Methods (H4)

Indicator

H4. Cancer incidence and mortality for children ages 0 to 19 years, 1992-2009.

Summary

Since 1973, the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute (NCI) has been collecting and publishing cancer incidence data from population-based cancer registries that currently cover a total of 18 geographical areas in the United States and one quarter of the population. Since the coverage area has expanded over time, the trend analyses for indicator H4 were based on the SEER 13 registries that cover 13.8% of the U.S. population for the years 1992 and later. Data include the type of cancer, age at diagnosis, year of diagnosis, sex, race, and ethnicity. For indicator H4, cancer incidence rates for malignant cancers in children ages 0 to 19 years were calculated using the SEER*Stat software provided by the NCI. SEER*Stat calculates incidence rates for each age group by dividing the number of new cancer cases in a given year by the total population for that age group in the cancer registry geographical area in that year. SEER*Stat then calculates age-adjusted incidence rates as a weighted average of the rates for each five-year age group, where the weights are the proportions of persons in each age group in the selected standard population; for these analyses the incidence rates were age-adjusted to the 2000 U.S. Standard Population. National cancer mortality rates were also obtained from SEER, which uses data from the National Vital Statistics System, administered by the National Center for Health Statistics. Data include the type of cancer, age at death, year of death, sex, race, and ethnicity. For indicator H4, cancer mortality rates for malignant cancers in children ages 0 to 19 years were calculated using the SEER*Stat software provided by the NCI. The mortality rate is calculated by dividing the number of deaths attributed to a malignant cancer in a given year by the total population for that age group in that year. The mortality rates were age-adjusted to the 2000 U.S. Standard Population. Tables H4a, H4b, and H4c provide the age-adjusted cancer incidence and mortality rates for children 19 and under, stratified by race/ethnicity, sex, and age for the period 2007–2009.

Data Summary

Indicator	H4. Cancer incidence and mortality for children ages 0 to 19 years.					
Time Period	1992-2009					
Data	Cancer incidence of malignant cancers in SEER 13 registries for children ages 0 to 19 years. U.S. mortality from malignant cancers in children ages 0 to 19 years.					
Years	1992	1993	1994	1995	1996	1997
Cancer incidence	1,649	1,722	1,646	1,679	1,766	1,707
Cancer mortality	2,417	2,419	2,354	2,275	2,224	2,261
Years	1998	1999	2000	2001	2002	2003

Indicator	H4. Cancer incidence and mortality for children ages 0 to 19 years.					
Cancer incidence	1,832	1,773	1,830	1,887	1,946	1,774
Cancer mortality	2,176	2,243	2,271	2,226	2,271	2,233
Years	2004	2005	2006	2007	2008	2009
Cancer incidence	1,907	1,994	1,800	1,973	1,983	2,018
Cancer mortality	2,223	2,183	2,035	2,068	2,039	1,964

Overview of Data Files

The following files are needed to calculate this indicator. All these files together with the SEER*Stat software used to calculate the incidence and mortality rates are available from the SEER website: http://seer.cancer.gov/data/access.html. The data and analyses used for this indicator were obtained using version 7.0.9 of SEER*Stat.

- SEER 13 Regs Research Data, Nov 2011 Sub, Vintage 2009 Pops (1992–2009) <Katrina/Rita Population Adjustment> - Linked To County Attributes - Total U.S., 1969–2010 Counties. Incidence data from the SEER 13 registries: Connecticut; Hawaii; Iowa; New Mexico; Utah; Atlanta, Georgia; Detroit, Michigan; San Francisco-Oakland, California; Seattle-Puget Sound, Washington; Los Angeles, California; San Jose-Monterey, California; Rural Georgia; and the Alaska Native Tumor Registry. Downloaded April, 2012.

- Mortality - All COD, Aggregated With State, Total U.S. (1969–2009) <Katrina/Rita Population Adjustment>. Mortality data from the entire United States for all causes of death. Mortality from All Malignant Cancers was selected. Downloaded April, 2012.

The databases incorporated into the SEER*Stat software include census population data used to compute the incidence and mortality rates and to age-adjust the rates to the 2000 U.S. Standard Population. These population data are obtained by NCI from the Census Bureau. Age-adjustment is explained below.

Calculation of Indicator

All the calculations for this indicator were carried out using SEER*Stat software, version 7.0.9. In this section we detail the menu options required for carrying out these analyses.

Note that any User-Defined variable previously created using the File and Dictionary menus in the same or any earlier session will not need to be recreated.

Incidence

Click the Rate Session button "Σ".

Data tab. Select "Incidence – SEER 13 Regs Research Data, Nov 2011 Sub, Vintage 2009 Pops (1992–2009) <Katrina/Rita Population Adjustment>"

Statistic tab. Check "Rates (Age – Adjusted)" and "Show Standard Errors and Confidence Intervals." For "Standard Population" select "2000 US Std Population (19 age groups – Census P25-1130)." For "Age Variable" select "Age recode with <1 year olds."

Selection tab. Check "Malignant Behavior." For "Age at Diagnosis" click the Edit button and then select the age group 0–19 using "Age recode with <1 year olds" = 00, 01-04, 05-09, 10-14, and 15-19. This restricts the data to malignant cancers in children ages 0 to 19 years.

Table tab. From the "Available Variables" panel, expand "Race, Sex, Year Dx, Registry, County" and select "Year of diagnosis." Click the Row button. This tabulates the rates by calendar year.

Output tab. For "Display Rates as Cases Per" select "1,000,000."

Click the Execute button to run the analyses and create the output Rate Matrix.

Use the Matrix and Export menus to output the data to a text file.

Mortality

Click the Rate Session button "Σ".

Data tab. Select "Mortality - All COD, Aggregated With State, Total U.S. (1969–2009) <Katrina/Rita Population Adjustment>."

Statistic tab. Check "Rates (Age – Adjusted)" and "Show Standard Errors and Confidence Intervals." For "Standard Population" select "2000 US Std Population (19 age groups – Census P25-1130)." For "Age Variable" select "Age recode with <1 year olds."

