Preface

Population studies facilitate the discovery of genetic and environmental determinants of cancer and the development of new approaches to cancer control and prevention. Furthermore, epidemiology studies play a central role in making health policies. Cancer epidemiology may address a number of research areas such as:

- familial predispositions to colon cancer and breast cancer study to determine whether families who carry a genetic predisposition to breast cancer may also be at risk of colon cancer, and vice versa;
- prospective examination of whether baseline dietary intakes and serum levels of carotenoids and vitamin A are associated with subsequent risk of lung cancer;
- analysis of the relationship between serum levels of sex-steroid hormones and genetic polymorphisms in biosynthesis enzymes in a prospective cohort of pre-menopausal women;
- analysis of the role of HLA-Class II similarity/dissimilarity between sexual partners and the role in HIV transmission, using the multicenter hemophilia cohort study population for the data set;
- multiple comparisons and the effect of stratifying data on study power.

This two-volume set compiles areas of research that cover etiological factors or determinants that contribute in the development of cancer as well as describe the latest technologies in cancer epidemiology. Emphasis is placed on translating clinical observations into interdisciplinary approaches involving clinical, genetic, epidemiologic, statistical, and laboratory methods to define the role of susceptibility genes in cancer etiology; translating molecular genetics advances into evidence-based management strategies (including screening and chemoprevention) for persons at increased genetic risk of cancer; identifying and characterizing phenotypic manifestations of genetic and familial cancer syndromes; counseling individuals at high risk of cancer; investigating genetic polymorphisms as determinants of treatment-related second cancers; and pursuing astute clinical observations of unusual cancer occurrences that might provide new clues to cancer etiology. All the chapters in these two books are divided into three categories:

Volume 1:

Cancer Incidence, Prevalence, Mortality, and Surveillance

Methods, Technologies, and Study Design in Cancer Epidemiology

Host Susceptibility Factors in Cancer Epidemiology

Volume 2:

Modifiable Factors in Cancer Epidemiology

Epidemiology of Organ-Specific Cancer

These chapters have been written in a way that allows readers to get the maximum advantage of the methods involved in cancer epidemiology. Several examples of specific organ sites would be helpful in understanding cancer etiology.

Mukesh Verma, Ph.D.

Contents

Chapter 2

Cancer Registry Databases: An Overview of Techniques of Statistical Analysis and Impact on Cancer Epidemiology

Ananya Das

Summary

Cancer registries provide systematically collected information on cancer incidence, prevalence, mortality, and survival of different cancers. Aggregated and de-identified patient-level information on cancer is available for analysis from individual cancer registries, nationally from the Surveillance, Epidemiology, and End Results program, the Centers for Diseases Control and Prevention, the North American Association of Central Cancer Registries; and internationally from the International Association of Cancer Registries. Over the past few decades, the type and extent of cancer-related information captured by different cancer registries have been greatly expanded by linkage with other population-based information sources, such as the census data and the Centers for Medicare and Medicaid Services claims data. In addition, sophisticated statistical analytical techniques have been developed that have greatly expanded the traditional purview of cancer registries focused on descriptive epidemiology and disease quantification to a much broader analytical horizon ranging from study of cancer etiology; rare cancers in specific demographic groups; interaction of environmental and genetic factors in causation of cancer; impact of co-morbidities, race, geographic, socioeconomic, and provider-related factors on access, diagnosis, and treatment; outcomes and end results of cancer treatment; and cancer control initiatives to diverse areas of cancer care disparity, public health policy, public health education, and importantly, cost-effectiveness of cancer care. Thus, it is not surprising that cancer registries have increasingly become indispensable parts of local, national, and international cancer control programs, and it is certain that cancer registries will continue to be extraordinary resources of information for clinicians, researchers, scientists, policy makers, and the public in our fight against cancer.

Key words: Cancer, cancer registry, epidemiology, SEER, SEER-Medicare database.

1. Introduction

Cancer is a leading cause of death worldwide, and it accounted for more than a quarter of all deaths in the United States in 2003. One of the essential tools in the fight against cancer is systemically collected information on all aspects of cancer. Early attempts at

Mukesh Verma (ed.), *Methods in Molecular Biology, Cancer Epidemiology, vol. 471*

Book doi: 10.1007/978-1-59745-416-2

systemic collection of information on cancer date back to the early eighteenth century; the first American cancer registry was established in the 1920s, and the first national cancer registry was the Danish cancer registry established in 1942 *(1)*. Since then, there has been explosive growth in the number of population based cancer registries and, currently approximately 450 member registries contribute to the International Association Cancer Registries (IACR) representing approximately 21% of the world population *(2)*.

2. Cancer Registry Databases

Although the original purpose of the data collected by cancer registries was to generate statistics on incidence of cancer, over the past several decades both the type of data that is being collected and the analytical information based on available data that is being reported have greatly expanded to examine nearly every aspect of cancer epidemiology. These aspects include but are not limited to statistical analysis of cancer incidence, mortality, survival, and risk factors; hypothesis generation regarding cancer etiology; studying rare forms of cancer, in specific patient groups; monitoring programs for screening and surveillance, and treatment outcomes; and identifying disparities across population subgroups with respect to race, socioeconomic, and demographic variables.

Cancer registry databases provide a wealth of data for research in cancer epidemiology. Aggregated and de-identified patient-level information on cancer is available for analysis from individual cancer registries, nationally from the Surveillance, Epidemiology, and End Results (SEER) Program, the Centers for Diseases Control and Prevention (CDC), the North American Association of Central Cancer Registries (NAACCR); and internationally from the International Association of Cancer Registries (IACR).

One of the most important requirements of cancer surveillance data is that the registry data need to be standardized in terms of coding practices, case identification, and conversion of medical terminology to appropriate categories for enabling comparison across registries, countries, and even over time for studying trend. Currently, most population based cancer registries have standardized operating procedures for collection of surveillance data that is reasonably accurate, complete, valid and timely; however, there is still considerable variability in the quality of data available from different cancer registries *(3)*.

The type and extent of cancer-related information captured by different cancer registries is quite variable and limited by available resources. The essential components of any cancer registry data

are personal identification components (for linkage); demographic variables; and incidence date of cancer with most valid diagnosis, site and morphology of cancer, tumor behavior, and source of information. Recommended variables are follow-up data, date and status at last contact, stage and extent of disease at diagnosis, and initial treatment provided *(4,5)*.

Because cancer registries are gradually becoming an essential component of national cancer control programs, an increasing amount of resources is being committed to the cancer registries; and correspondingly, the capture of information on cancer patients has been going beyond the confines of traditional registry data set and incorporating details of clinical treatment and also useful socioeconomic information.

