

# **CANCER IN NORTH AMERICA, 1993-1997**

## **VOLUME ONE: INCIDENCE**

### **PURPOSE**

The Data Evaluation and Publication Committee, a standing committee of the North American Association of Central Cancer Registries (NAACCR), produced this monograph. The NAACCR bylaws charge the Data Evaluation and Publication Committee to gather data from member registries, review and evaluate them, and compile them for publication. The year 2000 marks the tenth release of the publication of *Cancer in North America*. This monograph represents the largest registry participation with total coverage in Canada, 44 states, the District of Columbia, and five metropolitan areas in the United States. Nine registries submitted their data for the first time. It is the fourth monograph to include cancer mortality data. We hope this effort continues to improve the completeness, timely reporting, and quality of the data collected by member registries; to promote the use of their cancer registry data; and to provide cancer incidence statistics that are inclusive of all racial/ethnic groups and geographic coverage of entire North America.

### **MONOGRAPH FORMAT**

*Cancer in North America (CINA), 1993-1997* monograph includes two volumes: Volume One contains cancer incidence data and Volume Two contains cancer mortality data. An electronic copy of data from both volumes of the monograph is available either through the Internet (on the NAACCR web site at <http://www.naacrr.org>) for downloading or on 3.5-inch diskette by request. README text files are also available on the web site and diskette. The README files provide record layouts, instructions, and general information about the data files. A separate README file has been prepared as supporting documentation and for each of the data files (incidence, mortality, data quality indicators, population counts, and registry descriptions and technical documentation for the incidence, mortality, and pediatric files). The electronic data files are not incorporated into a software program and thus they require the use of other statistical or database software packages for data analysis. All participants in the *Call for Data 2000* consented to have their data reproduced in this electronic format.

The data in Volume One of this monograph pertain to cancer incidence from 1993 through 1997, unless indicated otherwise, for all participating registries. Volume One has seven major sections as described below.

#### **VOLUME ONE, SECTION I: Introduction and Technical Notes**

This section states the purpose of the monograph, describes the monograph format, and details the various data sources, data definitions, and data interpretation. It also provides some background information on NAACCR and lists member rosters of the Data Evaluation and Publication Committee and its subcommittees.

#### **VOLUME ONE, SECTION II: Registry-specific Cancer Incidence by Sex and Race, Age-adjusted to the 1970 U.S. Population and the World Standard**

Incidence data are presented in this monograph for 62 central population-based registries: 50 from the United States (44 states and 5 metropolitan areas, and District of Columbia) and 12 (10 provinces and 2 territories) from Canada. This represents total coverage in Canada and all but 6 states (Arkansas, Kansas, Missouri, Oklahoma, South Dakota and Vermont) in the U.S. The first page for each participating registry provides descriptive information about the registry, identifies the contact person(s), and summarizes several data quality and completeness indicators. This descriptive information about each registry is useful in evaluating the general data quality of the registry and the comparability of incidence rates among registries and in interpreting differences

among them. These indicators are the percents of death certificate only (DCO) cases, duplicate records (from the NAACCR protocol), and the NAACCR estimated completeness of case ascertainment adjusted for duplicates. The NAACCR adjusted estimate of completeness could not be calculated for registries that did not submit their results for the *Protocol for Assessing Duplicate Cases* (Ohio, Tennessee, and Quebec). To assist in data comparisons across registries, the percents of total cases that are of races other than white or black and of unknown race are also listed. The sources for case finding are included, with the diagnosis year that each was implemented (e.g., hospitals, death certificates, pathology laboratories, radiation therapy sites, interstate data exchanges, physicians, ambulatory surgical centers, and nursing homes or hospices).

In addition to these general quality indicators, several cancer site-specific data quality indicators are presented. These cancer sites were selected based on both their frequent incidence and their importance in cancer control and prevention activities. For each cancer, four data quality indicators are listed: the percent of DCO cases for that cancer site; the percent of microscopically confirmed cases; and the site-specific incidence-to-mortality rate ratios for whites and for blacks. If race-specific data were not provided for publication, incidence-to-mortality rate ratios are provided for all races combined. Unlike earlier releases, the incidence-to-mortality rate ratios for cancer types occurring in only one gender are based on the rates for that gender and not both genders combined (as done in previous years).

Following the registry description, annual age- and sex-specific population estimates for all races, whites, and blacks are reproduced for each registry in the U.S. For Canadian registries, population estimates are presented for all races combined.

