Editorial

Cancer Surveillance Research

William F. Anderson
Bioinformatics Branch (BB) DHHS/NIH/NCI Division of Cancer Epidemiology and Genetics (DCEG) EPS, Executive Blvd Bethesda, Maryland

The journal Cancer Epidemiology, Biomarkers & Prevention (CEBP) has launched a new manuscript section entitled Cancer Surveillance Research (CSR). The CSR section will consist of original reports using cancer case and population data to examine, test, and develop hypotheses for cancer prevalence, incidence, and mortality. The scope of CSR includes descriptive epidemiology, public health statistics for time trends in cancer burden, genetic, behavioral, and environmental risk factors, cancer disparities and geographic variations, screening and diagnostic practice patterns, and methodological developments for assessing cancer data.

CSR studies are the "eyes and ears" for the monitoring and assessment of cancer burden through the examination of vital health statistics. Evaluation of demographic, temporal, and geographic variations in cancer rates can suggest clues to genetic or environmental exposures, cultural influences, health behaviors, geographic variations, and racial/ethnic variations for subgroups at unexpected risk for certain cancers. Cancer rates may be used to verify the consistency of existing cancer-related hypotheses and/or to generate new ideas for future analytic research. CSR can estimate the external validity of a randomized clinical trial. A well-designed and controlled clinical trial has strong internal validity. However, the generalizability (or external validity) of the randomized study cannot be assumed because subject participation depends upon selection and inclusion criteria, which might not reflect the population at large. Generalizability concerns are further heightened by the fact that clinical trial participants are generally healthier, wealthier, younger, Caucasian, and urban dwellers (1, 2). The merging of population-based and clinical trial evidence can help to determine if an efficacious clinical trial is effective in the general population (3, 4). Additionally, when a disease is rare, cancer surveillance data might be the most reliable source of information. Finally, as the population ages and the cost of cancer-related services increase (5), CSR can aid health care planners and policy makers manage and direct limited resources.

CSR has been enhanced through advancements in computer hardware and software, bioinformatics, and statistical methodologies, local, national and international databases (6). However, cancer surveillance data are underutilized, largely owing to two mistaken impressions (7). First, there is a lack of recognition of the available resources for CSR. Second, descriptive epidemiology—the core methodology for CSR—is often viewed as simplistic, uncertain, and/or unreliable.

Contrary to some mistaken views, population-based resources, databases, and statistical tools are readily available from public-use websites such as the National Cancer Institute’s Surveillance, Epidemiology, and End Results program (SEER) (8), Cancer Mortality Maps & Graphs (9), Centers for Disease Control and Prevention (CDC) (10), North American Association of Central Cancer Registries (NAACCR) (11), and the International Agency for Research on Cancer (IARC) (12). SEER provides cancer incidence and survival data from 17 Tumor Registries, covering approximately 26% of the United States. The current SEER database has nearly 5 million cancer cases spanning more than 1 billion person-years from 1973 through 2005. The Cancer Mortality Maps & Graphs website has interactive charts, text, tables, and figures for more than 40 cancers from 150 through 1994. The CDC’s National Program of Cancer Registries (NPCR) supports the maintenance of high quality tumor registries for states in the United States. The CDC’s National Center for Health Statistics (NCHS) distributes cancer mortality information for the calendar period 1969 to 2005. The NAACCR promotes CSR as an umbrella organization for central cancer registries in the United States and Canada, government agencies, and professional organizations. IARC’s CANCER Mondial website provides information on the occurrence of cancer world-wide through five programs: (1) Cancer Incidence in Five Continents volumes I to IX, (2) ACCIS, incidence and survival data of children and adolescents in Europe, (3) mortality data from the World Health Organization (WHO), (4) GLOBOCAN 2002 for the incidence, prevalence, and mortality of 27 cancers for all countries in the world in 2002, and (5) NORDCAN project from 41 major cancers in Nordic countries.

The simplistic view of descriptive epidemiology partly reflects the reality that descriptive studies are secondary and/or retrospective analyses, dependent upon the observational method. Observational results are cross-sectional, capturing a “snap-shot” in time and are subject to uncontrollable chance, bias, or confounding. Population-based descriptive studies begin with a rate matrix (sometimes referred to as a Lexis diagram) (13), as illustrated in Table 1 for female breast cancer from the SEER database (1974-2005). Indeed, it would be imprudent for CSR to...
ignore the complex interactions associated with the Lexis figure.

