# TECHNICAL NOTES

## DATA SOURCES

**Cancer Incidence Data:** Cancer incidence data are from the Illinois Department of Public Health, Illinois State Cancer Registry (ISCR), the only source of population-based cancer incidence data for the state. Newly diagnosed cancer cases among Illinois residents are reported to ISCR by the health care facilities where the cancer is diagnosed and treated. Central cancer registries and facilities in other states also report data to ISCR on Illinois residents diagnosed and treated for cancer in their states. ISCR has agreements with the following central cancer registries to exchange cancer data: Arkansas, California, Florida, Indiana, Iowa, Kentucky, Michigan, Minnesota (Mayo Clinic, through October 2005), Mississippi, Missouri, Mississippi (through 2004), North Carolina, Washington, Wisconsin, and Wyoming (through February 2008).

Most out-of-state cases come from Florida, Indiana, Iowa, Kentucky, Minnesota, Missouri and Wisconsin. Completeness of out-of-state reporting depends upon the years of operation of these other central registries, the extent of their identification of out-of-state residents, and their standards of quality.[1] For these data, 5.4 percent of ISCR cases are reported from out-of-state agencies and organizations. However, two states did not report all cancer cases among Illinois residents diagnosed in 2009. This had little impact on the overall cancer incidence rate for Illinois but the rates among some rural counties along the border of Illinois may be affected.

Additionally, a death certificate clearance process has been employed since August 1993. The process involves active follow back of cancer deaths in an effort to identify the cases that are not reported to ISCR. For these data, 1.6 percent of reported cases were identified from death certificate clearance.

The preparation and release of these data are dependent on the completion of annual reporting by Illinois facilities. Although case reporting is mandated within six months of diagnosis, it has been the ISCR policy to keep database files open for late reporting of cases and to allow for the two-to four-year lag in case identification of Illinois residents from other state central cancer registries. This practice is consistent with data published nationally. For these data, the database files reflect the status of ISCR as November 2011.

**Cancer Mortality Data:** The SEER program of the National Cancer Institute (NCI) was the source of information on cancer mortality for 1986 through 2009 by race and ethnicity. The underlying cause of death was provided by the National Center for Health Statistics (NCHS). Data presented were released in April 2012.

**Data Use Agreement:** By using these data, you signify your agreement to comply with the following statutorily based requirements.

The Illinois Health and Hazardous Substances Registry Act (410 ILCS 525/12) provides data collected by the Illinois State Cancer Registry be made available to the public, however the identification or contact of individuals is prohibited.

In an effort to exclude identifying information on individual patients, the data (e.g. age, race, Hispanic ethnicity, year of diagnosis and type of cancer) have been aggregated into categories within individual records, the number of which depends on the size of the geographic area.

These data are provided as a public service for the purpose of statistical reporting and analysis only. There should be no attempt to learn the identity of any person included in these data. If the identity of any person is discovered inadvertently, no disclosure or other use of the identity will be made.

Uses of these data do not constitute an endorsement of the user's opinion or conclusions by the Department and none should be inferred.

**Population Estimates:** The population estimates of the sex- and race-specific, as well as sex- and ethnicity/race-specific groups, in five-year age categories were used as denominators for rate calculations in these data. These population estimates of Illinois and Illinois counties for all races, whites, blacks and Asian/other races from 1986 through 2009 and for Hispanics and non-Hispanics for 1990 through 2009, were obtained from the SEER program based on United States Bureau of Census population estimates.[2]

At the time these data were made public in September 2012, U.S. population data reflecting the 2010 census were not available. It is anticipated that the census 2010 population data will be available in late summer of 2012, at which point, these data will be updated. This data release uses the Vintage 2009 population estimates in the rate calculations.

