
Uses of Data to Plan Cancer Prevention and Control Programs

LESLIE P. BOSS, MPH, PhD
LUCINA SUAREZ, MS

Dr. Boss, formerly with the National Cancer Institute, is currently Chief, Cancer Prevention and Control Branch, Centers for Disease Control. Ms. Suarez is with the Bureau of Disease Control and Epidemiology of the Texas Department of Health. Tearsheet requests to Dr. Boss, Centers for Disease Control, Mail Stop F 11, 1600 Clifton Rd. NE, Atlanta, GA 30333.

Support for the preparation of this paper was provided by staff associated with the following health departments and National Cancer Institute grants: Illinois Department of Public Health, Dr. Charles Bennett, CA 46668; Nebraska Department of Health, Dr. Gregg Wright, CA 46574; New Jersey Department of Health, Dr. Lawrence Meinert, CA 46733; New York State Department of Health, Dr. Philip Nasca, CA 46586; North Carolina Department of Human Resources, Dr. Georjean Stoodt, CA 46582; Texas Department of Health, Ms. Lucina Suarez, CA 46587; and the Wisconsin Division of Health, Dr. Henry Anderson, CA 46883.

Synopsis

Seven State health departments, those in Illinois, Nebraska, New Jersey, New York, North Carolina,

Texas, and Wisconsin, have participated in an effort to utilize a variety of State-specific cancer-related data to describe the cancer burden in their State's population. The data were then used to develop a statewide cancer plan or supplement an existing plan to address the defined problems. Cancer data have not been well utilized in the planning of intervention programs in the past, and the efforts in these States can serve as models for data use in programs to prevent and control cancer and other chronic diseases. State-specific data can be used to rank needs and make a clear case that can influence decision makers regarding resource allocation.

The purpose of this report is to describe the data sources and additional statistics that were used to provide a broad picture of the cancer burden that will aid in targeting and defining intervention needs. Mortality, incidence, risk factor prevalence, and hospital discharge data appear to be the most accessible and potentially useful of the data sources examined, whereas insurance claims data, sources of treatment data, and environmental data bases were less useful in planning intervention strategies.

IN 1985, THE NATIONAL CANCER INSTITUTE (NCI) defined cancer prevention and control goals for the nation. The resulting document (*1*) includes specific objectives in smoking prevention and cessation, dietary modification, screening for breast and cervical cancer, control of occupational and environmental exposures, and access to state-of-the-art cancer treatment to all persons with cancer, objectives that are considered attainable given the knowledge and technology that exist today (see accompanying box page 356). Given these objectives, NCI staff instituted a grant program for State health departments to stimulate the development of activities focused on these goals. The program includes a statewide planning process which is based on State-specific cancer data.

The purpose of this paper is to describe the various data sources and the manner in which these sources were used by the program participants to accomplish the initial objectives of the grant program. Although the programs we describe were focused specifically on cancer, many of the lessons learned in data use are applicable to other chronic diseases.

The Problem

Although the NCI had defined cancer prevention and control goals for the nation, it is the responsibility of State and local officials to modify and prioritize the objectives in light of State-specific needs and competing priorities. Since interventions occur on the local or State level, an understanding of local needs and resulting local goals and objectives is necessary. A thorough review of State-specific data related to cancer is a first step.

Due primarily to the existence of many population-based incidence registries, more data exist to describe cancer than any other disease. Although the existing data have been well-utilized to understand further the etiology of cancer, in only a few situations have the data been systematically evaluated for the purpose of using the findings to target and plan public health programs in cancer prevention and control. Program planning has frequently been based on historical or political priorities, and resulting programs may not have been located where the need or potential impact could be the greatest.

Potential Solutions

In September 1987, the NCI funded the grant programs, Data-based Interventions for Cancer Control, in State health departments in seven States: Illinois, Nebraska, New Jersey, New York, North Carolina, Texas, and Wisconsin. The primary tasks for the first phase of the program were the use of existing, State-specific data to describe the cancer problem in the State and the preparation of a statewide cancer plan based on the findings from the data. Participants were urged to be creative in their identification of data sources with the hope of identifying those sources that would be helpful to the planning process as well as those that were not.