Selection tab. For "Age at Death" click the Edit button and then select the age group 0–19 using "Age recode with <1 year olds" = 00, 01-04, 05-09, 10-14, and 15-19. For "Other (Case Files)" click the Edit button and then follow the menus to select "Site and Morphology," "Cause of death recode," and "All Malignant Cancers." This restricts the data to malignant cancers in children ages 0 to 19 years.

Table tab. From the "Available Variables" panel, expand "Race, Sex, Year Dth, State, Registry" and select "Year of Death." Click the Row button. This tabulates the rates by calendar year.

Output tab. For "Display Rates as Cases Per" select "1,000,000."

Click the Execute button to run the analyses and create the output Rate Matrix.

Use the Matrix and Export menus to output the data to a text file.

Age adjustment

For indicator H4, all cancer incidence and mortality rates were age-adjusted to the 2000 U.S. Standard Population using 5 age groups for children. The calculations were carried out automatically by the SEER*Stat software. The age-adjusted rates are often preferred when comparing different populations (such as for different calendar years or different race groups) since differences in the age-adjusted cancer rates are mainly attributable to factors other than age. In general, very young children and adolescents will tend to have a higher crude (unadjusted) cancer rate than children ages 5 to 14. The age adjustment replaces the distribution of ages in each given population by the same standard age distribution, in this case the age distribution of the U.S. population in 2000.

The calculation is shown by the following example which gives the age-adjusted cancer incidence rate for the year 2009 for children ages 0 to 19 years.

Age-adjusted Cancer Incidence for 2009.

Age Group [1]	Cases [2]	SEER 13 Population [3]	Crude Rate (Cases per Million)[4] = [2]/[3] ×1,000,000	U.S. 2000 Standard Population [5]	U.S. 2000 Proportion of Age Group 0-19 [6]	Adjusted Rate (Cases per Million) = [4] ×[6]
00 years	129	593,032	217.526	3,794,901	0.048	10.478
01-04 years	492	2,379,384	206.776	15,191,619	0.193	39.873
05-09 years	341	2,847,163	119.768	19,919,840	0.253	30.283
10-14 years	383	2,700,635	141.818	20,056,779	0.255	36.105
15-19 years	673	2,885,813	233.210	19,819,518	0.252	58.669
00-19 years	2,018	11,406,027	176.924	78,782,657	1.000	175.407

The first three columns give the number of new malignant cancer cases in 2009 and the total population for each children's age group within the counties included in the SEER 13 cancer registries. The number of cases divided by the SEER 13 population and multiplied by 1 million gives the crude rate shown in the fourth column. For example, for the age group 00 years, there were 129 cases in a population of 593,032 giving a crude rate of $(129/593032) \times 1000000 = 217.526$ cases per million. The crude rate for all children ages 0 to 19 years is 176.924 (bottom row).

The U.S. 2000 Standard Population for children is shown in the fifth column. This is the age distribution of children for the entire United States in 2000, in five age groups (0, 1–4, 5–9, 10–14, and 15–19). For analyzing children's populations only five age groups are needed (0, 1–4, 5–

9, 10–14, and 15–19). However, for analyzing children and adults the U.S. 2000 Standard Population has 19 age groups, since the U.S. 2000 Standard Population also includes five-year age groups up to 80–84 and the 85+ age group. The standard age distribution for children ages 0 to 19 years is shown in the sixth column. The U.S. 2000 Standard Population in each age group is divided by the total U.S. 2000 Standard Population for children ages 0 to 19 years, i.e., 78,782,657.

The age-adjusted population is a weighted average of the crude rates (column 4) weighted by the U.S. 2000 Standard Population proportions (column 6). Thus, the first five rows of column 7 contain the products of columns 4 and 6, and the age-adjusted rate is given by the total for column 7, shown in the bottom row as 175.407.

For this example, the crude rate for ages 0 to 19 years, 176.924, is very close to the age-adjusted rate, 175.407, reflecting the fact that in 2009 the age distribution for children 0 to 19 years in the SEER 13 regions closely matches the 2000 U.S. population age distribution. For other years or for different race/ethnicity/sex groups, crude and age-adjusted rates can differ by a much greater percentage.

Relative Standard Error

The uncertainties of the incidence and mortality rates were computed by the SEER*Stat software under the assumption that the counts have Poisson distributions. The relative standard error is the standard error of the rate divided by the estimated rate:

Relative Error (%) = [Standard Error (Rate) / Rate] × 100%

Rates with a relative error less than 30% were treated as being reliable and were tabulated. For indicator H4, all the relative standard errors were at most 30% and so all values were tabulated.

Rates by Race, Ethnicity, Sex, and Age

For the Tables H4a, H4b, and H4c, cancer incidence and mortality rates are tabulated by race, ethnicity, sex, and age.

The following SEER*Stat data files were used for the analyses in Tables H4a, H4b, and H4c:

- SEER 13 Regs Research Data, Nov 2011 Sub, Vintage 2009 Pops (1992–2009) <Katrina/Rita Population Adjustment> - Linked To County Attributes - Total U.S., 1969–2010 Counties. Incidence data from the SEER 13 registries: Connecticut; Hawaii; Iowa; New Mexico; Utah; Atlanta, Georgia; Detroit, Michigan; San Francisco-Oakland, California; Seattle-Puget Sound, Washington; Los Angeles, California; San Jose-Monterey, California; Rural Georgia; and the Alaska Native Tumor Registry. Downloaded April, 2012.

- Mortality - All COD, Aggregated With State, Total U.S. (1990–2009) <Katrina/Rita Population Adjustment>. Mortality data from the entire United States for all causes of death. Mortality from All Malignant Cancers was selected. Downloaded April, 2012.

The supplementary tables by race, ethnicity, sex, and age for incidence and mortality in 2007-2009 were calculated as follows:

Incidence by Race/Ethnicity and Sex

Click the Rate Session button "Σ".

Data tab. Select "Incidence – SEER 13 Regs Research Data, Nov 2011 Sub, Vintage 2009 Pops (1992–2009) <Katrina/Rita Population Adjustment."

