

CANCER IN NORTH AMERICA, 1997 - 2001

VOLUME ONE: INCIDENCE

INTRODUCTION

The North American Association of Central Cancer Registries, Inc. (NAACCR) annually produces this report on cancer in North America. The purpose of this report is to provide cancer incidence and mortality statistics for as much of the population of North America as possible and to promote uniform data standards for cancer registration, improvement of data quality, and promote the use of the cancer registry data. The report provides a means of accessing incidence and mortality data in a common format for most states and provinces of North America, as well as providing estimates of cancer incidence rates for the entire United States, based on a select set of registries with high quality data.

This volume includes data from 64 central population-based registries: 51 from the United States (45 states, 5 metropolitan areas and the District of Columbia) and 13 from Canada (10 provinces and 3 territories). This represents total coverage in Canada and all but 5 states in the U.S. (Kansas, Mississippi, Pennsylvania, Tennessee, and Vermont). For the United States, 90% of the population is covered in the registry specific data and 55% is covered in the combined rates.

This monograph was produced by the Data Evaluation and Publication Committee, a standing committee of the NAACCR. The NAACCR bylaws charge the Data Evaluation and Publication Committee (DEPC) to gather data from member registries, review, evaluate, and compile the information for publication. The year 2004 marks the 14th release of the annual publication of *Cancer in North America (CINA)* series and the 8th monograph to include cancer mortality data. This monograph reflects the much-appreciated participation of NAACCR member registries in this monograph. We hope this effort continues to improve the completeness, the timeliness and the quality of data collected by member registries; to promote the use of their cancer registry data; and to provide cancer statistics that are inclusive of all racial/ethnic groups and geographic coverage of North America.

MONOGRAPH FORMAT

The *Cancer in North America (CINA), 1997-2001* monograph includes three volumes:

- Volume One contains cancer incidence data for individual member registries.
- Volume Two contains cancer mortality data for individual Canadian provinces and the individual United States.
- Volume Three contains cancer incidence data combined across registries that meet certain criteria for high quality data (NAACCR Combined Incidence Data).

CONTENTS OF VOLUME ONE

Volume One presents incidence data for individual central cancer registries within Canada and the United States who have agreed to participate in the CINA monograph. For most cancer registries, data are presented for the years 1997 to 2001. If the full five years of data were not available, however, as many years as possible are presented. In order to meet the needs of users throughout North America and the world, this volume contains incidence data age-adjusted to several standards. The contents are as follows:

- Section I: Introduction and Technical Notes
- Section II: Registry Contact List
- Section III: Registry-specific Cancer Incidence by Race and Sex, Age-adjusted to the 2000 U.S. and World Population Standards. This section also includes a registry information page with select registry-specific demographic and data quality indicators.
- Section IV: Registry-specific Cancer Incidence by Sex, Age-adjusted to the 1996 Canadian and World Population Standards
- Appendix A: SEER Site Groups for Primary Site Based on ICD-O-2
- Appendix B: SEER Site Groups for Primary Site Based on ICD-O-3
- Appendix C: Data Quality Indicators by Registry
- Appendix D: Demographic and Population Information by Registry

TECHNICAL NOTES

Data Sources

Incidence. Each member registry provided its own incidence data for 1997 to 2001; if all five years were not available, data were provided for as many of the five years as possible. Cancer incidence data for registries in the U.S. SEER program were obtained from the SEER public use data file (November 2003 submission) produced by the National Cancer Institute (NCI). Statistics Canada provided data for all Canadian registries except for Alberta, Ontario, and Manitoba that submitted their own data. The 1997-2001 incidence data for Northwest Territories and Nunavut reflect current geographic boundaries.

Cancer registries reported invasive cancers only, with the exception of in-situ cancers of the bladder and breast. Squamous and basal cell carcinomas of the skin were not reportable, except those of the lip and genital organs (see Appendices A and B).

Population Estimates – United States. Estimates of the population for the U.S., individual U.S. states, and all SEER areas for 1997 through 2001 were obtained from the SEER program, based on U.S. Bureau of Census population estimates for these years. These population estimates represent a modification of the annual time series of population estimates produced by the Population Estimates Program of the Bureau of the Census 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 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.¹

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 has been vastly undercounted in previous censuses. The "Hawaii-adjustment" to the Bureau of the Census estimates has the net result of reducing the estimated white population and increasing the Asian and Pacific Islander population for the state. The Bureau of the Census estimates for the total population, black population, and American Indian and Alaska Native populations in Hawaii are unaffected. Please refer to the SEER Cancer Statistics Review, 1975-2000² and its methodologies for specific documentation regarding modifications made by the NCI to the Census Bureau estimates.