The cancer incidence tables show the total five-year (unless indicated otherwise) incidence counts by site and sex and the average annual incidence rates age-adjusted to the 1970 U.S. population and to the World standard for all races, whites, and blacks for each registry in the U.S. For Canadian registries, rates are presented only for all races combined because race identifiers are not collected by Canadian cancer registries.

### **VOLUME ONE, SECTION III: Registry-specific Cancer Incidence by Sex, Age-adjusted to the 1991 Canadian Population and the World Standard**

Age-adjusted incidence rates for all races combined were also calculated for males and females separately using the 1991 Canadian standard population (Cdn.) and the World standard for all U.S. and Canadian participating registries.

### **VOLUME ONE, SECTION IV: Combined Cancer Incidence Rates for the United States, Canada, and North America**

The inclusion of registry data in the combined rates is reserved for those registries of highest quality, as demonstrated by meeting all criteria for completeness and quality as part of the *Call for Data* submission. The standards for such selection are described below under Technical Notes. All cases from qualified registries were included in the calculation of combined rates. However, in states where a SEER program is located within their boundaries and when both SEER area and state (e.g., Los Angeles and Greater Bay Area in California) were qualified for inclusion, only the state (California) data were included in the combined rates.

In this section, combined rates are presented for the United States, Canada, and North America. A general description regarding the population covered and several data quality indicators are provided for each of these three geographic areas with a map showing the registries meeting inclusion criteria for the combined rates. Combined rates for the major and minor cancer groups are presented, using the same table format as those in Sections II and III for individual registries. For the United States, combined rates are provided for all races, whites, and blacks and are age-adjusted to the 1970 U.S. population and the World standards. For Canada, combined rates are presented for all races combined and age-adjusted to the 1991 Canadian population and the

World standards. Rates for North America are provided for all races combined and are age-adjusted to the 1970 U.S., 1991 Canadian, and World standard populations.

In addition, summary tables of sex-specific incidence rates for selected major cancer sites are presented for all races, whites (U.S. only), and blacks (U.S. only) for individual registries qualified for inclusion in the combined rates. However, incidence data are not presented when fewer than six cases were reported for the specified sites, although the counts are included in the total. Rates are age-adjusted to the 1970 U.S. population for U.S. registries, the 1991 Canadian population for Canadian registries, and both the 1970 U.S. and 1991 Canadian populations for the North American combined rates. These summary tables facilitate comparison among high quality registries. Before comparing rates among these registries, however, it is important to consider whether the populations are comparable by race (i.e., *percent other races* or *percent unknown race* on the registry description page in Section II). It is also important to evaluate the quality and completeness of case ascertainment, as differences in rates may be attributable to case completeness and data quality, as well as to actual regional differences in cancer incidence.

For the first time this year, a table that summarizes the five most common cancer types for eight race groups and for persons of Hispanic origin has been added to the combined section for the United States. NAACCR committees have begun to assess the reliability of identifiers for race groups other than white and black in the United States. Our intent is that this effort will enable us to expand the presentation of cancer incidence rates in future monographs for more race and ethnic groups. With the availability of national, annual population estimates for American Indians/Alaskan Natives, Asian/Pacific Islanders, and Hispanic persons, evaluation of these population identifiers will be a primary focus during the next year.

This section also presents combined rates for cancer in children in the U.S., Canada, and North America. These tables present age-adjusted rates for 0-14 year olds and 0-19 year olds as well as 5-year, age-specific rates for selected pediatric cancer types. The cases were grouped into pediatric cancer groupings using the International Classification of Childhood Cancer ("ICCC" - see Appendix B) published by the International Agency for Research on Cancer (IARC) in 1996.<sup>1</sup>

## **VOLUME ONE, SECTION V:     Variation in Colorectal Incidence among States and Metropolitan Areas in the United States, 1991-1995**

Data available from population-based central cancer registries and other sources were used to measure gross geographic variation in age-adjusted colorectal cancer incidence across states and metropolitan areas in the United States based on cancer incidence data aggregated by NAACCR for its annual publication, *Cancer in North America, 1991-1995*, including data from 11 of the 14 SEER areas (five states, six metropolitan areas), and data from another 25 states across the United States. Selected essential variables were used to explore potential determinants of geographic variation, suggesting directions for future surveillance and analysis..