With the widespread availability and use of sophisticated computerized databases, it has become possible to link cancer registry information to other population-based information sources that not only provide population counts as cancer rate denominators but also provide patient-level data on co-morbidities, risk factors, treatment outcomes, and access to care; and linkage to census data at the neighborhood level (e.g., zip code, census track, block-group); and other geographical information systems have enabled researchers to answer complex issues such socioeconomic disparity in cancer care to and test etiologic hypotheses. For example, SEER uses the Population Estimates Program data of the U.S. Census Bureau and U.S. mortality data, collected and maintained by the National Center for Health Statistics, for population counts to be used as denominator for calculation of cancer incidence and mortality *(6)*. Similarly, linkage of SEER data to the Medicare database and to AIDS registries has enabled researchers to study cancer in Americans eligible for Medicare *(7)* and AIDS-related cancers *(8)*, respectively. Similarly, population based databases, such as, Behavioral Risk Factor Surveillance System of the CDC *(9)*, the National Health Interview Survey *(10)*, the National Hospital Discharge survey *(11)*, and the National Ambulatory Medical Care Surveys *(12)* are available for characterization of risk factors and cancer screening behaviors.

3. Cancer Registry Databases in the United States

In the United States, the National Cancer Institute's (NCI) SEER Program, which started collecting data from 1973, currently collects and publishes cancer incidence and survival data from population-based cancer registries covering approximately 26% of the U.S. population *(13,14)*. The National Program of Cancer Registries (NPCR), which was established by an act of

Congress in 1992, and which is administered by the CDC, supports central cancer registries in 45 states, the District of Columbia, and other U.S. territories *(15)*. These data represent 96% of the U.S. population. Together, NPCR and SEER collect data for the entire U.S. population.

3.1. SEER Database

The SEER Program is one of the best sources of information for evaluating cancer epidemiology in the United States. SEER currently collects and publishes data on patient demographics, primary tumor site, tumor morphology and stage at diagnosis, first course of treatment, and follow-up for vital status from 18 geographical areas in the United States, and it is considered representative of the U.S. population including minorities. SEER Program registries cumulatively have information on 6 million cancer cases, and >350,000 cases are added to the database each year. This database is updated annually and provided free of charge as a public service in print and web-based electronic formats, for use by researchers, physicians, public health officials, policy makers, and the public. The SEER program is considered to be the gold standard for all population based cancer registries with rigorous quality control measures.

3.2. SEER-Medicare Database

To provide patient-level information on different types of cancer in the United States, in a collaborative effort of the NCI, SEER registries, and the Centers for Medicare and Medicaid Services (CMS), the SEER database have been linked to claims-based measures of co-morbidities; screening and evaluation tests; and detailed treatment and treatment outcomes data, including cost data from the CMS *(7,16)*. Using a matching algorithm based on unique patient identifiers, such as social security numbers and date of birth, cancer data on individual patients available from the SEER registries was linked to a master Medicare enrollment first in 1991; and since then, it has been updated in 1995, 1999, and 2003. The SEER-Medicare data are available to outside investigators for research purposes. The SEER data included as part of the SEER-Medicare files are in a customized file known as the Patient Entitlement and Diagnosis Summary File. This file contains one record per person for individuals in the SEER data who have been matched with Medicare enrollment records with clinical information available for up to 10 diagnosed cancer cases, selected variables pertaining to Medicare enrollment information for that patient, and information about the median household economic and educational status for the census tract or zip code where the person resides. In general, all Medicare files have fields for race, sex, and date of birth or age, the date(s) of service, diagnostic codes, and procedure codes (either International Classification of Diseases [ICD]-9 codes for procedures and diagnoses or Health Care Financing Administration Common Procedure

Coding System codes for procedures) in addition to the amounts for charges and reimbursement. In addition, every Medicare file contains a provider identification number for the hospital or physician. Medicare files included as part of the SEER-Medicare data contain the SEER case number on each claim, which is the unique nonidentifiable number assigned to each cancer patient by the registries. To allow comparison studies with a control group without cancer, there are Medicare files containing similar information for a random sample of 5% of Medicare beneficiaries residing in the SEER areas persons who do not have cancer. The availability of the SEER-Medicare linked data provides researchers with a unique resource for extracting information on cancer with a patient-level focus and a longitudinal perspective before, during and after diagnosis of a particular cancer.

4. Overview of Statistical Techniques of Analyzing Cancer Registry Databases

Statistical analysis of cancer registry databases can be broadly grouped into two categories: descriptive and analytical. Traditionally, one of the basic functions of cancer registries have been to provide the public, scientists, researchers, and policy makers with descriptive data on incidence, mortality, prevalence, and survival of different cancers.

4.1. Incidence

Cancer incidence rate, a basic index of cancer epidemiology, is the number of newly diagnosed cancers of a specific site/type occurring in a specified population during a defined period. Age-adjusted incidence rates, and also temporal trends in incidence, are commonly derived statistics from population registry-based data. An age-adjusted rate is the weighted average of the age-specific rates where the weights are the proportion of persons in the corresponding age-specific group of a standard population. It should be noted that the unique characteristics of registry-based data require specialized statistical techniques for correct analysis and interpretation. For example, there is always an inevitable delay of certain period between the diagnosis of the cancer and its eventual reporting to a cancer registry. To adjust the current case count with the anticipated future corrections considering delay in reporting, the delay distribution of cancer cases has been modeled in precisely determining the current cancer trends, and also in monitoring the timeliness of cancer data collection by the registries. It has been shown that ignoring reporting delay and reporting error may produce downwardly biased cancer incidence trends, particularly in the most recent years of diagnosis *(17,18)*.

A risk-adjusted incidence rate can be calculated from registry databases that use first primary cancer as the numerator and the population who never had that particular cancer as denominator and help in understanding the actual transition rate of a healthy population to the cohort with a particular cancer *(19)*. Not surprisingly these risk adjusted rates are often different than the standard incidence rates that are derived from reported count of multiple instances of primaries of the same cancer in the numerator, and use the total population as the denominator.

Calculation of population-based measures of lifetime and age-conditioned probability of developing cancer, and also dying of cancer in the general population from a particular cancer, has been extensively described and is being increasingly used as practical and easily interpretable measures of cancer epidemiology *(20,21)*. The identification of changes in the temporal trend is an important issue in the analysis of cancer mortality and incidence data. Given the limitations of traditional linear and Poisson regression models, newer statistical techniques such as joinpoint models have been described to analyze registry-based data for temporal trends. The point in time where a trend changes direction is called a joinpoint. The joinpoint regression model describes continuous changes in rates, and by using a grid-search method, it fits a series of joined straight lines on a log scale to the expected annual percentage change in the incidence rate of a particular cancer over a defined number of years to fit the regression function with a number of joinpoints. Commonly, it uses a Monte Carlo permutation-based significance testing to determine the points in time when the direction of trends changes significantly *(22)*.