Statistic tab. Check "Rates (Age – Adjusted)" and "Show Standard Errors and Confidence Intervals." For "Standard Population" select "2000 US Std Population (19 age groups – Census P25-1130)." For "Age Variable" select "Age recode with <1 year olds."

Selection tab. Check "Malignant Behavior." For "Age at Diagnosis" click the Edit button and then select the age group 0–19 using "Age recode with <1 year olds" = 00, 01-04, 05-09, 10-14, and 15-19. For "Race, Sex, Year Dx, Registry, County" click the Edit button and select the years 2007-2009 using "Year of Diagnosis" = 2007, 2008, and 2009. This restricts the data to malignant cancers in children ages 0 to 19 years diagnosed in 2007-2009.

Table tab. From the "Available Variables" panel, expand "Race, Sex, Year Dx, Registry, County" and select "Sex." Click the Row button. Use the menus "File," "Dictionary," and "Merge" to create a User-Defined Merged variable Raceeth from the Race recode (W, B, AI, API) and Origin recode NHIA (Hispanic, Non-Hisp) variables in the database:

- Raceeth = All if Race recode = White, Black, American Indian/Alaskan Native, Asian or Pacific Islander, Other Unspecified (1991+) or Unknown
- Raceeth = WhiteNH if Race recode = White and Origin recode NHIA = Non-Spanish-Hispanic-Latino
- Raceeth = BlackNH if Race recode = Black and Origin recode NHIA = Non-Spanish-Hispanic-Latino
- Raceeth = AIANNH if Race recode = American Indian/Alaskan Native and Origin recode NHIA = Non-Spanish-Hispanic-Latino
- Raceeth = APINH if Race recode = Asian or Pacific Islander and Origin recode NHIA = Non-Spanish-Hispanic-Latino
- Raceeth = Hispanic if Origin recode NHIA = Spanish-Hispanic-Latino

In the Table tab, from the "Available Variables" panel, expand "Merged" and select "Raceeth." Click the Column button. These options tabulate the rates by sex and race/ethnicity.

Output tab. For "Display Rates as Cases Per" select "1,000,000."

Click the Execute button to run the analyses and create the output Rate Matrix.

Use the Matrix and Export menus to output the race/ethnicity/sex data to a text file.

Incidence by Age

Assuming you are in the same session as above, modify the selections as follows:

Table tab. In the Table tab, from the "Display Variables" panel, select the Row variable "Sex" and click Remove. Also select the Column variable "Raceeth" and click Remove. Use the menus "File" and "Dictionary" to create a User-Defined variable AgesH4 that takes the values 0-4 years, 5-9 years, 10-14 years, 15-19 years, and 0-19 years if the "Age at Diagnosis" is in the selected age group. Age at Diagnosis is defined by the values of "Age recode with <1 year olds." In the Table tab, from the "Available Variables" panel, expand "User-Defined" and select "AgesH4." Click the Row button. This tabulates the rates by the selected age groups.

Click the Execute button to run the analyses and create the output Rate Matrix.

Use the Matrix and Export menus to output the data to a text file. Tabulate the rates for selected age groups.

Mortality by Race/Ethnicity and Sex

Click the Rate Session button "Σ".

Data tab. Select "Mortality - All COD, Aggregated With State, Total U.S. (1990–2009) <Katrina/Rita Population Adjustment>."

Statistic tab. Check "Rates (Age – Adjusted)" and "Show Standard Errors and Confidence Intervals." For "Standard Population" select "2000 US Std Population (19 age groups – Census P25-1130)." For "Age Variable" select "Age recode with <1 year olds."

Selection tab. For "Age at Death" click the Edit button and then select the age group 0 to 19 years using "Age recode with <1 year olds" = 00, 01-04, 05-09, 10-14, and 15-19. For "Race, Sex, Year Dth, State, Registry" click the Edit button and select the years 2007–2009 using "Year of Death" = 2007, 2008 and 2009. For "Other (Case Files)" click the Edit button and then follow the menus to select "Site and Morphology," "Cause of death recode," and "All Malignant Cancers." These options restrict the data to malignant cancers in children ages 0 to 19 years who died in 2007–2009.

Table tab. From the "Available Variables" panel, expand "Race, Sex, Year Dth, State, Registry" and select "Sex." Click the Row button. Use the menus "File," "Dictionary," and "Merge" to create a User-Defined Merged variable Raceeth2 from the Race recode (W, B, AI, API), Origin recode 1990+ (Hispanic, Non-Hisp), and State variables in the database. As explained below, the

mortality rates for Hispanic and Non-Hispanic subgroups excluded data from two States with insufficiently complete ethnicity data:

- Raceeth2 = All if Race recode = White, Black, American Indian/Alaskan Native, or Asian or Pacific Islander
- Raceeth2 = WhiteNH if Race recode = White, Origin recode 1990+ (Hispanic, Non-Hisp) = Non-Spanish-Hispanic-Latino or Unknown, and State ≠ North Dakota or South Carolina
- Raceeth2 = BlackNH if Race recode = Black, Origin recode 1990+ (Hispanic, Non-Hisp) = Non-Spanish-Hispanic-Latino or Unknown, and State ≠ North Dakota or South Carolina
- Raceeth2 = AIANNH if Race recode = American Indian/Alaskan Native, Origin recode 1990+ (Hispanic, Non-Hisp) = Non-Spanish-Hispanic-Latino or Unknown, and State ≠ North Dakota or South Carolina
- Raceeth2 = APINH if Race recode = Asian or Pacific Islander, Origin recode 1990+ (Hispanic, Non-Hisp) = Non-Spanish-Hispanic-Latino or Unknown, and State ≠ North Dakota or South Carolina
- Raceeth2 = Hispanic if Origin recode 1990+ (Hispanic, Non-Hisp) = Spanish-Hispanic-Latino, and State ≠ North Dakota or South Carolina

Hispanic Index. Following the recommendations of the National Cancer Institute (NCI), the mortality rates for all the groups except for "All" excluded data from two states that had large numbers with unknown ethnicity: North Dakota and South Carolina. As explained at the url

http://seer.cancer.gov/seerstat/variables/mort/origin_recode_1990+/index.html,

NCI defined a Hispanic Index to measure the reliability of the Hispanic origin data for the mortality data from each year and state. The Hispanic Index is defined as

Hispanic Index = ([Hispanic Population / Total Population] × Unknown Origin Count / Hispanic Count) × 100%

When this index gives a value >= 10.00, data on Hispanic and non-Hispanic mortality are deemed unreliable. The Excel file of Hispanic Indexes, available at the same url, shows that for the years 2007 to 2009, North Dakota and South Carolina had Hispanic Index values >= 10.00 for one or more of those years. We therefore defined Raceeth2 to exclude those states from the 2007–2009 mortality rate calculations for groups that included Hispanics and non-Hispanics in their definitions.

