Cancer in North America: 2012-2016

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, Canada, and North America.

This year marks the 29th annual release of the series of Cancer in North America (CiNA) monographs. The 2019 monograph includes data from 68 central population-based registries: 57 from the United States (50 states, the District of Columbia, and Puerto Rico, plus 5 regions (California--excluding the Greater Bay area and Los Angeles and 4 metropolitan areas) and 11 from Canada (9 provinces and 2 territories).

The submitted data reflect nearly 100% population coverage of Canada and the United States (U.S.). Combined, five-year average annual incidence statistics for the United States, for Canada and for North America are created from registries that met or exceeded the NAACCR standards of "fitness for use" in aggregated incidence data for all years, 2012-2016. Note: This year the "fitness for use" standard does not fully align with NAACCR Certification for diagnosis year 2016 cases. Specifically, the New Mexico Tumor Registry did not meet the completeness standard for certification in every calendar year of the evaluation. However, there is emerging evidence that the NAACCR completeness algorithm may not work well in geographic areas with unique demographic profiles, though this issue is still under review. Based on review of multiple submissions, the years for which New Mexico did not meet the NAACCR completeness standard remained below the threshold even with subsequent years of data collection and case ascertainment. In addition, a review of incidence rates of historically hard to capture cancer types indicates these cancers in New Mexico have rates at or higher than the U.S. average. This is a striking difference from registries with data collection problems that result in low case ascertainment. Therefore, we have included New Mexico based on the assumption that the demographic profile of the state precludes the completeness model from accurately calculating estimated cases. This inclusion aligns with the standards for inclusion in the SEER Program public-use dataset.

Volume One presents the combined U.S. and Canadian statistics for cancer incidence rates across North America. The combined statistics presented in Volume One reflect 98.9% of the U.S. population (49 states and Puerto Rico) and includes statistics from the 9 Canadian registries that met the CiNA inclusion requirements, reflecting 38.2% of the Canadian population.

NAACCR issues an annual call for data to member registries to create the CiNA monograph and other useful CiNA products for data analyses and cancer incidence information (e.g., NAACCR FastStats on-line query system, NACCCR Maps on-line query system, CiNA Public Use Dataset, and CiNA Deluxe Research Files). For more information on these products and the process for gaining access to them, please visit www.naaccr.org and select *Cancer in North America CiNA Data Products* under the *Research & Analytic Tools* tab.

This monograph is a testament to the continuing commitment by population-based cancer registries throughout the United States and Canada. The efforts of the NAACCR membership to collect timely, complete, and accurate data make this publication possible. 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 in production of the monograph. The objective of this work is 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.

The publication includes U.S. cancer incidence and mortality for white, black, Hispanic/Latino and Asian/Pacific Islander (API) as combined and registry specific rates and American Indian and Alaskan Native (AIAN)

populations presented for U.S. counties located in Indian Health Services Purchased/Referred Care Delivery Areas (PRCDA) of the United States (excepting only for Kansas and Minnesota due to limitations in county-level reporting from those states).

In the 2012-2016 Cancer Incidence in North America monograph, registry specific delay adjusted rates are provided to adjust for reporting delay. Each year, registries submit cases approximately two years after the close of a calendar year (e.g. cases for 2016 were first submitted in the December 2018 NAACCR Call For Data submission). These case counts are then revised in each subsequent submission. Delay adjusted rates represent the observed rate adjusted to estimate what will be the "eventual" case count, providing more accurate trends and levels of incidence rates.

Starting in 2015, a joint effort by NCI, CDC, and NAACCR was mounted to develop a unified approach for estimating and reporting delay-adjusted rates across all of the U.S. and Canada. The delay-adjusted rates for NAACCR registries, including SEER registries reported in the Cancer Statistics Review, is based on historic data submitted to NAACCR. Statistical models to estimate delay are estimated for each registry by age (<50, 50-69, 70+), gender (male and female) and race/ethnicity (all races combined, white, black, API, AIAN PRCDA region, Hispanic, non-Hispanic white and non-Hispanic black), and delay time (for the 2018 submission a delay of 2 years for diagnosis year 2016, 3 years for diagnosis year 2015, etc.) The statistical models estimate reporting delay factors by age, race and gender for each registry, and calculating delay adjusted rates for any grouping is done in SEER*Stat by weighting each group appropriately.