Although many sources of information were used by the grantees, four kinds of data were used most consistently and effectively: mortality data, incidence data, risk factor data, and hospital discharge data (see accompanying box).

Mortality data. The number and crude rates of site-specific cancer deaths have been readily available since mortality statistics were computerized several decades ago. Age-adjusted mortality data have been mapped to the county level for the nation, again with the primary purpose being the understanding of etiologic factors related to cancer occurrence. Mortality data were used in two forms, the raw data from each State's vital statistics unit and published site-specific numbers, rates, and trends by county for the years 1950–79 (2).

Ways that mortality data were analyzed. All items in the following list of data categories were analyzed by age, race, sex, and geographic area as appropriate and as the number of deaths allowed:

- mortality rates, crude and age-adjusted,
- mapping of rates,
- numbers of deaths by geographic areas,
- standardized mortality ratios (SMRs) for analyzing data involving small numbers of deaths,
- rate ratios (ratio of two age-adjusted rates) to compare regions within a State,
- cause-specific years of life lost,
- trends in State mortality rates over time,
- projected reductions in cancer mortality given proposed intervention goals,
- comparison of county rates to the State rate and to the State with lowest rates,
- number of deaths per county in relation to county's mortality rate,
- smoking-attributable mortality and years of life lost,
- average number of deaths per year compared to an

- expected reduction in deaths per year given a defined intervention,
- ranking of counties with regard to mortality over multiple decades, and
- rank order of States with (site-specific) mortality rates exceeding the U.S. rate, by race.

Health department staff were able to use mortality data very effectively to describe the extent of the cancer problem in their States and the extent of progress toward NCI's cancer control objectives for the nation. For example, the mortality data document the considerable number of preventable cancer deaths still occurring, such as the deaths caused by cancer of the cervix. The preventable cervical cancer deaths observed indicated a possible public health system failure even without an evaluation of incidence data. Many geographic areas do not have access to accurate incidence data, and mortality data can be used as a proxy in those situations where incidence and mortality are similar (for example, for lung cancer).

Incidence data. In the attempt to understand the etiology of cancer and monitor treatment variables and survival of cancer patients, numerous cancer incidence registries have been developed; the oldest, which has existed for more than 50 years, is that of the State of Connecticut. The NCI funds the collection of cancer incidence, stage, treatment, and survival data on geographically discrete segments of the population representing more than 10 percent of the total U.S. population through its Surveillance, Epidemiology, and End Results (SEER) Program. The majority of States without SEER registries have incidence registries, though these are of varying maturity and levels of completeness and accuracy. In addition, incidence registries from other geographic areas where the population's demographic, economic, and lifestyle factors were known to be similar were used as a rough estimate of the completeness of the registry covering the population under consideration.

Ways that incidence data were analyzed:

- incidence rates by selected clinical, pathologic, and sociodemographic variables;
- tabulations for selected sites of the extent of disease (stages);
- mapping of incidence data by county or other geographic area;
- numbers of new cases by geographic areas;
- calculation of the probability of developing (site-specific) cancer by sex and age;
- annual number of new cases prevented if intervention goals are reached;

Summary of Cancer Control Objectives for the Nation, National Cancer Institute

| <i>Action</i> | <i>Target and rationale</i> | <i>Year 2000 objectives</i> |
|---------------|---|---|
| Prevention | Smoking. The causal relationship between smoking and cancer has been scientifically established. | Reduce the percentage of adults who smoke from 34 percent in 1983 to 15 percent or less. Reduce the percentage of youths who smoke by age 20 from 36 percent in 1983 to 15 percent or less. |
| Prevention | Diet. Research indicates that consumption of high-fat and low-fiber food may increase the risk for various cancers. In 1983, NAS reviewed research on diet and cancer and recommended a reduction in fat; more recent studies lead NCI to recommend an increase in fiber. Research is under way to verify the causal relationships and to test the impact on cancer incidence. | Reduce average consumption of fat from 37–38 percent to 30 percent or less of total calories. Increase average consumption of fiber from 8–12 grams to 20–30 grams per day. |
| Screening | Breast. The effectiveness of breast screening in reducing mortality has been scientifically established. | Increase the percentage of women ages 50–70 who have an annual breast examination coupled with mammography to 80 percent from 45 percent for physical examination alone and 15 percent for mammography. |
| Screening | Cervix. The effectiveness of cervical screening in reducing mortality has been scientifically established. | Increase the percentage of women who have a Papanicolaou smear every 3 years to 90 percent from 79 percent (ages 20–39) and to 80 percent from 57 percent (ages 40–70). |
| Treatment | Transfer of research results to practice. Review of the NCI clinical trial and SEER Program data indicates that, for certain cancer sites, mortality as shown by SEER data is greater than that experienced in clinical trials. | Increase adoption of state-of-the-art treatment. |

