Methods

Indicator

E5. Percentage of children ages 0 to 6 years regularly exposed to environmental tobacco smoke in the home, by family income, 1994, 2005, and 2010.

Summary

Since 1957, the National Center for Health Statistics, a division of the Centers for Disease Control and Prevention, has conducted the National Health Interview Survey (NHIS), a series of annual U.S. national surveys of the health status of the noninstitutionalized civilian population. This indicator shows the percentage of children ages 0 to 6 years who are exposed regularly (four or more days per week) to environmental tobacco smoke (ETS) in the home. For each household, the NHIS survey includes survey weights and demographic information for all members of the household. The responses from all the adults in the household were combined to give an overall household answer to whether or not there was regular exposure to ETS in the home. Percentages are calculated by combining positive responses for each household with the survey weights for each child in the survey. The survey weights are the annual numbers of children in the noninstitutionalized civilian population represented by each child. Table E5 reports percentages for all children and by family income for the years 1994, 2005, and 2010. Table E5a reports percentages by race/ethnicity and family income for 2010.

Data Summary

Indicator: E5. Percentage of children ages 0 to 6 years regularly exposed to environmental tobacco smoke in the home, by family income, 1994, 2005, and 2010

Time Period	1994, 2005, and 2010		
Data	Prevalence of exposure in the home to environmental tobacco smoke (ETS) for four or more days in a week in children ages 0 to 6 years		
Year	1994	2005	2010
Children	5,438	10,090	9,344
ETS exposure non-missing responses	5,387 (99%)	7,765 (77%)	6,890 (74%)
ETS exposure missing responses	51 (1%)	2,325 (23%)	2,454 (26%)

Overview of Data Files

The following files are needed to calculate this indicator. The files together with the survey documentation and SAS programs for reading in the data are available at the NHIS website: http://www.cdc.gov/nchs/nhis.htm.

- NHIS 1994: Person file personsx.asc, Year 2000 Objectives Supplement file year2000.asc. The personsx.asc file is an ASCII file containing interview data for all persons. For children ages 0 to 17 years, the responses were obtained from a knowledgeable adult family member residing in the household. The year2000.asc file is an ASCII file that contains supplementary interview data including household smoking variables. The variables needed for these analyses are age (AGE), survey weight (WTFA), whether or not someone smokes inside the home (SMOKEHOM), and the number of days per week that residents smoke in the home (NDSMOKHM).

- NHIS 2005: Person file personsx.dat, Sample Adult Cancer file cancerxx.dat, Imputed income files incmimp1.dat, incmimp2.dat, incmimp3.dat, incmimp4.dat, and incmimp5.dat. The personsx.dat file is an ASCII file containing demographic and other data for all persons living in the sampled households. For children ages 0 to 17 years, the responses were obtained from a knowledgeable adult family member residing in the household. The cancerxx.dat file is an ASCII file that contains supplementary cancer-related interview data for sampled adults including the household smoking variable LVDYSMOK that gives the number of days per week with smoking in the home. The two files were sorted using the identifier variable for the household (HHX). The values of LVDYSMOK for the sampled adult household members were combined to create a summary smoking variable SMK4DYWK for each household. The SMK4DYWK variable was merged with the personsx.dat file using the household identifier, HHX. From each of the imputed income files we need the imputed poverty income ratio (RAT_CATI), which gives the poverty income ratio category calculated from the reported exact family income, if available, or else gives the imputed category randomly generated by multiple imputation using regression models. The Person and Imputed Income files are sorted and merged using the identifiers HHX, FMX, and FPX. The other variables needed for these analyses are age (AGE_P), person survey weight (WTFA), the race (RACERPI2), the Hispanic origin (ORIGIN_I), the specific Hispanic origin (HISPAN_I), the pseudo-stratum (STRATUM), and the pseudo-PSU (PSU).

- NHIS 2010: Person file personsx.dat, Sample Adult Cancer file cancerxx.dat, Imputed income files incmimp1.dat, incmimp2.dat, incmimp3.dat, incmimp4.dat, and incmimp5.dat. The personsx.dat file is an ASCII file containing demographic and other data for all persons living in the sampled households. For children ages 0 to 17 years, the responses were obtained from a knowledgeable adult family member residing in the household. The cancerxx.dat file is an ASCII file that contains supplementary cancer-related interview data for sampled adults including the household smoking variable LVDYSMOK that gives the number of days per week with smoking in the home. The two files were sorted using the identifier variable for the household (HHX). The values of LVDYSMOK for the sampled adult household members were combined to create a summary smoking variable SMK4DYWK for each household. The SMK4DYWK variable was merged with the personsx.dat file using the household identifier, HHX. From each of the imputed income files we need the imputed poverty income ratio (POVRATI3), which gives the poverty income ratio value in thousandths calculated from the reported exact family income, if available, or else gives the imputed value randomly generated by multiple imputation using regression models. The Person and Imputed

Income files are sorted and merged using the identifiers HHX, FMX, and FPX. The other variables needed for these analyses are age (AGE_P), person survey weight (WTFA), the race (RACERPI2), the Hispanic origin (ORIGIN_I), the specific Hispanic origin (HISPAN_I), the pseudo-stratum (STRAT_P), and the pseudo-PSU (PSU_P).

National Health Interview Survey (NHIS)

Since 1957, the National Center for Health Statistics, a division of the Centers for Disease Control and Prevention, has conducted the National Health Interview Survey (NHIS), a series of annual U.S. national surveys of the health status of the noninstitutionalized civilian population.