Age-adjusted colorectal cancer incidence rates were computed for all races combined by sex and year (1991-1995) for each of the 36 geographic areas. Values of estimated annual percent change (EAPC) in age-adjusted colorectal cancer incidence rates were computed for the period 1991-1995, by sex. Data from the U.S. census bureau and CDC's Behavioral Risk Factor Surveillance System (BRFSS) were used to create six additional variables to be tested as potential determinants of age-adjusted incidence and EAPC, including completeness of case reporting (two variables), racial composition, age structure, urbanization, and colorectal cancer

---

<sup>1</sup>Kramárová E, Stiller CA, Ferlay J, Parkin DM, Draper GJ, Michaelis J, Neglia J, and Qureshi S (eds). *International Classification of Childhood Cancer 1996*. Lyon, France: International Agency for Research on Cancer, IARC Technical Report No. 29, 1996.

screening. Geographic distributions of incidence and EAPC were described by computing ranges, medians, means, and standard deviations, and by constructing histograms. Pearson correlation coefficients were computed for all pairs of variables to explore statistical relationships.

Geographical variation in colorectal cancer incidence was substantial across the 36 states and metropolitan areas, suggesting the possibility of large future declines in areas with a high colorectal cancer burden. Some of this variation may be due to residual (uncontrolled) age effects in the rates. Geographic variation in EAPC ran the gamut from positive to negative, although the central tendency was clearly negative. The downward trend in colorectal cancer incidence (EAPC) was moderately correlated with level of screening.

The data suggest that colorectal cancer screening may be related to reduced cancer incidence in the period 1991-1995. The level of screening appears to have been more intense in more urbanized areas than less urbanized areas, suggesting greater access to screening services in the former. However, the present exploratory study analyzes *ecological* relationships, and as such is merely suggestive, not conclusive. Relationships observed among urbanization, screening, and incidence deserve closer attention. Attempting to replicate the present analysis with county data may be a good place to start.

## **VOLUME ONE, SECTION VI: Exploring the Internal Consistency of Registry Data on Stage of Disease at Diagnosis: Part III: Suggestions for an Edits Metafile**

In this report the Data Quality Indicator Subcommittee (“the Subcommittee”) completes the study of inconsistencies in data on stage of disease at diagnosis begun in 1998. In Parts I and II, the Subcommittee focused on cancer of the female breast and cancer of the prostate. In Part III, the Subcommittee explores the internal consistency of registry data on stage of disease at diagnosis for all cancers combined, suggesting outputs for an EDITS metafile. GenEDITS is a software package designed by the Centers for Disease Control and Prevention (CDC) for performing internal data consistency edits.

Using 1996 incidence data for all cancers combined from four central cancer registries, the Subcommittee analyzed the internal consistency of data on stage of disease at diagnosis. The objectives were to develop an analytical framework with which to identify consistencies and inconsistencies between staging schemes *for all cancers combined*, to analyze case reports with combinations of staging data not evaluated as “consistent,” to measure the yield of errors from combinations of staging data not evaluated as “consistent,” and to propose EDITS outputs for different combinations of staging data, based on the yield of errors.

Summary stage (SS) was selected as the standard for comparison, because of its pervasive use among NAACCR registries to categorize stage of disease at diagnosis. The other two indicators of stage were the SEER Program’s EOD system and the AJCC TNM system. Cancer registries in four states (Atlanta SEER, Arizona, California - Bay Area, and Illinois) contributed data for this study. Information on stage of disease at diagnosis was extracted on case reports for all cancers diagnosed in 1996. Combinations of SS and other codes for stage were evaluated using coding manuals and the expertise of committee members in tumor classification, and one of three EDITS outputs was originally proposed for each data combination (pass, warning, or error), based on the percentage of true staging errors expected among cases with that combination.

When cases in “error” cells were examined, most were found to have obvious coding errors. Similarly, when cases in “warning” cells were examined, many were found to have obvious coding errors. Minor adjustments were made to proposed EDITS outputs, based on empirical results.

The Subcommittee has shared its work with other NAACCR subcommittees: the Data Evaluation and Publication Committee’s Staging Subcommittee and the Uniform Data Standards’ EDITS Subcommittee.

## VOLUME ONE, SECTION VII: Adopting the HL7 Standard for Cancer Registry Work: Clarifying Unresolved Issues

Health Level Seven (HL7) is an American National Standards Institute (ANSI) approved standard for electronic data exchange in the health care industry which allows disparate computer systems to talk to each other. It combines codes and variable-length text fields, offering its users great flexibility. HL7's flexible structure is virtually ideal for cancer registration, where modifications to reporting formats occur frequently. HL7 has the potential to save time, money, and effort for its users. With HL7 on the horizon, the practical and policy significance of its use in central registries needs to be considered.

