



Section I

Introduction and Technical Notes



CANCER IN NORTH AMERICA: 2004-2008

INTRODUCTION

The North American Association of Central Cancer Registries, Inc. (NAACCR) is a professional organization that develops and promotes uniform data standards for cancer registration; provides education and training; certifies population based registries; aggregates and publishes data from central cancer registries; and promotes the use of cancer surveillance data and systems for cancer control and epidemiologic research, public health programs, and patient care to reduce the burden of cancer in North America. NAACCR annually produces this statistical monograph on cancer in North America to provide cancer incidence and mortality statistics for the United States and Canada.

This year marks the 21st annual release of the series of Cancer in North America (CINA) monographs. The 2011 monograph includes data from 70 central population-based registries: 57 from the United States (50 states, the District of Columbia, Puerto Rico and 5 metropolitan areas) and 13 from Canada (10 provinces and 3 territories). The submitted data reflect 100% population coverage of Canada and the United States. Combined, five-year average annual incidence statistics are created from registries that met or exceeded the NAACCR standards of “fitness for use” in aggregated incidence data for all years, 2004-2008. The combined statistics presented in Volume One reflect 96% of the U.S. population (47 states). Volume One includes statistics from the Canadian registries that met the CINA inclusion requirements. However, because these registries combined represent less than 50% of the Canadian population, the combined statistics for Canada and North America as a whole are not presented.

This monograph would not be possible without the commitment by population-based cancer registries throughout the United States and Canada to collect timely, complete, and accurate data. The NAACCR standing committee, Data Use and Research (DURC), is charged with the annual call for data from member registries to create the CINA monograph and other useful CINA products for data analyses and cancer incidence information (e.g., CINA+ Online, CINA in SEER*Stat, CINA Plus in SEER*Stat, CINA Monograph data in SAS Data set and CINA Deluxe). An Editorial Board of member volunteers, with support from NAACCR staff and staff from Information Management Services, takes on the responsibility to oversee the call for data, evaluate all submissions, and make all editorial decisions and activities in production of the monograph. Our efforts are to provide timely and useful information on cancer incidence and mortality for all geographic areas and race/ethnic groups in the United States and Canada. In recent years, the publication has been expanded to include cancer incidence and mortality on U.S. Hispanic/Latino and Asian/Pacific Islander populations and combined rates U.S. for U.S. American Indian and Alaskan Native populations located in Indian Health Service Contract Health Service Delivery Areas (CHSDA) of the United States. This year we are pleased to provide, for the first time, registry specific statistics for U.S. American Indian and Alaskan Native populations located in Indian Health Service Contract Health Service Delivery Areas (CHSDA) counties by submitting U.S. registry for those states with CHSDA counties (excepting only for Kansas and Montana).

CONTENTS OF THE CINA MONOGRAPH

The CINA monograph is comprised of three volumes as described below. Rates are presented as average-annual, age-adjusted rates using three different standard million populations: the 2000 U.S., the 1996 Canadian, and the World standard populations. We believe that presenting cancer rates adjusted to all three standards enhances the meaning and utility of the rates for users in the United States, Canada, and international settings. All volumes of the CINA publication, including sections containing incidence rates standardized to the World Standard, and the population counts used in the calculation of cancer rates are available for download on the NAACCR website, <http://www.naacr.org/DataandPublications/CINAPubs.aspx>.

Volume One, Combined Cancer Incidence for the United States, Canada, and North America presents cancer incidence data that have been combined to create five-year, average-annual data for cancers occurring in all persons and for children, aged 0 to 14, and 0 to 19. Registries submitting the data must meet the NAACCR criteria of high quality incidence data that are fit for use at the time of data submission. The criteria for the standard are described in detail below in the Data Quality Indicator section.

In the United States, cancer counts and incidence rates are presented for all races combined, black, white Asian/Pacific Islander, American Indian/Alaskan Native populations and for Hispanic/Latino, non-Hispanic white, non-Hispanic black. Volume 1 presents the combined rates with the incidence statistics for Canada and North America suppressed because the qualifying provincial registries represent less than 50% of the Canadian population. Finally, due to the large migration of both cancer patients and populations from Louisiana into Texas, Alabama, and Mississippi in the aftermath of Hurricane Katrina in September 2005, and some hospital closures and loss of many patient records in affected areas, data for the four states (Alabama, Louisiana, Mississippi and Texas) are presented in 4.5 years, from 2004 through June 2005 and all of 2006 and 2008 to avoid miscalculation of cancer rates due to the impact of these unavoidable events on the accuracy of the cancer rates.

