Using Hospital Discharge Data for Disease Surveillance

Linda Huff
Greg Bogdan, DrPH
Kenneth Burke, MD
Edward Hayes, MD
William Perry, MS
Lani Graham, MD, MPH
Harold Lentzner, PhD

Three authors are with the Maine Bureau of Health. Ms. Huff is an Epidemiologist in the Year 2000 Assessment Project, Dr. Bogdan is the Director of the Division of Disease Control, and Dr. Graham is the Director of the Bureau of Health. Dr. Burke was an Epidemiologist with the Bureau of Health during work on this project. Mr. Perry is the former Director of the Division of Research and Data Management, Maine Health Care Finance Commission. Two authors are with the National Center for Health Statistics, Public Health Service, Centers for Disease Control and Prevention. Dr. Hayes is a Medical Epidemiologist assigned to the Maine Bureau of Health from the Division of Health Promotion Statistics, and Dr. Lentzner is an Epidemiologist in the Office of Analysis, Epidemiology, and Health Promotion.

Tearsheet requests to Linda Huff, Division of Disease Control, Maine Bureau of Health, State House Station 11, Augusta, ME 04333; tel. 207-287-5559; fax 207-287-6865.

SYNOPSIS

THE AUTHORS EXAMINE the effectiveness of using hospital discharge data in assessing trends and geographic variations in the occurrence of selected chronic diseases.

The Chronic Disease Surveillance System, in place from 1987 to 1991, used hospital discharge data, mortality data, and Cancer Registry data to track selected chronic diseases. The authors reviewed data on three diseases: breast cancer, cervical cancer, and lung cancer.

A computerized algorithm was used to link multiple records representing a single disease occurrence. To estimate disease occurrence rates from hospital discharge data, repeat admissions for the same disease in any given calendar year were discounted. All rates were directly age-adjusted to the 1985 Maine state population.

For all three diseases, the rates obtained from hospital discharge data were higher than Cancer Registry rates. Possible causes for the discrepancies and suggestions for improving the utility of hospital discharge data for chronic disease surveillance are discussed.

Chronic diseases account for more than 70% of all deaths in the United States each year[1]. While trends in mortality due to chronic diseases can be monitored using death certificate data, sources of data that can be used in tracking chronic disease incidence are scarce. Cancer registries are used to track cancer incidence in many states. Hospital discharge data have been used for chronic disease surveillance but have several limitations. These limitations include: (*a*) difficulties in differentiating hospital admission patterns for a disease from patterns of actual disease occurrence and (*b*) errors or biases associated with disease coding[2].

In July 1987, the Maine Department of Human Services joined with the Maine Health Care Finance Commission under a cooperative agreement with the Centers for Disease Control to develop a chronic disease surveillance system using hospital discharge data, the state cancer registry, and mortality data from death certificates. Integrating the three data sources allowed a better assessment of the total public health burden for the selected diseases being evaluated and provided a more complete view of trends and geographic variations than could be obtained from a single data source. Much of the funding for the project was provided by the federal Agency for Toxic Substances and Disease Registry in cooperation with the Centers for Disease Control. This paper describes the Maine Chronic Disease Surveillance System which was in place from 1987 to 1991, and assesses the effectiveness of using state hospital discharge data to evaluate trends and geographic variations in the occurrence of selected chronic diseases.

Methods

The Maine Chronic Disease Surveillance System utilized the hospital discharge database maintained by the Maine Health Care Finance Commission; the Maine Cancer Registry, maintained by the Maine Bureau of Health, Division of Disease Control; and the mortality database drawn from death certificates, maintained by the state Office of Data, Research and Vital Statistics. The system extracted data from the hospital discharge and mortality databases from 1980 through 1988 and from the Cancer Registry beginning with its inception in 1983 through 1988.

For this paper, we reviewed state- and county-level data on three diseases of high public health impact: breast cancer (ICD-9 code 174.XX), cervical cancer (ICD-9 code 180.XX), and lung cancer (ICD-9 code 162.XX). We did not include in situ cancers.