Results are calculated from responses to the following survey questions:

In NHIS 1994:

- SMOKEHOM: "Does anyone who lives here smoke cigarettes, cigars, or pipes anywhere inside this home?" If yes, then the following question was asked:
- NDSMOKHOM: "On the average, about how many days per week do people who live here smoke anywhere inside this home?"

In NHIS 2005 and NHIS 2010:

- LIVINTRO: "In a usual week, does ANYONE who lives here, including yourself, smoke cigarettes, cigars, or pipes anywhere inside this home?" If yes, then the following question was asked:
- LVDYSMOK: "Usually, about how many days per week do people WHO LIVE here smoke anywhere INSIDE this home?"

For all three years, the questionnaire was designed so that if the first question was not answered positively, the second question about the numbers of days of smoking per week was skipped, and thus given a missing value for the response.

For each home surveyed, we assumed that there was regular exposure to ETS if one or more of the adult respondents answered the second question about the number of days of smoking per week and said that there were four or more days of smoking per week.

The NHIS uses a complex multi-stage, stratified, clustered sampling design. Certain demographic groups have been deliberately over-sampled. Oversampling is performed to increase the reliability and precision of estimates of health status indicators for these population subgroups. In 1994, Blacks were over-sampled. In 2005, Blacks and Hispanics were over-sampled. In 2010, Blacks, Hispanics, and Asians were over-sampled. The publicly released data includes survey weights to adjust for the over-sampling, non-response, and non-coverage. The statistical analyses used the survey weights (WTFA) to re-adjust the responses to represent the total national population for each year.

Race/Ethnicity and Family Income

For this indicator, the prevalence percentages were calculated for demographic strata based on family income and race/ethnicity. Family income strata were used for the main table; the supplementary table gives results for the combined stratification of family income and race/ethnicity.

The family income was characterized based on the RAT_CATI variable for 2005 and the POVRATI3 variable for 2010, which give the level of the ratio of the family income to the poverty level. The National Center for Health Statistics obtained the family income for the respondent's family during the family interview. The U.S. Census Bureau defines annual poverty level money thresholds varying by family size and composition. The poverty income ratio (PIR) is the family income divided by the poverty level for that family. For 2005, the public release variable RAT_CATI gives the value of the PIR for various ranges, Under 0.5, 0.5-0.74, 0.75 to 0.99, …, 4.50-4.99, 5.00 and Over. For 2010, the RAT_CATI category was computed from the public release variable POVRATI3 which gives the numerical value of PIR multiplied by 1,000.

Family income was stratified into the following groups:

- Below Poverty Level: PIR < 1, i.e., RAT_CATI = 1, 2, or 3.
- Between 100% and 200% of Poverty Level: 1 ≤ PIR < 2, i.e., RAT_CATI = 4, 5, 6, or 7.
- Above 200% of Poverty level: PIR ≥ 2, i.e., RAT_CATI = 8, 9, 10, 11, 12, 13 or 14.
- Above Poverty Level: PIR ≥ 1 (combines the previous two groups).
- Unknown Income: PIR is missing ("undefinable"), i.e., RAT_CATI = 96.

Approximately 30% of families did not report their exact family income in 2010. In 2010, families not reporting an exact income were first asked to report their income as the two categories above or below $50,000, and were then asked appropriate additional questions to refine the income range as either 0-$34,999, $35,000-$49,999, $50,000-74,999, $75,000-$99,999, or $100,000 and above.. In 2010, over 90% of families either gave the exact income or a categorical response.

NCHS reports[i] evidence that the non-response to the income question is related to person-level or family-level characteristics, including items pertaining to health. Therefore, treating the missing responses as being randomly missing would lead to biased estimates. To address this problem, NCHS applied a statistical method called "multiple imputation" to estimate or "impute" the family income based on the available family income and personal earnings information and on responses to other survey questions. A series of regression models were used to predict the exact family income from the available responses. Five sets of simulated family income values were generated for each family that did not report their exact family income. In this manner, NCHS generated five data sets, each containing a complete set of family income values (either the reported or the imputed values). The poverty income ratio categories or values were calculated from the income values and the family size and composition variables. An estimated

[i] "Multiple imputation of family income and personal earnings in the National Health Interview Survey: Methods and Examples," http://www.cdc.gov/nchs/data/nhis/tecdoc_2010.pdf. August, 2011.

prevalence percentage was computed for each of the five data sets. The overall estimated prevalence percentage is the arithmetic mean of the five estimates.

The poverty income ratios were calculated by NCHS using the exact family income, if available, or otherwise were calculated from the imputed family income. Among the sampled children ages 6 years and under for the year 2010, the weighted percentage of children with imputed poverty income ratios was 20%.

Race was characterized using the race variable for the 1997 OMB standards,[ii] RACERPI2. The possible values of this variable are:

- 1. White only
- 2. Black / African American only
- 3. American Indian Alaska Native (AIAN) only
- 4. Asian only
- 5. Race group not releasable
- 6. Multiple race

The Native Hawaiian or Other Pacific Islander (NHOPI) race group is not specified in the public release version due to confidentiality concerns. Respondents with the single race NHOPI have RACERPI2 = 5 and respondents of multiple races including NHOPI have RACERPI2 = 6.

The ORIGIN_I variable indicates whether or not the ethnicity is Hispanic or Latino. ORIGIN_I = 1 if the respondent is Hispanic or Latino. ORIGIN_I = 2 if the respondent is not Hispanic or Latino.

The HISPAN_I variable indicates the specific Hispanic origin or ancestry.