In addition, Volume One has three sections: section one includes the five most common types of cancer by race and ethnicity, based on the combined United States data by race and ethnicity. Section two provides cancer incidence rates for the United States, including rates for all races combined, white, black, Asian or Pacific Islander, American Indian/Alaska Native, Hispanic, non-Hispanic white, and non-Hispanic black populations for all ages. These tables are followed by tables of cancer rates for children, aged 0-14 and 0-19; section three provides cancer incidence rates for adults for qualifying Canadian provinces.

Volume Two, Registry-specific Cancer Incidence in the United States and Canada presents cancer incidence data for NAACCR-member, population-based central cancer registries in Canada and the United States that have agreed to participate in the CINA monograph. For most cancer registries, five-year averages of data are presented for the years 2004 to 2008. If all five years of data were not available, an average of all available years is presented. Finally, due to the large migration of both cancer patients and populations from Louisiana into Texas, Alabama, and Mississippi in the aftermath of Hurricane Katrina in September 2005, and some hospital closures and loss of many patient records in affected areas, data for these four states are presented in 4.5 years, from 2004 through June 2005 and all of 2006 and 2008 to avoid miscalculation of cancer rates due to the impact of these unavoidable events on the accuracy of the cancer rates.

Each set of data tables includes demographic and data quality information and registry descriptions to help interpret the statistics reported. In the United States, cancer incidence rates are calculated for the following race/ethnic populations: all races combined, white, black, Asian or Pacific Islander, American Indian/Alaskan Native for CHSDA counties, Hispanic/Latino (all races), non-Hispanic white, and non-Hispanic black. Race and ethnicity information is not collected in Canada, and thus is not presented here.

Volume Three, Registry-specific Cancer Mortality in the United States and Canada presents cancer mortality data for fifty states, the District of Columbia, five SEER metropolitan areas, Puerto Rico, nine Canadian provinces and three Canadian territories. The United States combined mortality data are presented as well. Unlike the incidence statistics, mortality statistics from Louisiana into Texas, Alabama and Mississippi are presented for the full five-year period, 2004 to 2008. Canadian cancer mortality for the years 2004 through 2008 is also presented for all Canadian provinces with the exception of Quebec. Cancer mortality data for Quebec were not available at the time of this publication.

SOURCES OF DATA

Cancer Incidence. A cancer registry must be population-based and a NAACCR member in good standing to be included in NAACCR data publications. All cancer registries in the United States and Canada are NAACCR members, including the three territories in Canada and three U.S. territories. An annual request is sent to all members to submit voluntarily a data file for use in CINA, CINA research and data information products, the U.S. *Annual Report to the Nation*, and the American Cancer Society's *Cancer Facts and Figures* annual publication.

All NAACCR member registries receive support from the state, province or territory. In the U.S., they also participate in the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program or the Centers for Disease Control and Prevention's National Program of Cancer Registries (NPCR) or both. In Canada, all registries submit data to the Canadian Cancer Registry maintained by Statistics Canada.

Mortality Data. Mortality data for 2004 to 2008 for United States registries were obtained from the National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention (CDC), as provided to the National Cancer Institute (NCI). Mortality data for Puerto Rico were provided by Puerto Rico. Canadian cancer mortality for the years 2004 through 2008 were obtained from Statistics Canada.

Population Estimates. To assist the reader, all population estimates by age, sex, and race/ethnicity (where relevant) used in the calculation of rates in the monograph are available for each central registry and for the United States. These are available for download from the NAACCR website at <http://www.naacr.org/DataandPublications/CINAPubs.aspx>.

Population estimates for 2004 through 2008 for the United States, individual states, and SEER metropolitan areas were obtained from the SEER program. These estimates were based on United States Census Bureau population estimates for these years and represent a modification of the annual time series population estimates produced by the Population Estimates Program of the Census Bureau with support from the NCI. The population estimates incorporate bridged, single race estimates that are derived from the original multiple race categories in the 2000 Census. These bridged estimates are consistent with the four race groups enumerated in the 1990 Census and were produced under a collaborative arrangement between the National Center for Health Statistics (NCHS) and the Census Bureau. The methodology implemented by the Census Bureau to develop these county estimates is comparable to that used to produce national and state 1990-2000 intercensal estimates and is described on the Census Bureau's website (National Center for Health Statistics 2003). In 2007, the U.S. Census revised their methodology for population estimation. This revision impacts all population denominators for 2004 to 2008. The main effect of this approach was on specific age groups, particularly, two of the oldest age groups, ages 65-74 and 75 and older. In many states, the new method results in a larger increase in the younger age category and fewer in the older age category than estimated in previous years. Subsequently, this revision contributed to a larger decrease in age-adjusted cancer mortality rates in some states than what was expected based on statistics from last year.

