

# Sample Text for Describing CDC's National Program of Cancer Registries (NPCR) and NCI's Surveillance, Epidemiology, and End Results (SEER) Program Incidence – U.S. Cancer Statistics Data Methods in Manuscripts

---

## Methods

### Cancer Incidence

Data about cancer incidence in this report come from the two federally funded population-based source of cancer cases in the United States, the Centers for Disease Control and Prevention's (CDC's) National Program of Cancer Registries (NPCR) dataset and the National Cancer Institute's Surveillance, Epidemiology, and End Results Program dataset (1). This dataset includes cancer incidence data from central cancer registries reported to NPCR in 46 states, the District of Columbia, and [IF APPLICABLE] Puerto Rico (2) and to SEER in 4 states. Data about all new diagnoses of cancer from patient records at medical facilities such as hospitals, physicians' offices, therapeutic radiation facilities, freestanding surgical centers, and pathology laboratories are reported to central cancer registries, which collate these data and use state vital records to collect information about any cancer deaths that were not reported as cases. The central cancer registries use uniform data items and codes as documented by the North American Association of Central Cancer Registries. These data are submitted annually to CDC and NCI and combined into one dataset (3). Cancer registries demonstrate that data were of high quality by meeting U.S. Cancer Statistics publication criteria (1); during [YEARX–YEARY], data from [X] cancer registries met these criteria, covering [X%] of the United States population. This report includes new cases of primary invasive [CANCER TYPE] cancer (International Classification of Diseases for Oncology, Third Edition code [CXX.X–CXX.X]) (4) diagnosed during [YEARX–YEARY]; [IF APPLICABLE] excluding histology codes 9050–9055, 9140, and 9590–9992 [OR] restricted to histology codes [XXXX–XXXX].

### [IF APPLICABLE] Race and Ethnicity

Data were analyzed by five major racial/ethnic groups: white, black, American Indian and Alaska Native (AI/AN), Asian/Pacific Islander (A/PI), and Hispanic. Information about race and Hispanic ethnicity were collected separately. An algorithm was applied to Hispanic ethnicity data to reduce misclassification of Hispanic persons as being of unknown ethnicity (5). To reduce misclassification of AI/AN race, some central cancer registries link case data with the Indian Health Service (IHS) patient registration database, which contains records of individuals who are members of federally recognized tribes; cases linked with the IHS database were coded as AI/AN (6).

Because states can opt not to present state-specific counts and rates for [AS APPLICABLE: A/PI, Hispanic, and AI/AN populations], these data are not shown for the following states [CHECK STATE LIST AT [www.cdc.gov/cancer/npcr/uscs/technical\\_notes/interpreting/race.htm](http://www.cdc.gov/cancer/npcr/uscs/technical_notes/interpreting/race.htm)]. FOR EXAMPLE, Because states can opt not to present state-specific counts and rates for AI/AN populations, these data are not shown for Delaware, Illinois, Kansas, Kentucky, New Jersey, and New York.]

### [IF APPLICABLE] Histology

Analyses by histology included only cases that were microscopically confirmed ([X%] of cases).

### [IF APPLICABLE] Stage

Stage was classified using a merged variable that spanned the two time periods when two different staging schemes were used, the SEER Summary Stage 2000 for cases diagnosed 2001–2003 and the Derived Summary Stage 2000 for cases diagnosed in 2004 or later. [IF APPLICABLE: SEER Summary Stage 2000 for cases diagnosed between 2001 and 2003.] OR [IF APPLICABLE: Stage was classified using Derived Summary Stage 2000 for cases diagnosed in 2004 or later.] OR [IF APPLICABLE: Stage was classified using a variable that combined SEER Summary Stage 2000 (for cases diagnosed from 2001 to 2003) and Derived Summary Stage 2000 (for cases diagnosed in 2004 or later).] The staging criteria characterize cancers as localized, regional, distant, or unknown stage; localized cancer is confined to the primary site; regional cancer has spread directly beyond the primary site (regional extension) or to regional lymph nodes; and distant cancer has spread to other organs (distant extension) or remote lymph nodes (7). Analyses by stage excluded cases that were diagnosed only by death certificate or autopsy ([X%] of cases).

