



Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2009 June ; 18(6): 1669–1671. doi:10.1158/1055-9965.EPI-09-0318.

CANCER SURVEILLANCE RESEARCH (CSR)

William F. Anderson, MD, MPH

Biostatistics Branch (BB), DHHS/NIH/NCI/Division of Cancer Epidemiology and Genetics (DCEG),
EPS, Room 8036, 6120 Executive Blvd, Bethesda, MD 20892-7244, Phone: 301 594-9125, Fax:
301 402-0081

William F. Anderson: wanderso@mail.nih.gov

Keywords

descriptive epidemiology; age-period-cohort (APC) models; non-identifiability issues

The journal *Cancer Epidemiology, Biomarkers, and 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 from patterns 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 since 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 rise (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 due to two mistaken impressions (7). First, there is a lack of recognition of the available resources for CSR. Second, descriptive

Disclaimer: The author does not have a financial conflict of interest that would have affected this research. This research was supported in part by the Intramural Research Program of the National Institutes of Health, National Cancer Institute. The authors had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

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 with 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 1950 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 1969–2005. 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 *CANCERmondial* 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 from 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. All 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 2-dimensional geometry of the Lexis diagram demonstrates 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 twenty-three birth-cohorts (1893, 1897, ... 1981, referred to by mid-year of birth) that are derived from eighteen 4-year age groups (21–24, 25–28, ... 81–84 years) and eight 4-year time periods (1974–1977, 1978–1981, ... 2002–2005). Birth-cohort reflects time trends that impact all age groups for a given generation. 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 “non-identifiability” issue.

Given the uncertainty associated with these non-identifiability 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. Non-linear 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 over emphasizing 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).

Two very useful APC parameters for descriptive studies are the “*drifts*” (linear trends) (19, 20) and the “*fitted*” age-specific curves (21). Net drift equals the sum of the linear trends in period and cohort effects for all age groups included during the study period (figure 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 during the study period, and is closely related to the estimated annual percentage change (EAPC) in the age standardized rates (ASR). Figure 1A shows an EAPC of 2.08% per year of calendar time and a net drift of 1.5% per year. In other words, female breast cancer incidence rates rose at a rate of approximately 1.5% to 2% for each yearly increment in the calendar time.

Another type of drift is the longitudinal age trend or LAT (figure 1B). The LAT is the sum of the linear trends in the age and period effects. It provides an estimate of the average annual percentage per year of attained age for the “*fitted*” age at onset curve (figure 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 (21). Figure 1B shows a LAT of 9.7% per year (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 (non-identifiability) for the age, period, and cohort effects in table 1 and figure 1. The 1973 birth-cohort in figure 1B corresponds to ages 21–24 to 29–32 years in table 1; the 1961 birth-cohort corresponds to the age-groups 21–24 to 41–44, 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 3000 words or less (not counting abstract, references, or legends), have a total of 6 or fewer tables and/or figures, a structured abstract of 250 words or less (with background, methods, results, and conclusion), and no more than 40 references. Supplemental data can be provided if needed.

References

1. Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Threats to applicability of randomised trials: exclusions and selective participation. *J Health Serv Res Policy* 1999;4:112–21. [PubMed: 10387403]
2. Hutchins LF, Unger JM, Crowley JJ, Coltman CA Jr, Albain KS. Underrepresentation of patients 65 years of age or older in cancer-treatment trials. *N Engl J Med* 1999;341:2061–7. [PubMed: 10615079]
3. Cochrane, AL. Effectiveness and Efficiency: Random Reflections on Health Services. London: Nuffield Provincial Hospitals Trust; 1972.
4. Berry DA, Ravdin PM. Breast cancer trends: a marriage between clinical trial evidence and epidemiology. *J Natl Cancer Inst* 2007;99:1139–41. [PubMed: 17652274]
5. Meropol NJ, Schulman KA. Cost of cancer care: issues and implications. *J Clin Oncol* 2007;25:180–6. [PubMed: 17210937]
6. Fraumeni JF Jr, Rimer BK. Cancer surveillance series: inauguration. *J Natl Cancer Inst* 1999;91:1004.
7. Glaser SL, Clarke CA, Gomez SL, O’Malley CD, Purdie DM, West DW. Cancer surveillance research: a vital subdiscipline of cancer epidemiology. *Cancer Causes Control* 2005;16:1009–19. [PubMed: 16184466]
8. SEER. Surveillance Epidemiology and End Results. 2009. [cited 2009; Available from: <http://seer.cancer.gov/>]