The NCI modifies the Census data for the population estimates for the State of Hawaii. The Epidemiology Program of the Hawaii Cancer Research Center has developed its own set of population estimates, based on sample survey data collected by the Hawaii Department of Health. This effort grew out of a concern that the native Hawaiian population had been vastly undercounted in previous censuses. The "Hawaii adjustment" to the Census Bureau estimates has the net result of reducing the estimated white population and increasing the Asian and Pacific Islander population for the state. The Census Bureau estimates for the total population, black population, and American Indian and Alaska Native populations in Hawaii are unaffected. Refer to the SEER *Cancer Statistics Review, 1975-2008* and its methodologies for specific documentation regarding modifications made by the NCI to the Census Bureau estimates (Horner 2009).

Finally, the NCI applied an algorithm to the 2005 population estimates to account for the massive out-migration from Louisiana following Hurricane Katrina and the subsequent unprecedented in-migration into specific counties in neighboring and other states (Texas, Alabama, Mississippi, and Louisiana). However, one-half of the unadjusted annual estimates were used in the calculation of rates since the incidence data excluded cases reported from July 2005 through December 2005 for these four states.

For **Canada**, Statistics Canada provided the estimates of the Canadian population for all Canadian provinces and territories, adjusted for census under coverage and non-permanent residents. Race and ethnicity information is not collected in Canada, and thus is not presented here.

CANCER SITE CODING, 2004-2008

Incidence Data. All cancer registries use the International Classification of Diseases for Oncology, third edition (ICD-O-3) to code the anatomic site and morphology. Cancer incidence statistics include invasive cancers only, with the exception of *in situ* cancer of the bladder. Although tables include incidence statistics for breast cancer *in situ*, these cases are not included in any counts or rates of total cancer incidence.

The SEER program site recode groups were used for classifying types of cancer, using anatomic site and morphology. Using this standard ontology, only squamous and basal cell carcinomas of the lip and genital organs are included in the data reported. All categories used to present pediatric cancer rates in Volume One are based on the International Classification of Childhood Cancer, third edition (ICCC-3) (Steliarova-Foucher 2005).

Summary tables of all codes and site groups for incidence can be found in the electronic copy of the CINA Monograph Appendices on the NAACCR website, <http://www.naacr.org/DataandPublications/CINAPubs.aspx>.

Mortality Data. Underlying cause of death was coded using the International Classification of Diseases (ICD). In the United States, ICD-10 was used for all deaths from 1999 and later. In Canada, ICD-9 was used through 2000, and ICD-10 was used for 2001 and later. Cancer deaths were defined as those coded 140.0 through 208.9 in ICD-9 and C00 through C97 in ICD-10.

The SEER mortality recode scheme was used to classify cancer deaths into the groupings used in the volume (see Appendix B on the NAACCR website, <http://www.naacr.org/DataandPublications/CINAPubs.aspx>).

CANCER CODING CHANGES DURING 2004-2008

Cancer Incidence. Several definitional changes occurred in some histologies and behaviors in ICD-O-3 that affected the inclusion and exclusion of reportable cancers diagnosed beginning in 2001. These changes may affect the comparability of data reported here with previous CINA monographs. The changes predominately affected leukemias, lymphomas, and cancer of the ovary. One category of change between ICD-O-2 and ICD-O-3 is the manner in which leukemias and lymphomas are classified and coded. Although conversion of histology codes from ICD-O-2 to ICD-O-3 for cases diagnosed prior to 2001 will help to minimize these differences, some minor differences may still exist, particularly with respect to some relatively rare lymphocytic cancers that can be coded to either leukemia or lymphoma. Leukemias that represent a disease progression from one of the myelodysplastic diseases or syndromes diagnosed in 2001 and forward are no longer reportable. It is unlikely that this change will have much impact on the counts or rates for leukemia in this monograph, but the effect may be larger in subsequent years.

Starting with ICD-O-3, several myelodysplastic diseases and syndromes are considered malignant, and therefore are now reportable for cases diagnosed in 2001 and later. Because these cancers were reportable for the entire time period covered by this monograph, they have been included in the tables. Prior to the 18th edition of this report, we had only included cases that were reportable in both ICD-O-2 and ICD-O-3. Most of the cases of myeloproliferative (9980, 9982-9987, 9989) and myelodysplastic (9960-9962) diseases reportable only under the ICDO-3 rules are grouped in the category of "Ill-defined." Thus, the rates reported in the past 4 years are noticeably higher than what has been reported in *Cancer in North America* in the 17th and earlier editions, and this change is due to the changes in reportability.