We first searched the three source databases for all records with any mention of the selected diseases either as a principal or secondary diagnosis (up to four secondary diagnoses for each record were searched) or as a contributing or underlying cause of death. Records from the Cancer Registry and mortality databases represented individual disease events, except for occasional duplicate entries that were found and corrected when the data were processed for the surveillance system. However, in the hospital discharge database, each record represented a single hospital admission, and multiple records existed for any person admitted to a hospital more than once (for the same or different diagnosis). Names and social security numbers were not listed in the hospital discharge database. Although a medical record number or patient admission number was available, the entry for this field was not always unique to each patient, either because the number changed with each admission or because hospitals had changed their medical record numbering systems over the years. We used a computerized algorithm to link multiple records representing a single disease occurrence (for example, multiple admissions to a single hospital or to different hospitals of the same person with lung cancer or duplicate entries for the same event) within each data source. To obtain disease rates from hospital discharge data, we discounted repeat admissions for the same disease in any given calendar year, but initially, not across different years. A geographic area was assigned based on the patient's place of residence on the first hospital record. All rates were directly age-adjusted to the 1985 Maine population.

The ability to link hospital records across years, which could not be done in the early stages of the project, improved as the project developed and the linking process became fully automated. As we analyzed the data, we noted that hospital discharge rates for most cancers were higher than Cancer Registry rates. We felt that this might be due to the failure to exclude from the hospital discharge data those patients who had been admitted for the same conditions in years prior to the year of analysis. For example, any patients who were discharged in 1986 with a diagnosis of breast cancer, and then readmitted and discharged in 1987 with the same diagnosis would have contributed to the breast cancer count for both years. With the development of a more automated linking process we were subsequently able to link records across years. We analyzed data for 1987 and for 1988 after eliminating all records of patients who had been previously discharged in any prior year back to 1980. This resulted in an estimate of the number of individuals with each disease under consideration who were discharged for the first time in either 1987 or 1988.

> An integrated analysis of surveillance data from different sources provides a more complete picture of trends and variations in disease occurrence than a review of any one of the data sources alone.

Results

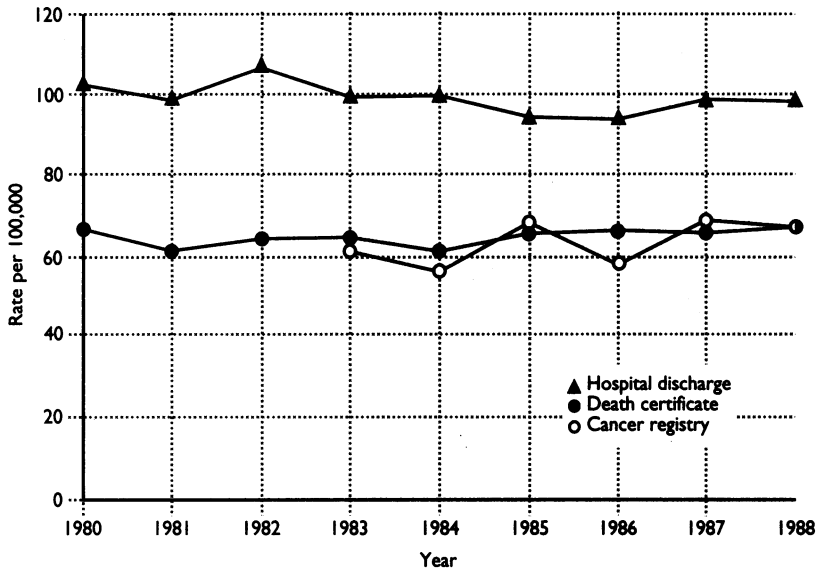
Lung cancer rates are shown in figure 1. Cancer Registry rates fluctuated but suggest a slight increase from 1983 through 1988. The lung cancer rates obtained from hospital discharge data do not show any obvious trend over time. Mortality associated with lung cancer appears relatively stable from 1980 to 1988. (A slight upward trend in lung cancer mortality between 1981 and 1991 has been noted in other analyses of Maine data using a different standard for age-adjustment)[3].

Figure 2 shows the breast cancer rate from each of the three data sources for each year. Breast cancer rates obtained from Cancer Registry data increased from 1983 through 1988. Breast cancer rates from hospital discharge data did not increase until 1987. The mortality rate associated with breast cancer remained relatively stable over the nine-year period.

Cervical cancer rates from each data source are shown in figure 3. Cervical cancer rates from both Cancer Registry and hospital discharge data appear to have decreased from 1984 to 1988, although the rates of decrease are quite different. Mortality associated with cervical cancer appears to have been relatively stable over the nine-year period.

When we linked records across years and discounted repeat admissions, the 1987 and 1988 rates for hospital discharge remained higher than Cancer Registry rates. The proportion of "new" hospital discharge records for 1987 and

Figure 1. Maine Chronic Disease Surveillance System lung cancer rates, 1980–88



1988 (after discounting repeat admissions from previous years) that matched Cancer Registry records were: 77% (1159 of 1501) for breast cancer, 66% (1202 of 1818) for lung cancer, and 40% (63 of 157) for cervical cancer.