For the third year, an improved statistical model has been used with the goal of producing more stable estimates. The details of the new modeling are available in the Development of the Delay Model of the website http://www.surveillance.cancer.gov/delay/.

For some registries, there were insufficient historical data available to calculate registry specific delay factors. In order to calculate delay adjusted national statistics, delay factors were applied to registries with missing data using an average of the U.S. or Canadian delay factors.

CONTENTS OF THE CiNA MONOGRAPH

This year the CiNA monograph is comprised of four volumes as described below. Rates are presented as average-annual, age-adjusted rates using three different standard million populations: the 2000 U.S. (Census P25-1130), the 2011 Canadian, and the World (WHO 2000-2025) 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. Recent 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, https://www.naaccr.org/cancer-in-north-america-cina-volumes/#Population.

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 indicating that the data 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, and for Hispanic/Latino, non-Hispanic white, non-Hispanic black populations.

It should be noted that cancer incidence data for two of the largest provinces in Canada, Ontario and Quebec, are not included in Volume One. Ontario submitted data that did not meet NAACCR standards for the December 2018 data and Quebec did not complete a data submission for the year. These two provinces represent over 60% of new cancer cases in Canada. Only 9 of 14 Canadian provinces met the fitness for use criteria for this report. As a result, the statistics for Canada should be interpreted with this in mind.

The Introduction and Technical Notes is the first section of every volume.

Volume One sections two through four are data sections:

Section two includes the five most common types of cancer by sex in North America, the United States and Canada based on the combined United States and the combined Canadian data. The section also includes the top five most common types of cancer by sex, race and ethnicity in the United States based on the combined United States data.

Section three provides cancer incidence rates for the United States, Canada and North America. The rates for the United States are presented for all races combined, white, black, Asian or Pacific Islander, American Indian/Alaskan Native, Hispanic, non-Hispanic white and non-Hispanic black populations for all ages combined. The rates for Canada and North America are presented for all races combined and for all ages combined. Also included in section three are tables of cancer rates for children, aged 0-14 and 0-19 for the United States, Canada and North America. As with the tables for all ages, the pediatric tables are presented by race and ethnicity for the United States and for all races combined for Canada and North America.

Section four contains the frequency counts, percent distribution and age-adjusted rates of cancer by summary stage at diagnosis (derived from and following the Collaborative Stage staging system) for female breast, and for prostate, lung, uterine cervix and colorectal cancers. Combined United States data are provided for white, black, Asian or Pacific Islander, and American Indian/Alaskan Native populations with American Indian/Alaskan Native data limited to PRCDA counties. In addition, combined United States data, combined Canadian data and combined North America data are provided for all races combined. U.S. data are also provided by Hispanic ethnicity.

Section five provides cancer counts and rates for the time period 2012-2016 that are "delay-adjusted." This adjustment is done to estimate the number of incident cancer cases that were diagnosed in the time period but whose reporting to a central registry is delayed. These estimates are derived from historic patterns of reporting lag adjusted by various factors including cancer type, sex, age and race/ethnicity. A U.S. average and a Canadian average delay factor was calculated and applied to excluded registries to assure the combined delay-adjusted data are consistent with the incidence rates otherwise presented in Volume One. For more information on the importance of this methodology see the articles by Clegg, Midthune and Zou listed in the references at the end of this section or visit http://surveillance.cancer.gov/delay/.

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. Five-year averages of data are presented for the years 2012 to 2016.

Each set of data tables includes demographic and data quality information and registry descriptions to help interpret the statistics reported. The 2019 publication includes registry specific data on stage at diagnosis for female breast, lung, colon and rectum, prostate and uterine cervix cases for all registries submitting stage data for 2012 through 2016. 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 PRCDA 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.

Volume Three, Registry-specific Cancer Mortality in the United States and Canada presents cancer mortality data for all geographic areas in the United States and Canada.