For example, the two-dimensional geometry of the Lexis diagram shows that three fundamental descriptive variables (age, period, and cohort) are in a single plane and are linearly dependent (Table 1), i.e., age at diagnosis in rows, year of diagnosis (calendar-period) in columns, and year of birth (birth-cohort) in the diagonals. Given the relationship \( C = P/A \) (birth-cohort = calendar-period / age at diagnosis), Table 1 has 23 birth-cohorts (1893, 1897, ..., 1981, referred to by mid-year of birth) that are derived from 16 four-year age groups (21-24, 25-28, ..., 81-84 years) and eight four-year time periods (1974-1977, 1978-1981, ..., 2002-2005). Birth-cohort reflects time trends that impact all age groups for a given generation, i.e., risk factor exposures. Calendar-period effects reflect secular trends that affect all age groups at a certain point in time, i.e., changing screening or diagnostic practice patterns. Age is a surrogate for age-related biological factors. Because age, period, and cohort are collinear, it is not possible to completely separate calendar-period effects from age effects or the birth-cohort effects from calendar-period effects, giving rise to the so-called “nonidentifiability” issue.

Given the uncertainty associated with these nonidentifiability issues, CSR requires a close interface between data resources and statistical techniques (14). Age standardization attempts to minimize the impact of different age distributions when comparing rates over time and across populations. Nonlinear regression models have been applied to a sophisticated analyses of time trends (15). Careful attention to plotting techniques facilitates temporal comparisons, fairly conveying the data without overemphasizing the results (16). Multivariate analyses allow for the simultaneous study of two or more dependent variables. Poisson regression can assess cancer-specific hazard rates that are adjusted for any number of covariates such as age at diagnosis, year of diagnosis, stage, grade, etc. (17). Age-period-cohort (APC) models estimate a number of identifiable parameters adjusted for age, period, and cohort effects (18).

For example, two very useful APC parameters for descriptive studies are the “drifts” (linear trends) (refs. 19, 20) and the “fitted” age at onset curve (21). Net drift equals the sum of the linear trends in period and cohort effects for all age groups (Fig. 1A). The net drift quantifies the average annual percentage change in the logarithm of the rates adjusted for period and cohort deviations. It is a summary measure of the overall trend in the disease.
trend during the study period, and is closely related to the estimated annual percentage in the age standardized rates. Figure 1A shows a net drift of 1.5% per year (95% CI, 1.3-1.6). In other words, female breast cancer incidence rates rose at a rate of 1.5% for each yearly increment in the calendar time.

Another type of drift is the longitudinal age trend or LAT (Fig. 1B). The LAT quantifies the average annual percentage change in the logarithm of the rates adjusted for age and period deviations, and estimates the average percentage change of the "fitted" age at onset curve (Fig. 1B). The fitted curve is an extrapolation of the age-specific rates for the mid birth-cohort based upon the age-specific rates for all other cohorts in the study (16). Figure 1B shows a LAT of 9.7% per year of attained age (95% CI, 9.4-10.0). In other words, female breast cancer age-specific rates rose at a rate of nearly 10% for each yearly increment in the age at diagnosis. Note the collinearity (nonidentifiability) for the age, period, and cohort effects in Table 1 and Fig. 1. The 1973 birth-cohort in Fig. 1B corresponds to the age-specific temporal trend for ages 21-24 to 29-32 years in Fig. 1A, whereas the 1961 birth-cohort in Fig. 1B corresponds to the age-specific temporal trend for ages 21-24 to 41-44 in Fig. 1A, etc.

In launching the new CSR section, the journal hopes to publicize and highlight both the resources and methodology for cancer surveillance; and through CSR, to assess emerging cancer trends and cancer-related hypotheses. CSR manuscripts should be 3,000 words or less (with background, methods, results, and conclusion), and no more than 40 references. Supplemental data can be provided if needed.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

References
Correction

Correction: Editorial on Cancer Surveillance Research

In the article (1) on Cancer Surveillance Research in the June 2009 issue of Cancer Epidemiology, Biomarkers & Prevention, there are errors in the first full paragraph on page 1671 and in the Fig. 1A legend. Reference 16 in that paragraph should be Reference 21. The last full sentence should read as follows: The 1973 birth-cohort in Fig. 1B corresponds to ages 21-24 to 29-32 in Table 1, whereas the 1961 birth-cohort corresponds to ages 21-24 to 41-44, etc. The correct Fig. 1A legend should read as follows from top to bottom: ages 77-80 yr, ages 65-68 yr, ages 53-56 yr, ages 41-44 yr, and ages 29-32 yr.

Reference