- 00 Multiple Hispanic
- 01 Puerto Rico
- 02 Mexican
- 03 Mexican-American
- 04 Cuban/Cuban American
- 05 Dominican (Republic)
- 06 Central or South American
- 07 Other Latin American, type not specified
- 08 Other Spanish
- 09 Hispanic/Latino/Spanish, non-specific type
- 10 Hispanic/Latino/Spanish, type refused
- 11 Hispanic/Latino/Spanish, type not ascertained
- 12 Not Hispanic/Spanish origin

[ii] Revised race standards were issued by the Office of Management and Budget in 1997 and were to be fully implemented across the federal statistical system by January 2003. Under the new standards, the minimum available race categories include: White, Black, AIAN, Asian, and Native Hawaiian or Other Pacific Islander (NHOPI). A very important change was that under the new standards, respondents may select more than one race category.

The race/ethnicity was defined based on RACERPI2, ORIGIN_I, and HISPAN_I:

Race/ethnicity:

- White non-Hispanic: RACERPI2 =1, ORIGIN_I = 2
- Black or African-American non-Hispanic: RACERPI2 = 2, ORIGIN_I = 2
- Asian non-Hispanic: RACERPI2 = 4, ORIGIN_I = 2
- Hispanic: ORIGIN_I = 1
    - Mexican: ORIGIN_I = 1 and HISPAN_I = 02, 03
    - Puerto Rican: ORIGIN_I = 1 and HISPAN_I = 01
- All Other Races: RACERPI2 = 3, 5 or 6, ORIGIN_I = 2
    - American Indian / Alaska Native, non-Hispanic: RACERPI2 = 3, ORIGIN_I = 2

The "All Other Races" category includes all other races not specified, together with those individuals who report more than one race.

Some respondents gave missing or incomplete answers to the race/ethnicity questions. In those cases NCHS applied a statistical method called "hot-deck imputation" to estimate or "impute" the race or ethnicity based on the race/ethnicity responses for other household members, if available, or otherwise based on information from other households. The NHIS variables ORIGIN_I, HISPAN_I, and RACERPI2 use imputed responses if the original answer was missing or incomplete. Among the sampled children ages 6 years and under for the year 2010, the weighted percentage of children with an imputed race or ethnicity was 11%. Among the sampled Hispanic (defined by ORIGIN_I) children ages 6 years and under for the year 2010, the weighted percentage of children with an imputed specific Hispanic origin was 2%.

Calculation of Indicator

Indicator E5 is the percentage of children ages 0 to 6 years regularly exposed to ETS inside the home. "Regularly" is interpreted as an average of four or more days per week. For the year 2005 and year 2010 data, the following calculations were applied to the publicly released data. For the year 1994 data, CDC staff performed the calculations using similar methods applied to their unreleased version of the database.

For 2005 and 2010, the NHIS question LIVINTRO asked sampled adults if anyone living in the residence smokes anywhere inside the home and, if the answer was Yes, the NHIS question LVDYSMOK asked for the average number of days per week there is smoking anywhere inside the home by anyone living in the residence. LVDYSMOK has the following values:

- Missing if LIVINTRO $\neq$ 1 ("Yes") or if LVDYSMOK not asked.
- 0, 1, 2, 3, 4, 5, 6, or 7 if 0, 1, 2, 3, 4, 5, 6, or 7 days smoking per week
- 97 if "refused"
- 98 if "not ascertained"
- 99 if "don't know"

For each sampled adult in the adult cancer file, the following ETS variable was calculated:

- ETS = No (0-3 days smoking per week) if LVDYSMOK = Missing, 0, 1, 2, 3
- ETS = Yes (4-7 days smoking per week) if LVDYSMOK = 4, 5, 6, 7
- ETS = Missing (unknown) if LVDYSMOK = 97, 98, 99

Since the responses from different adults in the household sometimes differ, these ETS responses were combined to give a summary smoking variable SMK4DYWK for each household, identified by the variable HHX. SMK4DYWK indicates whether there is smoking anywhere inside the home on four or more days per week:

- SMK4DYWK = 1 (Yes) if ETS = Yes for one or more adults in the household
- SMK4DYWK = 2 (No) if ETS = Yes for zero adults in the household and ETS = No for one or more adults in the household
- SMK4DYWK = Missing if ETS = Missing for all adults in the household

The value of SMK4DYWK for each household was merged into the personsx.dat Person file using the household identifier variable HHX.

The rest of the calculation uses the Person file data for every child ages 0 to 6 years. Note that this sample of children includes all children ages 0 to 6 years in each sampled household. This is a larger sample than the children in the NHIS Sample Child file, which has only one child per family. To illustrate the calculations we will apply them to children of all incomes in 2010. We have rounded all the numbers to make the calculations easier:

We begin with all the non-missing responses to the SMK4DYWK question for children ages 0 to 6 years. Assume for the sake of simplicity that Yes or No responses were available for every sampled child. Each sampled child has an associated survey weight that estimates the annual number of children represented by that sampled child. For example, the first response for a child aged 6 or under was No with a survey weight of 4,000, and so represents 4,000 children ages 6 years or under. A second child aged 6 years or under had a No response with a survey weight of 2,000, and so represents 2,000 children ages 6 years or under. A third child aged 6 years or under had a Yes response with a survey weight of 5,000, and so represents 5,000 children ages 6 years or under. The total of the survey weights for the sampled children equals 30 million, the total U.S. population of children ages 6 years or under in 2010.

To calculate the proportion of children exposed to ETS, we can use the survey weights to expand the data to the total 2010 U.S. population of 30 million children ages 0 to 6 years. We have 4,000 No responses from the first child, 2,000 No responses from the second child, 5,000 Yes responses from the third child, and so on. Of these 30 million responses, a total of 1.8 million responses are Yes and the remaining 28.2 million are No. Thus 1.8 million of the 30 million children were exposed to ETS more than four days per week, giving a percentage of 6% (1.8/30).

