Section I

Introduction and Technical Notes

NAACCR Cancer in North America (CiNA) Prevalence
Introduction and Technical Notes

CiNA Volume Five: Cancer Prevalence in the United States and Canada 2006-2015 includes data from 59 registries on more than 8 million cases diagnosed among North Americans between 2006 and 2015.

Volume Five is comprised of two data sections:

Section two includes two, five, and ten year limited-duration prevalence counts and prevalence percentages 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 and specific age groups for each region. In the United States combined, prevalence statistics for non-Hispanic whites, non-Hispanic Blacks, and Hispanics are also presented.

Section three includes two, five, and ten year limited-duration prevalence counts and prevalence percentages and confidence intervals by registry, sex and select cancer sites. Prevalence statistics are available for the United States and Canadian registries by all races and for the United States registries for non-Hispanic whites, non-Hispanic Blacks, and Hispanics.

Special thanks go to Steve Scoppa and Rick Firth, Information Management Services, Inc., who prepared the analytic dataset used for this report; members of the NAACCR Prevalence Volume Task Force, who authored this first edition of the CiNA Prevalence Volume; and all of the registry staff who contributed data.

Cancer in North America: Prevalence

INTRODUCTION

The North American Association of Central Cancer Registries (NAACCR) has been producing reports of cancer incidence statistics for over 27 years and survival statistics since March 2016. This is the inaugural edition of the Cancer in North America (CiNA) Prevalence Volume.

Cancer prevalence is the number of persons alive on a certain date who have a history of cancer, so is a function of both incidence and survival. Prevalence includes new (incidence) and pre-existing cases and information on prevalence can be used for health planning, resource allocation, and an estimate of cancer survivorship.[1] Limited-duration prevalence represents the number, or proportion, of people alive on a certain day who had a diagnosis of the disease within the past x years (e.g. x = 2, 5 or 10 years).[2] This is in contrast to complete prevalence, which represents the proportion of people alive on a certain day who were diagnosed with the disease, regardless of how long ago the diagnosis was made. Complete prevalence can be estimated from self-reported population-based surveys, although one must be concerned with underreporting and misclassification of disease. Direct computation (the counting method) of complete cancer prevalence requires registry data that have been collected over a sufficiently long period of time to capture all prevalent cases of the disease. Historically, data from the National Cancer Institute's Surveillance, Epidemiology, and End Results Program (SEER) have been used to estimate U.S. national complete prevalence, but these estimates have generally been based on data from SEER-9 registries, which cover about 9.4% of the U.S. population.[3] We are unable to estimate complete prevalence using CiNA data. The CiNA Prevalence Volume presents two, five, and ten year limited-duration prevalence counts and prevalence percentages and confidence intervals for individual registries and the United States, Canada and North America combined.

TECHNICAL NOTES

Registry Inclusion. For registries to be included in *CiNA Prevalence* volume, they needed to: (1) provide consent, (2) meet CiNA incidence criteria for all relevant years, and (3) either meet the SEER standards for follow-up[4] or ascertain deaths through the study cutoff date (December 31, 2015). For the November 2018 Call for Data, the publication of prevalence estimates in CiNA was included as a Primary Use of Data.

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. These follow-up dates could have been the result of either passive or active patient follow-up mechanisms.[5] We used the November 2018 Call for Data file to measure the follow-up rate and the vital status follow-up activities form that is part of the annual call for data to learn what registries performed to ascertain deaths. 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. National death clearance in Canada is on hold pending legal agreements among the provinces and Statistics Canada, which impacts follow-up for 2009 and later. However, because the number of deaths that occur out of province is a small proportion of total deaths, we believe that Canadian follow-up data are negligibly influenced by the lack of national death linkage.

The *CiNA Prevalence* publication includes diagnosis years 2006-2015 with follow-up through the end of 2015. Using information from the NAACCR Call for Data Follow-Up Activities Forms, follow-up through 2015 was deemed to be the best balance between including the most current data and including the most registries. In terms of national coverage, *CiNA Prevalence* includes data from 9 of 13 Canadian provinces/territories and 45 of 51 states/District of Columbia. National population coverage by *CiNA Prevalence* is about 92% for the United States and 38% for Canada. The NAACCR U.S. combined and NAACCR Canadian combined statistics may not be representative of the total national populations. NAACCR North American prevalence statistics are a combination of U.S. and Canadian data. Data from 59 state, sub-state, and provincial or territorial registries are included in the registry-specific tables. To avoid double counting, data from sub-state registries in California, Michigan and Washington were not included in the NAACCR U.S. combined or NAACCR North American statistics. The sub- state registries in California, Michigan and Washington were included in the registry-specific statistics.