The U.S. Census Bureau revised their population estimation methodology for Vintage 2009. The Vintage 2009 population estimates used in this report incorporate four improvements in methodology: 1) changes in the estimation of net internal migration, 2) changes in the estimation of the distribution of deaths to people aged 70 and older by age, sex, race, and Hispanic origin, 3) changes in the estimation of domestic migration of the population 65 years and older, and 4) changes in the estimation of the age distribution of migrations to and from counties. These changes are in addition to changes made in Vintage 2008 and Vintage 2007.[3] The net impact of these changes was a downward shift in the latest post-censal population estimates covering 2000 through 2009. In general, this shift caused a small upswing in rates.[4]

Illinois population estimates from Vintage 2008 data for 2003 through 2007 were compared to those produced using the Vintage 2009 data for the same time period. Use of new methods in the Vintage 2009 file resulted in a 0.2 percent decrease in Illinois' total population. The new Vintage 2009 methods put Illinois' cancer rate 0.8 percent higher than what was calculated previously using Vintage 2008. Differences in incidence rates may not solely be the result of changes in population estimates. Changes in rates also could be attributable to the addition of cases reported late. [4]

## DEFINITIONS

**Cancer Site Coding for Incidence Data:** Although the anatomic site and morphology for cancer cases diagnosed prior to 2001 were coded using the *International Classification of Diseases for Oncology* version 2 (ICD-O-2)[5] and for cancer cases diagnosed in 2001 through 2009 the version 3 (ICD-O-3),[6] all ICD-O-2 coded cases were converted to version 3 codes and grouped according to the site group definitions established by the SEER program of NCI and also used by NAACCR and CDC (for complete listing please see Appendix B of the state report (see Illinois State Cancer Incidence and Mortality Review and Update, 1986-2009). Data at a state level are available on request for these site groups.

Data at the state and county level in this application are aggregated into 24 major groups and at the ZIP codes level into 11 groups. These standardized classification schemes allow direct comparisons of Illinois data with international, national and state publications.

Several definitional changes occurred in some histologies and behaviors in ICD-O-3 that affected the inclusion and exclusion of reportable cancers diagnosed beginning in 2001. These changes may affect the comparability of data between rates prior to 2001 and 2001 or later. The changes predominately affected leukemias, lymphomas, and cancer of the ovary. Several cancers that previously were not coded as malignant in ICD-O-2 are coded as malignant in ICD-O-3. For example, Myolodysplastic syndrome (MDS) and chronic myeloproliferative disease (CMPD) are considered malignant cancer in ICD-O-3, as are papillary ependymomas and papillary meningiomas which, according to ICD-O-3, are included in the "Brain and Nervous System" and "All Cancers Combined" categories. Some endometrial tumors also are classified as malignant in ICD-O-3. Conversely, some low malignant potential tumors of the ovary and pilocytic astrocytomas are no longer coded as malignant in ICD-O-3. Overall, these changes would have a slight impact on incidence of a specific cancer site; however, it might result in a noticeable increase in cancer incidence rates for "All Cancers Combined."

In addition, beginning with data reported in 2005, both Kaposi sarcoma and mesothelioma are classified as separate site groups within the SEER recode. This change has a slight impact on cancer incidence rates for a few specific cancers, compared to using the previous site grouping method.

Counts and rates were calculated for invasive cancers only, with the exception of *in situ* cancer of the bladder. Although counts and rates for breast cancer *in situ* are displayed in a separate table, these cases were not included in any counts or rates of all sites combined incidence.

Confidentiality of the incidence data is maintained by aggregating data within individual records into categories, the number of which depend on the size of the geographic area. Individual year of diagnosis is available for the Illinois data files, however, for the county and ZIP code files, the diagnosis year is a five-year aggregate (1990-1994, 1995-1999, 2000-2004 and 2005-2009).

**Cancer Site Coding for Mortality Data.** Underlying cause of death was coded using the *International Classification of Diseases* (ICD-9)[7] for all deaths for years 1986 through 1998 and the *International Classification of Diseases* (ICD-10)[8] for all deaths for year 1999 and later. In the present data, the SEER mortality recode scheme based on ICD-9 and ICD-10 was used to classify cancer deaths sites.

