volunteers in the community.[15] The large sample size has provided interesting information regarding risk factors for both cancer, heart disease, and total mortality. The results of the present study are consistent with other previous reports. Perhaps the most important observation is the apparent lack in never smokers of any increase in lung cancer mortality over time. This might suggest that such environmental factors as indoor and outdoor air pollution, and especially exposure to radon, or occupational hazards have had relatively little effect on lung cancer. On the other hand, volunteers in the study may be self-selected for low risk from environmental exposures, and the sample size of never smokers may not be large enough to identify relatively small changes in risks over time from environmental or occupational exposures.

Puro and colleagues' study of the risks of occupational hepatitis C in a health care provider population is a good example of the use of an existing database for a defined population (i.e., health workers).[16] The results of this study will be reassuring except for the fact that the population of patients treated in the particular study hospitals may have had a much lower prevalence of hepatitis C and likelihood of transmission to hospital personnel than in other hospitals in the United States.

Bucher and Ragland's analysis of Western Collaborative Study data[17] and Schwartz and colleagues' follow-up of the Terman Life Cycle Study of Children[18] are examples of the use of originally collected longitudinal cohorts to test new hypotheses. The observation that low income, less education, and having divorced parents are not particularly good for your health is probably not surprising to any of the readers of the Journal. The observation that not being a cheerful child is associated with longevity seems counterintuitive and may be an example of how one may find something significant if one digs deep enough into almost any data set.

The uses of traditional hospital and mortality statistics are reflected in several other morbidity and mortality papers in this issue.[19–22] The evaluation of mortality statistics has been a long-standing interest of many epidemiologists.[23] The complete ascertainment of deaths, the availability of data for the entire country, comparisons across communities both within the United States and with other countries, and the ability to follow trends over long periods of time—these all have made

Editor's Note. See related articles by Thun et al. (p 1223), Puro et al. (p 1272), Bucher and Ragland (p 1231), and Schwartz et al. (p 1237) in this issue.

mortality statistics the cornerstone of descriptive epidemiology of chronic disease. Unfortunately, in spite of many efforts to improve the quality of the information on death certificates, the use of death certificates for diseases with relatively low case-fatality rates and for deaths that occur outside of the hospital still arouses substantial concerns, especially in older age groups. It is very unlikely that alcohol- and nonalcohol-related cirrhosis of the liver can be separated on the death certificate without review of other records or interview with next-of-kin. Even with these other sources of information, the data regarding alcohol consumption may be inaccurate.[24]

Traditional morbidity statistics, such as hospital records, present some of the same problems as death certificates. However, a further problem is the absence of good national databases, the lack of longitudinal data and the problem of counting hospitalizations rather than individuals so that one individual hospitalized three or four times for the same disease may be counted as four separate events. We anticipate in the future that morbidity statistics will play a larger role in the evaluation of chronic diseases.[25,26]

The use of routinely collected data such as morbidity or mortality statistics has a great appeal. Major pitfalls in the use of these data can have adverse effects on public health policy. Overinterpretation of the information can lead to poor decisions regarding prevention and treatment of diseases. The identification of spurious associations found in large data sets can also result in needless expenditures of time and money for further evaluation of these associations. An unfortunate tendency is to generate hypotheses from these epidemiology studies even if the initial observations are based on inadequate or inaccurate information. What begins as an inexpensive study can often result in very expensive secondary studies. Furthermore, these spurious observations have the potential for generating ill-founded policy decisions.

In epidemiological studies, as in most other situations, you often get what you pay for. Existing databases are obviously of considerable value as long as the investigator recognizes the limitation of the data, tries to improve the quality of the original data measurements, and deals with the biases that are associated with identification of disease, selection for specific treatments, and measurements of outcome. It is important to remember the five key truisms. First, as the quality and amount of health services in a community increase, the measured morbidity will probably also increase. Second, poor, less educated, unemployed individuals in the community are almost certain to have more morbidity, mortality, and social problems and be depressed and unhappy about their situation. Third, individuals in the community who adhere to various treatments, diet, pill taking, etc., will be healthier to begin with. Fourth, individuals like to live in pleasant environments. People who live in trailer parks are more likely not only to be killed by a tornado, but also to have a higher prevalence of cancer and of acute, infectious, and chronic diseases, higher alcohol consumption, and to live in close proximity to chemical waste dumps, which are sources of air pollution and other environmental exposures. Finally, biologically implausible associations remain unlikely to be true causal associations in epidemiological studies, even if the confidence levels are greater than 1 and the P value less than .05.