Ten full members of NAACCR were contacted and interviewed about their perceptions of the unresolved practical and policy issues related to the adoption of HL7 by central registries and by NAACCR. Participants were selected to represent a wide variety of central cancer registries and because their interest in HL7 had been recognized by NAACCR leadership. Although participants came from the United States only, they represented registries from both federal funding programs, large and small, and from all regions of the country. Some participants had attended a 1998 conference on the use of HL7 by central registries, and some continue to study issues related to the use of HL7 in registry work. Participants were asked: "On the basis of what you know about HL7 and NAACCR's discussion of its potential adoption by central cancer registries, what important issues or problems concerning its use in cancer registration remain unresolved?" Participants' responses were followed-up with a set of specific probing questions to clarify responses.

Eight key groups of issues emerged: How do we develop new central registry software at reasonable cost? How do we avoid disruption of ongoing cancer registration? How do we avoid backlash from reporting institutions? How do we assure the confidentiality of data? How universal will HL7 become? How can we maximize HL7's potential as a case finding tool? How accurate is SNOMED coding at present? How can we maximize HL7's potential as an editing query tool?

When we weigh the appropriateness of its application to cancer registries, we must go beyond the technical aspects of translation and implementation, considering practical, administrative, and managerial issues related to its successful implementation, as well. NAACCR needs to study these issues further, with the goal of formulating a strategic plan for the use of HL7 in cancer registry systems.

### TECHNICAL NOTES

#### Data Sources

**Incidence.** Each member registry provided its own incidence data for 1993 to 1997; 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 SEER program, with the exception of Los Angeles and Seattle/Puget Sound, were obtained from the SEER public use data tape (August 1999 submission) produced by the National Cancer Institute (NCI). Statistics Canada provided data for all Canadian registries; the only exceptions are Manitoba and Ontario, for which the data were provided by the provincial registry.

**Mortality.** Mortality data for 1993 to 1997 for U.S. registries were obtained from the National Center for Health Statistics of the Centers (NCHS) for Disease Control and Prevention (CDC), as provided to NCI. For Canadian registries, the mortality data were obtained from Statistics Canada.

**Population Estimates.** Estimates of the population for the entire U.S., individual U.S. states, and all SEER areas for 1993 through 1997 were obtained from the SEER program, based on U.S. Bureau of Census 1999 population estimates for these years. Statistics Canada provided the estimates of the Canadian population for

all Canadian provinces and territories, adjusted for census under-coverage and non-permanent residents.

## Definitions

**Cancer Sites.** SEER rules were used to define cancer sites (see Appendix A)<sup>2</sup>. Numbers and rates were calculated for invasive cancers only, with the exception of cancer of the bladder. Registries that did not submit bladder *in situ* cases were noted in their tables by omitting the phrase “incl. *in situ*.” Numbers and rates for carcinoma *in situ* of the breast were listed separately, when supplied by the registry. *In situ* cases of the breast were not included in the “All Sites” category. Squamous and basal cell carcinomas of the skin were excluded, except those of the lip and genital organs. Cancers in non-residents of the area and cases of unknown sex or age were omitted from all calculations, but cases of unknown race were included in the computation of “all races” cancer rates.

**Incidence Rates.** Rates were calculated per 100,000 population and age-adjusted by the direct method to the 1970 U.S. and the World<sup>3</sup> standards for the U.S. registries, and the 1991 Canadian (Cdn.) and the World standards for Canadian Registries. The incidence rates are five-year averages for the period 1993 through 1997. The age distributions of the three population standards are presented below:

AGE GROUP	WORLD	1970 U.S.	1991 CDN.
0-4	12,000	8,441.6	6,946.5
5-9	10,000	9,820.4	6,945.4
10-14	9,000	10,230.4	6,803.4
15-19	9,000	9,384.5	6,849.5
20-24	8,000	8,056.1	7,501.6
25-29	8,000	6,632.0	8,994.4
30-34	6,000	5,624.9	9,240.0
35-39	6,000	5,465.6	8,338.8
40-44	6,000	5,895.8	7,606.3
45-49	6,000	5,962.2	5,953.6
50-54	5,000	5,464.3	4,764.9
55-59	4,000	4,907.7	4,404.1
60-64	4,000	4,240.3	4,232.6
65-69	3,000	3,440.6	3,857.0
70-74	2,000	2,678.9	2,966.0
75-79	1,000	1,887.1	2,212.7
80-84	500	1,124.1	1,359.5
85+	500	743.5	1,023.7
Total	100,000	100,000.0	100,000.0

<sup>2</sup> Ries LAG, Kosary CL, Hankey BF, Miller BA, Hurray A, Edwards BK (eds). *SEER Cancer Statistics Review 1973-1994*. Bethesda, MD: National Cancer Institute, NIH Pub. No. 97-2789, 1997.