Because of many changes in ICD-10 as compared to ICD-9, discontinuities in trends for many causes of death, including cancer, may arise. According to a study, compared to using ICD-9 coding, overall, approximately 0.7 percent more deaths are assigned to cancer when ICD-10 is used, leading to a higher mortality rate for all cancers combined.[9] But this pattern does not hold for specific cancer sites, whose rates may be higher or lower using ICD-10. These discontinuities are relatively small, and the changes in mortality rates across the years of the ICD-9/ICD-10 boundary are still interpretable, especially for major cancer sites.[10]

Cancer mortality rates are available by single year for Illinois only. Deaths among non-residents and deaths of unknown sex or age were omitted from all calculations. Due to NCHS policy, statistics were not calculated for cells containing less than 10 deaths.

**Incidence and Mortality Rates:** The SEER*Stat® software package,[11] developed by the Information Management Services Inc. for NCI, was used to calculate both incidence and mortality rates. Rates are expressed per 100,000 population. Age-adjustment of rates was calculated by the direct method adjusting to the 2000 U.S. standard million population. Rates are rounded to the nearest 10$^{th}$ and very small rates (e.g., 0.04) are shown as 0.0. They are presented with the lower and upper confidence intervals computed at the 95 percent level using Tiwari method.[12] Algorithms used for the calculation of standard errors and 95 percent confidence intervals are displayed in Appendix D of the state incidence report (see Illinois State Cancer Incidence and Mortality Review and Update, 1986-2009). As mentioned above, due to NCHS policy, mortality statistics were not calculated for cells containing less than 10 deaths. This is not the case for incidence rates associated statistics.

**Race-specific Rates**

**Incidence:** At the state level, the race-specific categories include "White," "Black," and "Asian/Other Races." Cases reported as unknown race were included in the "All races" category, but not in any race-specific group. At the county level, race-specific data are available for whites and blacks for 15 counties (Champaign, Cook, DuPage, Kane, Kankakee, Lake, Macon, Madison, Peoria, Rock Island, Sangamon, St. Clair, Vermilion, Will and Winnebago) that have sufficiently large black population to allow meaningful statistics for the race groups. No race-specific data are available at the ZIP code level.

To improve the identification and surveillance of American Indians and Alaska Natives diagnosed with cancer and to be consistent with the national data, cancer incidence data since 1995 were linked to the Indian Health Service (IHS), which provides medical services to an estimated 55 percent of the American Indian/Alaska Native

population.[13] If a race code in the ISCR database is white, black, other, or unknown and the IHS link is positive, then the race code is re-categorized to American Indian/Alaskan Native; otherwise the race code stays unchanged. This practice has minimal impact on the incidence rates for whites or blacks due to the small number of cases affected.

**Mortality**: The race-specific categories in these data are “All races” combined, “White.” “Black” and “Asian/Other Races.” Cases reported as unknown race were included in the "All races" category, but not in any race-specific group.

### Ethnicity/Race Rates

**Incidence:** For the incidence data at the state level, Hispanic ethnicity was derived according to the NAACCR Hispanic identification algorithm (NHIA).[14] NHIA is a generally reliable method to enhance the ethnic identification of the Latino population in the United States.[15] In consistency with national or state data, categories are reported as "Hispanic (any race)," "non-Hispanic (any race)," "non-Hispanic White" and "non-Hispanic Black." Cases that meet certain criteria around race and birthplace, and who are also identified as non-Hispanic, Hispanic not otherwise specified, Spanish surname only, and unknown ethnicity are examined. Through the use of race, birthplace, last name, first name and maiden name, NHIA assigns a more specific and sometimes different ethnicity to these cases.

**Mortality:** Hispanic ethnicity was used as defined in the database. Because there were a considerably large number of cancer deaths with unknown Hispanic ethnicity in the mortality database, the mortality rates calculated for Hispanics may be underestimated. To be consistent with all national reports, categories are reported as "Hispanic (any race)," "non-Hispanic (any race)," "non-Hispanic White" and "non-Hispanic Black."