For pediatric cancers, differences in incidence rates from previously published rates may be due to changes between the second and third edition of the International Classification of Childhood Cancers (ICCC). For example, incidence rates on non-Hodgkin lymphoma cancers presented in this monograph are much higher than those presented based the previous version of ICCC. Two changes in the ICCC-3 classification are main contributors to this change. 1) Burkitt lymphoma and unspecified lymphoma, which were separated from non-Hodgkin lymphoma in previous monographs, are combined with non-Hodgkin lymphoma in this monograph; 2) Some lymphomas, which were grouped in the miscellaneous lymphoreticular neoplasms in previous monographs, are included in the non-Hodgkin lymphoma category of this monograph.

Pilocytic astrocytoma is considered to have uncertain behavior in the published version of ICD-O-3, but is reportable as a malignant cancer in North America. Including the childhood astrocytomas in the category of malignant brain tumors may introduce differences between childhood brain cancer rates in North America compared to other areas of the world that may not include these tumors as malignant.

In addition, mesothelioma and Kaposi sarcoma cases were reported as separate categories (see Appendices A and B). This change has little or no impact on most rates for specific cancers.

Multiple Primary Determination Rules. During 2007 a significant revision was made to the decision rules followed to determine when a multiple primary cancer is present. These new rules may have an affect on the number of subsequent primaries identified with these effects varying by cancer site and type. Preliminary data suggest that overall incidence rates will vary only marginally for most cancer sites. This new rule will affect data for incidence years 2007 and 2008 within the current publication. For more information on these new rules please see <http://www.seer.cancer.gov/tools/mphrules/>

Cancer Mortality. Among the many changes in ICD-10 were increases in classification detail, the shift to an alphanumeric classification system, and a number of changes in the coding rules by which a single cause of death is selected from among the multiple causes reported by physicians as causing or contributing to the death. The change from ICD-9 to ICD-10 caused discontinuities in trends for many causes of death, including cancer. The extent of these discontinuities has been measured by comparability studies in which death records are double coded using both the Ninth and Tenth Revisions, and the results compared. Overall, approximately 0.7% more deaths are assigned to cancer when ICD-10 is used than when ICD-9 is used (Anderson, et al, 2001). For some cancers, the differences are larger. Accordingly, the death rate for all cancers combined is higher when ICD-10 is used than when ICD-9 is used. This general rule does not hold for specific cancer sites, whose rates may be higher or lower using ICD-10. However, as discontinuities are small, changes in death rates across the years of the ICD-9/ICD-10 boundary are still interpretable, especially for major cancer sites.

Cancer deaths among non-residents and deaths of unknown sex or age were omitted from all calculations.

Hispanic/Latino Ethnicity Identification. The ethnicity available in medical records and reported to cancer registries is enhanced by the use of the NAACCR Hispanic Identification Algorithm, version 2 (NHIAv2). NHIAv2 uses a combination of NAACCR variables including direct identification of ethnicity from the Spanish/Hispanic Origin variable (NAACCR data element 190 values 1-6), and information indirectly derived based on an evaluation of the strength of the birthplace, race, and surname (including maiden name when available) associations with Hispanic ethnicity status. After applying NHIAv2, cases not ultimately classified as Hispanic are classified as non-Hispanic, leaving no cases with “unknown” Hispanic status. The NHIAv2 algorithm allows the user to run the program using one of three possible options:

- 1) All Records: Surname portion is run on 0 (non-Hispanic), 7 (surname only), and 9 (unknown) for all counties. After the algorithm, item 190 codes (0, 7, 9) are either NHIA - Hispanic or non-Hispanic.
- 2) Option 1: (The default) For counties with 5% or more Hispanics - as for all records option. For counties with < than 5% Hispanic, surname portion is run on 7 (surname only) and 9 (unknown). After the algorithm, item 190 codes (7, 9) are either NHIA - Hispanic or non-Hispanic.
- 3) Option 2: For counties with 5% or more Hispanics - as for all records option. For counties with < than 5% Hispanic, surname portion is run on 7 (surname only). These become either NHIA - Hispanic or non-Hispanic. All cases coded to 9 on item 190 are converted to NHIA - non-Hispanic.

A comparison of the direct and indirect categorization of Hispanic ethnicity is provided in the Appendix F available from the NAACCR Website, <http://www.naacr.org/DataandPublications/CINAPubs.aspx>.