For coding and other general technical information, please see the technical section of Cancer in North America: 2012-2016, Volume One: Combined Cancer Incidence for the United States, Canada and North America.[6]

Statistical Methods. We used 2006-2015 incidence cases and survival from CiNA data based on the November 2018 NAACCR submission to estimate two, five, and ten year limited-duration prevalence (also known as partial prevalence) on January 1, 2016. We used the counting method to estimate prevalence from incidence and follow-up data. The counting method estimates prevalence by counting the number of persons who are known to be alive at a specific calendar time and adjusting for those lost to follow-up.[1] For registries meeting SEER follow-up standards (SEER-18 registries,[3] Montana, and Wyoming), survival estimates were used to adjust for loss to follow-up. For other registries, it was assumed that all deaths were ascertained through the study cutoff date and remaining persons were presumed to be alive, which may slightly overestimate prevalence. For cases lost to follow-up in registries meeting SEER follow-up standards, prevalence was estimated using observed survival on the remaining cases, by year of diagnosis (2006-2010 and 2011-2015), registry, sex, age group (<60, 60-69, 70+), race/ethnicity (non-Hispanic white, non-Hispanic Black, non-Hispanic American Indian and Alaska Native, non-Hispanic Asian and Pacific Islander, and Hispanic), and SEER Site Recode.[5]

For *CiNA Prevalence*, we included malignant cases per the SEER behavior recode for analysis[7] during 2006-2015 with follow-up/death ascertainment through the study cutoff date of December 31, 2015. Cases reported solely via death certificates or autopsy were excluded. For registries meeting SEER follow-up standards, alive cases with no survival time were excluded from analysis. Using SEER 2007 Multiple Primary and Histology Coding Rules,[8] we allowed for multiple primary cancers to be included for each patient, but only one record per patient was included in each prevalence estimate. For example, if a person had three primary tumors during the period 2006-2015, in the order of breast → colon → breast, then the first breast cancer case was used for breast cancer prevalence, the colon case was used for colon cancer prevalence, and the first breast cancer case was used for "all sites." Thus, the "All Sites" totals are less than the sum of the individual cancer types because each person is counted only once for each statistic, and some persons have more than one cancer type.

SEER*Stat (version 8.3.6) was used to perform the prevalence calculations.[9] Staff at Information Management Services, Inc. (Calverton, MD) prepared a SEER*Stat database for the purpose of calculating *CiNA Prevalence* statistics.[10] Prevalence statistics are suppressed when prevalence estimate counts are less than 10. For U.S., Canadian, and NAACCR North American statistics, we estimated prevalence counts by multiplying prevalence percentages for the respective registries included in CiNA Prevalence by the total population estimates by age (19 age groups: <1, 1-4, 5-9, 10-14… 80-84, 85+), sex (male, female), and race (white/unknown, black, other).

Incidence and prevalence statistics calculated using registry data do not include information on persons with a history of cancer who move to a new state (net-migration). For states with high population growth, the prevalence would be underestimated. For some cancers (e.g., hematopoietic, melanoma of the skin) there are known issues with reporting delay and potentially missed incidence cases when the person is diagnosed and treated in a physician's office, but not seen in a hospital, which may underestimate prevalence.

Variation in prevalence by registry catchment area can be due to several factors, including but not limited to: (1) differences in demographic characteristics related to race, ethnicity, and SES; (2) cancer screening rates; (3) access to and quality of care; and (4) cancer registration practices that impact case ascertainment, date of diagnosis and follow-up.

REFERENCES

1. Understanding cancer prevalence statistics https://seer.cancer.gov/statistics/types/prevalence.html.

2. Limited-duration prevalence https://surveillance.cancer.gov/prevalence/limited.html.

3. Registry Groupings in SEER Data and Statistics https://seer.cancer.gov/registries/terms.html.

4. Current SEER follow-up standards https://seer.cancer.gov/tools/seer.nov19instructions.pdf, page B-1.

5. Weir HK, Johnson CJ, Mariotto AB, Turner D, Wilson RJ, Nishri D, Ward KC. Evaluation of NAACCR Cancer in North America data for use in population-based cancer survival studies. J Natl Cancer Inst Monogr. 2014 Nov;2014(49):198-209. doi: 10.1093/jncimonographs/lgu018.

6. For coding and other general technical information, Cancer in North America: 2012-2016, Volume One: Combined Cancer Incidence for the United States, Canada, and North America, https://www.naaccr.org/cancer-in-north-america-cina-volumes/#Vol1

7. SEER Behavior Code for Analysis http://seer.cancer.gov/behavrecode/.

8. SEER 2007 Multiple Primary and Histology Coding Rules http://seer.cancer.gov/tools/mphrules/.

9. SEER*Stat version 8.3.6; produced by the Surveillance Research Program of the Division of Cancer Control and Population Sciences, National Cancer Institute, and Information Management Services, Inc., Calverton, MD.

10. SEER*Stat Database: NAACCR Incidence Data - CiNA Analytic File, 1995-2016, for Expanded Races, Custom File With County, Johnson - Prevalence WG (which includes data from CDC's National Program of Cancer Registries (NPCR), CCCR's Provincial and Territorial Registries, and the NCI's Surveillance, Epidemiology and End Results (SEER) Registries), certified by the North American Association of Central Cancer Registries (NAACCR) as meeting high-quality incidence data standards for the specified time periods, submitted December 2018. (SEER*Stat Database ID 41415).