## Population Estimates

Population estimates for rate denominators were a modification of annual county population estimates by age, sex, bridged race, and ethnicity produced by the U.S. Census Bureau in collaboration with CDC and with support from NCI (8). Modifications incorporated bridged, single-race estimates that were derived from multiple-race categories in the Census and accounted for known issues in certain counties (8). The modified county-level population estimates, summed to the state and national levels, were used as denominators in rate calculations (8).

## Statistical Analysis

### Incidence and Death Rates

Average annual rates for [YEARX–YEARY] per 100,000 population were age-adjusted (using 19 age groups) by the direct method to the 2000 U.S. standard population (9). Corresponding 95% confidence intervals (CIs) were calculated as modified gamma intervals (10). Rates based on fewer than 16 cases tend to have poor reliability and were not presented. To determine differences between subgroups, rate ratios were calculated; rates were considered statistically different if the 95% CIs of the rate ratios excluded 1 (11). Rates were calculated using SEER\*Stat software version X.X.X (12).

### [IF APPLICABLE] Trends in Rates

Annual percentage change (APC) was used to quantify the change in rates during [YEARX–YEARY] and was calculated using weighted least squares regression (13). A two-sided *t*-test was used to test whether the APC was statistically different from zero ( $P < .05$ ). Rates were considered to increase or decrease if  $P < .05$ ; otherwise rates were considered stable. APCs were calculated using SEER\*Stat software version X.X.X (12).

[OR]

Change in rates during [YEARX–YEARY] was calculated using joinpoint regression, which involves fitting a series of joined straight lines on a logarithmic scale to the trends in the annual age-standardized rates (14); up to [X] joinpoints ([X] line segments) were allowed. The trend of the line segment was used to quantify the annual percentage change (APC). A two-sided *t*-test was used to test whether the APC was statistically different from zero ( $P < .05$ ). The average annual percentage change (AAPC) for [YEARX–YEARY] was calculated using a weighted average of the slope coefficients of the underlying joinpoint regression line with the weights equal to the length of each segment over the interval. To determine whether the AAPC was statistically different from zero ( $P < .05$ ), a two-sided *t*-test was used for 0 joinpoints, and a two-sided *z*-test was used for 1 or more joinpoints. Rates were considered to increase or decrease if  $P < .05$ ; otherwise rates were considered stable. Trends were calculated using Joinpoint regression program version X.X.X (15).

## FOOTNOTES for Tables

It is recommended that standard footnotes from U.S. Cancer Statistics or slight derivations be used for tables and figures.

### [FOR POPULATION COVERAGE]

Data are from population-based registries that participate in CDC's National Program of Cancer Registries and/or NCI's Surveillance, Epidemiology, and End Results Program and meet high-quality data criteria. These registries cover approximately [XX]% of the United States population.

### [FOR AGE-ADJUSTED RATES]

Rates are per 100,000 persons and are age-adjusted to the 2000 U.S. standard population (19 age groups – Census P25–1130).