The NHIAv2 method is described in detail elsewhere. (Howe et al, 2003; NAACCR Expert Panel in Hispanic Identification, 2003; NAACCR Latino Research Work Group, 2005)

Asian/Pacific Islander Identification The information on Asian and Pacific Islander populations is also enhanced by submitting registries using a NAACCR developed algorithm (NAACCR, 2010). The NAACCR Asian Pacific Islander Identification Algorithm version 1 (NAPIIA v1.2.1) uses a combination of NAACCR variables to classify cases directly or indirectly as Asian/Pacific Islander for analytic purposes. It is focused on coding cases with a race code of Asian NOS (race code 96) or Pacific Islander NOS (race code 97) to a more specific Asian or Pacific Islander race category, using the birthplace and name fields (first, last, and maiden names). Birthplace can be used to indirectly assign a specific race to one of eight Asian groups (Chinese, Japanese, Vietnamese, Korean, Asian Indian, Filipino, Thai, and Cambodian) and three Pacific Islander groups (Samoan, Micronesian, and Polynesian). Names can be used to indirectly assign a specific race to one of seven Asian groups (Chinese, Japanese, Vietnamese, Korean, Asian Indian, Filipino, and Hmong) and three Pacific Islander groups (Hawaiian, Guamanian, and Samoan). The algorithm uses the following NAACCR standard variables: Race 1 through Race 5 (NAACCR Items 160 through 164), Name – Last (NAACCR Item 2230), Name – First (NAACCR Item 2240), Name – Maiden (NAACCR Item 2390), Birthplace (NAACCR Item 250), Sex (NAACCR Item 220)

Misclassification of American Indian and Alaskan Native Populations. The collection of accurate information on the cancer incidence in American Indian and Alaskan Native populations is hampered by misclassification of this population within cancer case reports. In order to address this problem, cancer registries in the United States coordinate with the Centers for Disease Control and the Bureau of Indian Affairs to link their case files to the enrollment files of the Indian Health Service (Espey, et al., 2008). This process identifies significant numbers of American Indians and Alaskan Natives that were otherwise classified within the cancer case report. The improved information on this population enables calculation of much more accurate cancer incidence rates. As presented in the incidence volumes, the data on American Indians and Alaskan natives are restricted to those counties where Indian Health Service Clinics are located, namely Contract Health Service Delivery Area (CHSDA) counties. This restriction assures that the incidence data presented is the best available data for this population.

DATA QUALITY INDICATORS

NAACCR assesses the quality of cancer incidence data from individual registries for a number of data quality indicators, which are described in detail below. Results for these indicators can be found in the *Cancer in North America, 2004-2008* Appendices on the NAACCR website. A dash in the tables indicates that data were not submitted for the year or were not able to be calculated.

In order to be included in the NAACCR Combined rates presented in Volume One, the data had to meet the criteria for high quality incidence data. These criteria were applied to each year of data individually, except for the estimate of duplicate reports, which was calculated for the years 2004 to 2008 as a whole.

- Data for 2004 through 2008 had to be submitted to NAACCR by December 1, 2010.
- The estimate of duplicate case reports had to be less than 2 duplicate reports per 1000 cases.
- All cases had to pass the edits on variables needed to compute cancer incidence rates by site, sex, race, and tumor descriptions on the NAACCR Call for Data EDITS metafile.
- Fewer than 3% of cases had missing information for sex, county of residence at diagnosis, and age at diagnosis, and fewer than 5% of the cases had an unknown race.
- Death clearance must be completed for 2004 through 2008 mortality with the percent of all cases derived from death certificates only (DCOs) of less than 5%.
- The NAACCR method to estimate completeness of case ascertainment yielded an estimate of 90% or higher.
- All cases had to pass the inter-record edits on cases reporting multiple primary cancers, as determined by the standard Inter-record EDITS program developed for the NAACCR Call for Data submissions.

Duplicate Case Records. Most central cancer registries rely on multiple reporting sources for cancer case reports. At the central cancer registry, multiple reports for the same patient must then be matched and the information from all records consolidated. In addition to determining whether multiple reports refer to the same individual, central cancer registry staff must also determine whether the tumor represents a new primary tumor or a duplicate report for a tumor already recorded. Failure to eliminate duplicate cases and duplicate tumors results in over-counting cancer cases. As a part of routine cancer registry operations, a variety of tools are used to ensure accurate case linkage and

case consolidation. As part of the preparation of the data submission to NAACCR, each registry uses the NAACCR protocol to determine the adequacy of case linkage and consolidation operations in identifying duplicate records.

The NAACCR protocol for assessing duplicate cases can be found on the NAACCR website, <http://www.naacccr.org>.