In reality, the calculations need to take into account that Yes or No responses were not reported for every respondent, and they need to use exact rather than rounded numbers. There were non-

missing SMK4DYWK responses for 6,890 of the 9,344 sampled children ages 0 to 6 years. The survey weights for all 9,344 sampled children add up to 29.97 million, the total 2010 U.S. population of children ages 0 to 6 years. The survey weights for the 6,890 sampled children with non-missing responses add up to 22.14 million. Thus the available data represent 22.14 million children, which is 74% of the 2010 U.S. population of children ages 0 to 6 years. The survey weights for the Yes responses add up to 1.36 million, which is 6.1% of the population with responses (1.36 million/22.14 million = 6%). Thus we divide the sum of the weights for participants with Yes responses by the sum of the weights for participants with non-missing responses. These calculations assume that the sampled children with non-missing responses are representative of the children with missing responses.

For calculation of prevalence by income group, we use the five sets of imputed income values, which each give different results. Suppose we wish to estimate the percentage of all children ages 0 to 6 years below the poverty level that were regularly exposed to ETS in the home. Using the above calculation method applied for children ages 0 to 6 years below the poverty level, the proportions for the five sets of imputed values are: 10.24%, 10.33%, 10.21%, 10.27%, and 10.12%. The estimated proportion of children ages 0 to 6 years below the poverty level regularly exposed to ETS in the home is given by the average of the five estimates, (10.24 + 10.33 + 10.21 + 10.27 + 10.12) / 5 = 10.23%.

Equations

The following equations give the mathematical calculations for the example of all children ages 0 to 6 years below the poverty level. Let w(i) denote the survey weight for the i'th surveyed child of ages 0 to 6 years. Exclude any surveyed children with a response other than Yes or No to the SMK4DYWK variable. Let the response indicator c(i) = 1 if the i'th surveyed child had a Yes response and let c(i) = 0 if the i'th surveyed child had a No response. Let the income indicator d(i, j) = 1 if the i'th surveyed child was below the poverty level according to the j'th set of imputed values and let d(i, j) = 0 if the i'th surveyed child was not below the poverty level according to the j'th set of imputed values.

1. Fix j = 1, 2, 3, 4 or 5. Sum (over i) all the survey weights multiplied by the income indicators to get the total weight W(j) for set j:

$$W(j) = \Sigma\, w(i) \times d(i, j)$$

2. Fix j = 1, 2, 3, 4 or 5. Sum (over i) all the survey weights multiplied by the response indicators and multiplied by the income indicators to get the total weight D(j) for set j for children below the poverty level with a Yes response:

$$D(j) = \Sigma\, w(i) \times c(i) \times d(i, j)$$

3. Divide D(j) by W(j) to get the percentage of children regularly exposed to ETS in the home in set j:

$$\text{Percentage } (j) = (D(j) / W(j)) \times 100\%$$

4. Average the percentages across the 5 sets to get the estimated percentage of children regularly exposed to ETS in the home:

$$\text{Percentage} = [\text{Percentage (1)} + \text{Percentage (2)} + \text{Percentage (3)} + \text{Percentage (4)} + \text{Percentage (5)}] / 5$$

If the demographic group of interest includes all incomes, then the percentages will be equal for all five sets of imputed values, so the calculation in steps 1 to 3 need only be done for j =1, and step 4 is not required.

Relative Standard Error

The uncertainties of the percentages were calculated using SUDAAN® (Research Triangle Institute, Research Triangle Park, NC 27709) statistical survey software. SUDAAN was used to calculate the estimated percentages and the standard errors of the estimated percentages. The standard error is the estimated standard deviation of the percentage, and this depends upon the survey design. The standard error calculation also incorporates the extra uncertainty due to the multiple imputations of the income variables (based on the variation between the estimated percentages from each of the five sets of imputations). For this purpose, the public release version of NHIS includes the variables STRATUM and PSU, which are the Masked Variance Unit pseudo-stratum and pseudo-primary sampling unit (pseudo-PSU). For approximate variance estimation, the survey design can be approximated as being a stratified random sample with replacement of the pseudo-PSUs from each pseudo-stratum; the true stratum and PSU variables are not provided in the public release version to protect confidentiality.

The relative standard error is the standard error divided by the estimated percentage:

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

Percentages with a relative standard error less than 30% were treated as being reliable and were tabulated. Percentages with a relative standard error greater than or equal to 30% but less than 40% were treated as being unstable; these values were tabulated but were flagged as being unstable. Percentages 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.

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

For this indicator, the question of interest for each child is whether or not they were regularly exposed to ETS in the home. Statistical analyses of the percentages of children with a positive response to the question of interest were used to determine whether the differences between percentages for different demographic groups were statistically significant. Using a logistic regression model, the logarithm of the odds that a given child has a positive response is assumed to be the sum of explanatory terms for the child's age group, sex, income group and/or race/ethnicity. The odds of a positive response is the probability of a positive response divided by the probability of a negative response. Thus if two demographic groups have similar (or equal) probabilities of a positive response, then they will also have similar (or equal) values for the logarithm of the odds. Using this model, the difference in the percentage between different demographic groups is statistically significant if the difference between the corresponding sums of explanatory terms is statistically significantly different from zero. The uncertainties of the regression coefficients were calculated using SUDAAN® (Research Triangle Institute, Research Triangle Park, NC 27709) statistical survey software to account for the survey weighting and design. 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 these statistical analyses we used two income groups, below poverty level, and at or above poverty level. The small number of children with unknown (and unimputed) incomes were included in the at or above poverty level group. For the main analyses we also used five race/ethnicity groups: White non-Hispanic; Black non-Hispanic; Asian non-Hispanic; Hispanic; All Other Races. In addition, for specific comparisons between the Mexican and Puerto Rican subgroups, we applied a similar statistical analysis using three ethnicity groups: Mexican; Puerto Rican; Other Hispanic or Non-Hispanic. We did not include the age group in these analyses.