In the Table tab, from the "Available Variables" panel, expand "Merged" and select "Raceeth2." Click the Column button. These options tabulate the rates by sex and race/ethnicity.

Output tab. For "Display Rates as Cases Per" select "1,000,000."

Click the Execute button to run the analyses and create the output Rate Matrix.

Use the Matrix and Export menus to output the race/ethnicity/sex data to a text file.

Questions and Comments

Questions regarding these methods, and suggestions to improve the description of the methods, are welcome. Please use the "Contact Us" link at the bottom of any page in the America's Children and the Environment website.

Statistical Comparisons

Statistical analyses of the cancer incidence and mortality rates were used to determine whether the differences between rates for different demographic groups were statistically significant. For these analyses, the incidence counts and populations for each combination of calendar year, age group, sex, and race/ethnicity were obtained from the SEER database using the same race/ethnicity groups described above (excluding the All category), and the age groups 0-4 years, 5-9 years, 10-14 years, and 15-19 years. Incidence data for malignant cancers in the years 1992 to 2009 were obtained from the SEER data set "SEER 13 Regs Research Data, Nov 2011 Sub, Vintage 2009 Pops (1992–2009) <Katrina/Rita Population Adjustment> - Linked To County Attributes - Total U.S., 1969–2010 Counties." U.S. mortality data for 1992 to 2009 were obtained from the SEER data set "Mortality - All COD, Aggregated With State, Total U.S. (1990–2009) <Katrina/Rita Population Adjustment>." For the US mortality data, the analysis excluded data from two regions which had large numbers with unknown ethnicity in one or more of the years 2007-2009: North Dakota and South Carolina.[i]

Using a Poisson regression model adjusted for dispersion, the number of cases for each demographic subgroup was assumed to have a Poisson distribution with a mean equal to the population of that demographic subgroup multiplied by the rate, and a variance equal to the mean multiplied by a dispersion parameter. The logarithm of the rate was assumed to be the sum of explanatory terms for age, sex, and race/ethnicity. Using this model, the difference in the value of a rate between different demographic groups is statistically significant if the difference between the corresponding sums of explanatory terms is statistically significantly different from zero. A p-value at or below 0.05 implies that the difference is statistically significant at the 5% significance level. No adjustment is made for multiple comparisons.

For each type of comparison, we present unadjusted and adjusted analyses. The unadjusted analyses compare a rate between different demographic groups. The adjusted analyses add other demographic explanatory variables to the statistical model and use the statistical model to account for the possible confounding effects of these other demographic variables. For example, the unadjusted race/ethnicity comparisons compare the rates between different race/ethnicity

---

[i]NCI provides a Hispanic Index for each state and calendar year at the url: http://seer.cancer.gov/seerstat/variables/mort/origin_recode_1990+/index.html. NCI recommends excluding mortality data from all states with a Hispanic Index at or above 10.0. Applying this rule to individual years would introduce a bias by excluding different states in different years. Excluding all states if they had a Hispanic Index at or above 10 in any calendar year would be too data restrictive. Thus we decided to only exclude data from North Dakota and South Carolina, which had high Hispanic Indexes in one or more of the years 2007-2009.

pairs. The adjusted analyses add age and sex terms to the statistical model and compare the rates between different race/ethnicity pairs after accounting for the effects of the other demographic variables. For example, if Hispanic children tend to be younger than White non-Hispanics, and if the rate strongly depends on age only, then the unadjusted differences between these two race/ethnicity groups would be significant but the adjusted difference (taking into account age) would not be significant.

Comparisons of incidence rates between pairs of race/ethnicity groups are shown in Table 1. Comparisons of mortality rates between pairs of race/ethnicity groups are shown in Table 2. In Tables 1 and 2, for the "Unadjusted" comparisons, the only explanatory variables are terms for each race/ethnicity group. For these unadjusted comparisons, the statistical tests compare the rates for each pair of race/ethnicity groups. For the "Adjusted for age, sex" comparisons, the explanatory variables are terms for each race/ethnicity together with terms for each age group and sex. For these adjusted comparisons, the statistical test compares the pair of race/ethnicity groups after accounting for any differences in the age and sex distributions between the race/ethnicity groups.

Additional comparisons are shown in Table 3 for incidence rates and in Table 4 for mortality rates. The Against = "age" unadjusted p-value compares the rates between all the age groups. The adjusted p-value includes adjustment terms for sex and race/ethnicity in the model. The Against = "sex" unadjusted p-value compares the rates between the two sexes. The adjusted p-value includes adjustment terms for age and race/ethnicity in the model. The Against = "year" unadjusted p-value compares the trends in the rates by regressing against the calendar year. The adjusted p-value includes adjustment terms for age, sex, and race/ethnicity in the model. To evaluate whether the difference in rates for Males and Females varies according to the race/ethnicity, p-values for comparing the Male and Female rates by race/ethnicity group are found in the rows where Against = "sex" and Subset = the race/ethnicity group. The adjusted p-value includes an adjustment term for age.

For more details on these statistical analyses, see the memorandum by Cohen (2011).[ii]

Table 1. Statistical significance tests comparing cancer incidence rates for children ages 0 to 19 years, between pairs of race/ethnicity groups, for 2007-2009.