NOTE: NAS = National Academy of Science; NCI = National Cancer Institute. SOURCE: Reference 1.

percent of patients treated at American College of Surgeons (ACOS) approved facilities; and extent of treatment provided or not provided.

A useful measure for registries that collect information on in situ cancers is an in situ-to-malignant ratio. This measure can indicate apparent late detection of the cancer in subgroups of the population under consideration. It is also useful as a tool for evaluating screening programs. Similarly, an incidence-to-mortality ratio can suggest survival differences in subgroups of the population if registries do not collect survival data or their populations are too small to calculate valid survival statistics. When interpreted with care, such substitute statistics can be useful in expanding the potential of a basic incidence registry.

Risk factor data. Various types of risk factor data, such as smoking prevalence of the population, have also existed for considerable periods. Of risk factor sources evaluated, the Centers for Disease Control's

(CDC) Behavioral Risk Factor Surveillance System was the most frequently used for these projects.

Ways that risk factor data were analyzed:
prevalence of tobacco use over time and by a variety of demographic variables;
comparison of prevalence rates to those of other States that collect similar data;
extent of knowledge about mammography;
percent reporting a recent mammogram by age and race, resulting in estimates of number of persons being screened and those not being screened; and
percent reporting obesity by age, sex, and race.

Risk factor data are important for program planning and as intermediate outcome measures since disease or death may not occur for many years. They can also serve to indicate future disease risk. For example, in many Hispanic populations, comparatively low rates of smoking-related diseases and deaths suggest that resources might be better utilized in other populations.

Cancer-related Data Bases Used by Seven State Health Departments

| <i>Data base</i> | <i>Cancer-related data provided</i> |
|--|--|
| <i>Mortality</i> | |
| Death certificates | Information on all deaths, demographic variables, cause of death, occupation, industry. |
| National Cancer Institute (NCI)—Environmental Protection Agency data, 1950–79 | Number of deaths, age-adjusted cancer and county-specific mortality rates by decade, percent changes in rates over decades by race (white, nonwhite) and sex, significant difference from U.S. data. |
| <i>Incidence</i> | |
| Population-based cancer incidence registries; statewide, regional specialized cancer registries (for example, pediatric, lymphoma) | Variable, but generally demographics, cancer site, and stage. Less often, histology, followup, occupation, smoking history. Information on all cases. |
| <i>Risk factor</i> | |
| Behavioral Risk Factor Surveillance System | Variable from State to State. Demographic variables, use of tobacco (smoking and smokeless) and alcohol, knowledge and past use of mammography, obesity. |
| <i>Hospital discharge</i> | |
| Hospital discharges | Demographic variables, (no patient identifiers) diagnosis, costs, principal source of payment, limited number of diagnoses and procedures related to hospital stay. |
| <i>Facilities</i> | |
| American College of Surgeons (ACOS) | Listing of ACOS-approved hospitals, cancer centers. |
| Local American Cancer Society office | Location of prevention-early detection programs and resources. |
| Cancer Information Service (NCI-funded offices only) | Location of prevention-early detection treatment programs and resources. |
| Radiologic health unit of State health department | Location of dedicated mammography units in State. |
| <i>Personnel</i> | |
| State medical societies, organizations State board of medical examiners | } Number and location of oncologists, pediatric oncologists, hematologists, radiation oncologists, and so forth. |
| <i>Environmental</i> | |
| State environmental protection agencies | Abandoned landfill listings, water and air quality monitoring, industrial and chemical surveys, radon testing. |
| SUPERLOG | Listing of Superfund sites. |
| <i>Treatment information</i> | |
| American College of Surgeons | Pattern of care surveys. |
| Cancer centers | Level of participation in clinical trials and treatment groups. |
| Public health clinics | Level of direct service care through public services. |
| <i>Taxation records</i> | |
| State department of taxation | Cigarette and smokeless tobacco sales taxes. |

However, high rates of smoking prevalence and sharply increasing prevalence over time suggest that smoking prevention and cessation programs should be targeted to those populations.