For each type of comparison, we present unadjusted and adjusted analyses. The unadjusted analyses directly compare a percentage 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 use and compare the percentages between different race/ethnicity pairs. The adjusted analyses add sex and income terms to the statistical model and compare the percentages between different race/ethnicity pairs after accounting for the effects of the other demographic variables. For example, if White non-Hispanics tend to have higher family incomes than Black non-Hispanics, and if the prevalence of exposure to ETS strongly depends on family income only, then the unadjusted differences between these two race/ethnicity groups could be significant but the adjusted difference (taking into account income) may not be significant.

Comparisons of the prevalence of regular exposure to ETS in the home in children ages 0 to 6 years between pairs of race/ethnicity groups are shown in Table 1. For the unadjusted "All incomes" comparisons, the only explanatory variables are terms for each race/ethnicity group. For these unadjusted comparisons, the statistical tests compare the percentage for each pair of race/ethnicity groups. For the adjusted "All incomes (adjusted for sex, income)" comparisons, the explanatory variables are terms for each race/ethnicity group together with terms for each sex

and income group. For these adjusted comparisons, the statistical test compares the pair of race/ethnicity groups after accounting for any differences in the age, sex and income distributions between the race/ethnicity groups.

In Table 1, for the unadjusted "Below Poverty Level" and "At or Above Poverty Level" comparisons, the only explanatory variables are terms for each of the 10 race/ethnicity/income combinations (combinations of five race/ethnicity groups and two income groups). For example, in row 1, the p-value for "Below Poverty Level" compares White non-Hispanics below the poverty level with Black non-Hispanics below the poverty level. The same set of explanatory variables are used in Table 2 for the unadjusted comparisons between one race/ethnicity group below the poverty level and the same race/ethnicity group at or above the poverty level. The corresponding adjusted analyses include extra explanatory variables for sex, so that race/ethnicity/income groups are compared after accounting for any differences due to sex.

Additional comparisons are shown in Table 3. The Against = "sex" unadjusted p-value compares the percentages for boys and girls. The adjusted p-value includes adjustment terms for income, and race/ethnicity in the model. The Against = "income" unadjusted p-value compares the percentages for those below poverty level with those at or above poverty level. The adjusted p-value includes adjustment terms for sex, and race/ethnicity in the model. For more details on these statistical analyses, see the memorandum by Cohen (2010).[iii]

Table 1. Statistical significance tests comparing the percentages of children ages 0 to 6 years with regular exposure to environmental tobacco smoke in the home, between pairs of race/ethnicity groups, for 2010.

			P-VALUES					
Variable	First race/ethnicity group	Second race/ethnicity group*	All incomes	All incomes (adjusted for sex, income)	Below Poverty Level	Below Poverty Level (adjusted for sex)	At or Above Poverty Level	At or Above Poverty Level (adjusted for sex)
ETS	White non-Hispanic	Black non-Hispanic	0.454	0.270	0.006	0.005	0.204	0.207
ETS	White non-Hispanic	Asian non-Hispanic	0.001	0.001	0.033	0.033	0.015	0.015
ETS	White non-Hispanic	Hispanic	< 0.001	< 0.001	< 0.001	< 0.001	0.002	0.002
ETS	White non-Hispanic	Other	0.347	0.568	0.306	0.297	0.177	0.176
ETS	Black non-Hispanic	Asian non-Hispanic	< 0.001	0.004	0.229	0.229	0.005	0.005
ETS	Black non-Hispanic	Hispanic	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
ETS	Black non-Hispanic	Other	0.685	0.249	0.500	0.506	0.662	0.658
ETS	Asian non-Hispanic	Hispanic	0.430	0.903	0.588	0.589	0.349	0.349
ETS	Asian non-Hispanic	Other	< 0.001	0.001	0.139	0.140	0.004	0.004
ETS	Hispanic	Other	< 0.001	< 0.001	0.001	0.001	0.001	0.001

[iii] Cohen, J. 2010. *Selected statistical methods for testing for trends and comparing years or demographic groups in ACE NHIS and NHANES indicators.* Memorandum submitted to Dan Axelrad, EPA, 21 March, 2010.

			P-VALUES					
Variable	First race/ethnicity group	Second race/ethnicity group*	All incomes	All incomes (adjusted for sex, income)	Below Poverty Level	Below Poverty Level (adjusted for sex)	At or Above Poverty Level	At or Above Poverty Level (adjusted for sex)
ETS	Mexican	Puerto Rican	0.106	0.125	0.333	0.333	0.166	0.162

* "Other" represents the "All Other Races" category, which includes all other races not specified, together with those individuals who report more than one race.

Table 2. Statistical significance tests comparing the percentages of children ages 0 to 6 years with regular exposure to environmental tobacco smoke in the home, between those below poverty level and those at or above poverty level, for 2010.

		P-Values for difference between income level	
Variable	Population*	Unadjusted	Adjusted (for sex)**
ETS	All	< 0.001	< 0.001
ETS	White non-Hispanic	< 0.001	< 0.001
ETS	Black non-Hispanic	0.175	0.172
ETS	Asian non-Hispanic	0.231	0.231
ETS	Hispanic	0.696	0.694
ETS	Other	0.284	0.287
ETS	Mexican	0.536	0.534
ETS	Puerto Rican	0.811	0.816

* "Other" represents the "All Other Races" category, which includes all other races not specified, together with those individuals who report more than one race.
** Comparison for "All" is adjusted for sex and race/ethnicity; comparisons for race/ethnicity categories are adjusted for sex.