Patients with a first primary cancer are more likely than the average person to develop a subsequent malignancy, because of genetic susceptibility, a shared etiology, or even as a consequence of treatment of the first cancer. Also, effective screening and treatment regimens, coupled with more cancer diagnoses because of an aging population in the western world, have resulted in increasing numbers of cancer survivors who are at risk for subsequent cancers *(23)*. Cancer registry databases provide a unique opportunity to study the association of multiple primary cancers and to test hypotheses that explore plausible links in the etiology of different cancers, such as effect of smoking. The statistical methods used to investigate multiple primary malignant neoplasms in large populations, such as the SEER cohort, are well established *(24,25)*. A defined cohort of persons previously diagnosed with a certain cancer is followed through time to compare their subsequent cancer experience to the number of cancers that would be expected based on incidence rates for the general population. The standardized incidence ratio (SIR) is calculated as the ratio of the observed number of second primary malignancies to the expected number of second primary malignancies. The

statistical significance is usually assessed based on the assumption that the observed number of cases follows a Poisson distribution. In examining any two primary malignancies (A and B), two relevant statistical parameters must be evaluated: the SIR of A following B (SIR A/B) and the SIR of B following A (SIR B/A). Biologic plausibility of a significant association between a pair of primary malignancies is better established if the association is bidirectional. Mathematical modeling has demonstrated that under relatively general assumptions regarding the number of common risk factors, the prevalence of these factors, and the interaction (synergism) between them, the two SIRs should be nearly equal, provided that the lifetime risk profiles of the individuals in the study do not change.

4.2. Mortality

Mortality rates are another group of basic statistical indices commonly reported by analysis of cancer registry-based data. A cancer mortality rate is the number of reported cancer deaths of a specific site or type occurring in a specified population during a year (or group of years), usually expressed as the number of cancers per 100,000 population at risk. Several statistical methods and software tools have been developed for the analysis and reporting on cancer mortality statistics from cancer registry databases. These methods include age-adjusted rates with gamma confidence intervals *(26)*, trends in rates over time based on frequencies (such as percentage of change, annual percentage of change), and incidence-based mortality, which allows a description of mortality by selected variables associated with the cancer onset *(27)*.

4.3. Prevalence

Prevalence of cancer represents new and preexisting cases alive on a certain date in contrast to incidence, which reflects new cases of cancer diagnosed during a given period. Thus, prevalence is a function of both the incidence and survival. In cancer epidemiology, prevalence of a cancer is a statistical parameter of utmost importance because it truly reflects the burden of a particular cancer in the population, and it is important for policy makers in making decisions regarding healthcare resource allocation. One of the emerging prevalence measures is care prevalence, which is the measure of prevalent cases under care *(28)*. Although cancer registries typically do not provide such information, with the availability of linked databases (such as, the SEER-Medicare–linked database) allowing longitudinal tracking of patients with cancer, such measures are being increasingly reported as a more refined quantification of the burden of cancer. Another nontraditional measure of cancer prevalence is noncure prevalence, which is an estimate of prevalent cases that have not been cured of disease, and it may be a more specific and practical prevalence measure in terms of economics of cancer care *(29,30)*.

Prevalence of a particular cancer may be described as limited duration or complete prevalence *(31,32)*. Limited duration prevalence represents the proportion of people alive on a certain day that had a diagnosis of the disease within a defined period of past years. Complete prevalence represents the proportion of people alive on a certain day that previously had a diagnosis of the disease, regardless of how long ago the diagnosis was, or whether the patient is still under treatment or is considered cured. Registries of shorter duration (such as the SEER) with <40 or 50 years of data collection, can only estimate limited duration prevalence. In the United States, the only registry with sufficient length of follow-up data (since at least 1940) for reasonable prediction of complete cancer prevalence is the Connecticut Tumor Registry *(33)*. However, it should be noted that projecting estimates of national prevalence from a single regional registry has inherent issues with representativeness.

To derive complete prevalence from data from registries of shorter duration, a statistical modeling technique known as completeness index has been described and used by cancer registries, including SEER, in reporting the cancer statistical reviews *(30,34)*. Completeness index, which has been validated by SEER for selected cancer sites by using the Connecticut cancer registry, uses an estimation technique where complete prevalence is described as function of the observation time of a particular registry; incidence and survival indices before 1975 are predicted by modeling the SEER data. Advantages of the modeled completeness index are its stability even for rare cancers and importantly, that it permits estimation by SEER-derived race and ethnicity data. One of its disadvantages is that this technique generally cannot be applied for estimating complete prevalence of childhood cancers. Other approaches to estimate complete prevalence are cross-sectional population surveys *(35)*, the transition method rate *(36)*, and back calculation *(37,38)*. Cross-sectional population surveys use self-reporting for identification of cancer cases, but they obviously have limitations of underreporting and misclassification of cancer. A second approach is to use data from disease registries to estimate the various intensity (hazard or transition rate) functions that determine point prevalence *(36)*. As described by Keiding *(39)*, a person at calendar time t in the healthy state H may transit to the chronic disease state (e.g., cancer), I, with intensity a(t, z) that may depend on calendar time t or age z. Alternatively, the individual may die (state D) with intensity p(t, z) directly from state H. A person in state I is at risk of death with intensity X (t, x, d), which may depend on duration d in state I as well as on t and x. These intensities determine the prevalence of the chronic disease if one assumes hat the numbers of births at calendar time t is governed by process with intensity p(t) that is independent of the subsequent life histories. Derived from earlier statistical modeling

techniques applied for estimation of incidence and prevalence of chronic diseases, such as human immunodeficiency virus, techniques using back calculation methods have been described that allow estimation and projection of cancer prevalence patterns by using cancer registry incidence and survival data. As a first step, the method involves the fit of incidence data by an age, period, and cohort model to derive incidence projections. Prevalence is then estimated from modeled incidence and survival estimates. Cancer mortality is derived as a third step from modeled incidence, prevalence, and survival *(37, 38)*.

The counting method, which is commonly used to estimate prevalence, uses tumor registry data to count cases alive on a particular prevalence date, whereas adjustments are made to account for cases that are lost to follow-up who would otherwise have made it to the prevalence date. The expected number of cases lost to follow-up who make it to the prevalence date is computed using conditional survival curves for specified cohorts. Depending on the research question in the counting method, it is important to clarify which method was adopted in counting the tumor; often, only the first malignant primary tumor recorded in a particular registry is counted; in other instances, the first malignant tumor per site in a defined observation period is counted *(32,36)*.

