CANCER STATISTICS REVIEW 1975-2009 (Vintage 2009 Populations): INTRODUCTION

The annual *SEER Cancer Statistics Review* (***CSR***) contains incidence, mortality, prevalence, and survival statistics from 1975 through the most recent year for which data are available. This report is published by the Surveillance Research Program of the National Cancer Institute, which manages the Surveillance, Epidemiology, and End Results (SEER) Program. The scope and purpose of the ***CSR*** follow a report to the Senate Appropriations Committee (Breslow, 1988), which recommended that a broad profile of cancer be presented regularly to the American public.

The SEER program is an authoritative source of information on cancer incidence and survival in the United States. SEER collects and publishes these statistics from population-based registries covering 28% of the US population. The 18 SEER registries routinely collect data on patient demographics, primary tumor site, tumor morphology, extent of disease, first course of treatment, and active follow-up for vital status. Detailed information describing these fields can be found at *http://seer.cancer.gov/resources/*.

This report presents statistics on 29 primary sites and subsites, organized into site-specific chapters. Detailed statistics on cancer incidence, mortality, survival, and prevalence are reported by sex, race and ethnicity, age, stage at diagnosis, and geographic area. Information on tumor morphology is also presented. In addition, the ***CSR*** features a chapter on adolescent and young adult cancers and a chapter on childhood cancers. Information on some rare cancers can be found in the summary tables of section I. For a detailed list of primary sites, the summary tables provide incidence and death rates for the most recent 5-year period, trends from 1975 to the most recent year, median age at diagnosis, median age at death, and survival rates.

Delay-adjusted cancer incidence rates are a distinctive feature of the ***CSR***. Delay-adjustment corrects the current case count to account for underreporting and corrections to the data. The final delay-adjusted rates are valuable in more precisely estimating trends.

New features recently added to the ***CSR*** include detailed histology breakdowns for lymphomas and for cancers of the oral cavity and pharynx, soft tissue, and pancreas; cause-specific survival by expanded race and ethnic groups; SEER 13 delay-adjustment; and adjustments for Veterans' Administration (VA) underreporting. Starting with patients diagnosed in 2007, the new multiple primary and histology coding rules may impact their incidence data for some cancer sites (e.g., female breast). However, the impact of the new rule on observed incidence is negligible for a majority of the cancer sites. To learn more about the multiple primary rules, visit: http://seer.cancer.gov/tools/mphrules/.

The ***CSR*** files are provided in both PDF and HTML formats. The HTML format is provided as an alternative and accessible version of the *SEER Cancer Statistics Review*. The current edition of

the ***CSR*** is available on the web at *http://seer.cancer.gov/csr/*. Statistics from SEER may also be obtained via ***FastStats*** (*http://seer.cancer.gov/faststats/*) or ***Cancer Query Systems*** (*http://seer.cancer.gov/canques/*), which allow the user to access over 10,000,000 cancer statistics. The SEER Research Data file (*http://seer.cancer.gov/data/*) may be accessed by the public, either through ***SEER*Stat*** software or in an ASCII text format that can be analyzed with standard statistical software.

While most of the rates in this publication have been age-adjusted to the 2000 US standard population, some previous SEER publications have used the 1970 US standard million population. Therefore, rates given in this publication cannot be compared to rates given in those publications. This change conforms to a federal policy for reporting disease rates; it allows for the age-adjusted rate to more accurately reflect the current age distribution and burden of cancer.

INTERPRETATION OF CANCER STATISTICS

A number of factors may affect the interpretation of cancer incidence, mortality, and survival statistics provided in this report.

Survival rates for all cancers combined: The mix of cancers changes over time as the incidence of some cancers increases and the incidence of others decreases. The overall cancer survival rate can fluctuate even when the survival rates for site-specific cancers remain unchanged. (While it is possible to adjust the survival rate for all cancers combined on the basis of the relative frequencies of the component cancers, rates adjusted in this manner differ by only a small amount from unadjusted rates. In the future, such an adjustment may become more important if there are substantial changes in the incidence of various cancers.)