Table 3. Other statistical significance tests comparing the percentages of children ages 0 to 6 years with regular exposure to environmental tobacco smoke in the home, for 2010.

		P-VALUES	
Variable	Against	Unadjusted	Adjusted*
ETS	income	< 0.001	< 0.001
ETS	sex	0.578	0.453

*For Against = "sex," the comparison is between boys and girls, and the p-values are adjusted for race/ethnicity and income.
For Against = "income," the comparison is between those below the poverty level and those at or above the poverty level, and the p-values are adjusted for sex and race/ethnicity.

Methods

Indicator

E6. Percentage of children ages 0 to 5 years living in homes with interior lead hazards, 1998-1999 and 2005-2006.

Summary

The United States Department of Housing and Urban Development (HUD) has conducted two nationally representative surveys of housing in the United States to assess children's potential household exposure to lead and other contaminants. The American Healthy Homes Survey (AHHS) was conducted from 2005–2006 to update the National Survey of Lead and Allergens in Housing (NSLAH), which was conducted from 1998–1999. AHHS also included measurements of arsenic, pesticides, and mold.

This indicator gives the percentages of children ages 0 to 5 years living in homes with interior lead hazards, either interior lead dust or interior deteriorated lead-based paint. Under the Lead Safe Housing Act, a significant lead-based paint hazard is the presence of deteriorating lead-based paint, lead-contaminated dust, or lead-contaminated soil above federal standards. For lead-contaminated dust, there are separate standards for dust on the floor and dust on windowsills. Floor dust samples should not have more than 40 micrograms of lead per square foot **(**$\mu g/ft^2$) and window dust samples should not have more than 250 $\mu g/ft^2$. Current federal health-based standards qualify a significantly deteriorated lead-based paint hazard as the deterioration of an area of lead-based paint greater than 20 square feet (exterior) and 2 square feet (interior) for large-surface items, such as walls and doors; or damage to more than 10% of the total surface area of small-surface components—such as windowsills, baseboards, and trim—with lead-based paint.

For each home, the NSLAH and AHHS surveys include information on the dust lead loadings of interior lead dust measured on surface wipes, and X-ray fluorescence measurements of lead in paint. The surveys also include survey weights and demographic information for all persons living in that home. For each home, the presence or absence of interior lead dust or interior deteriorated lead-based paint was determined. Percentages of children ages 0 to 5 years living in homes with interior lead hazards are calculated by combining the interior lead hazard indicators for each home with the numbers of children ages 0 to 5 years and the survey weights for each home in the survey. The survey weights are the numbers of U.S. homes represented by each home surveyed.

Data Summary

Indicator: E6. Percentage of children ages 0 to 5 years living in homes with interior lead hazards, 1998-1999 and 2005-2006.

Data	Prevalence of exposure in the home to interior lead hazards in children ages 0 to 5 years.	
Time Period	1998-1999 or 2005-2006.	
Years	1998-1999 (NSLAH)	2005-2006 (AHHS)
Homes with non-missing data	831	1,131
Homes with non-missing data and one or more children ages 0 to 5 years	184 (22%)	206 (18%)

Overview of Data Files

The following files are needed to calculate this indicator. The files were obtained directly from the U.S. Department of Housing and Urban Development (HUD).[iv]

- NSLAH: Derived data file blenplay.sd2, Resident file RES03_A.sd2, Jackknife weight file jknfac.dat. The blenplay.sd2 file is a SAS dataset file with home measurement data including the housing unit ID code (HUID), interior lead dust indicator (LD99INT), interior deteriorated lead-based paint indicator (DLP99INT), home survey weight (FINDUWT), and the 99 jackknife survey weights (FINDUW1, FINDUW2, … , FINDUW99). The RES03_A.sd2 file is a SAS dataset file with resident demographic information including the HUID and the age (Q25C) of all residents. The jknfac.dat file is an ASCII file that lists the 99 jackknife factors used for estimating uncertainties.

- AHHS: Laboratory wipe data file wipe_lab.sas7bdat, X-ray fluorescence data file xrf_lbp.sas7bdat, People file people_tab.sas7bdat, Weights file weights.sas7bdat, Jackknife weight file jknfactors.txt. The wipe_lab.sas7bdat file is a SAS dataset file with surface wipe home measurement data including the dwelling unit ID code (DUID), the location code (LOCATION), and the dust lead loading (LEAD_RESULT_BYAREA). The xrf_lbp.sas7bdat file is a SAS dataset file with X-ray fluorescence data on paint including the DUID, lead level (PBL), room type (ROOMTYPE), and level of deterioration (DET). The people_tab.sas7bdat file is a SAS dataset file with resident demographic information including the DUID and the age (P38C) of all residents. The weights.sas7bdat file is a SAS dataset file with the home survey weight (RPL000), and the 116 jackknife survey weights (RPL001, RPL002, … , RPL116). The jknfactors.txt file is an ASCII file that lists the 116 jackknife factors used for estimating uncertainties.