Completeness of Case Ascertainment. In order to evaluate case completeness for all geographic areas included in this monograph, the NAACCR Method to Estimate Completeness is used. The method is described in detail elsewhere (Wu, et al., 2002; Howe, 2007). A data analysis tool is available on the NAACCR website, <http://www.naacccr.org/Research/DataAnalysisTools.aspx>, that documents and calculates sex-, race-, and site-specific estimates based on observed incidence counts and rates, observed local cancer death rates, and a standard rate ratio of incidence to mortality. In addition, adjustments are made to the calculations to account for some variation in sex-, race- and site-specific variation in case fatality. The method assumes that the relationship between incidence rates and mortality rates are stable. However, with the recent annual decreases in cancer death rates, and the awareness that the decreases are not the same in all parts of the United States and Canada, it appears that this assumption is being challenged with a noticeable bias in an upward migration of completeness estimates for many locales. Due to the late availability of United States mortality data for 2008, the completeness estimates provided in this publication used mortality data for the years 2003-2007 to determine completeness. The percent completeness presented for each registry and used to determine inclusion of individual registry data in the combined rates presented in volume one were based upon rate ratios of 2004-2008 SEER 11 incidence rates divided by 2003-2007 U.S. mortality rates.

Missing Case Information. NAACCR has developed standards for completeness of data on key data items that are needed to produce meaningful cancer incidence statistics. These key data items include race, sex, county of residence at diagnosis, and age at diagnosis. Cases with unknown sex or age are omitted from all calculations. Cases with unknown race and county of residence were included in the cancer counts and rates for all races and/or all counties combined.

Death Certificate Only Cases. The proportion of cases identified by death certificate only (DCO) is an indicator of data quality and completeness. Central cancer registries use death certificates to identify potentially missed cases and to conduct follow back on cases that have cancer on the death certificate but who are not incident cases in the registry. Cases without follow back information are considered to be DCO cases and may have incomplete or missing information, including date and stage of diagnosis. For DCO cases, the date of death is used as the date of diagnosis. Registries that did not use death certificates as a source of case ascertainment have “na” listed in the death certificate only row on the registry description page.

Data Reliability and Accuracy. All data submitted to NAACCR are evaluated for reliability and accuracy using the EDITS program and a specific set of edits incorporated into a Call for Data Edits Metafile. An edit reviews the internal consistency between and among data elements, such as anatomic site and morphology. Cases that are identified as having errors are reviewed by registry staff and resolved prior to their NAACCR submission. The NAACCR Call for Data EDITS metafile is available on the NAACCR website.

[URL <http://www.naacccr.org/DataandPublications/CallforData.aspx>, Last accessed April 24, 2011].

An inter-record edits program, developed by the Centers for Disease Control and Prevention, is used to identify errors among the records reported for persons who have had multiple tumors diagnosed. For example, an inter-record edit will identify whether birthplace is the same on all reports of multiple tumors that are reported for one individual.

Timeliness of Data. The NAACCR standard defines timely data as data that are available for use in incidence statistics within 23 months of the close of a diagnostic year (i.e., December 1, 2010 for all cases diagnosed in 2008).

Site-Specific Microscopic Confirmation. This criterion is not used by NAACCR to determine high quality for the purposes of this publication. However, it is a useful indicator of quality of data collection. A proportion of microscopically confirmed cases that is higher or lower may suggest problems in case ascertainment or abstracting. However, this proportion varies by cancer site. For sites that are more likely to rely on a clinical or radiological

diagnosis, e.g., cancers of the pancreas and brain, confirmation rates that are too high may suggest that some clinically diagnosed cases are missing. Also, registries that do not use death certificates for case finding have an artificially high proportion of microscopically confirmed cases, because DCO cases by definition do not have information on whether the tumor was microscopically confirmed. While no NAACCR standard has been determined for microscopic confirmation, the guideline we use is that the proportion should fall between 92 and 96 percent of all cancer cases, based on the experience of the SEER program.

CALCULATION OF STATISTICS

Rates. Rates are per 100,000 population and are age-adjusted by five-year age groups to the 2000 U.S. standard population based on single years of age (Ries, et al. 2005), the 1996 Canadian population standard (Statistics Canada, 2003), and to the World population standard (Parkin, et al., 2002). Rates for childhood and adolescents in Volume One were expressed per million population. The incidence and death rates in this monograph are annual averages for the period 2004 through 2008. (Note: Not all registries submitted incidence data for all five years; their rates are annual averages for the years submitted. Further, rates for Alabama, Louisiana, Mississippi, and Texas are truncated at June 2005, including data for 2004 through June 2005 and all of 2006 and 2008 to avoid potential misclassification of cancer rates, as noted above). The age distributions of the three population standards are as follows:

