INCIDENCE DATA TECHNICAL NOTES

DATA USE AGREEMENT

By using these data, you signify your agreement to comply with the Illinois Health and Hazardous Substances Registry Act (410 ILCS 525/12). Data collected by the Illinois State Cancer Registry (ISCR) are made available to the public, however the identification or contact of individuals is prohibited.

In an effort to exclude identifying information on individual patients, these 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 Illinois Department of Public Health and none should be inferred.

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), 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, 5.6 percent of ISCR cases are reported from out-of-state agencies and organizations.

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. The Quality Control section below contains a table with annual percentages for death certificate only cases.

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 of November 2014.

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 in the formulation of rates.[2] These population estimates of Illinois for all races, whites, blacks and Asian/other races from 1986 through 2012, and for Hispanics, non-Hispanics, non-Hispanic white and non-Hispanic black for 1990 through 2012 were obtained from both the intercensal and Vintage 2013 bridged-race post censal population estimates files. Population estimates by age, sex, race and Hispanic origin were produced by the U.S. Bureau of Census Population Estimates Program (http://www.census.gov/popest/index.html), in collaboration with the National Center for Health Statistics, and with support from the National Cancer Institute (NCI) through an interagency agreement. The population estimates incorporate intercensal (for 2000-2009) and Vintage 2013 (for 2010-2012) bridged single-race estimates are derived from the original multiple race categories in the 2000 census (as specified in the 1997 Office of Management and Budget standards for the collection of data on race and ethnicity http://www.whitehouse.gov/omb/inforeg_statpolicy). The bridged single-race estimates and a description of the methodology used to develop them appear on the National Center for Health Statistics website (http://www.cdc.gov/nchs/nvss/bridged_race.htm).

The intercensal estimates provide an adjustment of previous population estimates based on the actual 2010 census results (http://www.census.gov/popest/methodology/2013-natstcopr-meth.pdf).[3,4] Previous estimates utilized prior to the availability of the 2010 census data were prone to increased error as the time from the actual 2000 census increased. At the national level, estimates using both the 2000 census and the 2010 census are not very different from the previous estimates. However, there are more significant differences at the state and county levels that may result in changes to cancer incidence rates when one compares this report to earlier versions. Changes in rates also could be attributable to the addition of cases reported late.

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. The current version of the histologies Web-based query data utilizes the ICD-O-3 recode with adjustment for WHO 2008 hematopoietic. SEER-NCI recommends this site recode scheme (Site Recode ICD-O-3/WHO 2008) be used for any data containing cases diagnosed in 2010 or later years. In the interests of comparability to other national, state and registry specific data, subsequent versions of the Web-based query data containing cases diagnosed in 2010 or later will indeed use the SEER Site Recode ICD-O-3/WHO 2008. For a complete listing, see Appendix B of the annual state report (see Illinois State Cancer Incidence Review and Update).

Data at the county level in this application are aggregated into 24 major site groups and at the ZIP codes level into 11 site 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, 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, ZIP code, Cook County and stage files, the diagnosis year is a five-year aggregate (1988-1992, 1993-1997, 1998-2002 (this and prior groups not available for the Cook County file), 2003-2007 and 2008-2012).

Incidence Rates: The SEER*Stat® software package,[7] developed by the Information Management Services Inc. for NCI, was used to calculate incidence 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 10th 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.[8] 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 Review and Update).

Percent Distribution by Stage at Diagnosis: Stage at diagnosis of cancer for counties is expressed as percentage localized, regional, distant and unstaged for cancers of the oral cavity and pharynx (both sexes); colon and rectum (both sexes); melanoma and the skin (both sexes); and for invasive cervix and prostate. Female breast cancer incidence data are displayed with in situ stage, in addition to localized, regional, distant and unstaged stage categories. Data by race are available 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 populations to allow meaningful statistics for the race groups.

Race-specific Rates: 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). These same race grouping are available on the Cook County specific file. 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.[9] 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.

Ethnicity/Race Rates: For the incidence data at the state level, Hispanic ethnicity was derived according to the North American Association of Central Cancer Registries (NAACCR) Hispanic identification algorithm (NHIA).[10] NHIA is a generally reliable method to enhance the ethnic identification of the Latino population in the United States.[11] 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.

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,[12] ISCR has been awarded gold certification for all diagnosis years from 1996 through 2011.

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 2014, ISCR quality control data for each diagnosis year are as follow:

Year	Completeness (NAACCR Method)[a] (% As of 11-14)	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	87	~	~	0.04	0.0	0.0	0.0	0.3
1989	88	~	~	0.04	0.0	0.0	0.0	0.2
1990	89	100	~	0.04	0.0	0.0	0.0	0.3
1991	88	100	~	0.04	0.0	0.0	0.0	0.6
1992	91	100	~	0.04	0.0	0.0	0.0	0.4
1993	92	100	2.2	0.04	0.0	0.0	0.0	0.3
1994	97	100	6.1	0.06	0.0	0.0	0.0	0.3
1995	99	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	0.9
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	0.01	0.0	0.0	0.0	1.2
2005	100	100	1.9	0.00	0.0	0.0	0.0	1.4
2006	100	100	2.0	0.00	0.0	0.0	0.0	1.0
2007	100	100	1.2	0.00	0.0	0.0	0.0	1.1
2008	100	100	1.7	0.07	0.0	0.0	0.0	1.3
2009	100	100	1.6	0.03	0.0	0.0	0.0	1.5
2010	100	100	1.8	0.03	0.0	0.0	0.0	1.4
2011	100	100	1.8	0.00	0.0	0.0	0.0	1.8
2012	100	100	0.9	0.02	0.0	0.0	0.0	1.6

~ 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.

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.

Population estimates used for denominators may be inaccurate or lack precision. The population data for 1990, 2000 and 2010, 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, 2000 or 2010.

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.[13]

Acknowledgements

The production of these data would not have been possible without the diligent work of the Illinois Department of Public Health's Illinois State Cancer Registry 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 5U58DP003883-03 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 v22, 1986-2012, data as of November 2014.

Copyright information

If you have questions about these data, contact Lori Koch, Illinois Department of Public Health, Division of Epidemiologic Studies, by phone at 217-785-1873 or e-mail lori.koch@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. (1969-2013) and (1990-2013), Katrina/Rita Adjustment - Linked To County Attributes - Total U.S., 1969-2013 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released October 2014.

3. The Population Estimates Program, U.S. Census Bureau. *Release notes Vintage 2013*. U.S. Census Bureau . http://www.census.gov/popest/methodology/2012-nat-st-co-meth.pdf . Published April 22, 2014. Accessed March 23, 2015.

4. Surveillance Epidemiology and End Results, National Cancer Institute, U.S. National Institutes of Health. SEER – website. http://seer.cancer.gov/popdata/ . Accessed March 23, 2015.

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. Surveillance Research Program, National Cancer Institute SEER*State software (seer.cancer.gov/seerstat) Version 8.1.5, March 26, 2014.

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

9. 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.

10. NAACCR Race and Ethnicity Workgroup. *NAACCR Guideline for Enhancing Hispanic-Latino Identification: Revised NAACCR Hispanic/Latino identification Algorithm [NHIA v2.2.1]*. Springfield, Ill.: North American Association of Central Cancer Registries, September 2011.

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

12. 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.

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