[iv] Peter Ashley, U.S. Department of Housing and Urban Development, Office of Healthy Homes and Lead Hazard Control, 202-402-7595, peter.j.ashley@hud.gov

National Survey of Lead and Allergens in Housing (NSLAH)

In 1998-1999, the U.S. Department of Housing and Urban Development (HUD) and other sponsors conducted the National Survey of Lead and Allergens in Housing (NSLAH), a U.S. national survey of lead dust, lead-based paint, lead in soil, and other contaminants in homes. (An augmentation of the soil sampling was carried out in 2000.) The survey included the determinations of the presence or absence of interior lead dust and the presence or absence of interior deteriorated lead-based paint defined as follows:

- Interior Lead Dust, LD99INT: Are there one or more floor wipe samples taken inside the home that have a dust lead loading of 40 $\mu g/ft^2$ or greater? Are there one or more windowsill wipe samples taken inside the home that have a dust lead loading of 250 $\mu g/ft^2$ or greater? These criteria are from the Lead Safe Housing Rule of 1999. If the answer to one or both questions is positive, then LD99INT = 1, indicating the presence of interior lead dust, otherwise LD99INT = 0, indicating the absence of interior lead dust.

- Interior Deteriorated Lead-Based Paint, DLP99INT: Are there one or more X-ray fluorescence readings taken inside the home that have a reading of 1.0 mg/ft^2 of lead or greater and have a non-zero measured percentage of deterioration? This criterion is from the Lead Safe Housing Rule of 1999. If the answer to this question is positive, then DLP99INT = 1, indicating the presence of interior deteriorated lead-based paint, otherwise DLP99INT = 0, indicating the absence of interior deteriorated lead-based paint.

For these analyses, we also computed a lead hazard indicator for the presence or absence of either interior lead dust or interior deteriorated lead-based paint:

- Either Interior Lead Dust or Interior Deteriorated Lead-Based Paint: Either the home has interior lead dust, or the home has interior deteriorated lead-based paint, or both.

The NSLAH used a complex multi-stage, stratified, clustered sampling design to select the homes. The data include home survey weights to adjust for the sampling design. The statistical analyses used the home survey weights (FINDUWT) to readjust the response indicators to represent the total national population of homes. The statistical analysis also adjusted the data by weighting each home by the number of resident children ages 0 to 5 years, using the resident age data for that home. Using both the survey weight and the number of children to adjust the data readjusts the response indicators to represent the total national population of children ages 0 to 5 years.

American Healthy Homes Survey (AHHS)

In 2005-2006, the U.S. Department of Housing and Urban Development (HUD) and other sponsors conducted the American Healthy Homes Survey (AHHS), a U.S. national survey of lead dust, lead-based paint, lead in soil, and other contaminants in homes. The survey included the determinations of the presence or absence of interior lead dust and the presence or absence of interior deteriorated lead-based paint defined as follows:

- Interior Lead Dust: Are there one or more floor wipe samples taken inside the home that have a dust lead loading of 40 µg/ft^2 or greater? This criterion holds if LOCATION = "F" and LEAD_RESULT_BYAREA ≥ 40 are both true for one or more wipe samples. Are there one or more windowsill wipe samples taken inside the home that have a dust lead loading of 250 µg/ft^2 or greater? This criterion holds if LOCATION = "WS" and LEAD_RESULT_BYAREA ≥ 250 are both true for one or more wipe samples. These criteria are from the Lead Safe Housing Rule of 1999. If the answer to one or both questions is positive, then interior lead dust is present in the home. Otherwise interior lead dust is absent.

- Interior Deteriorated Lead-Based Paint: Are there one or more X-ray fluorescence readings taken inside the home that have a reading of 1.0 mg/cm^2 of lead or greater and have a non-zero measured percentage of deterioration? This criterion holds if ROOMTYPE ≠ "EXT" (external), PBL ≥ 1, and DET ≠ "0%" all apply for one or more readings. This criterion is from the Lead Safe Housing Rule of 1999. If the answer to this question is positive, then interior deteriorated lead-based paint is present in the home. Otherwise interior deteriorated lead-based paint is absent.

- Either Interior Lead Dust or Interior Deteriorated Lead-Based Paint: Either the home has interior lead dust, or the home has interior deteriorated lead-based paint, or both.

The AHHS used a complex multi-stage, stratified, clustered sampling design to select the homes. The data includes home survey weights to adjust for the sampling design. The statistical analyses used the home survey weights (RPL000) to readjust the response indicators to represent the total national population of homes. The statistical analysis also adjusted the data by weighting each home by the number of resident children ages 0 to 5 years, using the resident age data for that home. Using both the survey weight and the number of children to adjust the data readjusts the response indicators to represent the total national population of children ages 0 to 5 years.

Calculation of Indicator

Indicator E6 is the percentage of children ages 0 to 5 years living in homes with interior lead hazards. The percentages were computed for the following interior lead hazards:

- Interior Lead Dust
- Interior Deteriorated Lead-Based Paint
- Either Interior Lead Dust or Interior Deteriorated Lead-Based Paint

For each home surveyed in NSLAH or AHHS, the presence or absence of an interior lead hazard was determined as described above.

To illustrate the calculations, we will apply them to the NSLAH surveyed homes in 1998-1999 for the interior lead hazard of Interior Lead Dust. A Yes response for a home is when interior lead dust is present in the home. A No response for a home is when interior lead dust is absent in the home.

Each sampled home has an associated home survey weight that estimates the national number of homes (in thousands) represented by that sampled home. For example, the first response was No with a survey weight of 160, and so represents 160 thousand homes. There were zero (0) children ages 0 to 5 years residing at that home. Thus the first home represents 160 × 0 = 0 thousand children ages 0 to 5 years. A second home had a No response with a survey weight of 245, and so represents 245 thousand homes. There was 1 child ages 0 to 5 years residing at that home. Thus the second home represents 245 × 1 = 245 thousand children ages 0 to 5 years. A third home had a Yes response with a survey weight of 188, and so represents 188 thousand homes. There were 2 children ages 0 to 5 years residing at that home. Thus the third home represents 188 × 2 = 376 thousand children ages 0 to 5 years. The total of the survey weights for the sampled homes equals 95,688, so that the data represent a total of 95,688 thousand U.S. homes in 1998-1999.