Hospital discharge data. Most patients with certain types of cancer are hospitalized, and hospital discharge records provide a wealth of information, particularly

that related to the cost of care.

Ways that hospital discharge data were utilized: rates of hospitalization associated with a diagnosis of cancer, inpatient costs associated with cancer care by type of cancer, use of hospital resources for cancer patients,

'Although the existing data have been well-utilized to understand further the etiology of cancer, in only a few situations have the data been systematically evaluated . . . to target and plan public health programs in cancer prevention . . .'

prevalence of certain types of medical procedures related to cancer care,
patient's mean and median length of stay for various cancer sites,
average annual number of hospitalizations,
analysis of patterns of care among mastectomy patients,
smoking-attributable health care costs, and
rate of admissions for chemotherapy.

Even though the relatively young hospital discharge data bases are not yet complete in many States, these inpatient data systems provide information on cancer-specific costs and the burden on the medical care system that is impressive in magnitude and effective in educating policy makers. As with all data, especially those being used in a manner different from the purpose for which they were collected, the data must be thoroughly understood and properly presented to be truly useful.

Additional data sources. Various additional data sets were also used. Many of these are available within State government, and often within the health department itself. Data sources used to describe the facilities within the State included the ACOS listing of hospitals with approved cancer programs and information from the local officials of the American Cancer Society (ACS), Cancer Information Service, and the radiologic health unit of the State health department. Information on personnel resources came from State medical organizations and the State board of medical examiners. Environmental data bases included lists of abandoned landfills and results of water and air monitoring. Industrial sources of chemical carcinogens have been surveyed in some States, and results of radon testing were often available. Limited treatment information could be obtained from ACOS patterns of care surveys, cancer centers, and public health clinics. State taxation records provided information on cigarette and smokeless tobacco sales and tax income.

Uses of these data sources:

mapping of various types of facilities, for example, radiation treatment facilities, ACOS approved hospi-

tals, location of dedicated mammography units;
estimation of percent of State's population not within X miles of a facility;
ratio of maximum number of women who can be screened given existing units to number of women needing screening;
mapping of personnel resources, for example, oncologists, pediatric oncologists, radiologists;
comparison of rates of various cancers around Superfund or other hazardous sites with other rates;
location of programs, for example, smoking cessation programs in industries that use carcinogens that react synergistically with tobacco;
likelihood of public service clients to access adequate treatment following diagnosis;
number of patients participating in clinical trials or treatment groups;
volume of tobacco tax revenues and how those monies are spent;
annual tobacco sales by State and by region (limited value for prevalence information because of export to and import from other jurisdictions); and
comparison of State tax rate with that of surrounding States.

In addition to these sources, others were used. Various types of insurance claims data were examined, for example, Medicare, Medicaid, Blue Cross, and State Employees Insurance records; however, such sources did not prove to be helpful, given the level of effort required to access the information.

National data sources were used primarily to compare with some local data. These national sources included the National Health and Nutrition Examination Survey (NHANES), Hispanic Health and Nutrition Examination Survey (HHANES), Nationwide Food Consumption Survey, National Health Interview Survey, ACS Cancer Prevention Study II, and the ACS Mammography Survey. The Texas sample from HHANES was sufficiently large to provide State-specific data on the Hispanic population. Finally, there were special surveys and epidemiologic studies unique to each State. Examples include the New York State Teenage Health Survey, Wisconsin Health Status Survey, North Carolina Citizens Survey, and the Nebraska Annual Social Indicators Survey.