			P-VALUES	
Variable	First race/ethnicity group	Second race/ethnicity group	Unadjusted	Adjusted for age, sex
Cancer incidence	White non-Hispanic	Black non-Hispanic	< 0.001	< 0.001
Cancer incidence	White non-Hispanic	API non-Hispanic	0.030	< 0.001
Cancer incidence	White non-Hispanic	AIAN non-Hispanic	0.228	0.009
Cancer incidence	White non-Hispanic	Hispanic	0.104	0.001
Cancer incidence	Black non-Hispanic	API non-Hispanic	0.346	0.040
Cancer incidence	Black non-Hispanic	AIAN non-Hispanic	0.912	0.875
Cancer incidence	Black non-Hispanic	Hispanic	0.027	< 0.001
Cancer incidence	API non-Hispanic	AIAN non-Hispanic	0.739	0.419
Cancer incidence	API non-Hispanic	Hispanic	0.316	0.036

[ii] Cohen, J. 2011. *Selected statistical methods for testing for trends and comparing years or demographic groups in other ACE health-based indicators.* Memorandum submitted to Dan Axelrad, EPA, 16 June, 2011.

			P-VALUES	
Variable	First race/ethnicity group	Second race/ethnicity group	Unadjusted	Adjusted for age, sex
Cancer incidence	AIAN non-Hispanic	Hispanic	0.448	0.088

Table 2. Statistical significance tests comparing cancer mortality rates for children ages 0 to 19 years, between pairs of race/ethnicity groups, for 2007-2009.

			P-VALUES	
Variable	First race/ethnicity group	Second race/ethnicity group	Unadjusted	Adjusted for age, sex
Cancer mortality	White non-Hispanic	Black non-Hispanic	0.706	0.598
Cancer mortality	White non-Hispanic	API non-Hispanic	0.003	< 0.001
Cancer mortality	White non-Hispanic	AIAN non-Hispanic	< 0.001	< 0.001
Cancer mortality	White non-Hispanic	Hispanic	0.202	0.035
Cancer mortality	Black non-Hispanic	API non-Hispanic	0.004	< 0.001
Cancer mortality	Black non-Hispanic	AIAN non-Hispanic	< 0.001	< 0.001
Cancer mortality	Black non-Hispanic	Hispanic	0.531	0.261
Cancer mortality	API non-Hispanic	AIAN non-Hispanic	< 0.001	< 0.001
Cancer mortality	API non-Hispanic	Hispanic	0.001	< 0.001
Cancer mortality	AIAN non-Hispanic	Hispanic	< 0.001	< 0.001

Table 3. Other statistical significance tests comparing cancer incidence rates for children ages 0 to 19 years for 2007 to 2009 (trends for 1992-2009).

					P-VALUES	
Variable	From	To	Against	Subset	Unadjusted	Adjusted*
Cancer incidence	2007	2009	age		< 0.001	< 0.001
Cancer incidence	2007	2009	race		0.005	< 0.001
Cancer incidence	2007	2009	sex		0.055	< 0.001
Cancer incidence	2007	2009	sex	White non-Hispanic	0.349	0.001
Cancer incidence	2007	2009	sex	Black non-Hispanic	0.644	0.511
Cancer incidence	2007	2009	sex	API non-Hispanic	0.643	0.453
Cancer incidence	2007	2009	sex	AIAN non-Hispanic	0.315	0.299
Cancer incidence	2007	2009	sex	Hispanic	0.192	0.001
Cancer incidence	1992	2009	year		0.014	< 0.001

*For Against = "age," the p-values are adjusted for sex and race/ethnicity.
For Against = "sex," the p-values are adjusted for age and race/ethnicity.
For Against = "year," the p-values are adjusted for age, sex, and race/ethnicity.

Table 4. Other statistical significance tests comparing cancer mortality rates for children ages 0 to 19 years for 2007 to 2009 (trends for 1992-2009).

					P-VALUES	
Variable	From	To	Against	Subset	Unadjusted	Adjusted*
Cancer mortality	2007	2009	age		< 0.001	< 0.001
Cancer mortality	2007	2009	race		0.036	0.001
Cancer mortality	2007	2009	sex		0.004	< 0.001
Cancer mortality	2007	2009	sex	White non-Hispanic	0.006	< 0.001
Cancer mortality	2007	2009	sex	Black non-Hispanic	0.370	0.245
Cancer mortality	2007	2009	sex	API non-Hispanic	0.018	0.012

					P-VALUES	
Variable	**From**	**To**	**Against**	**Subset**	**Unadjusted**	**Adjusted***
Cancer mortality	2007	2009	sex	AIAN non-Hispanic	NA**	NA**
Cancer mortality	2007	2009	sex	Hispanic	0.475	0.222
Cancer mortality	1992	2009	year		< 0.001	< 0.001

*For Against = "age," the p-values are adjusted for sex and race/ethnicity.
For Against = "sex," the p-values are adjusted for age and race/ethnicity.
For Against = "year," the p-values are adjusted for age, sex, and race/ethnicity.
NA**. Statistical model failed to converge.

Methods (H5)

Indicator

H5. Cancer incidence for children ages 0 to 19 years, by type, 1992-2006.

Summary

Since 1973, the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute (NCI) has been collecting and publishing cancer incidence data from population-based cancer registries that currently cover a total of 18 geographical areas in the United States and one quarter of the population. Since the coverage area has expanded over time, the trend analyses for indicator H5 were based on the SEER 13 registries that cover 13.8% of the U.S. population for the years 1992 and later. Data include the type of cancer (cancer site), age at diagnosis, and year of diagnosis. For indicator H5, cancer incidence rates for specific malignant cancers in children ages 0 to 19 years for three-year periods 1992-1994 to 2004-2006 were calculated using the SEER*Stat software provided by the NCI. SEER*Stat calculates incidence rates for each age group by dividing the number of new cancer cases by the total population for that age group in the cancer registry geographical area. SEER*Stat then calculates age-adjusted incidence rates as a weighted average of the rates for each five-year age group, where the weights are the proportions of persons in each age group in the selected standard population; for these analyses the incidence rates were age-adjusted to the 2000 U.S. Standard Population. Table H5a provides the age-adjusted cancer incidence rates for children ages 19 years and under, stratified by age group and cancer type, for the period 2004-2006.