## QUALITY CONTROL

Ongoing quality control procedures are integral components of ISCR operations that assure high quality cancer incidence data. In addition to these activities, in 1997, NAACCR developed a certification process that reviews registry data for completeness, accuracy and timeliness of reporting (starting with cases diagnosed in 1995). Since then, ISCR has submitted data each year to the NAACCR *Call for Data* and for NAACCR registry certification. Based on the certification criteria shown in the following table,[16] ISCR has been awarded gold certification for all diagnosis years from 1996 through 2009.

| Completeness (NAACCR Method) | Pass EDITS | DCO | Timeliness | Unresolved Duplicate | Missing Data Fields | | | | Certification Status |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | Sex | Age | County | Race | |
| ≥ 90% | ≥ 97% | ≤ 5% | Within 23 months | ≤ 2/1000 | ≤ 3% | ≤ 3% | ≤ 3% | ≤ 5% | SILVER |
| ≥ 95% | 100% | ≤ 3% | Within 23 months | ≤ 1/1000 | ≤ 2% | ≤ 2% | ≤ 2% | ≤ 3% | GOLD |

Constantly updating registry data is a standard operation in ISCR. As of November 2011, ISCR quality control data for each diagnosis year are as follow:

| Year | Completeness (NAACCR Method)[a] (% As of 11-10) | Pass EDITS (%) | DCO[b] (%) | Unresolved Duplicate[c] (%) | Missing Data Fields | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Sex (%) | Age (%) | County (%) | Race (%) |
| **1986** | 88 | ~ | ~ | ~ | 0.0 | 0.0 | 0.0 | 0.2 |
| **1987** | 90 | ~ | ~ | ~ | 0.0 | 0.0 | 0.0 | 0.2 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **1988** | 88 | ~ | ~ | 0.04 | 0.0 | 0.0 | 0.0 | 0.3 |
| **1989** | 88 | ~ | ~ | 0.04 | 0.0 | 0.0 | 0.0 | 0.3 |
| **1990** | 89 | 100 | ~ | 0.04 | 0.0 | 0.0 | 0.0 | 0.3 |
| **1991** | 88 | 100 | ~ | 0.04 | 0.0 | 0.0 | 0.0 | 0.5 |
| **1992** | 91 | 100 | ~ | 0.04 | 0.0 | 0.0 | 0.0 | 0.2 |
| **1993** | 92 | 100 | 2.2 | 0.04 | 0.0 | 0.0 | 0.0 | 0.2 |
| **1994** | 97 | 100 | 6.1 | 0.06 | 0.0 | 0.0 | 0.0 | 0.3 |
| **1995** | 100 | 100 | 2.7 | 0.03 | 0.0 | 0.0 | 0.0 | 0.4 |
| **1996** | 100 | 100 | 1.8 | 0.02 | 0.0 | 0.0 | 0.0 | 0.5 |
| **1997** | 100 | 100 | 1.8 | 0.09 | 0.0 | 0.0 | 0.0 | 0.7 |
| **1998** | 100 | 100 | 1.5 | 0.03 | 0.0 | 0.0 | 0.0 | 1.0 |
| **1999** | 100 | 100 | 1.8 | 0.02 | 0.0 | 0.0 | 0.0 | 1.0 |
| **2000** | 100 | 100 | 2.4 | 0.03 | 0.0 | 0.0 | 0.0 | 1.0 |
| **2001** | 100 | 100 | 2.4 | 0.00 | 0.0 | 0.0 | 0.0 | 0.9 |
| **2002** | 100 | 100 | 2.6 | 0.00 | 0.0 | 0.0 | 0.0 | 1.1 |
| **2003** | 100 | 100 | 1.5 | 0.02 | 0.0 | 0.0 | 0.0 | 1.2 |
| **2004** | 100 | 100 | 1.7 | 0.04 | 0.0 | 0.0 | 0.0 | 1.1 |
| **2005** | 100 | 100 | 1.9 | 0.00 | 0.0 | 0.0 | 0.0 | 1.3 |
| **2006** | 100 | 100 | 2.0 | 0.00 | 0.0 | 0.0 | 0.0 | 0.9 |
| **2007** | 100 | 100 | 1.2 | 0.00 | 0.0 | 0.0 | 0.0 | 1.0 |
| **2008** | 100 | 100 | 1.7 | 0.06 | 0.0 | 0.0 | 0.0 | 1.0 |
| **2009** | 100 | 100 | 1.7 | 0.01 | 0.0 | 0.0 | 0.0 | 1.3 |