The registry information page of volume one provided for each participating registry in Section III includes the estimated population of each registry by race and year, as well as the percent of the population in each race group for the 1997-2001 time period. Appendix D also summarizes this information for all registries.

Population Estimates – 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. The 1997-2001 populations for Northwest Territories and Nunavut reflect current geographic boundaries.

The registry information page of volume one provided for each participating registry in Section III includes the estimated population of each registry by year. Appendix D also summarizes this information for all registries. Canadian data are not stratified by race.

¹ National Center for Health Statistics. U.S. Census Populations with Bridged Race Categories(On-line). Available: <http://www.cdc.gov/nchs/about/major/dvs/popbridge/popbridge.htm>.

² Available: <http://seer.cancer.gov/popdata/>

Data Quality Indicators

NAACCR assesses the quality of the data from individual registries based on the following criteria:

- Duplicates from NAACCR Protocol
- Completeness of Case Ascertainment
- Missing Race, Sex, County, Age
- Death Certificate Only Cases
- Passing EDITS
- Site Specific Microscopic Confirmation

These data quality indicators are described in the following sections. Where appropriate, the NAACCR standard for each data quality indicator is provided. NAACCR has two levels of standards: Gold level and Silver level certification standards. If a registry did not submit data for all the years covered in this report, a dash is shown on the registry description page for the data quality indicators for the years not submitted.

Duplicates from NAACCR Protocol. 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 be matched and consolidated. In addition to determining if a subsequent report is for the same individual, the central cancer registry must also determine if the tumor represents a new primary tumor, or a subsequent report for a tumor already recorded. Failure to eliminate duplicate cases and duplicate tumors will result in over counting cancers. As part of the preparation of the data file for submission to NAACCR for this monograph, each participating registry examines a sample of cases to determine if there are unresolved duplicates on their file. The Protocol for Assessing Duplicate Cases used for this assessment can be found on the NAACCR Web site. The NAACCR standard for duplicates is fewer than 1 per 1000 cases for Gold and fewer than 2 per 1000 cases for Silver. If the registry did not report the results of the NAACCR duplicate protocol, “na” is listed as the duplicate prevalence on the registry information page at the beginning of each registry’s portion of Section III.

Completeness of Case Ascertainment. The gold standard for measuring case ascertainment for cancer registries is to conduct case-finding audits to determine the percent of cases missed. Because of the time and expense involved in the audits, however, it is not feasible to obtain annual case ascertainment completeness estimates for each registry annually. In order to provide a criterion by which to evaluate registry case completeness, NAACCR has developed a completeness measure based on incidence-to-mortality rate ratios. The NAACCR standards for completeness based on the incidence to mortality rate ratio method adjusted for duplicates are 95% for Gold and 90% for Silver. The incidence-to-mortality rate ratio method is described in full detail in Holly L. Howe, Conclusions of the Work Group for High Quality Criteria for Data Use, NAACCR Narrative, 2001; Winter.³

Briefly, the method uses the age-adjusted incidence rates observed in the SEER program and the age-adjusted mortality rates for the entire United States to determine a standard incidence-to-mortality rate ratio. This standard is then applied to the registry specific age-adjusted mortality rates to estimate an expected age-adjusted incidence rate. The expected incidence rate is then compared to the observed age-adjusted incidence rate to estimate completeness. This method provides a completeness measure that is relative to the completeness of the SEER program.

This method assumes that the ascertainment of cancer deaths in the United States is complete and that the ratio of incidence to mortality does not vary by geographic area. Because of differences in survival, the method is applied separately by cancer site, age, and sex. For registries within the United States (other than Hawaii), the method is also applied separately for blacks and whites. For Hawaii and Canadian registries the SEER incidence-to-mortality rate ratio for white is used as the standard. The completeness measure is a summary measure based on the weighted estimates by cancer site, sex, and race.

Missing Race, Sex, County, Age. NAACCR has developed standards for completeness of key demographic data items collected by cancer registries, including race, sex, county of residence at diagnosis, and age at diagnosis. To achieve the Gold level standard, not more than 2% of cases can be missing information on sex, county, and age; and not more than 3% of cases from United States registries can be missing race

³ http://www.naacccr.org/index.asp?Col_SectionKey=6&Col_ContentID=9

information. For Silver level, not more than 3% of cases can be missing sex, county, and age; and not more than 5% can be missing race. The percent of cases missing these data items can be found on the registry information page at the beginning of each registry's portion of Section III.