Data Summary

Indicator	H5. Cancer incidence for children ages 0 to 19 years, by type, 1992-2006.					
Time Period	1992-2006					
Data	Cancer incidence of malignant cancers in SEER 13 registries for children ages 0 to 19 years.					
Years	1992	1993	1994	1995	1996	1997
Cancer incidence	1, 646	1,723	1,644	1,679	1,764	1,708
Years	1998	1999	2000	2001	2002	2003
Cancer incidence	1,829	1,768	1,822	1,882	1,942	1,771
Years	2004	2005	2006			
Cancer incidence	1,901	1,983	1,773			

Overview of Data Files

The following files are needed to calculate this indicator. All these files together with the SEER*Stat software used to calculate the incidence rates are available from the SEER website: http://seer.cancer.gov/data/access.html. The data and analyses used for this indicator were obtained using version 6.6.2 of SEER*Stat.

- SEER 13 Regs Research Data, Nov 2008 Sub (1992–2006) <Katrina/Rita Population Adjustment> - Linked To County Attributes - Total U.S., 1969–2006 Counties. Incidence data from the SEER 13 registries: Connecticut; Hawaii; Iowa; New Mexico; Utah; Atlanta, Georgia; Detroit, Michigan; San Francisco-Oakland, California; Seattle-Puget Sound, Washington; Los Angeles, California; San Jose-Monterey, California; Rural Georgia; and the Alaska Native Tumor Registry. Downloaded June, 2010.

The databases incorporated into the SEER*Stat software include census population data used to compute the incidence rates and to age-adjust the rates to the 2000 U.S. Standard Population. These population data are obtained by NCI from the Census Bureau. Age-adjustment is explained below.

Calculation of Indicator

All the calculations for this indicator were carried out using SEER*Stat software, version 6.6.2. In this section we detail the menu options required for carrying out these analyses.

Note that any User-Defined variable previously created using the File and Dictionary menus in the same or any earlier session will not need to be recreated.

ICCC site recode ICD-O-3.

Click the Rate Session button "Σ".

Data tab. Select "SEER 13 Regs Research Data, Nov 2008 Sub (1992–2006) <Katrina/Rita Population Adjustment>."

Statistic tab. Check "Rates (Age – Adjusted)" and "Show Standard Errors and Confidence Intervals." For "Standard Population" select "2000 US Std Population (19 age groups – Census P25-1130)." For "Age Variable" select "Age recode with <1 year olds."

Selection tab. Check "Malignant Behavior." For "Age at Diagnosis" click the Edit button and then select the age group 0 to 19 years using "Age recode with <1 year olds" = 00, 01-04, 05-09, 10-14, and 15-19. This restricts the data to malignant cancers in children ages 0 to 19 years.

Table tab. Use the menus "File" and "Dictionary" to create a User-Defined variable YearsH5 that takes the values 1992–1994, 1995–1997, 1998–2000, 2001–2003 and 2004–2006 if the "Year of Diagnosis" is in the selected year group. In the Table tab, from the "Available Variables" panel, expand "User-Defined" and select "YearsH5." Click the Row button. From the

"Available Variables" panel, expand "Site and Morphology" and select "ICCC site recode ICD-O-3." Click the Column button. This tabulates the rates by the selected calendar year groups and cancer types.

Output tab. For "Display Rates as Cases Per" select "1,000,000."

Click the Execute button to run the analyses and create the output Rate Matrix.

Use the Matrix and Export menus to output the data to a text file. Tabulate the rates for selected cancer types and year groups.

AYA Site Recode.

The cancer types acute lymphocytic leukemia and Wilms' tumor are not specifically included in the ICCC site recode ICD-O-3 categories. Thus the incidence rate for these two cancers is calculated separately using the "AYA site recode," which codes cancer types for adolescents and young adults.

Modify the above selections as follows:

Table tab. In the Table tab, from the "Display Variables" panel, select the Column variable "ICCC site recode ICD-O-3" and click Remove. From the "Available Variables" panel, expand "Site and Morphology" and select "AYA site recode." Click the Column button. This tabulates the rates by the selected calendar year groups for the cancer types coded in the AYA site recode.

Click the Execute button to run the analyses and create the output Rate Matrix.

Use the Matrix and Export menus to output the AYA site recode data to a text file. Tabulate the rates for selected cancer types and year groups.

Age-adjustment

For indicator H5, all cancer incidence rates were age-adjusted to the 2000 U.S. Standard Population using five age groups for children. The calculations were carried out automatically by the SEER*Stat software. The age-adjusted rates are often preferred when comparing different populations (such as for different calendar years or different race groups) since differences in the age-adjusted cancer rates are mainly attributable to factors other than age. In general, a younger population will tend to have a lower crude (unadjusted) cancer rate. The age-adjustment replaces the distribution of ages in each given population by the same standard age distribution, in this case the age distribution of the U.S. population in 2000.

The calculation is shown by the following example which gives the age-adjusted cancer incidence rate for the years 2004–2006 for acute myeloid leukemia in children ages 0 to 19 years.

Age-adjusted Acute Myeloid Leukemia Incidence for 2004-2006.

Age Group [1]	Cases [2]	SEER 13 Population [3]	Crude Rate (Cases per Million)[4] = [2]/[3]	U.S. 2000 Standard Population [5]	U.S. 2000 Proportion of Age Group 0-19 [6]	Adjusted Rate (Cases per Million) = [4] × [6]
00 years	37	1,753,659	21.099	3,794,901	0.048	1.016
01-04 years	76	6,769,698	11.226	15,191,619	0.193	2.165
05-09 years	38	8,220,990	4.622	19,919,840	0.253	1.169
10-14 years	64	8,760,297	7.306	20,056,779	0.255	1.860
15-19 years	77	8,577,973	8.976	19,819,518	0.252	2.258
00-19 years	292	34,082,617	8.567	78,782,657	1.000	8.468

The first three columns give the number of new malignant acute myeloid leukemia cases in 2004–2006 and the total population for each children's age group within the counties included in the SEER 13 cancer registries. The number of cases divided by the SEER 13 population and multiplied by 1 million gives the crude rate shown in the fourth column. For example, for the age group 00 years, there were 37 cases in a population of 1,753,659 giving a crude rate of $(37/1753659) \times 1000000 = 21.099$ cases per million. The crude rate for all children ages 0 to 19 years is 8.567 (bottom row).