~ not applicable a. For data prior to1995, the NAACCR's completeness estimating algorithm (version 1) was used. For data on or after 1995, the NAACCR's completeness estimating algorithm (version 2) was used.
b. DCO follow-back not started until end of 1993 reporting year
c. NAACCR's duplicate protocol was run for each year at the time of data submission for registry certification.

Completeness of reporting for the five-year groups in the county and ZIP code files is as follows: 1990-1994 = 91 percent, 1995-1999 = 100 percent, 2000-2004 = 100 percent, 2005-2009 = 100 percent.

## DATA INTERPRETATION

Observed variations and differences over years and across sex and race groups in cancer incidence may be real, reflecting modifications in the risk factor status of the population or the consequence of participation in screening and early detection programs. Such changes or differences, however, may not be real, but instead may be the result of random fluctuations and other factors related to the estimation process. Any conclusions should be made only after carefully considering the following factors that influence annual incidence rates.

Random fluctuations in annual rates are usual and may be substantial, especially for rates based on small numbers of incidence counts (i.e., less than 16).

Differences in registry database completeness and data quality will influence the magnitude of estimated cancer incidence rates. It should be noted that, because years prior to 1994 are less than 95 percent complete (see above table), some rates for those years, especially for all sites combined, would be underestimates of the "true" rates for the Illinois population. The rates presented here have not been adjusted for completeness differences across the database.

Due to the fact that some out-of-state data exchange was incomplete for 2009, rates for some border counties in Illinois, as well as ZIP codes within these counties, should be interpreted with caution. This depends upon the percent of a county's cases that are reported through out-of-state data exchanges. Any changes in these rates as compared with earlier years should be evaluated cautiously.

Population estimates used for denominators may be inaccurate or lack precision. The population data for 1990 and 2000, the years of the U.S. decennial census, are the most accurate for all age-, race-, ethnicity- and sex-specific categories and would, therefore, produce the most accurate incidence and mortality rates. Those for other years are not based on actual population counts but rather on interpolation or extrapolation of estimates based on

demographic characteristics of the population. Incidence rates based on these population estimates would be expected to be less accurate than those for 1990 or 2000.

The 95 percent confidence intervals are included with reported rates to help put the rate in perspective and to facilitate rate comparisons over years and across sex, race and ethnicity/race groups. Observed differences may not be statistically significant. The range between the lower confidence interval and the upper confidence interval defines with 95 percent probability where the "true" rate may fall. The comparison of two sets of confidence intervals is approximately equivalent to statistical significance tests for differences between two rates and is more conservative than the standard significance test when the null hypothesis is true.[17]

**Acknowledgements**

The production of these data would not have been possible without the diligent work of the Department's Illinois State Cancer Registry (ISCR) staff, the personnel at the reporting facilities that diagnose or treat cancer patients throughout Illinois, and the staff members at other state central cancer registries with data exchange agreements.

This work was supported by Cooperative Agreement Number 5U58DP000805-05 from the U.S. Centers for Disease Control and Prevention (CDC) National Cancer Prevention and Control Program. The contents are solely the responsibility of the authors and do not necessarily represent the official views of CDC.

**Suggested citation for Incidence Data**

Please reference the source of these data in any published document as follows: Illinois Department of Public Health, Illinois State Cancer Registry, public data set v19, 1986-2009, data as of November 2011.