<sup>3</sup> 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.

**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<sub>*j*</sub> category,  $n_j$  = number of cases in that age category, and  $p_j$  = person-years denominator.<sup>4</sup> For many registries, the S.E. of the rates are small, as the population covered is large. However, for registries that cover a small population, the S.E. may be substantial.

**Comparison of Rates.** The S.E. 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  $S.E._1$  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  $x = (R_1 - R_2) / \sqrt{(S.E._1)^2 + (S.E._2)^2}$  and  $z = 1.96$  for 95% confidence limits.<sup>5</sup> 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 10 cases, and it should not be used for rates based on fewer than six cases.

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.

While it is possible to compare incidence rates among populations in various individual registries, it is important to consider whether the groups are comparable by race (*i.e.*, *percent other races* or *percent unknown race* from Section II). One should also consider the registries' quality and completeness as differences can be related to both true underlying differences and differences in reporting completeness and data quality. Interpretation without consideration of these factors may contribute to misleading or inaccurate conclusions.

**Combined Rates for the United States, Canada, and North America.** For inclusion in the combined rates, registries had to meet the following four criteria:

The first criterion was that data for all five years, 1993 to 1997, were submitted.

The second criterion was that a registry had to complete and submit the results for the *Protocol for Assessing Duplicate Cases*, developed by the Data Evaluation and Publication Committee. If a registry had an estimate of duplicates that exceeded one per 1,000 records or 0.1 percent, given its required sample size, the registry was

---

<sup>4</sup>Breslow NE and Day NE. *Statistical Methods in Cancer Research, vol. II*, Lyon, France: IARC, 1987, p. 59.

<sup>5</sup>Parkin DM, Muir CS, Whelan SL (eds). *Cancer Incidence in Five Continents, Volume VI*. Lyon, France: International Agency for Research on Cancer, IARC Scientific Pub No 120, 1992, p.869.

not eligible for inclusion in the combined rates. The estimate for duplicate reports was also used to adjust the estimate of the completeness of case ascertainment described below.

The third criterion for inclusion was that a registry had to run all of its case records for 1993 to 1997 against the *Call for Data* metafile prepared for the EDITS software and to have made all corrections. This metafile identified records that did not pass all intra-field and inter-field edits on data elements used in preparation of the monograph. All identified errors and warnings had to be resolved or reconciled before the final data file was submitted.

The fourth criterion for inclusion was the completeness of case ascertainment. This was computed using incidence-to-mortality rate ratios for whites that were standardized for age, sex, and site of primary tumor. Thus, the estimates were calculated using the following formulae:

$$\text{Expected } I_s = (M_s) \left( \frac{I_{SEER}}{M_{U.S.}} \right)$$

$$\text{Expected } T_s = \sum_{i=1}^2 \sum_{j=1}^N \text{Expected } I_s$$

where  $I$  = Age-adjusted (1970 U.S.) incidence rate in gender  $i$  and site  $j$  category for 1993 to 1997;  $M$  = Age-adjusted (1970 U.S.) mortality rate in gender  $i$  and site  $j$  category for 1993-1997;  $s$  = State, SEER area, province or territory;  $SEER$  = Combined nine historical areas in SEER program;<sup>6</sup>  $U.S.$  = United States;  $T$  = Age-adjusted (1970 U.S.) incidence rate for total sites.<sup>7</sup>

The percent completeness was calculated by dividing the observed (or reported) age-adjusted (1970 U.S.) incidence rates for total sites<sup>7</sup> for both genders (Observed  $T_s$ ) by the expected age-adjusted (1970 U.S.) incidence rate for total sites for both genders (Expected  $T_s$ ).