4.4. Survival

Cancer survival is the proportion of patients alive at some point subsequent to the diagnosis of their cancer, or from some point after diagnosis (conditional survival). It is usually represented as the probability of a group of patients "surviving" a specified amount of time. It is important to understand that unlike incidence or mortality parameters, where the total population constitutes the denominator, only patients diagnosed with cancer are taken into account in calculating the survival parameter. Commonly used survival measures are observed all cause survival, and net cancer-specific survival, which is the probability of surviving cancer in the absence of other causes of death. Net cancer-specific survival rate does not take into account the impact of mortality from other causes; thus, it is considered a better parameter to understand temporal trends in survival or comparing survival in different racial and ethnic groups or even amongst different registries *(40)*. Conversely, crude probability of death (which is the probability of dying of cancer in the presence of other causes of death) is a better measure to assess cancer survival at a patient-level focus because mortality from other causes practically does play an important role in determining cancer survival for an individual patient.

Net cancer-specific survival and crude probability of death have two methods in which they can be estimated: using cause of death information or expected survival tables. Cause of death information in the registry data typically comes from death

certificates, which are often incorrect *(41)*. For example, in a patient dying from a metastatic cancer, the death certificate often cites the metastatic cancer is the cause of death rather than the primary cancer. One way of circumventing the errors inherent in the death certificates is to use expected survival rates from population based life expectancy tables, with an assumption that the general population dies of causes other than cancer at the same rate as the cancer population *(42,43)*. Besides being a relatively strong assumption, this method may be problematic in that such tables may not be available for defined cohorts in a particular geographic area. If life tables are used for estimating survival measures, then there are two basic measures. One measure, relative survival, is defined as the ratio of the proportion of observed survivors (all causes of death) in a cohort of cancer patients to the proportion of expected survivors in a comparable cohort of cancer-free individuals. The formulation is based on the assumption of independent competing causes of death. Because a cohort of cancer-free individuals is difficult to obtain, practically expected life tables are used assuming that the cancer deaths are a negligible proportion of all deaths. The second measure is crude probability of death by using expected survival, which uses expected survival (obtained from the expected life tables) to estimate the probability of dying from other causes in each interval.

There are a few issues that are important to understand in using survival estimates based on cancer registry data. There is always a lag between current year and available follow-up data in a particular registry, and it is important to specify the cohort of patients in terms of year of diagnosis and length of available follow-up data when making survival estimates. The other inherent issue in estimating long-term survival is that such estimates are only available for only those cohorts who were diagnosed a long time ago and have enough follow-up. Thus, direct estimates of long-term survival may not be very relevant for newly diagnosed patients, especially with respect to those cancers where there have been tremendous improvements in management and survival *(44)*. To provide more up-to-date estimates of survival for newly diagnosed patients, in the projection method that has been developed by SEER, a regression model is fit to interval relative survival and includes a parameter associated with a trend on diagnosis year *(45,46)*. The cumulative relative survival rate in a target year is calculated by multiplying the projected interval survival rates for that year.

4.5. Spatial

One of the newest applications of population-based cancer registry data is spatial representation and analysis of cancer data by using geographic information system (GIS). *(47,48)*. Such spatial representation or mapping of cancer data is becoming an invaluable tool in the exploration, analysis, and communication

of cancer data in understanding relationships between cancer and other health, socioeconomic, and environmental variables; and importantly, in enhancing computer- and Internet-based public health education *(49)*.

5. Statistical Software for Analyzing Cancer Registry Databases

The vast amount of data available in cancer registries and the complexities of the statistical analytical techniques involved require use of dedicated statistical software for any meaningful extraction and analysis of registry data. Fortunately, the SEER Program has taken a lead role in developing these statistical software packages, and several very useful software packages are now available. The foremost of the analytical software is the powerful SEER*Stat, which enables researchers analyze the entire SEER database and compute data on frequency distribution, incidence rates, temporal trends, and survival rates, including conditional survival *(50)*. Also, advanced statistical measures, such as limited duration prevalence, incidence-based mortality, and standardized incidence ratio for multiple primary cancers, can be calculated using this software. There is an accompanying software called The SEER*Prep software, which converts ASCII text data files to the SEER*Stat database format, thus allowing researchers analyze data from other cancer registries by using the SEER*Stat program *(51)*. The DevCan software computes probabilities of developing or dying from cancer for a hypothetical population for specific cancers, by using robust statistical techniques; to derive these probabilities, population estimates of incidence rates are obtained using cross-sectional counts of incident cases from the standard areas of the SEER Program and mortality counts for the same areas from data collected by the National Center for Health Statistics *(52)*. Joinpoint is a statistical software for the analysis of trends by using joinpoint models, and it takes trend data (e.g., cancer rates) and fits the simplest joinpoint model that the data allow *(53)*. ComPrev software estimates incidence and survival models using SEER cancer data for specific cancer sites, sex, and races to calculate the completeness index *(54)*. Complete prevalence is calculated by dividing limited duration prevalence by the completeness index as a proportion. Limited duration prevalence statistics can be generated from SEER*Stat software and imported into ComPrev. ProjPrev is a software made available by the SEER Program that is primarily used to derive U.S. prevalence by projecting SEER prevalence onto U.S. populations *(55)*. CanSurv is a powerful statistical software for analyzing population-based survival data *(56)*. For grouped survival data, it can fit both the

standard survival models and the mixture cure survival models, and it provides various graphs for model diagnosis. It also can fit parametric (cure) survival models to individually listed data. For geospatial analysis, Headbang is a software based on a smoothing algorithm for identifying a geographical patter by using the SEER data *(57)*. SaTScan is another software that allows to test for randomness of space distribution, time distribution, or both of a particular cancer, and it allows recognition of disease (cancer) clusters *(58)*. Most of these software packages are freely available for downloading from the SEER website, with appropriate tutorials, and they are backed by technical support by the SEER team. Besides SEER, other internet interfaces, such as CDC (state Cancer profiles) *(59)*, CINA + online Cancer in North America by NAACCR *(60)*, and Globocan by the IARC *(61)*, are frequently used by researchers to access aggregated cancer surveillance data generated from population-based cancer registries.