9. Cancer Mortality Maps & Graphs. 2009. [cited 2009; Available from: <http://www3.cancer.gov/atlasplus/>]
10. Centers for Disease Control and Prevention (CDC). 2009. [cited; Available from: www.cdc.gov]
11. North American Association of Central Cancer Registries (NAACCR). 2009. [cited; Available from: www.naacr.org]
12. International Agency for Research on Cancer (IARC). 2009. [cited; Available from: www.iarc.fr]
13. Vandeschrick C. The Lexis diagram, a misnomer. *Demographic Research* 2001;4:97–124.
14. Hankey BF, Ries LA, Edwards BK. The surveillance, epidemiology, and end results program: a national resource. *Cancer Epidemiol Biomarkers Prev* 1999;8:1117–21. [PubMed: 10613347]
15. Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. *Stat Med* 2000;19:335–51. [PubMed: 10649300]
16. Devesa SS, Donaldson J, Fears T. Graphical presentation of trends in rates. *Am J Epidemiol* 1995;141:300–4. [PubMed: 7840107]
17. Breslow NE, Lubin JH, Marek P, Langholz B. Multiplicative models and cohort analysis. *Journal of the American Statistical Association* 1983;78:1–12.
18. Holford, TR. Age-period-cohort analysis. In: Armitage, P.; Colton, T., editors. *Encyclopedia of Biostatistics*. 1. Chichester, England: John Wiley & Sons; 1998. p. 82-99.
19. Clayton D, Schifflers E. Models for temporal variation in cancer rates. I: Age-period and age-cohort models. *Stat Med* 1987;6:449–67. [PubMed: 3629047]
20. Clayton D, Schifflers E. Models for temporal variation in cancer rates. II: Age-period-cohort models. *Stat Med* 1987;6:469–81. [PubMed: 3629048]
21. Anderson WF, Rosenberg PS, Menashe I, Mitani A, Pfeiffer RM. Age-related crossover in breast cancer incidence rates between Black and White Ethnic Groups. *J Natl Cancer Inst* 2008;100:1804–14. [PubMed: 19066264]