The new era of better methods of technology transfer and of the ability to link key databases offer new opportunities for improving the morbidity databases and potentially reducing the costs of epidemiological studies. A major effort must be made to improve the quality of the data.[27,28] □

Lewis H. Kuller
Department of Epidemiology
Graduate School of Public Health
University of Pittsburgh
Pittsburgh, Pa

References

1. *SEER Cancer Statistics Review, 1973–1990.* Bethesda, Md: National Cancer Institute; 1993. NIH publication 93-2789.
2. LaPorte RE, McCarty D, Tull E, Libman I, Matsushima M. Beyond insulin-dependent diabetes mellitus registries: capture-recapture approaches for monitoring incidence. In: Dorman J, ed. *Standardization of Epidemiologic Studies of Host Susceptibility.* New York, NY: Plenum Press; 1994.
3. Data systems of the National Center for Health Statistics: programs and collection procedures. *Vital Health Stat [1].* 1982; no. 16. DHHS publication PHS 82-1318.
4. Dawber TR. *The Framingham Study. The Epidemiology of Atherosclerotic Disease.* Cambridge, Mass: Harvard University Press; 1980.
5. Comstock GW, Abbey H, Lundin FE Jr. The nonofficial census as a basic tool for epidemiologic observations in Washington County, Maryland. In: Kessler II, Levin ML, eds. *The Community as an Epidemiologic Laboratory. A Casebook of Community Studies.* Baltimore, Md: The Johns Hopkins Press; 1970:73–99.
6. Commission on Chronic Illness. *Chronic Illness in a Large City, Chronic Illness in the United States.* Cambridge, Mass: Harvard University Press; 1957: chap 14, append F, G.
7. Commission on Chronic Illness. *Chronic Illness in a Rural Area.* Cambridge, Mass: Harvard University Press; 1959.
8. Sanders BS. Have morbidity surveys been oversold? In: Lilienfeld AM, Gifford AJ, eds. *Chronic Diseases and Public Health.* Baltimore, Md: The Johns Hopkins Press; 1966:89–99.
9. Klatsky AL, Armstrong MA, Friedman GD. Risk of cardiovascular mortality in alcohol drinkers, ex-drinkers and nondrinkers. *Am J Cardiol.* 1990;66:1237–1242.
10. Manson JE, Stampfer MJ, Hennekens CH, Willett WC. Body weight and longevity. A reassessment. *JAMA.* 1987;257:353–358.
11. *Special Report: Hospital Data by Geographic Area for Aged Medicare Beneficiaries: Selected Diagnostic Groups, 1986.* Vol. 1. Washington, DC: Health Care Financing Administration, Office of Research and Demonstrations; 1986.
12. *Medicaid Source Book: Background Data and Analysis.* Washington, DC: Congressional Research Service; 1988.
13. Watzlaf VJM, Kuller LH, Ruben FL. The use of the medical record and financial data to examine the cost of infections in the elderly. *Top Health Inform Manag.* 1992;13: 65–76.
14. Young LC, Kuller LH. Tracking of blood pressure from adolescence to middle age: the Dormont High School Study. *Prev Med.* 1994;23:418–426.
15. Thun MJ, Day-Lally CA, Calle EE, Flanders WD, Heath CW Jr. Excess mortality among cigarette smokers: changes in a 20-year interval. *Am J Public Health.* 1995; 85:1223–1230.
16. Puro V, Petrosillo N, Ippolito G, et al. Occupational hepatitis C virus infection in Italian health care workers. *Am J Public Health.* 1995;85:1272–1275.
17. Bucher HC, Ragland DR. Socioeconomic indicators and mortality from coronary heart disease and cancer: a 22-year follow-up of middle-aged men. *Am J Public Health.* 1995;85:1231–1236.
18. Schwartz JE, Friedman HS, Tucker JS, Tomlinson-Keasey C, Wingard DL, Criqui MH. Sociodemographic and psychosocial factors in childhood as predictors of adult mortality. *Am J Public Health.* 1995;85:1237–1245.
19. Hodge MJ, Dougherty GE, Pless IB. Pediatric mortality and hospital use in Canada and the United States, 1971 through 1987. *Am J Public Health.* 1995;85: 1276–1279.
20. Hurwitz ES, Holman RC, Strine TW, Chorba TL. Chronic liver disease mortality in the United States, 1979 through 1989. *Am J Public Health.* 1995;85:1256–1260.
21. Sonnenberg A, Wasserman IH. Associations of peptic ulcer and gastric cancer with other diseases in US veterans. *Am J Public Health.* 1995;85:1252–1255.
22. Langlois JA, Buechner JS, O'Connor EA, Nacar EQ, Smith GS. Improving the E coding of hospitalizations for injury: do hospital records contain adequate documentation? *Am J Public Health.* 1995;85: 1261–1265.
23. Dorn HF. Mortality. In: Lilienfeld AM,

Gifford AJ, eds. *Chronic Diseases and Public Health.* Baltimore, Md: The Johns Hopkins Press; 1966:23–54.
24. Shim YK, Perper JA, Kuller LH. Factors associated with the decline in cirrhosis death rates among young adults in Allegheny County, Pennsylvania, 1973–1985. *Am J Epidemiol.* 1993;138:531–543.
25. Ries P. Health of black and white Americans, 1985–87. *Vital Health Stat [10].* 1990; no. 171.
26. WHO DIAMOND Research Group. Global surveillance of childhood diabetes to the year 2000: the WHO multinational project. *Pact Diabetes.* 1995;8:94–95.
27. WHO Directors of Non-Communicable Disease, Collaborating Centers and Key Officials. Needed: universal monitoring of all serious diseases of global importance. *Am J Public Health.* 1993;83:941–943.
28. LaPorte RE. Assessing the human condition: capture-recapture techniques. *BMJ.* 1994;308:5–6.