AGE GROUP	2000 U.S.	1996 CDN.	WORLD
00 years	3,794,901	12,342	24,000
01-04 years	15,191,619	53,893	96,000
05-09 years	19,919,840	67,985	100,000
10-14 years	20,056,779	67,716	90,000
15-19 years	19,819,518	67,841	90,000
20-24 years	18,257,225	67,761	80,000
25-29 years	17,722,067	72,914	80,000
30-34 years	19,511,370	87,030	60,000
35-39 years	22,179,956	88,510	60,000
40-44 years	22,479,229	80,055	60,000
45-49 years	19,805,793	71,847	60,000
50-54 years	17,224,359	55,812	50,000
55-59 years	13,307,234	44,869	40,000
60-64 years	10,654,272	40,705	40,000
65-69 years	9,409,940	37,858	30,000
70-74 years	8,725,574	32,589	20,000
75-79 years	7,414,559	23,232	10,000
80-84 years	4,900,234	15,424	5,000
85+ years	4,259,173	11,617	5,000
Total	274,633,642	1,000,000	1,000,000

Standard Errors. Standard errors (S.E.) of the rates were calculated using the formula:

$$S.E. = \sqrt{\sum \frac{w_j^2 n_j}{p_j^2}}$$

where w_j = the fraction of the standard population in age group j , n_j = number of cases or deaths in that age group, and p_j = person years denominator (Breslow and Day, 1987). For many registries, the standard errors of the rates are small, as the population covered is large. However, for registries that cover a small population, the standard error may be substantial.

Confidence Intervals. Although not reproduced in the printed monograph, bar charts of the registry-specific, age-adjusted rates with 95 percent confidence intervals (Tiwari, et al., 2006) are available on the NAACCR website (<http://www.naacrr.org/DataandPublications/CINAPubs.aspx>). The confidence intervals allow the user to assess the precision of the estimate and is an approximate and conservative indicator of whether a registry's rate is statistically higher or lower than the rate for the combined United States, based on whether or not the upper or lower limit of the confidence interval overlaps the 95% confidence interval for the United States.

Comparison of Rates. In addition to true regional variation in cancer risk, differences in cancer incidence or mortality rates between areas may be due to either differences in the demographic make up of the population or differences in data quality. In making valid comparisons of cancer incidence rates among registries, it is important to review the data quality indicators for each registry before attributing rate differences to regional variation. In addition to data quality, it is important to consider differences in the racial composition of the populations being compared before conclusions are drawn about variations in regional rates. Interpretation without consideration of these factors may contribute to misleading or inaccurate conclusions.

The standard error of adjusted rates can be used to evaluate the statistical significance of rate differences among comparable populations. For example, if the adjusted rates in two populations are R_1 and R_2 and their standard errors are $S.E._1$ and $S.E._2$, an approximate confidence interval for the rate ratio can be calculated using the following formula:

$$(R_1/R_2)^{1+z/x}$$

where $x = (R_1 - R_2) / \sqrt{(S.E._1^2 + S.E._2^2)}$ and $z = 1.96$ for 95% limits (Parkin, et al., 1992). If this interval does not include one, the two rates are statistically significantly different at a p value of 0.05. This test can be inaccurate for rates based on fewer than 16 cases or deaths, and it should not be used for rates based on fewer than six cases or deaths. It should be emphasized that this kind of comparison of adjusted rates must be undertaken with caution as misleading conclusions may be drawn if the ratios of the age specific rates in the two populations are not constant in all age groups. In these circumstances, the ratios of the adjusted rates will vary according to the standard populations used (Esteve, et al., 1994).

Cell Suppression. Counts and rates were suppressed (indicated by “-”) in the tables if the race, gender, and site-specific number of cases or deaths was less than six. These counts are included in the calculation of all sites combined. A dash is also used to indicate not applicable, as in the gender specific cancers. If the rate was less than 0.05 per 100,000 then the rate is listed as 0.0.

REFERENCES

Anderson RN, Minino AM, Hoyert DL, Rosenberg HM. Comparability of cause of death between ICD-9 and ICD-10: Preliminary estimates. *National Vital Statistics Reports* 2001;49:1-32.

Breslow NE and Day NE. *Statistical Methods in Cancer Research*. Vol. II. 1987. International Agency for Research on Cancer: Lyon, France.

Espey DK, Wiggins CL, Jim MA, Miller BA, Johnson CJ, Becker TM. Methods for improving cancer surveillance data in American Indian and Alaska Native populations. *Cancer*. 2008 Sep 1;113(5 Suppl):1120-30.