Cases with unknown sex or age were omitted from all calculations, but cases with unknown race were included in the computation of "all races" cancer counts and rates. Cases with unknown county of residence are included in all calculations but cases that were not resident in the registry coverage area are excluded.

Death Certificate Only Cases. The proportion of cases identified by death certificate only (DCO) is a measure 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 for whom follow-back information is not available are DCO cases and may have incomplete or missing information, including date of diagnosis and stage of diagnosis. For DCO cases, the date of death is used as the date of diagnosis. The proportion of cases that are DCO for each year is listed in the registry information page at the beginning of each registry's portion of Section III. Only invasive cancer cases are included in the denominator of this proportion. The NAACCR standard for percent of cases ascertained by death certificate only is 3% for Gold level and 5% for Silver.

Many new registries postpone the use of death certificates for case finding until the registry has at least five years' data because deaths occurring in the early years of registration are likely to have been diagnosed before the registry's date of establishment. Were these deaths to be collected and registered as DCO cases, they would inflate cancer incidence during the first several years of registry operations because they are registered in the year of death. 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.

Passing EDITS. All data submitted to NAACCR is checked for quality using a standard set of quality control criteria termed EDITS (NAACCR EDITS metafile, Versions 9.1.10 or 10). EDITS contain checks of internal consistency between data elements, such as anatomic site and morphology or between morphology and age. Cancer case reports that do not meet these criteria are either flagged by the central registry as having been confirmed as correctly coded, or the case generates an error when the EDITS software is applied. For example, EDITS will cause an error if a cancer case is reported to have carcinoma in the brain. Carcinomas do not arise in the brain because the brain does not contain epithelial tissue. Registries use EDITS to identify cases that have potential data errors so that they can be resolved prior to submission to NAACCR. The NAACCR Gold level standard is for all cases (100%) to pass EDITS. The Silver level standard is for 97% cases to pass EDITS. The NAACCR EDITS metafiles are available on the NAACCR website.

Site Specific Microscopic Confirmation. The proportion of total cases with microscopic confirmation can also be used as an indicator of the quality of data collection. Between 92 to 96 percent of all cancer cases

registered in SEER program are confirmed microscopically. 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 have unknown microscopic confirmation. There is no NAACCR standard for microscopic confirmation.

Definitions

Primary Cancer Sites for Incidence Data. All cancer registries participating in the Monograph used International Classification of Diseases for Oncology (ICD-O) to report the anatomic site of cancer and morphology to NAACCR. For the cases diagnosed in 1997 through 2000, the second edition was used (ICD-O-2) and for cases diagnosed in 2001, the third edition was used (ICD-O-3). This Volume uses the SEER program site recode groups for classifying types of cancer, using anatomic site and morphology (see Appendices A and B).

There were several changes in coding effective with ICD-O-3 that may affect the comparability of the data provided in this Monograph compared to previous versions of CINA. These predominately affect the leukemias and cancers of the ovary.

One category of change between ICD-O-2 and ICD-O-3 is how leukemias are classified and coded. Changes have been made in the schema used to make leukemia subtype groupings for cancers coded to ICD-O-2 (ICD-O-2 SEER site recode, Appendix A). These changes were made so that the rates for leukemia sub-types will be consistent over time, and not be influenced by the edition of ICD-O used for coding. Small differences may still exist, particularly with respect to some relatively rare lymphocytic cancers that can be coded to either leukemia or lymphoma.

Starting with ICD-O-3, several myelodysplastic diseases and syndromes were changed to be considered malignant, and are therefore now reportable in North America. Because these cancers were not reportable for the entire time period covered by this monograph, they have been excluded from the tables. There were approximately 12,500 cases excluded from all registries which submitted data for CINA. A small percent of leukemias may no longer be reportable because they represent a progression of disease from one of the myelodysplastic diseases or syndromes. It is unlikely that this change will have a large impact on the counts or rates for leukemia in this Monograph, but the affect may be larger in subsequent years.