This method of estimating completeness assumes that site-specific incidence-to-mortality rate ratios are relatively stable within sex and race categories. The incidence-to-mortality rate ratio standard to which all registries were adjusted was the ratio of the combined cancer incidence rates among white persons for the 9 historical SEER areas<sup>6</sup> for 1993 to 1997 to the U.S. mortality rates in whites during the same time interval. This method is the NAACCR standard for estimating completeness of case ascertainment for the U.S. and Canada. The standard rate ratio is applied to death rates in whites for all registry jurisdictions except Hawaii and Canada, where the standard is applied to the entire population (all races combined). The District of Columbia, a predominantly black population, is the sole exception where the incidence-to-mortality rate ratio is based on SEER incidence and U.S. mortality rates for blacks as the standard.

---

<sup>6</sup>Excludes San Jose/Monterey and Los Angeles, California.

<sup>7</sup>The cancer sites included in this calculation were oral cavity, esophagus, stomach, colorectum, liver, pancreas, lung, melanoma, cervix, uterus, ovary, bladder, kidney, nervous system, Hodgkin's Disease, non-Hodgkin's lymphoma, myeloma, and leukemia. Cancers of the breast and prostate were not included due to the lack of stability in the incidence-to-mortality rate ratios caused by the dramatic increase in screening for these diseases in some regions.

Based on the percent completeness estimate, expected number of cancer cases for all sites (C) in the registry was computed as follows:

$$\text{Expected } C_s = \frac{\text{Observed } C_s}{\% \text{ Completeness}}$$

where  $C$  = cancer counts for all sites;  $s$  = State, SEER area, province or territory.

The number of duplicate records in the data set was calculated using the NAACCR estimate of duplicates, based on the registry's results from completing the *Protocol for Assessing Duplicate Cases*. The NAACCR completeness estimate was further adjusted for duplicates using the following equation:

$$\text{Adjusted \% completeness} = \left( \frac{\text{Observed } C_s - D_s}{\text{Expected } C_s} \right) \times 100\%$$

Where  $\text{Observed } C_s$  = number of cancer cases for all sites in the registry;  $\text{Expected } C_s$  = estimated number of cancer cases for all sites if completeness is 100%;  $D_s$  = number of duplicate records. For registries that did not complete the *Protocol for Assessing Duplicate Cases*, the NAACCR adjusted estimate for completeness is omitted from the registry description in Section II.

Every registry included in the combined rates had an adjusted completeness estimate of at least 90 percent. The computed completeness estimate for all registries included in the combined incidence rates for the U.S. was about 101 percent; for Canada, about 96 percent; and for North America, about 100 percent.

In the United States, 28 registries (23 states and 5 metropolitan areas in the SEER program) met all the criteria for inclusion in the U.S. combined rates. These were Arizona, California, the Greater Bay Area and Los Angeles in California, Colorado, Connecticut, Delaware, Florida, Metropolitan Atlanta in Georgia, Hawaii, Idaho, Illinois, Iowa, Kentucky, Louisiana, Metropolitan Detroit in Michigan, Minnesota, Montana, Nebraska, New Hampshire, New Jersey, New Mexico, North Carolina, Rhode Island, Utah, Seattle/Puget Sound in Washington, West Virginia, and Wisconsin. Since both California and its SEER areas of Los Angeles and Greater Bay Area were qualified for inclusion, only the state (California) data were included in the calculation of combined rates.

In Canada, six registries met all the criteria for inclusion in the Canadian combined rates. These were Alberta, British Columbia, Manitoba, New Brunswick, Prince Edward Island and Saskatchewan.

Although all cases from qualified registries were included in computation of combined rates, incidence rates are suppressed in the summary tables of selected major cancer sites if they had fewer than six cases. However, no suppression was used in the tables of pediatric incidence since data were not presented for individual registries.

## Data Interpretation

**Race-specific Rates.** Data are presented for all races and for whites and blacks when the number of cases was sufficiently large to produce stable rates. Race-specific (either white or black) incidence rates are presented only for those registries with both an annual population of 200,000 and a total of 2,500 cancer cases for the specified race in Sections II and III. Cases of unknown race are included in the all races category. Due to limitations and misclassification in both case counts and population estimates for races other than white or black, no race-specific rates

are presented for any other race groups for individual registries. As noted above, two new tables presenting the top five cancers among eight race/ethnic groups have been added to Section IV. Canadian data are presented for all races only, as are the data for Hawaii and Massachusetts. To facilitate comparisons among registries of similar race distribution, the proportion of *unknown race* and *races other than black or white* is reported in the Registry Description at the beginning of each registry's section.