The U.S. 2000 Standard Population for children is shown in the fifth column. This is the age distribution for the entire United States in 2000. For analyzing children's populations only five age groups are needed (0, 1–4, 5–9, 10–14, and 15–19). However, for analyzing children and adults the U.S. 2000 Standard Population has 19 age groups, since the U.S. 2000 Standard Population also includes five-year age groups up to 80–84 and the 85+ age group. The standard age distribution for children 0 to 19 years is shown in the sixth column. The U.S. 2000 Standard Population in each age group is divided by the total U.S. 2000 Standard Population for children ages 0 to 19 years, i.e., 78,782,657.

The age-adjusted population is a weighted average of the crude rates (column 4) weighted by the U.S. 2000 Standard Population proportions (column 6). Thus, the first five rows of column 7 contain the products of columns 4 and 6, and the age-adjusted rate is given by the total for column 7, shown in the bottom row as 8.468.

For this example, the crude rate for ages 0 to 19 years, 8.567, is very close to the age-adjusted rate, 8.468, reflecting the fact that the age distribution for children 0 to 19 years in the SEER 13 regions closely matches the 2000 U.S. population age distribution. For other years or for different cancer types, crude and age-adjusted rates can differ by a much greater percentage.

Relative Standard Error

The uncertainties of the incidence rates were computed by the SEER*Stat software under the assumption that the counts have Poisson distributions. The relative standard error is the standard error of the rate divided by the estimated rate:

$$\text{Relative Error (\%)} = [\text{Standard Error (Rate)} / \text{Rate}] \times 100\%$$

Rates with a relative error less than 30% were treated as being reliable and were tabulated. For indicator H5, all the relative standard errors in Table H5 were at most 30% and so all values were tabulated. For Table H5a (see below), some relative standard errors were greater than 30% and less than 40% and those rates were tabulated but flagged as to be interpreted with caution. Rates with a relative standard error greater than or equal to 40%, or without an estimated relative standard error, were treated as being unreliable; these values were not tabulated and were flagged as having a large uncertainty.

Rates by Age Group

For Table H5a, cancer incidence for 2004–2006 is tabulated by cancer type and age group.

The Table H5a was calculated as follows using SEER*Stat:

ICCC site recode ICD-O-3.

Click the Rate Session button "Σ".

Data tab. Select "SEER 13 Regs Research Data, Nov 2008 Sub (1992–2006) <Katrina/Rita Population Adjustment>."

Statistic tab. Check "Rates (Age – Adjusted)" and "Show Standard Errors and Confidence Intervals." For "Standard Population" select "2000 US Std Population (19 age groups – Census P25-1130)." For "Age Variable" select "Age recode with <1 year olds."

Selection tab. Check "Malignant Behavior." For "Age at Diagnosis" click the Edit button and then select the age group 0 to 19 years using "Age recode with <1 year olds" = 00, 01-04, 05-09, 10-14, and 15-19. For "Race, Sex, Year Dx, Registry, County" click the Edit button and select the years 2004–2006 using "Year of Diagnosis" = 2004, 2005 and 2006. This restricts the data to malignant cancers in children ages 0 to 19 years diagnosed in 2004–2006.

Table tab. Use the menus "File" and "Dictionary" to create a User-Defined variable AgesH5 that takes the values 0-4 years, 5-9 years, 10-14 years, 15-19 years, and 0-19 years if the "Age at Diagnosis" is in the selected age group. Age at Diagnosis is defined by the values of "Age recode with <1 year olds." In the Table tab, from the "Available Variables" panel, expand "User-Defined" and select "AgesH5." Click the Row button. From the "Available Variables" panel, expand "Site and Morphology" and select "ICCC site recode ICD-O-3." Click the Column button. This tabulates the rates by the selected age groups and cancer types.

Output tab. For "Display Rates as Cases Per" select "1,000,000."

Click the Execute button to run the analyses and create the output Rate Matrix.

Use the Matrix and Export menus to output the data to a text file. Tabulate the rates for selected cancer types and age groups.

AYA Site Recode.

The cancer types acute lymphocytic leukemia and Wilms' tumor are not specifically included in the ICCC site recode ICD-O-3 categories. Thus the incidence rate for these two cancers is calculated separately using the "AYA site recode," which codes cancer types for adolescents and young adults.

Modify the above selections as follows:

Table tab. In the Table tab, from the "Display Variables" panel, select the Column variable "ICCC site recode ICD-O-3" and click Remove. From the "Available Variables" panel, expand "Site and Morphology" and select "AYA site recode." Click the Column button. This tabulates the rates by the selected age groups for the cancer types coded in the AYA site recode.

Click the Execute button to run the analyses and create the output Rate Matrix.

Use the Matrix and Export menus to output the AYA site recode data to a text file. Tabulate the rates for selected cancer types and age groups.

Questions and Comments

Questions regarding these methods, and suggestions to improve the description of the methods, are welcome. Please use the "Contact Us" link at the bottom of any page in the America's Children and the Environment website.