Early detection/screening: The improved earlier detection and diagnosis of cancers—caused by new screening procedures--may produce an *increase* in both incidence rates and survival rates. These increases can occur as a result of the introduction of a new procedure to screen subgroups of the population for a specific cancer; they need not be related to whether use of the screening test results in a decrease in mortality from that cancer. As the proportion of cancers detected at screening increases, presumably as a result of increased screening of the population, patient survival rates will *increase*, because they are based on survival time *after diagnosis*. The interval between the time a cancer is diagnosed by a screening procedure and the time when the cancer would have been diagnosed in the absence of screening is called **lead-time** (Zelen, 1976). (Screening for breast cancer has been demonstrated to result in increased survival over and above that resulting from lead-time alone and to reduce breast cancer mortality. The benefit of screening is being studied for some other cancers.)

If a new screening procedure consistently detects cancer in a *preinvasive* phase, it may result in a *decrease* in survival rates for *invasive* cancer. In this case, **length-biased sampling** (Zelen, 1976) may be operating. Length-biased sampling would result in the preferential detection—in

a preinvasive phase—of those cancers that would have had a relatively good prognosis had they progressed to invasive disease; these potentially invasive cancers would be systematically eliminated. If this occurs, the mix of cancers that are not detected at screening and then progress to invasive behavior may become less prognostically favorable, resulting in a *decrease* in survival rates for patients with invasive cancers. (Length-biased sampling may at least partially explain survival trends for cervical cancer. Other cancers possibly affected include breast, colon, rectum, and prostate.)

Changes in diagnostic criteria: Early detection of cancer--resulting from either screening or earlier response to symptoms--may result in the increasing diagnosis of small tumors that are not yet life-threatening. This may have the effect of raising the incidence rates and survival estimates without changing the mortality rates. Breast, colon, prostate, cervix uteri, bladder, and skin (melanoma) are the cancer sites most likely to be affected.

Technological advances in diagnostic procedures: In this report, trends in survival by stage at diagnosis for specific cancers are not presented; trends in stage distributions are presented rarely. However, it is possible to compare survival by stage.

The assignment of a given stage to a particular cancer may change over time due to advances in diagnostic technology. Introduction of new technology can give rise to a phenomenon known as **stage migration**. Stage migration occurs when diagnostic procedures change over time, resulting in an *increase* in the probability that a given cancer will be diagnosed in a *more advanced* stage. For example, certain distant metastases that would have been undetectable a few years ago can now be diagnosed by a computer tomography (CT) scan or by magnetic resonance imaging (MRI). Therefore, some patients who would have been diagnosed previously as having cancer in a *localized* or *regional* stage are now diagnosed as having cancer in a *distant* stage. The likely result would be to remove the worst survivors—those with previously undetected distant metastases—from the localized and regional categories and put them into the distant category. As a result, the stage-at-diagnosis distribution for a cancer may become less favorable over time, but the survival for each stage may improve: the early stage will *lose* cases that will survive *shorter* than those remaining in that category, while the advanced stage will *gain* cases that will survive *longer* than those already in that category. However, *overall survival would not change* (Feinstein et al., 1985). Stage migration is an important concept to understand when examining temporal trends in survival by stage at diagnosis as well as temporal trends in stage distributions; it could affect the analysis of virtually all solid tumors.