Borderline serous, mucinous and papillary cystadenomas, which had been reportable as invasive malignancies using ICD-O-2, are no longer considered invasive malignancies in ICD-O-3. Most tumors with these histologies occur in the ovaries, and this change affects a relatively large proportion of ovarian tumors. Based on previous analysis of NAACCR combined data, this is about 13% of all ovarian cases, but the proportion may vary by registry⁴. These borderline tumors, approximately 17,200 cases from all registries which submitted data for CINA, have been excluded from this monograph for all years. The ovarian cancer rates provided here, therefore, are not comparable to the rates in previous editions 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.

⁴ Goodman MT, Howe HL, Tung KH, Hotes J, Miller BA, Coughlin SS, Chen VW. *Incidence of Ovarian Cancer by Race and Ethnicity in the United States, 1992-1997*. Cancer (Supplement) 2003 (May 15); 7/10:2676-85.

Rates. Rates were calculated per 100,000 population and age-adjusted by the direct method to the 2000 U.S., the 1996 Canadian (Cdn.) and the World population standards.⁵ Rates for childhood and adolescents in volume three were expressed per million. The incidence and mortality rates in this monograph are annual averages for the period 1997 through 2001. (Note: Not all registries submitted data for all five years; their rates are annual averages for the years submitted.) The age distributions of the three population standards are presented below:

AGE GROUP	2000 U.S.	1996 CDN.	WORLD
00 years	13,818	12,342	24,000
01-04 years	55,317	53,893	96,000
05-09 years	72,533	67,985	100,000
10-14 years	73,032	67,716	90,000
15-19 years	72,169	67,841	90,000
20-24 years	66,478	67,761	80,000
25-29 years	64,529	72,914	80,000
30-34 years	71,044	87,030	60,000
35-39 years	80,762	88,510	60,000
40-44 years	81,851	80,055	60,000
45-49 years	72,118	71,847	60,000
50-54 years	62,716	55,812	50,000
55-59 years	48,454	44,869	40,000
60-64 years	38,793	40,705	40,000
65-69 years	34,264	37,858	30,000
70-74 years	31,773	32,589	20,000
75-79 years	26,999	23,232	10,000
80-84 years	17,842	15,424	5,000
85+ years	15,508	11,617	5,000
Total	1,000,000	1,000,000	1,000,000

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

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

⁵ Waterhouse J, Muir C, Correa P, Powell J (eds). *Cancer Incidence in Five Continents, Volume III*. Lyon, France: International Agency for Research on Cancer, IARC Scientific Publications No.15, 1976.

where w_j = the fraction of the standard population in age_j group (5-year age interval), n_j = number of cases or deaths in that age group, and p_j = person-years denominator.⁶ For many registries, the standard error 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.

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 $z = 1.96$ for 95% confidence limits.⁷ 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.⁶

Cell Suppression. Counts and rates were suppressed (shown as a dash in the table “-”) 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.

NAACCR MISSION

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.

Please address all comments and suggestions about the monograph to:

NAACCR
Attention: Data Evaluation and Publication Committee
2121 West White Oaks Drive, Suite C
Springfield, IL USA 62704-6495
(217) 698-0800
(217) 698-0188 (FAX)

⁶ Breslow NE and Day NE. *Statistical Methods in Cancer Research, vol. II*, Lyon, France: IARC, 1987, p. 59.

⁷ Esteve J, Benhamou E, Raymond L. *Statistical Methods in Cancer Research, Volume V*. Lyon, France: IARC Publication No. 128, 1994.

For a copy of this monograph, please contact NAACCR at the above address. The monograph can also be downloaded or viewed from the NAACCR web site (<http://www.naacr.org>).

Data Evaluation and Publication Committee Members, 2003-2004

John P. Fulton, Rhode Island, *Chair*
Vivien W. Chen, Louisiana
Greer Gay, American College of Surgeons
Royale Anne Hinds, North American Association of Central Cancer Registries, Inc.
Joellyn Hotes, North American Association of Central Cancer Registries, Inc.
Holly L. Howe, North American Association of Central Cancer Registries, Inc.
Carol Kosary, National Cancer Institute
Andrew Lake, Information Management Services, Inc.
Steven D. Roffers, Emory University, School of Public Health
David Roney, Information Management Services, Inc.
Maria Schymura, New York
Sherri Stewart, Centers for Disease Control and Prevention
Anne-Marie Ugnat, Health Canada
Valerie Vesich, Commission on Cancer, ACOS
Ghislaine Villeneuve, Statistics Canada
Elizabeth Ward, Centers for Disease Control and Prevention
Dee West, Northern California
William Wright, California
Xiao-Cheng Wu, Louisiana