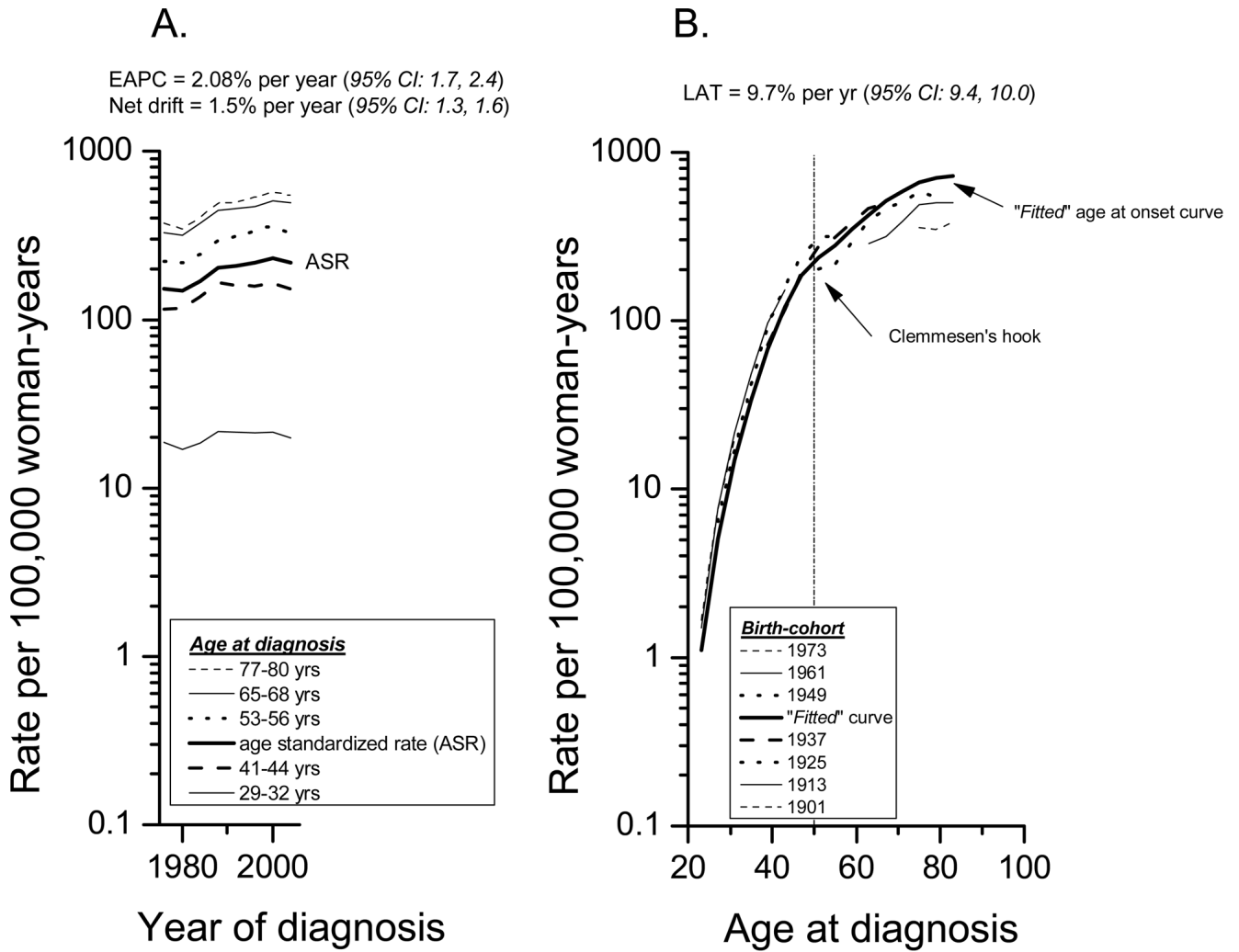


Figure 1. (A) Female breast cancer cases (*in situ* and invasive) in SEER's 9+13+17 Registries Databases, diagnosed during the years 1974 through 2005. (A) Age-specific incidence rate trends with the estimated annual percentage change in the age standardized rates (ASR) and the annual percentage change in the net drifts per year of calendar time. (B) Age-period-cohort (APC) "fitted" age-specific curves with the estimated annual percentage change per year of attained age in the longitudinal age trend (LAT).

Table 1

Female breast cancer cases (*in situ* and invasive) in SEER's 9+13+17 Registries Databases, diagnosed during the years 1974 through 2005

Time (yrs)	1974-1977	1978-1981	1982-1985	1986-1989	1990-1993	1994-1997	1998-2001	2002-2005	Birth Cohort
Age (yrs)	Rate	Rate	Rate	Rate	Rate	Rate	Rate	Rate	
21-24	1	1	2	2	2	2	2	2	1981
25-28	6	6	7	8	7	7	8	8	1977
29-32	19	17	19	22	22	21	22	20	1973
33-36	39	39	42	48	47	48	49	44	1969
37-40	72	72	83	94	91	92	96	86	1965
41-44	116	117	137	166	160	158	165	152	1961
45-48	167	164	193	238	246	241	245	226	1957
49-52	202	194	223	277	290	306	308	277	1953
53-56	222	217	245	296	312	335	363	323	1949
57-60	260	254	292	345	356	383	421	403	1945
61-64	287	286	328	396	399	419	464	451	1941
65-68	328	317	371	446	459	472	509	497	1937
69-72	344	341	386	474	486	510	538	513	1933
73-76	357	353	410	487	510	534	575	536	1929
77-80	374	348	405	493	499	533	572	545	1925
81-84	352	348	381	464	482	498	547	519	1921
									1917
									1913
									1909
									1905
									1901
									1897
									1893