Evolution of stage classifications: Every few years, the American Joint Committee on Cancer produces a new cancer-staging manual; the seventh edition is the most recent (Edge et al., 2010). The evolution of such classifications reflects the identification of new prognostic factors that may influence choice of treatment. Historically, the SEER Program has only collected data on **extent of disease** (**EOD**), rather than stage. EOD is *more specific* than stage and usually determines stage, even when stage definitions change. Thus, SEER easily adapts

to changes in stage definitions; moreover, trends in a newly redefined stage can usually be calculated. Recently the SEER Program has begun collecting **Collaborative Stage**. Collaborative Stage has the advantage of being a consolidated data collection system of three main staging systems (TNM, EOD, and Summary Stage) and allows combined pathological and clinical stage to be captured. For those cancers for which new prognostic variables are introduced into staging, so that previously collected EOD data cannot determine new stage categories, there can be problems in assessing trends in stage of disease. Only by reviewing the evolution of staging for a given cancer is it possible to determine what effects changes in stage definitions have had on stage-specific survival and on stage-at-diagnosis distributions. Stage migration (mentioned above) and EOD migration need also be taken into account. For some sites, the historic stage (*localized*, *regional*, or *distant*) is not shown, either because of inconsistencies in its definition over time or because stage isn't appropriate (such as for leukemias, which are all considered to be distant at diagnosis).

Interpreting relative survival: The relative survival estimate is the ratio of observed survival to expected survival for a given patient cohort. Expected survival is based on mortality rates for the entire population, taking into account, as appropriate, the age, sex, race, and year of diagnosis of the patients. Assuming that the presence of cancer is the only factor that distinguishes the cancer patient cohort from the general population, the relative survival estimate approximates the probability that a patient will *not* die of the diagnosed cancer within the given time interval. This is the same as the probability that the patient will either survive the interval or die of a different cause.

A factor related to the risk of a cancer may also be related to the risk of dying from causes unrelated to the cancer. An example of such a factor is smoking. Smoking is a major risk factor for lung cancer; therefore, a cohort of lung cancer patients will contain a much higher proportion of smokers than does the general population. However, smoking is also a risk factor for other diseases, resulting in smokers having a shorter life expectancy than nonsmokers. For this reason, expected survival estimates for lung cancer patients that are based on the life tables for the general population will be unrealistically high; since relative survival = observed / expected, this will result in relative-survival estimates that are *lower* than they would be if the population consisted only of smokers. The problem cannot be easily corrected because separate life tables for smokers and nonsmokers are not available. Moreover, amount of smoking (usually measured in pack-years) is clearly an important variable and can't be easily quantified. The possibility that expected survival may not be appropriate for a given patient cohort should also be considered when examining relative survival for patients with cancers of the cervix uteri or breast, because the risk of these cancers has been associated with socioeconomic status (Baquet et al., 1991), which may be related to life expectancy.

Previous to the *CSR* for 1973–1996, the expected survival tables used were for 1970 and 1980; there were separate tables for whites, blacks, American Indians, Chinese, Japanese, Filipinos, white Hispanics, and Hawaiians. In updating the tables for 1990, several problems emerged. The US life tables are based on age, race, and sex information from death certificates. The

information on race on the death certificate may not be accurate (Rosenberg et al., 1999). One reason is that funeral directors may inaccurately report race on a death certificate. Also, reported age at death, especially for those older than 85, may not be accurate because birth certificates were not issued with as much regularity in the early 1900s as they are today. Although race misclassification and age-at-death misreporting exist across all races, they may be more problematic for races other than white or black because of those races' smaller population sizes. Therefore, life tables were generated for 1970, 1980, 1990, and 2000 only for white, black, and other; these life tables were used to produce the relative survival estimates in this book. There may be small variations among survival estimates calculated in this *CSR* and those in *CSR*s prior to 1973–1996.