Editorial: The Timing of High-Dose Vitamin A Supplementation to Children

Improved low-light vision and ocular health, reduced severity of some illnesses, and increased chances for survival through the preschool years are among the recognized benefits of restoring the vitamin A status of deficient populations where there is some evidence of xerophthalmia. These health advantages obviously improve the quality of life of individuals and their families. But the larger society also benefits. A population with adequate vitamin A status reduces the drain on the use of expensive health personnel and facilities.

Among several intervention options, vitamin A supplements distributed to deficient children are a rapid means of improving their vitamin A status. Unless dietary intake is improved and other underlying causes of a deficiency are corrected, however, supplement distribution will have to be repeated periodically. Although the range in expected protective benefits from a specific intervention is difficult to predict, restoring vitamin A adequacy among deficient child populations increases survival chances by more than 20%.[1]

Large-scale community trials among preschool-age children consistently have resulted in enhanced survival among recipients of vitamin A supplements provided as a fortified product in the daily diet (45% protection),[2] at weekly (54% protection),[3] or 4-monthly intervals (19–30% protection).[4–6] Protection has not been demonstrated consistently when the distribution interval is every 6 months. Of the 3 studies of 6-monthly distribution, one reported significant survival enhancement (34% protection)[7] and two did not (0 and 7% protection[8,9]).

Two questions arise as a result of the completed community trials: Would the protective benefits from the 4-monthly distribution have been greater if the interval between supplementation had been shorter? Could other public health measures to control disease (e.g., immunization and deworming), or modifications in the timing of supplement distribution to correspond to peak disease occurrence or the seasonal availability of vitamin A-containing foods, consistently stretch the protective period to 6 months? Answers to these questions are important for designing affordable programs that use human and financial resources efficiently without diminishing their impact on health and survival.

High-dose vitamin A supplements are inexpensive, less than 3 cents per capsule. Because of the potential for misuse resulting in toxicity, however, periodic delivery of high-dose vitamin A supplements requires close supervision by the health system. This very requirement greatly increases program costs. Where universal prophylaxis is concerned, cost is dependent on the frequency of distribution. On the other hand, there are potential cost savings to be realized from adequate vitamin A prophylaxis, as a result of decreasing demands placed on costly clinic and hospital facilities, because of decreased incidence of severe disease.

The Ghana Vitamin A Study Team elsewhere reported details of a survival study and a health study performed in adjacent areas of Northern Ghana.[4] These were randomized, masked, placebo-controlled community trials involving, in the survival study, nearly 22 000 children 6 to 90 months of age and, in the health study, nearly 1500 children 6 to 59 months of age. High-dose (i.e., 30 mg) retinyl palmitate supplements (100 000 IU or 105 μmol) for infants 6 to 11 months of age, and 60 mg (200 000 IU or 210 μmol) from 12 months of age onward, were distributed at 4-month intervals. High coverage (90% or more) was achieved at each of the distribution rounds. Children could have received a maximum of 3 doses of vitamin A in the health study or 6 doses of vitamin A in the survival study. On average, children were followed for 9.8 months in the health study and for 18.2 months in the survival study. These interventions reduced death rates by about 19%, the frequency of vomiting by 13%, anorexia and dehydrating diarrhea each by 15%, and the utilization of clinic and hospital facilities by 12 and 38%, respectively.

In this issue of the Journal, Ross and colleagues[10] report on a further analysis of the Ghana data to examine questions of the temporal relationship between benefits and the time since dosing, the number of doses, and the season when administered. The last point was particularly relevant because the disease pattern in the area included a high prevalence of symptoms of subclinical infections,[11] malaria,[12] "hot body," cough, diarrheal disease, and respiratory symptoms.[4] The prevalence of many of these symptoms frequently shows a seasonal peak occurrence; it is thought to be accompanied by increased metabolic utilization of vitamin A and, when accompanied by fever, possibly an increased loss via urine.[13] For maximum effectiveness, it is often recommended that supplement distribution be linked to peak seasonal disease patterns and periods of food scarcity.

Ross and colleagues' analysis provided no evidence for a progressive reduction in protective benefits from severe morbidity or fatality with advance in time up to 150 days from receiving the dose. There also was no evidence for an accumulative effect from receiving sequential 4-monthly high doses or for a preferred time of the year for their distribution. These observations indicate that at the age-specific high dosages given, critical protective functions are sustained by a 4 to 5 month (120–150 days) dosing interval in a population with a minimal prevalence of xerophthalmia but high prevalence of low serum retinol levels, and high prevalence of morbidity and

Editor's Note. See related article by Ross et al. (p 1246) in this issue.