6. Impact of Cancer Registry Databases on Cancer Epidemiology

Cancer surveillance research based on cancer registry databases have over the past few decades expanded from its primary purview of descriptive epidemiology and disease quantification to a much broader analytical range, and it has made a significant contribution to every aspect of cancer epidemiology *(62)*. One direct and often underappreciated impact of cancer registry databases on cancer epidemiology is development of sophisticated and robust statistical techniques for solution in quantitative problems in cancer surveillance and control, population risk assessment, and development of methodology and relevant software for analyzing large databases with relative ease. The most visible and widely disseminated contributions of cancer registry databases remain descriptive periodic publications, such as the SEER Cancer Statistics Review, published annually by the Cancer Statistics Branch of the NCI; the annual report to the Nation on the Status of Cancer 1975–2003, jointly developed by several agencies; and internationally, the monograph on Cancer Incidence in Five Continents, published every fifth year by the IARC. These publications are statistical summaries that track the trend in cancer incidence, prevalence, survival, and mortality, and they serve as the most recognized references on cancer epidemiology globally. More detailed analysis of cancer registries have identified patterns that often point to specific etiologies. Landmark examples include the identification of perimenopausal shift in the age-specific incidence curve of breast cancer in women, implicating reproductive and hormonal factors in

etiology *(63)*; different rates of gastric cancer in second generation and immigrant U.S. Japanese, highlighting the interaction of genetic and environmental factors *(64)*; the role of pesticides in prostrate cancer as revealed in the Agricultural Health Study *(65)*; excess bladder cancer risk in truck drivers, workers exposed to motor exhaust, and workers within the chemical, rubber, and plastics industries, as suggested by the National Bladder Cancer Study *(66)*; several studies pointing to the carcinogenic effects of environmental tobacco smoke *(67)*; the relationship between oral contraceptive and menopausal estrogen use and breast, endometrial, and ovarian cancers among U.S. women, as studied in the Cancer and Steroid Hormone Study *(68)*; and role of diet *(69)*, physical activity *(70)*, and nonsteroidal anti-inflammatory drug use *(71)* in many common cancers, such as colon and reproductive cancers. A special example in studying causality of cancer is study of multiple primary cancers, which is practically impossible without using large cancer registry databases, and many studies have used these databases to study the association of multiple primary cancers and plausible etiological factors, such as shared risk or exposure, effect of treatment, or genetic susceptibility *(72–76)*. Familial cancer registries are rapidly emerging to be a powerful tool in studying genetic susceptibility and in identifying genetic factor in the etiopathogenesis of many common cancers, such as esophageal, pancreatic, colon, and breast cancers. The infrastructure of cancer registry databases such as the SEER has been critical both to the recruitment of these families and to the retrieval of related cancer data for conducting large multicenter, population-based studies such as the Women's Environment, Cancer, and Radiation Epidemiology study that is investigating gene–environment interactions that may influence susceptibility to breast cancer *(77–79)*. Surveillance data that form the foundation of the cancer registry databases provide unique glimpse into different aspects of cancer epidemiology. Important examples of such high-impact studies that were primarily derived from careful analysis of cancer registry databases included recent recognition of rising trend in the incidence of esophageal and gastroesophageal junctional adenocarcinoma in the western countries *(80)*; the association of Kaposi's sarcoma in patients with AIDS *(81)*, use of the GIS techniques in identifying statistically significant increase in childhood cancers for the period 1979 to 1995 in Dover Township in New Jersey, possibly related to exposure to environmental carcinogen *(82)*; and identification of geographically clustered neighborhoods with high rates of late-stage breast cancers *(83)*. Cancer registry databases also are unique resources for studying epidemiology of rare cancers such as the male breast cancer *(84)* and also cancer in small population groups such as the native Americans *(85)*.

Cancer registry-based analytical studies are becoming increasingly important in public health education, and in initiating and evaluating public health efforts. The lifetime risk of developing breast cancer, which is a commonly cited statistic and has been extensively used for health education and advocating targeted health interventions, was derived from studies conducted using the SEER database *(86,87)*. Health disparities in the minorities and socioeconomically deprived sections of the population is a stark, unpleasant reality, and researchers increasingly use the cancer registry databases to identify and highlight such disparities to bring about appropriate public heath and sociopolitical interventions *(88)*. Important studies based on the cancer registry database have highlighted the influence of socioeconomic factors on cancer incidence, treatment outcomes, and mortality by using the linkage of cancer registry data to other databases, such as the U.S. census, which provided information on selected socioeconomic variables at the neighborhood level *(89–93)*. Data from cancer registry databases also provide the final yardstick for measuring the long-term impact of implementation of public policy, as was demonstrated by the study that demonstrated rapid decline of lung cancer in association with the California Tobacco Control program *(94)*.

The SEER-initiated "pattern of care" studies effectively demonstrate the potential of cancer registry databases in studying complex issues in the area of treatment outcomes and end results, which are increasingly becoming an integral concept in understanding the epidemiology of cancer in its broadest purview *(95–97)*. Similarly, linkage with the Medicare database offers researcher innovative use of cancer registry databases in studying issues as diverse as cancer control practices and their effect on the cancer burden; patterns of access to cancer care; impact of co-morbidities, race, geographic, socioeconomic, and provider-related factors on access, diagnosis, treatment, and treatment outcomes; and importantly, cost-effectiveness of cancer care *(7,98–101)*. Despite the limitations of training and funding opportunities *(62)*, since 1974 >4,500 scientific publications have been published using the SEER and other linked databases in the U.S. alone, leaving no doubt regarding the enormity of the impact of cancer registry databases in cancer epidemiology.

In conclusion, over the past few decades cancer registry databases have evolved a long way in terms of number, coverage, technical sophistication, quantity, quality, and scope of information, and they are increasingly recognized as an indispensable part of local, national, and international cancer control programs. It is certain that cancer registry databases will continue to be an extraordinary resource of information for researchers, scientists, policy makers, and the public in our uphill and global fight against cancer.