Comparison with other databases: The SEER data are obtained from population-based cancer registries covering about 28 percent of the US population. It is sometimes of interest to compare cancer statistics for SEER areas with those from other registries both in the US and worldwide. In making such comparisons, one must carefully consider the factors considered above for both data sources. In addition, one should assess all of the following:(1)completeness of case ascertainment, (2) rules used to determine multiple primaries, (3) follow-up, (4) rules used in assigning and coding cause of death, and (5) the sources and procedures used in obtaining population estimates. Depending on the rates being compared, there could be other confounding factors which should be considered. The same standard or standard million population should be used for the age-adjustment of each group being compared; most statistics from outside the US are based on the 2000 world standard million population. Examples of other databases are US Cancer Statistics (*http://apps.nccd.cdc.gov/uscs*) and CINA+ Online (http://www.cancer-rates.info/naaccr/).

It is sometimes interesting to compare survival for cancer patients in SEER areas with data from clinical trials. *This must be done with great caution.* Survival data from clinical trials may have been obtained from a patient population that differs from that of SEER patients in prognostic factors for the given cancer; any survival comparisons would have to adjust for such differences. Also, it is necessary to verify that the methodology used in computing survival is the same for both data sources. Furthermore, clinical-trials patients may differ from SEER patients in characteristics that may be related to survival but are not recorded in either database. If this were true for a given cancer, it would not be possible to make valid comparisons of this type.

Errors in data collection: In the process of registering cancer patients, errors may be made in abstracting and coding the data, which include demographic information, cancer site, histology, extent of disease, treatment, and patient survival. Quality control studies are periodically carried out to detect and correct this type of error, but no attempt is made to incorporate this source of error into the variance estimates of cancer rates reported here.

Comparison of this report with previous reports: The cancer registries that participate in the SEER Program submit data on all cancers diagnosed in their coverage areas to the NCI each year. Because of the dynamic nature of the registries' databases, *the reported*

number of new cancer cases in a particular race, sex, age, cancer category in a given calendar year may change from that which has been reported in a previous publication. For a given diagnosis year, additional cancer cases that were previously overlooked may have been found and reported to the central registry. There may have been follow-back of cancers diagnosed by death certificate only; successful efforts to establish the dates of diagnosis for such patients will change the number of patients reported for a given diagnosis year. Code changes may occur when a patient dies; for example, information on race is generally available on the death certificate and may be used to update a previously unknown value. There may have been elimination of duplicate records for the same patient, often due to name changes or misspellings.

Thus, a recent report may have a different number of cases for a given diagnosis year than an earlier report, with resulting effects on incidence and possibly survival. Population estimates may also change from one report to another for some calendar years. This occurs because the NCI receives population estimates that are regularly revised and updated by the Bureau of the Census (**BOC**). Such changes may result in some differences between incidence and mortality rates for a given calendar period as published in different reports. See our website for the most current information about the population estimates (*http://seer.cancer.gov/popdata/*).

REFERENCES

Baquet CR, Horm JW, Gibbs T, Greenwald P. Socioeconomic factors and cancer incidence among blacks and whites. *J Natl Cancer Inst* 1991; 83:551-557.

Breslow L (Chairman, Extramural Committee to Assess Measures of Progress Against Cancer). Measurement of progress against cancer: Final report to the Senate Appropriations Committee. Bethesda: National Cancer Institute; 1988.

Edge SB, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti A, editors. *AJCC Cancer Staging Manual*, 7th ed. New York (NY): Springer; 2010.

Feinstein AR, Sosin DM, Wells CK. The Will Rogers phenomenon: Stage migration and new diagnostic techniques as a source of misleading statistics for survival of cancer. *New Engl J Med* 1985;312:1604-1608.

Rosenberg HM, Maurer JD, Sorlie PD, Johnson NJ, MacDorman MF, Hoyert DL, Spitler JF, Scott C. Quality of death rates by race and Hispanic origin: A summary of current research. Hyattsville (MD): National Center for Health Statistics; Vital and Health Statistics, Series 2, No. 128, 1999.

Zelen M. Theory of early detection of breast cancer in the general population. In: Heuson J-C, Mattheiem WH, Rozencweig M, editors. *Breast Cancer: Trends in Research and Treatment*. New York (NY): Raven Press; 1976. p. 287-299.