New data variables can be calculated or estimated from a variety of existing information sources. For example, one can estimate the number of persons in a defined population who need to have a Papanicolaou (Pap) smear, have a mammogram, stop smoking, and so forth if a particular goal is to be reached. If the goal is a mammogram for every woman ages 40–49 every other year and an annual mammogram for every woman

50 and older, then the number of mammograms that will have to occur annually in a population can be estimated readily. Such information is essential to plan programs that meet a particular goal. Once the number of tests needed to be done is estimated, the number and distribution of mammography units can be estimated and projections of the professional staff needed can be made. Other statistics useful for the planning process are years of life lost from premature mortality and numbers of lives to be saved through potential prevention programs.

Two micro computer software programs can assist in this data-based planning process. SAMMEC—Smoking Attributable Morbidity, Mortality, and Economic Costs—is software developed by staff of the Minnesota Department of Health that assists in the calculation of a number of smoking-related statistics for a defined geographic area (3). The statistics include the number of deaths, years of life lost, and direct and indirect health care costs for smoking-related deaths. The second program, CAN*TROL, allows for modeling the effect of proposed prevention and control programs on future mortality in a defined population (4). Results from CAN*TROL are being used to compare the potential effect on mortality of a variety of different possible intervention programs.

One State's use of data in decision making. Since each State varied in the choice of data sources, approach to data, and subsequent selection of interventions based on findings, the process can best be described by illustration. In one State, the percentage of excess mortality based on the ratio of the race or ethnic group- sex-specific regional rate to the State rate was used to identify regions with differing risks of cervical, breast, lung, and colon cancer. Maps specific for sex, race, and ethnicity were produced, and they showed wide variation in mortality risks not only by race and ethnicity but also by geographic region. Decisions to place screening programs were ultimately based on a relative measure of potential impact depending on the percentage excess or the attributable proportion and the number of deaths in each region.

Incidence data from a limited registry were used primarily to document racial and ethnic differences in the early detection of certain cancers and to support screening programs in these high-risk populations. Data on the prevalence of tobacco use, use of Pap smears and clinical breast examinations, and on insurance coverage from various surveys specific to the State were employed to emphasize racial and ethnic differences and to shed light on the possible reasons for excess mortality in certain groups.

Analyses of information from all of these sources led

to the identification of two regions of the State, one with a predominately Hispanic population and the second with a predominately black population, where screening women for breast and cervical cancer and smoking cessation programs targeted to minority youth would have the greatest impact.

Use in Defining Problems

Given this wealth of data and the past success of public health workers in using data to plan other public health programs, particularly in the field of infectious diseases, one would expect to have numerous models that use such data for the planning and evaluation of public health programs to prevent and control cancer. In fact, few models exist. Several incidence registries, for example, those in Pennsylvania and Missouri, have been developed with such uses as their primary objectives. Staff of several health departments have successfully used the SAMMEC approach along with other findings to generate interest in and resources for anti-tobacco programs. In contrast, most State health departments have experience in cervical cancer screening as part of regular entitlement programs. Some States also have incidence registries but have not used information on the proportions by stage of cervical cancer cases to plan and evaluate these screening programs.

Although adequate planning of programs for many chronic diseases is hampered by the lack of data, such is not true for cancer. For many agencies, having a cancer incidence registry is considered essential for the planning process; however, those health departments in States with high-quality incidence registries appear to be no more likely to have used cancer data for planning purposes than health departments without registries. Although problems with the accuracy of cause of death information on death certificates of cancer patients are well known (5), it is interesting to note that for the few participants in this project, stage and mortality data were consistent, and planning decisions based on mortality data alone would not have been different if they had been based on incidence and stage data. Clearly, there is an advantage in having multiple consistent sources of data to plan programs. Excessive mortality could be due to inadequate care following diagnosis or represent more aggressive disease rather than a later stage at diagnosis. In such situations, the information that can be provided by a complete and accurate incidence registry is invaluable.

Appropriate and accessible data were not uniformly available for each of the priority areas for cancer control listed in the year 2000 objectives. Adequate State-specific nutritional data were lacking in all cases, and health department staff had to depend almost totally on

national data sources for nutritional information. As a result, potential nutritional interventions were not fairly considered in this planning process that were required in the specifications of the NCI grants to be based on data findings. It is hoped that later rounds of grants, which include States with better State-specific nutritional data, will provide models in this area.