References

1. Clive, R. E. (2004) Introduction to cancer registries, in Cancer Registry Management: Principles and Practice (Hutchison, C. L., ed.), Kendall Hunt Publishing, Dubuque, IA, pp. 1–9.
2. Parkin, D. M. (2006) The evolution of the population-based cancer registry. Nat. Rev. Cancer 6(8), 603–12.
3. Parkin, D. M., Chen, V. W., Ferlay, J., Galceran, J., Storm, H. H., and Whelan, S. L. (1994) Comparability and quality control in cancer registration, in IARC Technical Report No. 19, International Agency for Research on Cancer, Lyon, France.
4. Jensen O. M. (ed.) (1991) Cancer Registration, Principles and Methods, No. 95, International Agency for Research on Cancer, Lyon, France.
5. Working Group of the International Association of Cancer Registries (2005) Guidelines for confidentiality in population-based cancer registration. Eur J Cancer Prev **14,** 309–327.
6. National Cancer Institute (2007) SEER statistical resources. http://seer.cancer.gov/resources/. Cited 4 April 2007.
7. Warren, J. L., Klabunde, C. N., Schrag, D., Bach, P. B., and Riley, G.F. (2002) Overview of the SEER-Medicare data, content, research applications, and generalizability to the United States elderly population. Med. Care **40,** 5–18.
8. Spittle, M. F. (1998) Spectrum of AIDS-associated malignant disorders. Lancet **351(9119),** 1833–9.
9. Centers for Disease Control and Prevention (2007) National Center for Chronic Disease Prevention and Health Promotion. Behavioral Risk Factor Surveillance System. http://www.cdc.gov/brfss/index.htm. Cited 4 April 2007.
10. National Center for Health Statistics (2007) National Health Interview Survey. http://www.cdc.gov/nchs/nhis.htm. Cited 4 April 4 2007.
11. National Center for Health Statistics (2007) National Hospital Discharge and Ambulatory Surgery Data. http://www.cdc.gov/nchs/about/major/hdasd/nhds.htm. Cited 4 April 2007.
12. National Center for Health Statistics (2007) Ambulatory Health Care Data. http://www.cdc.gov/nchs/about/major/ahcd/ahcd1.htm. Cited 4 April 2007.
13. Harlan, L.C., and Hankey, B.F. (2003) The surveillance, epidemiology, and end-results program database as a resource for conducting descriptive epidemiologic and clinical studies. J Clin Oncol **21(12),** 2232–3.
14. Hankey, B. F., Ries, L. A., and Edwards, B. K. (1999) The surveillance, epidemiology, and end results program: a national resource. Cancer Epidemiol Biomarkers Prev **8(12),** 1117–21.
15. Centers for Disease Control and Prevention (2007) National Program of Cancer Registries. http://www.cdc.gov/cancer/npcr/about.htm. Cited 4 April 2007.
16. National Cancer Institute (2007) SEER-Medicare Linked Database. http://healthservices.cancer.gov/. Cited 4 April 2007.
17. Clegg, L. X., Feuer, E. J., Midthune, D., Fay, M. P., and Hankey, B. F. (2002) Impact of reporting delay and reporting error on cancer incidence rates and trends. J Natl Cancer Inst **94,** 1537–45.
18. Midthune, D. N., Fay, M. P., Clegg, L. X., and Feuer, E. J. (2005) Modeling reporting delays and reporting corrections in cancer registry data. J Am Stat Assoc **100(469),** 61–70.
19. Merrill, R. M., and Feuer, E. J. (1996) Risk-adjusted cancer-incidence rates (United States). Cancer Causes Control **7(5),** 544–52.
20. Fay, M. P. (2004) Estimating age conditional probability of developing disease from surveillance data. Popul Health Metr **2(1),** 6.
21. Fay, M. P., Pfeiffer, R., Cronin, K. A.., Le, C., and Feuer, E. J. (2003) Age-conditional probabilities of developing cancer. Stat Med **22(11),** 1837–48.
22. Kim, H. J., Fay, M. P., Feuer, E. J., and Midthune, D. N. (2000) Permutation tests for joinpoint regression with applications to cancer rates. Stat Med **19,** 335–51
23. Curtis, R. E., Freedman, D. M., Ron, E., Ries, L. A. G., Hacker, D. G., Edwards, B. K., Tucker, M. A., and Fraumeni, J. F., Jr. (eds) (2006) New malignancies among cancer survivors: SEER Cancer Registries, 1973–2000. National Cancer Institute. National Institutes of Health Publ. No. 05-5302. National Institutes of Health, Bethesda, MD.
24. Schoenberg, B. S., and Myers, M. H. (1977) Statistical methods for studying multiple primary malignant neoplasms. Cancer **40,** 1892–1898.
25. Begg, C. B., Zhang, Z., Sun, M., Herr, H. W., and Schantz, S. P. (1995) Methodology for

evaluating the incidence of second primary cancers with application to smoking-related cancer from the Surveillance, Epidemiology, and End Results (SEER) Program. Am J Epidemiol **142,** 653–665.

26. Fay, M. P., and Feuer, E. J. (1997) Confidence intervals for directly standardized rates: a method based on the Gamma distribution. Stat Med **16,** 791–801.
27. Chu, K. C., Miller, B. A., Feuer, E. J., and Hankey, B. F. (1994) A method for partitioning cancer mortality trends by factors associated with diagnosis: an application to female breast cancer. J Clin Epidemiol **47(12),** 1451–61.
28. Mariotto, A., Warren, J. L., Knopf, K. B., and Feuer, E. J. (2003) The prevalence of colorectal cancer patients under care in the US. Cancer **98,** 1253–61.
29. Coldman, A. J., McBride, M. L., and Braun, T. (1992) Calculating the prevalence of cancer. Stat Med **11,** 1579–89.
30. Capocaccia, R., and De Angelis, R. (1997) Estimating the completeness of prevalence based on cancer registry data. Stat Med **16,** 425–40.
31. Clegg, L., Gail, M., and Feuer, E. J. (2002) Estimating the variance of disease prevalence estimates from population-based registries. Biometrics **58(3)**, 684–8.
32. Feldman, A. R., Kessler, L., Myers, M. H., and Naughton, M. D. (1986) The prevalence of cancer. N Engl J Med **315,** 1394–7.
33. Gershman, S. T., Flannery, J. T., Barrett, H., Nadel, R. K., and Meigs, J. W. (1976) Development of the Connecticut Tumor Registry. Conn Med **40,** 697–701.
34. Merrill, R. M., Feuer, E. J., Cappacaccia, R., and Mariotto, A. (2000) Cancer prevalence estimates based on tumor registry data in the SEER Program. Int J Epidemiol **29,** 197–207.
35. Byrne, J., Kessler, L. G., and Devesa, S. S. (1992) The prevalence of cancer among adults in the United States. Cancer **68,** 2154–9.
36. Gail, M. H., Kessler L, Midthune, D., and Scoppa S. (1999) Two approaches for estimating disease prevalence from population-based registries of incidence and total mortality. Biometrics **55,** 1137–44.
37. De Angelis, G. D., De Angelis, R., Frova, L., and Verdecchia, A. (1994) MIAMOD: a computer program to estimate chronic disease morbidity using mortality and survival data. Comput Methods Programs Biomed **44,** 99–107.
38. Verdecchia, A., De Angelis, G., and Capocaccia, R. (2002) Estimation and projections of cancer prevalence from cancer registry data. Stat Med **21(22),** 3511–26.
39. Keiding, N. (1991) Age-specific incidence and prevalence: a statistical perspective. J R Stat Soc Ser A **154,** 371–412.
40. Marubini, E., and Valsecchi, M. G (eds) (2004) Analyzing Survival Data from Clinical Trials and Observational Studies, Wiley, Hoboken, NJ.
41. Boer, R., Ries, L., van Ballegooijen, M., Feuer, E., Legler, J., and Habbema, D. (2003) Ambiguities in calculating cancer patient survival: the SEER experience for colorectal and prostate cancer. Technical Report 2003-05, Statistical Research and Applications Branch, National Cancer Institute, Bethesda, MD (http://srab.cancer.gov/reports).
42. Brown, C.C. (1983) The statistical comparison of relative survival rates. Biometrics **39,** 941–8.
43. Ederer, F., Axtell, L. M., and Cutler, S. J. (1961) The relative survival rate: a statistical methodology. Natl Cancer Inst Monogr **6,** 101–21.
44. Brenner, H., and Hakulinen, T. (2002) Advanced detection of time trends in long-term cancer patient survival: experience from 50 years of cancer registration in Finland. Am J Epidemiol **156(6),** 566–77.
45. Yu, B., Tiwari, R. C., Cronin, K. A., McDonald, C., and Feuer, E. J. (2005) CANSURV: a Windows program for population-based cancer survival analysis. Comput Methods Programs Biomed **80(3),** 195–203.
46. Yu, B., Tiwari, R. C., Cronin, K. A, and Feuer, E. J. (2004) Cure fraction estimation from the mixture cure models for grouped survival data. Stat Med **23(11),** 1733–47.
47. Brewer, C. A. (2006) Basic mapping principles for visualizing cancer data using geographic information systems (GIS). Am J Prev Med **30(2 Suppl),** S25–S36.
48. Bell, B. S, Hoskins, R. E., Pickle, L. W., and Wartenberg, D. (2006) Current practices in spatial analysis of cancer data: mapping health statistics to inform policymakers and the public. Int J Health Geogr **8,** 5:49.
49. Pickle, L. W., Hao, Y., Jemal, A., Zou, Z, Tiwari, R. C., and Ward, E., et al. (2007) A new method of estimating United States and state-level cancer incidence counts for the current calendar year. CA Cancer J Clin 57, 30–42.
50. National Cancer Institute (2006) SEER*Stat software, version 6.2. Statistical Research