Volume Four, Cancer Survival Rates in the United States and Canada includes data from 59 registries; 50 U.S. registries (45 states and 5 regions) and 9 Canadian registries (9 provinces) on more than 9 million cases diagnosed among North Americans between 2009 and 2015.

Volume Four is comprised of two data sections:

Section two includes counts, relative survival ratios and confidence intervals for the United States, Canada and North America combined. These statistics are presented for all races by sex and select cancer sites. The tables for each cancer site present statistics by a total, stage and specific age groups for each region. In the United States combined, survival statistics by white and black race are also presented.

Section three includes counts, relative survival ratios and confidence intervals by registry, sex and select cancer sites. Survival statistics are available for the United States and Canadian registries by all races and for the United States registries by white and black race.

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 2012 to 2016 for United States and Puerto Rico 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). The Canadian mortality data for 2012 to 2016 were provided by Statistics Canada. These data are available from Statistics Canada's Vital Statistics – Death Database (http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3233).

Population Estimates. To assist the reader, 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, the United States and Canada. These data are available online on the NAACCR website at https://www.naaccr.org/cancer-in-north-america-cina-volumes/.

Population estimates for 2012 through 2016 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 2010 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 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 2012 to 2016. 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 Alaskan Native populations in Hawaii are unaffected. Refer to the SEER *Cancer Statistics Review, 1975-2006* and its methodologies for specific documentation regarding modifications made by the NCI to the Census Bureau estimates (Horner 2009).

For **Canada**, Statistics Canada provided the estimates of the Canadian population for all Canadian provinces and territories, adjusted for census net under coverage (http://www5.statcan.gc.ca/cansim/a26?lang=eng&retrLang=eng&id=0510001&pattern=population+estimates&tabMode=dataTable&srchLan=-1&p1=1&p2=-1).

CODING, 2012-2016

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* and benign tumors of the brain and central nervous system, 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, https://www.naaccr.org/cancer-in-north-america-cina-volumes/.

Mortality Data. Underlying cause of death was coded using the International Classification of Diseases version 10 (ICD-10). Cancer deaths were defined as those coded C00 through C97 in ICD-10.

The SEER mortality recode scheme was used to classify cancer deaths into the groupings used in Volume III (see Appendix B on the NAACCR website: https://www.naaccr.org/cancer-in-north-america-cina-volumes/).

Behavior Recode Selection used in the CiNA Publication.
The CiNA publication uses the SEER Behavior Recode variable for selecting tumors. In ICD-O-2 and ICD-O-3 there are differences as to which tumors are considered malignant. These differences can affect which cases are reported and therefore need to be taken into account. In the CiNA publication all in situ bladders are considered malignant. Invasive cancers are presented based on the selection from the Behavior Recode variable values Malignant, Only malignant in ICD-O-3 and Only malignant 2010+. The exception to this selection is the delay adjusted incidence contained in volume 1, section 5. For these statistics, invasive cancers are based on the Malignant only. In some cases non-invasive breast (behavior recode variable equals in situ) and Brain and Other Central Nervous System (behavior recode variable equal to benign or borderline malignancy) statistics are available. For more details on the categories, how they are defined and how to create the behavior recode variable see https://seer.cancer.gov/behavrecode/ .

Summary Stage used in the CiNA Publication.
In this year's call for data, Derived Summary Stage 2000 (data item 3020) was reported for diagnosis years 2004-2014. The SEER Summary Stage 2000 field (data item 759) was reported for diagnosis years 2015-2016. Some registries continue to record and report this information for years outside of these ranges.

The CiNA publication presents data from 2012-2016, a timeframe that spans both reporting diagnosis years for Derived Summary Stage 2000 and SEER Summary Stage 2000. For diagnosis years 2012-2014, the Derived Summary Stage 2000 field was used. If the field was blank or unknown, then the SEER Summary Stage 2000 value was used as long as it was not blank or unknown. For diagnosis years 2015-2016, SEER Summary Stage 2000 was used. If that field was blank or unknown, then the Derived Summary Stage 2000 value was used, as long as it wasn't blank or unknown.

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, 8), 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, https://www.naaccr.org/cancer-in-north-america-cina-volumes/.