Although it is beyond the scope of this paper to comment in detail on the quality of each data source mentioned, awareness of the strengths and weaknesses of each source is critical. Few of the data sets were collected for the specific purpose of being used in the manner proposed in this paper, so the data must be well understood before they are used. There are advantages and disadvantages to every data source and to using the data in the manner planned, and failure to understand the issues may lead to misuse of the information.

Although many potential sources of information exist, a number of these sources did not provide enough information to justify the effort needed to access the information, and others provided overlapping information. For example, a number of surveys included information on smoking prevalence, and the sometimes considerable differences in rate estimates found may have been the result of the different methodologies used to collect the information and may result in confusion. Understanding the methodology of data collection is, therefore, a crucial first step in the use of the data.

Having the information accessible in published tables, charts, and graphs does not ensure that the data will be adequately and properly utilized. Within the health department, it is necessary for the data staff and the program staff to come together over the tables, charts, and graphs, collate the findings from various sources, and assess the significance of the findings in the light of existing and potential state-of-the-art programs. This working together increases the likelihood that findings will influence programs and helps ensure the proper interpretation and use of the data by program staff. In some situations sufficient information may not exist, and further epidemiologic studies may be required prior to planning actual programs. Program evaluations need to be planned with knowledge of existing data bases and with input from program staff as to what data are needed for the future. In addition, a feedback loop of evaluation results needs to be directed into the subsequent planning process. Although this describes a standard planning model, it has not always been well used by chronic disease epidemiologists and program staff in health departments.

In the past many health department staff have not been involved in cancer prevention and control and, indeed, chronic diseases in general. Chronic disease epidemiologists are relatively new to many health

departments, and some departments are still without such expertise. As a result, there can be reticence in approaching data for such a complex group of diseases as cancer. In some of the sites of this grant program, a variety of persons outside the health department who were expert in the cancer field contributed their knowledge, assessing the quality and value of various data sources, guiding health department staff in the correct questions to ask of the data, and in interpretation of data results. Such cooperation has been central to the success of the programs.

An additional use of State-specific data is to prioritize needs and make a clear case that will influence decision makers regarding resource allocation. This step requires that data be displayed in a format that is quickly absorbed and easily compared with other data, a skill not natural to many epidemiologists. Because this grant program requires the sharing of information with State legislators regarding the extent of the cancer problems within the State, the potential for responding to the problem, and the resources necessary to respond, models of such data presentation should be forthcoming.

The fact that a cancer plan exists does not automatically ensure acceptance of the plan by the health care community or guarantee adequate resources to address the plan. Nor does it ensure an eventual decrease in cancer incidence and mortality rates. However, an understanding of the problem and a plan to respond to it are essential preliminary steps. The process has appeared to be useful in each of the seven health departments involved in the project. Most, if not all, States have this type of data available and could consider using this process for cancer as well as for other chronic diseases and conditions.

References

1. Greenwald, P. G., and Sondik, E. J., editors: Cancer control objectives for the nation: 1985-2000. NCI Monograph No. 2, NIH Publication No. 86-2880, U.S. Department of Health and Human Services, 1986.
2. Riggan, W. B., et al.: U.S. cancer mortality rates and trends, 1950-1979. Vol. 1, EPA-600/1-83-015a, published by the NCI/EPA Interagency Agreement on Environmental Carcinogenesis. U.S. Government Printing Office, Washington, DC, September 1983.
3. Shultz, J. M.: SAMMEC: smoking attributable mortality, morbidity, and economic costs (computer software and documentation). Minnesota Department of Health, Center for Nonsmoking and Health, Minneapolis, 1986.
4. Eddy, D. M.: A computer-based model for designing cancer control strategies. *In* Cancer control objectives for the nation: 1985-2000, edited by P. G. Greenwald and E.J. Sondik. NCI Monograph No. 2, NIH Publication No. 86-2880, U.S. Department of Health and Human Services, 1986, pp. 75-82.
5. Percy, C., Stanek, E., and Gloeckler, L.: Accuracy of cancer death certifications and its effect on cancer mortality statistics. *Am J Public Health* 71: 242-250 (1981).