## References

- <sup>1</sup>U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999–[YEAR]. Incidence and mortality Web-based report. Atlanta, GA: U.S. Department of Health and Human Services, CDC, National Cancer Institute; [YEAR]. Available at <https://nccd.cdc.gov/uscs/>.
- <sup>2</sup>Singh SD, Henley SJ, Ryerson AB. Summary of notifiable noninfectious conditions and disease outbreaks: surveillance for cancer incidence and mortality—United States, 2012. *MMWR* 2016;63(55):17–58. Available at [www.cdc.gov/mmwr/volumes/63/wr/mm6355a4.htm](http://www.cdc.gov/mmwr/volumes/63/wr/mm6355a4.htm).
- <sup>3</sup>National Program of Cancer Registries and Surveillance, Epidemiology, and End Results SEER\*Stat Database: [ENTER DATABASE TITLE], United States Department of Health and Human Services, Centers for Disease Control and Prevention. Released [DATE], based on the November [YEAR] submissions. Available at [www.cdc.gov/cancer/public-use](http://www.cdc.gov/cancer/public-use).
- <sup>4</sup>Fritz A, Percy C, Jack A, Shanmugarathnam K, Sobin L, Parkin D, et al., editors. *International Classification of Diseases for Oncology*. 3rd edition. Geneva, Switzerland: World Health Organization; 2000. Available at [www.who.int/classifications/icd/adaptations/oncology/en/](http://www.who.int/classifications/icd/adaptations/oncology/en/).
- <sup>5</sup>NAACCR Race and Ethnicity Work Group. *NAACCR Guideline for Enhancing Hispanic/Latino Identification: Revised NAACCR Hispanic/Latino Identification Algorithm [NHIA v2.2.1]*. Springfield (IL): North American Association of Central Cancer Registries. September 2011. Available at [www.naacr.org/wp-content/uploads/2016/11/NHIA\\_v2\\_2\\_1\\_09122011.pdf](http://www.naacr.org/wp-content/uploads/2016/11/NHIA_v2_2_1_09122011.pdf).
- <sup>6</sup>Jim MA, Arias E, Seneca DS, Hoopes MJ, Jim CC, Johnson NJ, Wiggins CL. Racial misclassification of American Indians and Alaska Natives by Indian Health Service Contract Health Service Delivery Area. *American Journal of Public Health* 2014;104:S295–S302. Available at [www.ncbi.nlm.nih.gov/pmc/articles/PMC4035863/](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4035863/).
- <sup>7</sup>Young JL Jr, Roffers SD, Ries LAG, Fritz AG, AA H, editors. *SEER Summary Staging Manual - 2000: Codes and Coding Instructions*. Bethesda, MD: National Cancer Institute; 2001.
- <sup>8</sup>National Cancer Institute. Surveillance, Epidemiology, and End Results (SEER) Program. Population Estimates Used in NCI's SEER\*Stat Software. Available at <https://seer.cancer.gov/popdata/methods.html>.
- <sup>9</sup>Anderson R, Rosenberg H. Age standardization of death rates: implementation of the year 2000 standard. *National Vital Statistics Report* 1998;47:1–16. Available at [www.ncbi.nlm.nih.gov/pubmed/9796247](http://www.ncbi.nlm.nih.gov/pubmed/9796247).
- <sup>10</sup>Tiwari RC, Clegg LX, Zou Z. Efficient interval estimation for age-adjusted cancer rates. *Statistical Methods in Medical Research* 2006;15:547–569. Available at [www.ncbi.nlm.nih.gov/pubmed/17260923](http://www.ncbi.nlm.nih.gov/pubmed/17260923).
- <sup>11</sup>Fay MP. Approximate confidence intervals for rate ratios from directly standardized rates with sparse data. *Communications in Statistics: Theory and Methods* 2007;28(9):2141–2160.
- <sup>12</sup>National Cancer Institute. SEER\*Stat software. Bethesda, MD: National Cancer Institute, Surveillance Research Program; [YEAR]. Available at <https://seer.cancer.gov/seerstat/>.
- <sup>13</sup>Kleinbaum DG, Kupper LL, Muller KE. *Applied Regression Analysis and Other Multivariable Methods*. 2nd ed. Boston, Mass: PWS-Kent; 1988.
- <sup>14</sup>Kim H-J, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. *Statistics in Medicine* 2000;19:335–351. Available at [www.ncbi.nlm.nih.gov/pubmed/10649300](http://www.ncbi.nlm.nih.gov/pubmed/10649300).
- <sup>15</sup>National Cancer Institute. Joinpoint regression program. Bethesda, MD: National Cancer Institute, Surveillance Research Program, Statistical Methodology and Applications Branch; [YEAR]. Available at <https://surveillance.cancer.gov/joinpoint/>.