and Applications Branch, National Cancer Institute. http://seer.cancer.gov/seerstat/. Cited 4 April 2007.

51. National Cancer Institute (2006) SEER*Prep software, version 2.3.5. Statistical Research and Applications Branch, National Cancer Institute. http://srab.cancer.gov/seerprep/. Cited 4 April 2007.
52. National Cancer Institute (2006) DevCan 6.1.1 for Windows software, version 6.1.1. Statistical Research and Applications Branch, National Cancer Institute. http://srab.cancer.gov/devcan/. Cited 4 April 2007.
53. National Cancer Institute (2006) Joinpoint regression program, version 3.0. Statistical Research and Applications Branch, National Cancer Institute. http://srab.cancer.gov/joinpoint/. Cited 4 April 2007.
54. National Cancer Institute (2006) Complete Prevalence (ComPrev) software, version 1.0. Statistical Research and Applications Branch, National Cancer Institute. http://srab.cancer.gov/comprev/. Cited 4 April 2007.
55. National Cancer Institute (2006) Projected Prevalence (ProjPrev) software, version 1.0.1. Statistical Research and Applications Branch, National Cancer Institute. http://srab.cancer.gov/projprev/. Cited 4 April 2007.
56. National Cancer Institute (2006) CanSurv software, version 1. Statistical Research and Applications Branch, National Cancer Institute. http://srab.cancer.gov/cansurv/. Cited 4 April 2007.
57. National Cancer Institute (2006) HeadBang PC software, version 3.0. Hansen; Simonson and Statistical Research and Applications Branch, National Cancer Institute. http://srab.cancer.gov/headbang/. Cited 4 April 2007.
58. Kulldorff, M. and Information Management Services, Inc. (2007) SaTScan™ version 7.0: software for the spatial and space-time scan statistics. http://www.satscan.org/. Cited 4 April 2007.
59. National Cancer Institute (200X) State cancer profiles. http://statecancerprofiles.cancer.gov/index.html/. Cited 4 April 2007.
60. North American Association of Central Cancer Registries (200X) CINA + Online Cancer in North America. http://www.cancer-rates.info/naaccr/. Cited 4 April 2007.
61. Ferlay, J., Bray, F., Pisani, P., and Parkin, D. M. (2004) GLOBOCAN 2002: Cancer Incidence, Mortality and Prevalence Worldwide. IARC Cancer Base No. 5. Version 2.0. IARC Press, Lyon, France. (http://www-dep.iarc.fr/).
62. Glaser, S. L., Clarke, C. A., Gomez, S. L., O'Malley, C. D., Purdie, D. M., and West, D. W. (2005) Cancer surveillance research, a vital subdiscipline of cancer epidemiology. Cancer Causes Control **16,** 1009–1019.
63. Clemmesen, J. (1948) Carcinoma of the breast. Results from statistical research (symposium). Br J Radiol **21,** 583–590.
64. Buell, P., and Dunn, J. E., Jr. (1965) Cancer mortality among Japanese Issei and Nisei of California. Cancer **18,** 656–64.
65. Alavanja, M. C., Sandler, D. P., McMaster, S. B., Zahm, S. H., McDonnell, C. J., and Lynch, C. F., et al. (1996) The Agricultural Health Study. Environ Health Perspect **104 (4),** 362–369.
66. Silverman, D. T., Hartge, P., Morrison, A. S., and Devesa, S. S. (1992) Epidemiology of bladder cancer. Hematol Oncol Clin North Am **6(1),** 1–30.
67. National Cancer Institute Health (1999) Effects of Exposure to Environmental Tobacco Smoke, Smoking and Tobacco Control Monograph No. 10. U.S. Department of Health and Human Services, National Cancer Institute, National Institutes of Health, Bethesda, MD.
68. No authors listed (1986) Oral contraceptive use and risk of breast cancer. The Cancer and Steroid Hormone Study of the Centers for Disease Control and the National Institute of Child Health and Human Development. N Engl J Med **315,** 405–411.
69. Le Marchand, L., Yoshizawa, C. N., Kolonel, L. K., Hankin, J. H., and Goodman, M. T. (1989) Vegetable consumption and lung cancer risk: a population-based case-control study in Hawaii. J Natl Cancer Inst **81,** 1158–1164.
70. Irwin, M. L., Aiello, E.J., McTiernan, A., Bernstein, L., Gilliland, F. D., Baumgartner, R. N., et al. (2007) Physical activity, body mass index, and mammographic density in postmenopausal breast cancer survivors. J Clin Oncol **25(9),** 1061–6.
71. Cerhan, J. R., Anderson, K. E., Janney, C. A., Vachon, C. M., Witzig, T. E., and Habermann, T. M. (2003) Association of aspirin and other non-steroidal anti-inflammatory drug use with incidence of non-Hodgkin lymphoma. Int J Cancer **106(5),** 784–8.
72. Kleinerman, R. A., Boice, J. D., Jr., Storm, H. H., Sparen, P., Andersen, A., Pukkala, E.,

Lynch, C. F., et al. (1995) Second primary cancer after treatment for cervical cancer. An international cancer registries study. Cancer **76,** 442–452.