To calculate the proportion of children ages 0 to 5 years living in homes with interior lead dust, we can use the survey weights to expand the data to the total U.S. population of 95,688 thousand U.S. homes in 1998-1999. The first sampled home represents 160 thousand homes and 0 children ages 0 to 5 years. The second home represents 245 thousand homes and 245 thousand children ages 0 to 5 years. The third home represents 188 thousand homes and 376 thousand children ages 0 to 5 years. The entire sample of homes represents a total of 95,688 thousand U.S. homes and 22,638 thousand children ages 0 to 5 years. We have 0 children with Yes responses from the first home, 0 children with Yes responses from the second home, 376 thousand children with Yes responses from the third home, and so on. Of the 22,638 thousand children ages 0 to 5 years, there are a total of 3,661 thousand children with Yes responses. Thus 3,661 thousand of the 22,638 thousand children ages 0 to 5 years were living in homes with interior lead dust, giving a percentage of 16.2% (3,661/22,638).

In this calculation we included sampled homes with zero children ages 0 to 5 years, which each contribute 0 children with Yes responses and 0 children with No responses. Exactly the same calculation could be done using only the sampled homes with one or more children ages 0 to 5 years and the same result would be obtained.

Equations

The following equations give the mathematical calculations for the example of interior lead dust. Let w(i) denote the survey weight for the i'th surveyed home. Let c(i) denote the number of children ages 0 to 5 years for the i'th surveyed home. Let the response indicator d(i) = 1 if the i'th surveyed home had a Yes response and let d(i) = 0 if the i'th surveyed home had a No response.

1. Sum (over i) all the survey weights multiplied by the number of children ages 0 to 5 years to get the total number of children C (in thousands):

$$C = \Sigma\, w(i) \times c(i)$$

2. Sum (over i) all the survey weights multiplied by the number of children ages 0 to 5 years and multiplied by the response indicators to get D, the total number of children ages 0 to 5 years with a Yes response (in thousands):

$$D = \Sigma\, w(i) \times c(i) \times d(i)$$

3. Divide D by C to get the percentage of children living in homes with interior lead dust:

$$\text{Percentage} = (D / C) \times 100\%$$

Relative Standard Error

The uncertainties of the percentages were calculated using SUDAAN® (Research Triangle Institute, Research Triangle Park, NC 27709) statistical survey software. SUDAAN was used to calculate the estimated percentages and the standard errors of the estimated percentages. The standard error is the estimated standard deviation of the percentage, and this depends upon the survey design. For this purpose, the data sets from NSLAH and AHHS each include sets of jackknife weights and jackknife factors. For NSLAH, the data set was subdivided into 99 "variance units," each consisting of one or more primary sampling units or pseudo-primary sampling units. To use the jackknife method, one variance unit at a time is dropped from the sample and the weights of the remaining variance units are multiplied by a reweighting factor to get a set of jackknife weights for that replicate. Thus you get one replicate for each variance unit that gets dropped. The jackknife weights are used in place of the original survey weights to get 99 estimated percentages, one for each replicate. The 99 jackknife factors are used together with the original estimated percentage and the 99 jackknife estimated percentages to estimate the variance and standard error of the percentage using a standard formula. For AHHS the same approach was used with 116 replicates.

The relative error is the standard error divided by the estimated percentage:

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

Percentages with a relative error less than 30% were treated as being reliable and were tabulated. Percentages with a relative error greater than or equal to 30% but less than 40% were treated as being unstable; these values were tabulated but were flagged to be interpreted with caution. Percentages with a relative error greater than or equal to 40% or missing were treated as being unreliable; these values were not tabulated and were flagged as having a large uncertainty. For the NSLAH and AHHS data, the percentages for the indicator Dust 1 all had relative errors less than 30%.

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 percentages of children living in homes with interior lead hazards were used to determine whether the differences between the NSLAH and AHHS percentages were statistically significant, which is the same as determining whether the trend between the NSLAH and AHHS surveys was statistically significant. For this calculation, we used the estimated percentages and their standard errors, calculated as described above in the subsection "Relative Standard Error." A z-statistic was computed by dividing the difference between the percentages by the estimated standard error of the difference:

$$z = \frac{[\text{percentage (NSLAH)} - \text{percentage (AHHS)}]}{\sqrt{\{\text{standard error (NSLAH)}^2 + \text{standard error (AHHS)}^2\}}}$$

The p-value for z is calculated using the standard normal distribution as twice the probability that a standard normal variate exceeds |z|. A p-value at or below 0.05 implies that the difference is statistically significant at the 5% significance level.[v] No adjustment is made for multiple comparisons.

The p-values are tabulated in Table 1.

Table 1. Statistical significance tests comparing the percentages of children ages 0 to 5 years living in homes with interior lead hazards, between the NSLAH (1998-1999) and AHHS (2005-2006).

Interior Lead Hazard	**P-value**
Interior lead dust	0.396
Interior deteriorated lead-based paint	0.739
Either interior lead dust or interior deteriorated lead-based paint	0.157

[v] For this method it is assumed that the two surveys were statistically independent, that the differences are approximately normally distributed, and that the uncertainties in the standard errors can be treated as negligible. An adjustment for the degrees of freedom was not applied since the NSLAH survey had 52 degrees of freedom for estimating the standard error and the AHHS survey had 61 degrees of freedom.