**Percent DCOs.** The proportion of cases identified by Death Certificate Only has been used as a rough guide to assess completeness of case ascertainment. Only invasive cancer cases are included in the denominator used to calculate this rate. However, this statistic can be artificially lowered by intensive follow-back to identify the incident report for a death certificate only case. Many registries do not distinguish between DCOs following routine, pro-active case finding and DCOs following intensive follow-back.

For many new registries, DCO follow-back is postponed until the registry has matured and it has become reasonable that a DCO case had a diagnosis date following the onset of the registry program. This decision may be related to the expense and difficulty in conducting death clearance programs. DCOs may inflate cancer incidence during the first several years of registry operations. The data for registries that did not use death certificates as a source for case ascertainment in 1993 to 1997 are indicated in the Registry Description at the beginning of each registry section.

**Percent of Microscopic Confirmation.** The proportion of total cases with microscopic confirmation can also be used as an indicator of the quality of data collection. SEER registries ranged from 92 to 96 percent microscopic confirmation of all reported cases. A rate that is higher or lower could suggest problems in case ascertainment and registry quality. However, the proportion varies by cancer site. For sites that rely mostly on a clinical diagnosis, e.g., cancers of the pancreas and brain, confirmation rates that are too high may suggest that the clinically-diagnosed cases are missing. Also, registries that do not use death certificates for case finding may have a higher proportion of microscopically confirmed cases and yet not be the most complete in case ascertainment.

#### **Values of Zero ("0") and "-" in the Tables**

When the incidence rate or count for certain sex- race- and site-specific group is 0.0 or 0, this indicates either only a few cases existed in the specific group, producing a rate less than 0.05 per 100,000, or that no cases were reported for that category.

"-" is used when the data were not available or when the category was not applicable. An example would be breast *in situ* when a registry did not provide the data. Another example would be incidence rates for female cancers in the columns for males, and *vice versa*.

"-" is also used when rates for individual registries in the combined rates section were not computed due to case counts of less than six for the specified cancer sites and both count and the rate were suppressed.

**Comparisons Among Registries.** All registries responding to the 2000 *Call for Data* are included in Sections II and III. In making valid comparisons among registries, it is important to review the data quality indicators for each registry before attributing rate differences to regional variation. Data quality can be an important contributor to observed differences in rates. Selected site-specific rates from the registries of highest quality are included in the tables in Section IV. 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.<sup>8</sup>

## **NAACCR BACKGROUND**

Founded in 1987, the North American Association of Central Cancer Registries (NAACCR) is a professional association of cancer registries, programs, and organizations and of individuals interested in the development and application of cancer registration. In April 1994, the name of the organization was officially changed from the American Association of Central Cancer Registries (AACCR) to the North American Association of Central Cancer Registries by vote of the membership to include Canadian registries and others throughout North America. NAACCR promotes the use of central cancer registry data in studies of defined populations and in cancer control programs in the United States and Canada. NAACCR welcomes membership from all cancer registries interested in collaborative efforts to collect, analyze, and publish data in defined populations and to use registry information in identifying target populations for cancer control interventions and in conducting surveillance of cancer incidence and evaluation of cancer control programs. Membership is also desirable for national organizations involved in cancer control, prevention, and research and other organizations and individuals interested in cancer registration or cancer control activities.

### **Data Evaluation and Publication Committee Members, 1999-2000**

Vivien W. Chen, Louisiana, *Chair*  
Susan Carozza, Texas  
John P. Fulton, Rhode Island  
Robert Greenlee, American Cancer Society  
Irene Hall, Centers for Disease Control & Prevention  
Carol Kosary, National Cancer Institute  
Yang Mao, Laboratory Centre for Disease Control, Health Canada  
Kathleen McDavid, Centers for Disease Control & Prevention  
Ghislaine Villeneuve, Statistics Canada  
Richard S. Porter, Arizona  
Steven Roffers, Georgia  
David Roney, Information Management Services, Inc.  
Andrew K. Stewart, American College of Surgeons  
Nancy S. Weiss, Texas  
Dee West, California  
Xiao Cheng Wu, Louisiana  
Holly L. Howe, *ex officio*, Executive Director, NAACCR  
Joellyn Hotes, NAACCR Staff

### **Data Evaluation and Publication Subcommittees, 1999-2000**

#### ***Comparative Analysis of Incidence Rates Subcommittee***

Sue Carozza, Texas, *Chair*  
Vivien W. Chen, Louisiana  
John P. Fulton, Rhode Island  
Robert Greenlee, American Cancer Society

---

<sup>8</sup>Chen, VW. Should we or shouldn't we compare cancer incidence rates among registries? in Howe HL (ed.) *Cancer Incidence in North America, 1988-1991*. Sacramento, CA: North American Association of Central Cancer Registries, April 1995, V-1 to V-6.