73. Travis, L. B., Curtis, R. E., Storm, H., Hall, P., Holowaty, E., Van Leeuwen, F. E., et al. (1997) Risk of second malignant neoplasms among long-term survivors of testicular cancer. J Natl Cancer Inst **89,** 1429–1439.
74. Travis, L. B., Gospodarowicz, M., Curtis, R. E., Clarke, E.A., Andersson, M., Glimelius, B., et al. (2002) Lung cancer following chemotherapy and radiotherapy for Hodgkin's disease. J Natl Cancer Inst **94,** 182–192.
75. Zablotska, L. B., Chak, A., Das, A., and Neugut, A.I. (2005) Increased risk of squamous cell esophageal cancer after adjuvant radiation therapy for primary breast cancer. Am J Epidemiol **161(4),** 330–7.
76. Das, A., Neugut, A. I., Cooper, G. S., and Chak, A. (2004) Association of ampullary and colorectal malignancies. Cancer **100(3),** 524–30.
77. Brewster, D. H, Fordyce, A., and Black, R. J. (2004) Scottish clinical geneticists. Impact of a cancer registry-based genealogy service to support clinical genetics services. Fam Cancer **3(2),** 139–41.
78. Ziogas, A., and Anton-Culver, H. (2003) Validation of family history data in cancer family registries. Am J Prev Med **24(2),** 190–198.
79. Bernstein, J. L., Langholz, B. M., Haile, R.W., Bernstein, L., Thomas, D. C., and Stovall, M., et al. (2004) Study design: evaluating gene-environment interactions in the etiology of breast cancer—the WECARE Study. Breast Cancer Res **6,** R199–R214
80. Blot, W. J., Devesa, S. S., Kneller, R. W., and Fraumeni, J. F., Jr. (1991) Rising incidence of adenocarcinoma of the esophagus and gastric cardia. JAMA **265,** 1287–1289.
81. Reynolds, P., Layefsky, M. E., Saunders, L. D, George, F. L., and Payne, S. F. (1990) Kaposi's sarcoma reporting in San Francisco: comparison of AIDS and cancer surveillance systems. J Acquir Immune Defic Syndr **3(Suppl 1),** S8–S13.
82. Blumenstock, J., Fagliano, J., and Bresnitz, E. (2000) The Dover Township childhood cancer investigation. N J Med **97,** 25–30.
83. Roche, L. M., Skinner, R., and Weinstein, R. B. (2002) Use of a geographic information system to identify and characterize areas with high proportions of distant stage breast cancer. J Public Health Manag Pract **8,** 26–32.
84. Giordano, S. H., Cohen, D. S., Buzdar, A. U., Perkins, G., and Hortobagyi, G. N. (2004) Breast carcinoma in men: a population-based study. Cancer **101(1),** 51–7.
85. Swan, J., and Edwards, B. K. (2003) Cancer rates among American Indians and Alaska Natives: is there a national perspective. Cancer **98(6),** 1262–72
86. Gail, M. H., Brinton, L. A., Byar, D. P, Corle, D. K., Green, S. B., Schairer, C., et al. (1989) Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst **81,** 1879–1886.
87. Feuer, E. J., Wun, L. M., Boring, C. C., Flanders, W. D., Timmel, M. J., and Tong, T. (1993) The lifetime risk of developing breast cancer. J Natl Cancer Inst **85,** 892–897.
88. Harper, S., and Lynch, J. (2005) Methods for measuring cancer disparities: using data relevant to healthy people 2010 cancer-related objectives. NCI Cancer Surveillance Monograph Series, No. 6. National Cancer Institute, Bethesda, MD.
89. Humble, C. G., Samet, J. M., Pathak, D. R., and Skipper, B. J. (1985) Cigarette smoking and lung cancer in 'Hispanic' whites and other whites in New Mexico. Am J Public Health **75(2),** 145–148.
90. Chen, V. W., Fontham, E. T. H., Craig, J. F., Groves, F. D., Culley, P., Rainey, J. M., et al. (1992) Cancer in South Louisiana. Part I: tobacco-related cancers. J La Med Soc **144,** 149–155.
91. Friedell, G. H., Tucker, T. C., McManmon, E., Moser, M., Hernandez, C., and Nadel, M. (1992) Incidence of dysplasia and carcinoma of the uterine cervix in an Appalachian population. J Natl Cancer Inst **84,** 1030–1032.
92. Chen, V. W., Wu, X. C., Andrews, P. A., Fontham, E. T., and Correa, P. (1994) Advanced stage at diagnosis: an explanation for higher than expected cancer death rates in Louisiana? J La Med Soc **146,** 137–145.
93. Mills, P. K., and Kwong, S. (2001) Cancer incidence in the united farm workers of America (UFW) 1987–1997. Am J Ind Med **40,** 596–603.
94. Barnoya, J., and Glantz, S. (2004) Association of the California tobacco control program with declines in lung cancer incidence. Cancer Causes Control **15(7),** 689–95.
95. Harlan, L. C., Abrams, J., Warren, J. L., Clegg, L., Stevens, J., and Ballard-Barbash, R. (200) Adjuvant therapy for breast cancer:

practice patterns of community physicians. J Clin Oncol **20,** 1809–1817.

96. Potosky, A. L., Harlan, L. C., Kaplan, R. S., Johnson, K. A., and Lynch, C. F. (2002) Age, sex, and racial differences in the use of standard adjuvant therapy for colorectal cancer. J Clin Oncol **20,** 1192–1202.
97. Harlan, L. C., Clegg, L. X., and Trimble, E. L. (2003) Trends in surgery and chemotherapy for women diagnosed with ovarian cancer in the U.S. J Clin Oncol **21(18),** 3488–94
98. Potosky, A. L., Riley, G. F., Lubitz, J. D., Mentnech, R. M., and Kessler, L. G. (1993) Potential for cancer related health services research using a linked Medicare-tumor registry database. Med Care **31(8),** 732–748.
99. Brown, M. L., Riley, G. F., Schussler, N., and Etzioni, R. (2002) Estimating health care costs related to cancer treatment from SEER-Medicare data. Med Care **40(8 Suppl IV),** 104-17.
100. Doebbeling, B. N., Wyant, D. K., McCoy, K. D., Riggs, S., Woolson, R. F., Wagner, D., et al. (1999) Linked insurance-tumor registry database for health services research. Med Care **37(11),** 1105–15.
101. McClish, D. K., Penberthy, L., Whittemore, M., Newschaffer, C., Woolard, D., Desch, C. E., et al. (1997) Ability of Medicare claims data and cancer registries to identify cancer cases and treatment. Am J Epidemiol **145(3),** 227–33.