We examined the correlations between the county rates from hospital discharge data and Cancer Registry data for each disease. Breast cancer rates from the two sources were most highly correlated, with a correlation coefficient (r) of 0.87. Lung cancer rates were less well correlated, with $r = 0.79$, and the cervical cancer rates from the two sources showed the weakest correlation, with $r = 0.55$.

Discussion

An integrated analysis of surveillance data from different sources provides a more complete picture of trends and variations in disease occurrence than a review of any one of the data sources alone. For three types of cancer, the rates from hospital discharge data were higher than the rates derived from Cancer Registry data, even though we attempted to discount repeat admissions through an extensive record linking process. The discrepancies between the two data sources probably resulted from a number of factors. First, the annual rates from hospital discharge data may include prevalent as well as incident cases, since records of persons admitted in previous years were not discounted. The Cancer Registry should include only incident cases for any given year. However, this difference persisted in 1987 and 1988 even after we corrected for repeat admissions for the same disease. Furthermore, low survival rates for lung cancer would tend to limit the number of prevalent cases for that disease in the years after initial admission.

Second, the linking process may not have completely identified all repeat hospital admissions of any given person within the same year. This problem could be alleviated by assuring the use of unique individual identifiers in hospital discharge data or through the use of more recently developed linking techniques and software[4].

Third, there may have been incomplete reporting of cases to the Cancer Registry. The initial increase in breast cancer rates derived from Registry figures for 1983 through 1986 could have been due to an increase in reporting to the Registry during the early years of its operation. An audit by the Centers for Disease Control of the Cancer Registry, which examined reporting of cases in 1985 from nine of 42 Maine hospitals, found that 84% of cancer cases from the nine hospitals were reported but did not determine whether the accuracy of reporting had changed over time.

Fourth, even if all repeat admissions within a year were linked and discounted from the hospital discharge data, variations in admitting practices for a given disease over time and by area could cause discrepancies between the number of hospital discharges and the actual occurrence of the disease[2,5]. Hospital admissions for breast cancer may have remained sta-

Figure 2. Maine Chronic Disease Surveillance System breast cancer rates, 1980–88

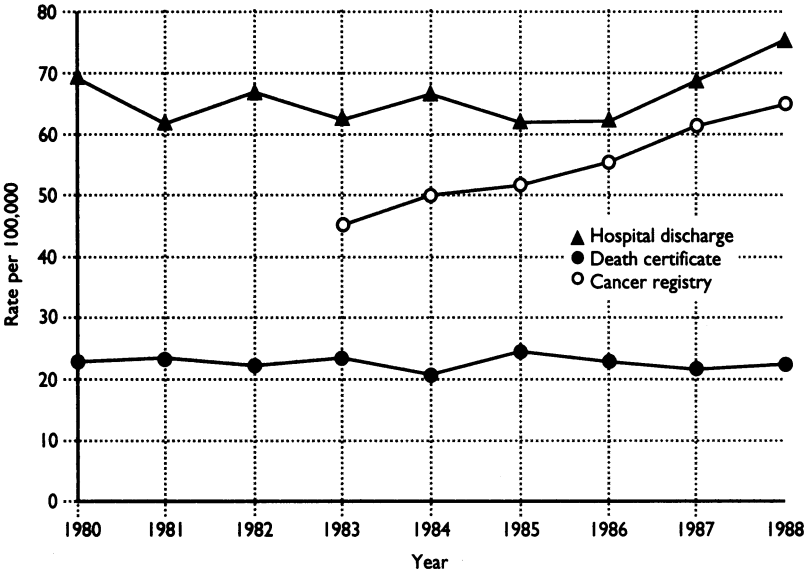
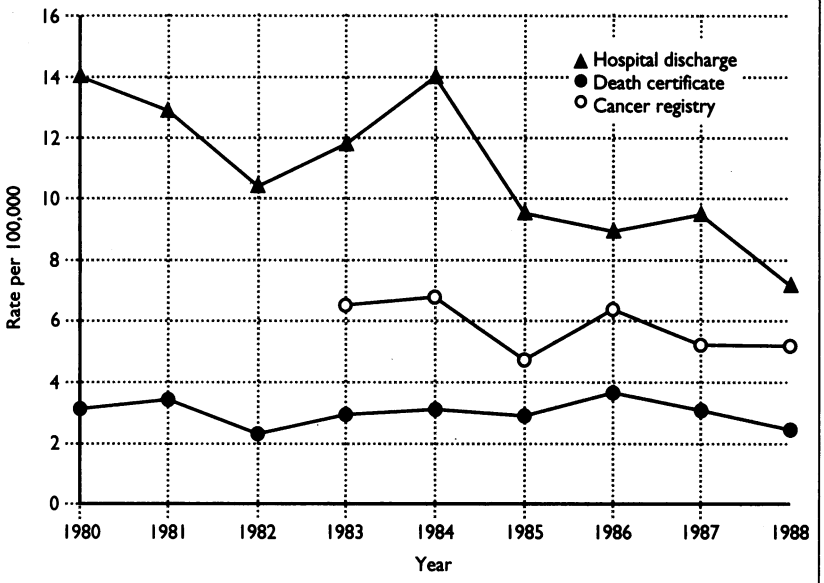