Mary D. Hutton, Centers for Disease Control and Prevention  
Judith B. Klotz, New Jersey  
Lihua Liu, California  
Brooke Steele, Centers for Disease Control and Prevention  
Xiao Cheng Wu, Louisiana

***Analysis of Data Quality Indicators Subcommittee***

John P. Fulton, Rhode Island, Chair  
Catherine N. Correa, Louisiana  
Nancy Jackson, California  
Chris Newton, Arizona  
Lilia O'Connor, California  
Pam Parrish, Illinois  
Carin Perkins, California  
Steven D. Roffers, Georgia  
Gayle Clutter, Centers for Disease Control and Prevention

***Canadian/U.S. Comparison of Cancer Statistics Subcommittee***

Kathleen McDavid, Centers for Disease Control and Prevention, Co-Chair  
Cyril Nair, Statistics Canada, Co-Chair  
Vivien W. Chen, Louisiana  
Glenn Copeland, Michigan Cancer Surveillance Program  
John Fulton, Rhode Island  
Rue-Nie Gao, Statistics Canada  
Eric Holowaty, Ontario  
Joelilyn Hotes, NAACCR  
Carol Kosary, National Cancer Institute  
Judy Lee, Statistics Canada  
Yang Mao, Laboratory Centre for Disease Control, Health Canada  
Mary McBride, British Columbia Cancer Agency  
Richard Porter, Arizona Cancer Registry  
Brenda Smith, Arizona Cancer Registry  
Jon Tonita, Saskatchewan Cancer Agency  
Ghislaine Villeneuve, Canadian Cancer Registry  
Lydia Voti, Florida Cancer Data System  
Brad Wohler-Torres, Florida Cancer Data System

***Public Use Data Sets Subcommittee***

Holly L. Howe, NAACCR, Chair  
Vivien W. Chen, Louisiana  
John P. Fulton, Rhode Island  
Carol Kosary, National Cancer Institute  
Sarah Landis, American Cancer Society  
Melinda Lehnher, Illinois  
Richard S. Porter, Arizona  
David Roney, Information Management Services, Inc.  
Hannah Weir, Centers for Disease Control & Prevention

***Monograph Production Subcommittee***

Vivien W. Chen, Louisiana, *Chair*  
Catherine N. Correa, Louisiana  
Joellyn Hotes, NAACCR  
Holly L. Howe, NAACCR  
Stephen Keith, Information Management Services, Inc.  
Carol Kosary, National Cancer Institute  
David Roney, Information Management Services, Inc.  
Xiao Cheng Wu, Louisiana

***Collaborative Research Working Group***

Vivien W. Chen, Louisiana, *Chair*  
Brenda Edwards, National Cancer Institute  
Holly L. Howe, NAACCR  
Betsy Kohler, New Jersey  
Kathleen M. McKeen, Iowa  
Lynn Ries, National Cancer Institute  
Thomas C. Tucker, Kentucky  
Hannah Weir, Centers for Disease Control and Prevention  
Dee West, California  
Phyllis Wingo, American Cancer Society

***Cancer Staging Work Group***

Steven D. Roffers, Georgia, *Chair*  
Mohammed Ahmed, Louisiana  
Patricia A. Andrews, Louisiana  
Vivien W. Chen, Louisiana  
Catherine N. Correa, Louisiana  
Gayle Clutter, Centers for Disease Control & Prevention  
Dan Curran, California  
April Fritz, National Cancer Institute  
Jennifer Jackson, Louisiana  
Nancy Jackson, California  
Carin Perkins, California  
Lynn Ries, National Cancer Institute  
Xiao Cheng Wu, Louisiana

***CINA Plus Online Subcommittee***

Holly L. Howe, NAACCR, *Chair*  
Vivien W. Chen, Louisiana  
Eric Durbin, Kentucky  
Kate Marie, Cancer Surveillance and Control Program/ Public Health Institute  
Long On, Health Canada  
Adam Probert, Health Canada  
David Roney, Information Management Services, Inc.  
Brian Serafini, Kentucky

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 62704-6495  
(217) 698-0800  
(217) 698-0188 (FAX)

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

April 2000