The NHIAv2 method is described in detail elsewhere. (NAACCR Race and Ethnicity Work Group, 2011; NAACCR Guideline for Enhancing Hispanic-Latino Identification: Revised NAACCR Hispanic/Latino Identification Algorithm [NHIA v2.2.1], Revised September 12, 2011 available here: https://www.naaccr.org/wp-content/uploads/2016/11/NHIA_v2_2_1_09122011.pdf).

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 (AI/AN) 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 PRCDA counties (Purchased/Referred Care Delivery Area counties previously known as the Contract Health Service Delivery Areas (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, 2012-2016* 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 2012 to 2016 as a whole.

- Data for 2012 through 2016 had to be submitted to NAACCR by December 1, 2018.
- 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 2012 through 2016 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. Note: New Mexico is an exception this year based on evidence that case ascertainment is high in the state and emerging evidence that the method does not appropriately estimate completeness for New Mexico's unique population profile.
- 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.

To be included in the combined stage data, the registry must meet the requirements for being included in the combined rates (above), have submitted stage data for all years 2012 through 2016 and provided the derived summary stage field in their submission.

To be included in the delay adjusted data, the registry must meet the requirements for being included in the combined rates, have submitted data for all years 2012 through 2016 and have provided the data needed to calculate the delay factors.

The selection criteria for including registry data in Volume Four on cancer survival included meeting the criteria for inclusion in the combined incidence rates and providing the follow up fields required for the survival calculations. In addition, the data either needed to meet the SEER standards for follow-up or the registry needed to have ascertained deaths through the study cutoff date (December 31, 2015). To meet the SEER standard for follow-up, a minimum of 90% of patients needed to have follow-up dates on or after January 1, 2016, or be deceased. For U.S. registries that did not meet the SEER standard for follow-up, it was necessary to conduct state death linkages and linkage with the National Death Index. For Canadian registries, it was necessary to conduct death linkages within the province or territory. Registry data also were required to have no irregularities in the variables that are key to the survival calculations.

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.naaccr.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, https://www.naaccr.org/analysis-and-data-improvement-tools/#COMPLETENESS 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. The completeness estimates provided in this publication used mortality data for the years 2012-2016 to determine completeness in U.S. registries and Canadian registries. 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 2012-2016 SEER 11 incidence rates divided by 2012-2016 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 of diagnosis and stage at 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 here: https://www.naaccr.org/call-for-data/#datatools. 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, 2018 for all cases diagnosed in 2016).

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 too high or low 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. (Census P25-1130) standard population based on single years of age (Day, et al., 1996), the 2011 Canadian population standard (Statistics Canada, CANSIM Table 051-0001), and to the World (WHO 2000-2025) population standard (Ahmad, et al., 2001). Rates for children and adolescents in Volume One were expressed per million population. The incidence and death rates in this monograph are annual averages for the period 2012 through 2016. The age distributions of the three population standards are as follows:

AGE GROUP	2000 U.S. (Census P25-1130)	2011 CDN.	WORLD (WHO 2000-2025)
00 years	3,794,901	376,321	17,917
01-04 years	15,191,619	1,522,743	70,625
05-09 years	19,919,840	1,810,433	86,870
10-14 years	20,056,779	1,918,164	85,970
15-19 years	19,819,518	2,238,952	84,670
20-24 years	18,257,225	2,354,354	82,171
25-29 years	17,722,067	2,369,841	79,272
30-34 years	19,511,370	2,327,955	76,073
35-39 years	22,179,956	2,273,087	71,475
40-44 years	22,479,229	2,385,918	65,877
45-49 years	19,805,793	2,719,909	60,379
50-54 years	17,224,359	2,691,260	53,681
55-59 years	13,307,234	2,353,090	45,484
60-64 years	10,654,272	2,050,443	37,187
65-69 years	9,409,940	1,532,940	29,590
70-74 years	8,725,574	1,153,822	22,092
75-79 years	7,414,559	919,338	15,195
80-84 years	4,900,234	701,140	9,097
85+ years	4,259,173	643,070	6,348
Total	**274,633,642**	**34,342,780**	**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 (https://www.naaccr.org/cancer-in-north-america-cina-volumes/). 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 makeup 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 and 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:

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

where $x = (R_1 \text{ B } R_2) / \sqrt{(S.E._1^2 + S.E._2^2)}$ 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 for incidence was less than six or when the number of deaths was less than ten. Complementary cell suppression was employed, when necessary, to prevent the contents of a single suppressed cell from being calculated. 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. The following SEER*Stat variables were used to create the statistics for the registry incidence pages: Year of Diagnosis, NAACCR Registry, Sex, Race Recode (W, B, AI, API) IHS Modified, PRCDA 2016, NHIAv2 Recode, Site recode ICD-O-3/WHO 2008 and Behavior recode for analysis.

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.

Clegg LX, Feuer EJ, Midthune D, Fay MP, Hankey BF. Impact of reporting delay and reporting error on cancer incidence rates and trends. J Natl Cancer Inst 2002;94:1537-45.

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, based on November 2008 SEER data submission, posted to the SEER web site, 2009. Available here: https://seer.cancer.gov/archive/csr/1975_2006/.

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.

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.

Midthune DN, Fay MP, Clegg LX, Feuer EJ. Modeling reporting delays and reporting corrections in cancer registry data. J Am Stat Assoc 2005;100(469):61-70.

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: https://www.naaccr.org/wp-content/uploads/2016/11/NHIA_v2_2_1_09122011.pdf Last Accessed May 10, 2019]

NAACCR Method to Estimate Completeness. A Data Analysis Tool for Calculations. Available here: https://www.naaccr.org/analysis-and-data-improvement-tools/#COMPLETENESS.

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. Available here: https://www.cdc.gov/nchs/nvss/bridged_race.htm.

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.

SEER. 2000 U.S. Standard Population vs. Standard Million. 13 December 2004. Available here: [ULR: 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.

Weir HK, Johnson CJ, Mariotto AB, Turner D, Wilson RJ, Nishri D, Ward KC. Evaluation of North American Association of Central Cancer Registries' (NAACCR) data for use in population-based cancer survival studies. J Natl Cancer Inst Monogr. 2014 Nov;2014(49):198-209.

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. Available here: https://www.naaccr.org/previous-volumes-of-cina/.

Zou J, Huang L, Midthune D, Horner MJ, Krapcho M, Feuer EJ. Effect of reporting year on delay modeling. Statistical Research and Applications Branch, National Cancer Institute; 2009. Technical Report #2009-01
Clegg LX, Feuer EJ, Midthune D, Fay MP, Hankey BF. Impact of reporting delay and reporting error on cancer incidence rates and trends. J Natl Cancer Inst 2002;94:1537-45. [Full Text]

Midthune DN, Fay MP, Clegg LX, Feuer EJ. Modeling reporting delays and reporting corrections in cancer registry data. J Am Stat Assoc 2005;100(469):61-70. [View Abstract]

Zou J, Huang L, Midthune D, Horner MJ, Krapcho M, Feuer EJ. Effect of reporting year on delay modeling. Statistical Research and Applications Branch, National Cancer Institute; 2009. Technical Report #2009-01. Available here: http://surveillance.cancer.gov/reports/.

Day, Jennifer Cheeseman, *Population Projections of the United States by Age, Sex, Race, and Hispanic Origin: 1995 to 2050*, U.S. Bureau of the Census, Current Population Reports, P25-1130, U.S. Government Printing Office, Washington, DC, 1996. Available here: https://www.census.gov/prod/1/pop/p25-1130/p251130.pdf.

2011 Canadian Standard Population. Statistics Canada, Table 051-0001 – Estimates of population, by age group and sex for July 1, Canada, provinces and territories, annual (persons unless otherwise noted). CANSIM (database).Available here: http://www5.statcan.gc.ca/cansim/a26?id=510001.

Ahmad O, Boschi-Pinto C, Lopez A, Murray C, Lozano R, Inoue M. Age Standardization of Rates: A New WHO Standard. GPE Discussion Paper Series: No. 31. World Health Organization 2001. Available here: http://www.who.int/healthinfo/paper31.pdf.