Figure 3. Maine Chronic Disease Surveillance System cervical cancer rates, 1980–88



ble, despite an increased number of cases reported to the Cancer Registry, if breast cancers were being detected at an earlier stage that did not require hospitalization. Some of these early detected cases could have resulted in increased admissions in later years. In addition, variations in admitting and reporting practices by area could cause discrepancies in county rates obtained from the two data sources.

Finally, errors or biases in coding of hospital records may result in misclassification of diseases or other discrepancies that could prevent accurate linking of records. Diagnostic coding bias should have been minimized somewhat in our system since both primary and secondary diagnostic codes were searched for any mention of the specific disease under surveillance.Other coding errors may be more widespread and difficult to detect. For example, the gender variable on hospital discharge records is coded with a "1" for male and "2" for female. These can easily be reversed during coding, which could prevent the subsequent linking of records belonging to the same individual.

The correlation between county hospital discharge and Cancer Registry rates varied directly by disease according to the percent of records of each disease that were linked between the two databases. Rates from hospital discharge data fit most closely with those from Cancer Registry data when the proportion of linked records was highest. This supports the conclusion that hospital discharge data are a better surveillance source for some diseases than for others. The reliability of hospital discharge data in estimating disease occurrence is affected by the likelihood that people with the disease under study will have multiple admissions, by the relative variability in provider practice with respect to hospital admissions for cases of disease, and by variability in access to care[2,5]. Consequently, the occurrence of a disease such as stroke should be more accurately measured with hospital discharge data than a disease such as asthma.

Utility of state and local hospital discharge data for chronic disease surveillance could be improved through the use of unique individual identifiers, better understanding of variations in and determinants of admitting practices, and by improvements in access to care. Where disease registries can be established, they may serve as more reliable surveillance sources than hospital discharge data. Unfortunately, registries do not exist for many chronic diseases, and hospital discharge data may serve as one of the few available data sources to track such conditions and monitor the public's health. Mortality data have historically been used as a surrogate for data on disease incidence. The comparisons presented here illustrate that hospital discharge data may indicate trends in incidence better than mortality data for some diseases. Finally, hospital discharge data is an important source for estimating the burden of health conditions on the health system.

Gene Stanton, Senior Programmer Analyst, Maine Health Care Finance Commission, developed the computer programming for the Chronic Disease Surveillance Project and provided assistance with data extraction; Richard Smith, Programmer Analyst, Maine Bureau of Health, assisted with the statistical analysis for this article. Melanie Lanctot, Manager, Maine Cancer Registry provided data and consultation.

This research was supported in part by funds from the Comprehensive Environmental Response, Compensation, and Liability Act trust fund through an interagency agreement with the Agency for Toxic Substances and Disease Registry, U.S. Public Health Service. Work on this report was also supported in part through a cooperative agreement with CDC, NCHS, Division of Health Promotion Statistics, the CDC Assessment Initiative.

References

1. Brownson RC, Remington PL, and Davis JR. Chronic disease epidemiology and control. Washington DC: APHA, 993.
2. Mendlein JM, Franks AL. Hospital discharge data. Using chronic disease data: a handbook for public health practitioners. Atlanta: Centers for Disease Control and Prevention, 1992.
3. Maine Department of Human Services, Bureau of Health: Maine's Health Status Indicators, 1980–1991. Augusta: Maine DHS, 1994.
4. Kilss B, Alvey W, editors. Record linkage technologies—1985. Proceedings of the Workshop on Exact Matching Methodologies; 1985 May 9–10; Arlington. Publication 1299 (2–86). Department of the Treasury, Internal Revenue Service, Statistics of Income Division, 1985.
5. Keller RB, et al. Dealing with geographic variations in the use of hospitals. J Bone Joint Surg 1990:72-A:1286–1293.