**Suggested citation for Mortality Data**

Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Mortality - All COD, Aggregated With State, Total U.S. (1969-2009) <Katrina/Rita Population Adjustment>, National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, April 2012. Underlying mortality data provided by NCHS (www.cdc.gov/nchs).



If you have questions about the data, please contact Melinda Lehnherr, Illinois Department of Public Health, Division of Epidemiologic Studies, by phone at 217-785-1873 or e-mail melinda.lehnherr@illinois.gov.

**REFERENCES**


1. Lehnherr M and Havener LA. Assessment of Interstate Exchange of Cancer Data, 1986-1998. *ISCR Quality Control Series 02:1.* Springfield, Ill.: Illinois Department of Public Health, January 2002.

2. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Populations - Total U.S. (1969-2009) and (1990-2009), Katrina/Rita Adjustment - Linked To County Attributes - Total U.S., 1969-2009 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released November 2010.

3. The Population Estimates Program, U.S. Census Bureau. *Release notes Vintage 2009.* U.S. Census Bureau . http://www.census.gov/popest/methodology/2009-est-relnotes.pdf. Published March 23, 2010. Accessed July 27, 2012.

4. Surveillance Epidemiology and End Results, National Cancer Institute, U.S. National Institutes of Health. SEER – website. http://seer.cancer.gov/popdata/impact_rates_2009.html. Accessed July 27, 2012.

5. Percy C, Van Holten V, and Muir C (eds). *International Classification of Diseases for Oncology*. Second edition. Geneva: World Health Organization, 1990.

6. Fritz A, Percy C, Jack A, et al (eds). *International Classification of Diseases for Oncology*. Third edition. Geneva: World Health Organization, 2000.

7. World Health Organization. *Manual of the International Statistical Classification of Diseases, Injuries and Causes of Death*, based on the recommendations of the Ninth Revision Conference, 1975. Geneva: World Health Organization, 1977.

8. World Health Organization. *International Statistical Classification of Diseases and Related Health Problems*. 10th revision. Geneva: World Health Organization, 1992.

9. Anderson RN, Minino AM, Hoyert DL, Rosenberg HM. Comparability of cause of death between ICD-9 and ICD-10: Preliminary estimates. National Vital Statistics Reports 2001;49:1-32.

10. Copeland G, Lake A, Firth R, et al (eds). *Cancer in North America, 2003-2007. Volume Three: Registry-specific Cancer Mortality in the United States and Canada.* Springfield, IL: North American Association of Central Cancer Registries, Inc., May 2010.

11. Surveillance Research Program, National Cancer Institute SEER*State software (seer.cancer.gov/seerstat) Version 7.0.9, April 12, 2012.

12. Tiwari RC, Clegg LX, Zou Z. Efficient interval estimation for age-adjusted cancer rates. *Stat Methods Med Res* 2006 Dec;15(6):547-69.

13. Espey DK, Wiggins CL, Jim MA, Miller BA, Johnson CJ, Becker TM. Methods for improving cancer surveillance data in American Indian and Alaska Native populations. *Cancer* 2008; 113(5 Suppl):1120-1130.

14. NAACCR Expert Panel in Hispanic Identification. *Report of the Expert Panel on Hispanic Identification 2003.* Springfield, Ill.: North American Association of Central Cancer Registries, October 2003.

15. Howe HL. *Evaluation of NHIA Submission for 1997-2001*. Springfield, Ill.: North American Association of Central Cancer Registries, October 2004.

16. Hofferkamp, J (Ed). *Standards for Cancer Registries Volume III: Standards for Completeness, Quality, Analysis, Management, Security and Confidentiality of Data*. Springfield (IL): North American Association of Central Cancer Registries, August 2008.

17. Schenke N, Gentleman JF. On judging the significance of differences by examining the overlap between confidence intervals. *The American Statistician* 2001;55:182-1