Espey DK, Wu XC, Swan J, Wiggins C, Jim MA, Ward E, Wingo PA, Howe HL, Ries LA, Miller BA, Jemal A, Ahmed F, Cobb N, Kaur JS, Edwards BK. Annual report to the nation on the status of cancer, 1975-2004, featuring cancer in American Indians and Alaska Natives. *Cancer*. 2007 Nov 15;110(10):2119-52.

Esteve J, Benhamou E, and Raymond L. *Statistical Methods in Cancer Research*. Vol. V. 1994. International Agency for Research on Cancer: Lyon, France.

Horner MJ, Ries LAG, Krapcho M, Neyman N, Aminou R, Howlader N, Altekruse SF, Feuer EJ, Huang L, Mariotto A, Miller BA, Lewis DR, Eisner MP, Stinchcomb DG, Edwards BK (eds). *SEER Cancer Statistics Review, 1975-2006*, National Cancer Institute. Bethesda, MD

[URL: http://www.seer.cancer.gov/csr/1975_2006/, based on November 2008 SEER data submission, posted to the SEER web site, 2009.]

Howe HL, Carozza S, O'Malley C, Dolecek TA, Finch JL, Kohler B, Wet D, Liu L, Schymura MJ, Williams M, Abe T, Agovino P, Chen VW, Firth R, Harkins D, Hotes, J, Lake A, Roney D, Suarez L (eds). *Cancer in U.S. Hispanics/Latinos, 1995-2000*. Springfield (IL): North American Association of Central Cancer Registries, December 2003.

Howe HL. The NAACCR Method to Estimate Completeness Training Webinar, March 15, 2007. [URL: <http://www.bizingo.com/m/NAACCR/p/subscribers-rateunderstanding.asp>, Last Accessed February 25, 2008].

Johnson CH, Peace S, Adamo P, Fritz A, Percy-Laurry A, Edwards BK. *The 2007 Multiple Primary and Histology Coding Rules*. National Cancer Institute, Surveillance, Epidemiology and End Results Program. Bethesda, MD, 2007.

NAACCR Expert Panel in Hispanic Identification. *Report of the NAACCR Expert Panel on Hispanic Identification 2003*. Springfield (IL): North American Association of Central Cancer Registries, October 2003.

NAACCR Latino Research Work Group. *NAACCR Guideline for Enhancing Hispanic/Latino Identification: Revised NAACCR Hispanic/Latino Identification Algorithm [NHIA v2]*. Springfield (IL): North American Association of Central Cancer Registries. September 2005.

[URL: <http://www.naacr.org/Research/DataAnalysisTools.aspx>. Last Accessed, May 12, 2011].

NAACCR Method to Estimate Completeness. A Data Analysis Tool for Calculations. [URL: <http://www.naacr.org/Research/DataAnalysisTools.aspx>. Last Accessed, May 12, 2011].

NAACCR Race and Ethnicity Work Group. *NAACCR Asian Pacific Islander Identification Algorithm [NAPIIA v1.2.1]*. Springfield, IL: North American Association of Central Cancer Registries, September 2010.

National Center for Health Statistics. *U.S. Census Populations with Bridged Race Categories*. September 2003. [URL: <http://www.cdc.gov/nchs/about/major/dvs/popbridge/popbridge.htm>. Last Accessed, April 25, 2009].

Parkin D.M., Whelan S.L., Ferlay J., Teppo L., Thomas D., Thomas D.B. *Cancer Incidence in Five Continents*. Vol. VIII. 2002. International Agency for Research on Cancer: Lyon, France.

Statistics Canada. 2003. *Annual Demographic Statistics*. Statistics Canada Catalogue no. 91-213-XIB-2002000. Ottawa: Minister of Industry. [URL: <http://www.statcan.ca/cgi-bin/downpub/listpub.cgi?catno=91-213-XIB2002000>].

SEER. 2000 U.S. Standard Population vs. Standard Million. 13 December 2004. http://seer.cancer.gov/stdpopulations/single_age.html

Steliarova-Foucher E, Stiller C, Lacour B, Kaatsch P. International Classification of Childhood Cancer, third edition. *Cancer* 2005;103:1457-67.

Tiwari RC, Clegg LX, Zou Z. Efficient interval estimation for age-adjusted cancer rates. *Stat Methods Med Res* 2006;15:547-569.

Wu XC, Hotes J, Fulton JP, Cormier M, Correa C, McLaughlin CC, Kosary C, Howe HL, Chen VW (Eds.) *Cancer in North America, 1996-2000, Volume One: Introduction and Technical Notes, Incidence*. April 2002. [URL: <http://www.naaccr.org/DataandPublications/CINA19962000.aspx>. Last Accessed, May 12, 2011].