Statistical Comparisons

Statistical analyses of the cancer incidence rates for each cancer type were used to determine whether the differences between rates for different demographic groups were statistically significant. Each cancer type was analyzed separately. Incidence data for malignant cancers in the years 1992 to 2006 were obtained from the SEER data set "SEER 13 Regs Research Data, Nov 2008 Sub (1992–2006) <Katrina/Rita Population Adjustment> - Linked To County Attributes - Total U.S., 1969–2006 Counties." For these analyses, the incidence counts and populations for each combination of cancer type, calendar year, age group, sex, and race/ethnicity were obtained from the SEER database using the age groups 0-4 years, 5-9 years, 10-14 years, and 15-19 years and the following race/ethnicity groups defined by the Race recode and Origin recode NHIA variables in the SEER database:

- Raceeth = WhiteNH if Race recode = White and Origin recode NHIA = Non-Spanish-Hispanic-Latino
- Raceeth = BlackNH if Race recode = Black and Origin recode NHIA = Non-Spanish-Hispanic-Latino
- Raceeth = AIANNH if Race recode = American Indian/Alaskan Native and Origin recode NHIA = Non-Spanish-Hispanic-Latino
- Raceeth = APINH if Race recode = Asian or Pacific Islander and Origin recode NHIA = Non-Spanish-Hispanic-Latino
- Raceeth = Hispanic if Origin recode NHIA = Spanish-Hispanic-Latino

Using a Poisson regression model adjusted for dispersion, the number of cases for each demographic subgroup was assumed to have a Poisson distribution with a mean equal to the population of that demographic subgroup multiplied by the rate, and a variance equal to the mean multiplied by a dispersion parameter. The logarithm of the rate was assumed to be the sum of explanatory terms for age, sex, and race/ethnicity. Using this model, the difference in the value of a rate between different demographic groups is statistically significant if the difference between the corresponding sums of explanatory terms is statistically significantly different from zero. A p-value at or below 0.05 implies that the difference is statistically significant at the 5% significance level. No adjustment is made for multiple comparisons.

For each type of comparison, we present unadjusted and adjusted analyses. The unadjusted analyses compare a rate between different demographic groups. The adjusted analyses add other demographic explanatory variables to the statistical model and use the statistical model to account for the possible confounding effects of these other demographic variables. For example, the unadjusted age comparisons compare the rates between the different age groups. The adjusted analyses add sex and race/ethnicity terms to the statistical model and compare the rates between different age groups after accounting for the effects of the other demographic variables. For example, if Hispanic children tend to be younger than White non-Hispanics, and if the cancer incidence rate strongly depends on race/ethnicity only, then the unadjusted differences between age groups would be significant because of their different race/ethnicity breakdowns but the adjusted difference (taking into account race/ethnicity) would not be significant.

The statistical comparisons for each cancer type are shown in Table 1. The Against = "age" unadjusted p-value compares the rates between all the age groups for the years 2004-2006. The adjusted p-value includes adjustment terms for sex and race/ethnicity in the model. The Against = "year" unadjusted p-value compares the trends in the rates for the period 1992 to 2006 by regressing against the calendar year. Instead of using the three-year rates given in the tables and graphs, the statistical trend analysis used the annual rates to improve the precision of the trend estimates and p-values. The adjusted p-value includes adjustment terms for age, sex, and race/ethnicity in the model.

For more details on these statistical analyses, see the memorandum by Cohen (2011).[iii]

[iii] Cohen, J. 2011. *Selected statistical methods for testing for trends and comparing years or demographic groups in other ACE health-based indicators.* Memorandum submitted to Dan Axelrad, EPA, 16 June, 2011.

Table 1. Statistical significance tests comparing cancer incidence rates for children ages 0 to 19 between age groups for 2004 to 2006 and analyzing trends for 1992-2006.

Type of Cancer	From	To	Against	P-VALUES Unadjusted	Adjusted*
Acute lymphoblastic leukemia	2004	2006	age	< 0.001	< 0.001
Acute lymphoblastic leukemia	1992	2006	year	0.485	0.033
Central nervous system tumors	2004	2006	age	< 0.001	< 0.001
Central nervous system tumors	1992	2006	year	0.517	0.825
Germ cell tumors	2004	2006	age	< 0.001	< 0.001
Germ cell tumors	1992	2006	year	0.268	0.212
Soft tissue sarcomas	2004	2006	age	< 0.001	< 0.001
Soft tissue sarcomas	1992	2006	year	0.018	0.009
Hodgkin's lymphoma	2004	2006	age	< 0.001	< 0.001
Hodgkin's lymphoma	1992	2006	year	0.658	0.081
Acute myeloid leukemia	2004	2006	age	< 0.001	< 0.001
Acute myeloid leukemia	1992	2006	year	0.163	0.147
Non-Hodgkin's lymphoma	2004	2006	age	< 0.001	< 0.001
Non-Hodgkin's lymphoma	1992	2006	year	0.004	0.001
Neuroblastoma	2004	2006	age	< 0.001	< 0.001
Neuroblastoma	1992	2006	year	0.504	0.727
Malignant melanoma	2004	2006	age	< 0.001	< 0.001
Malignant melanoma	1992	2006	year	0.025	< 0.001
Thyroid carcinoma	2004	2006	age	0**	0**
Thyroid carcinoma	1992	2006	year	0.472	0.436
Osteosarcoma	2004	2006	age	< 0.001	NA***
Osteosarcoma	1992	2006	year	0.862	0.568
Wilms' tumor	2004	2006	age	< 0.001	< 0.001
Wilms' tumor	1992	2006	year	0.028	0.003
Other and unspecified carcinomas†	2004	2006	age	< 0.001	< 0.001
Other and unspecified carcinomas†	1992	2006	year	0.731	0.189
Ewing's sarcoma	2004	2006	age	< 0.001	< 0.001
Ewing's sarcoma	1992	2006	year	0.986	0.964
Burkitt's lymphoma	2004	2006	age	0.479	0.368
Burkitt's lymphoma	1992	2006	year	0.249	0.115
Hepatoblastoma	2004	2006	age	0**	0**
Hepatoblastoma	1992	2006	year	0.063	< 0.001

*For Against = "age," the p-values are adjusted for sex and race/ethnicity.
For Against = "year," the p-values are adjusted for age, sex, and race/ethnicity.
** P-value = zero because one age group has exactly zero cases.
NA***. Statistical model failed to converge.
†""Other and unspecified carcinomas" represents all carcinomas and other malignant epithelial neoplasms other than thyroid carcinoma and malignant melanoma.