Methods

Indicator

B1. Lead in children ages 1 to 5 years: Median and 95th percentile concentrations in blood, 1976-2012

B2. Lead in children ages 1 to 5 years: Median concentrations in blood, by race/ethnicity and family income, 2007-2012

Summary

Since the 1970s, the National Center for Health Statistics, a division of the Centers for Disease Control and Prevention, has conducted the National Health and Nutrition Examination Surveys (NHANES), a series of U.S. national surveys of the health and nutrition status of the noninstitutionalized civilian population. The National Center for Environmental Health at CDC measures environmental chemicals in blood and urine samples collected from NHANES participants.[i] Indicators B1 and B2 use blood lead measurements in children ages 1 to 5 years. NHANES II (1976-1980) included blood lead data for children from six months to 5 years. NHANES III (1988-1994) and the NHANES 1999-2000, 2001-2002, 2003-2004, 2005-2006, 2007-2008, 2009-2010, and 2011-2012 surveys included blood lead data for children ages 1 to 5 years.

Indicator B1 is the median and 95th percentile concentrations of blood lead for children ages 1 to 5 years for each NHANES survey period. The median is the estimated concentration such that 50% of all noninstitutionalized civilian children ages 1 to 5 years during the survey period have blood lead concentrations below this level. The 95th percentile is the estimated concentration such that 95% of all noninstitutionalized civilian children ages 1 to 5 years during the survey period have blood lead concentrations below this level. Indicator B1 gives the median and 95th percentile concentrations of blood lead for children ages 1 to 5 years for each NHANES survey period.

Indicator B2 is the median concentrations of blood lead for children ages 1 to 5 years for 2009-2012, stratified by race/ethnicity and family income. Table B1a presents the median and 95th percentile concentrations of blood lead for children ages 1 to 17 years for 2011-2012, stratified by age group. Table B2a presents the 95th percentile concentrations of blood lead for children ages 1 to 5 years for 2009-2012, stratified by race/ethnicity and family income. Table B2b presents the median concentrations of blood lead for children ages 1 to 5 years for 1991-1994, stratified by race/ethnicity and family income. The survey data were weighted to account for over-sampling, non-response, and non-coverage.

[i] Centers for Disease Control and Prevention. 2009. Fourth National Report on Human Exposure to Environmental Chemicals. Atlanta, GA. Available at: www.cdc.gov/exposurereport.

Data Summary

Indicator	B1. Lead in children ages 1 to 5 years: Median and 95th percentile concentrations in blood, 1976-2012. B2. Lead in children ages 1 to 5 years: Median concentrations in blood, by race/ethnicity and family income, 2009-2012.								
Time Period	1976-2012								
Data	Blood lead								
Years (1976-2010)	1976-1980	1988-1991	1991-1994	1999-2000	2001-2002	2003-2004	2005-2006	2007-2008	2009-2010
Limits of Detection (µg/dL)*	Not reported	1	1	0.3	0.3	0.3	0.3 or 0.25	0.25	0.25
Number of Values	3,770	3,008	2,714	1,085	1,330	1,267	1,410	1,231	1,295
Number of Non-missing Values**	2,345 (62%)	2,203 (73%)	2,367 (87%)	723 (67%)	898 (68%)	911 (72%)	968 (69%)	817 (66%)	836 (65%)
Number of Missing Values**	1,425 (38%)	805 (27%)	347 (13%)	362 (33%)	432 (32%)	356 (28%)	442 (31%)	414 (34%)	459 (35%)
Percentage Below Limit of Detection***		4	8	0	1	0	0	0	0
Years (2011-2012)	2011-2012								
Limits of Detection (µg/dL)*	0.25								
Number of Values	1,203								
Number of Non-missing Values**	713 (59%)								
Number of Missing Values**	490 (41%)								
Percentage Below Limit of Detection***	1								

* The Limit of Detection (LOD) is defined as the level at which the measurement has a 95% probability of being greater than zero.

**Non-missing values include those below the analytical LOD, which are reported as LOD/√2. Missing values are the number of sampled children ages 1 to 5 years in the Mobile Examination Center (MEC) sample that have no value reported for the particular variable used in calculating the indicator.

***This percentage is survey-weighted using the NHANES MEC survey weights for the given period.

Overview of Data Files

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

- NHANES II: Hematology and Biochemistry file DU5411.txt. This text file contains the measured blood lead (N2LB0409), age in years (N2LB0190), sex (N2LB0055), race/ethnicity (N2LB0060), poverty income ratio (N2LB0210), the lead final examined weight (N2LB0300), the pseudo-stratum (N2LB0324), and the pseudo-PSU (N2LB0326).

- NHANES III: Laboratory file LAB.DAT. This text file contains the measured blood lead (PBP), age in months (MXPAXTMR), sex (HSSEX), race/ethnicity (DMARETHN), poverty income ratio (DMPPIR), NHANES III Phase (SDPPHASE), the Final Examination (Mobile Examination Center (MEC) only) weights (WTPFEX1 for Phase I and WTPFEX2 for Phase 2), the pseudo-stratum codes (SDPSTRA1 for Phase 1 and SDPSTRA2 for Phase 2), and the pseudo-PSU codes (SDPPSU1 for Phase 1 and SDPPSU2 for Phase 2).

- NHANES 1999-2000: Demographic file demo.xpt. Laboratory file lab06.xpt. The demographic file demo.xpt is a SAS transport file that contains the subject identifier (SEQN), age (RIDAGEYR), sex (RIAGENDR), race/ethnicity (RIDRETH1), poverty income ratio (INDFMPIR), pseudo-stratum (SDMVSTRA), pseudo-PSU (SDMVPSU), and the two year MEC weight (WTMEC2YR). The laboratory file lab06.xpt contains SEQN and the blood lead (LBXBPB). The two files are merged using the common variable SEQN.

- NHANES 2001-2002: Demographic file demo_b.xpt. Laboratory file l06_b.xpt. The demographic file demo_b.xpt is a SAS transport file that contains the subject identifier (SEQN), age (RIDAGEYR), sex (RIAGENDR), race/ethnicity (RIDRETH1), poverty income ratio (INDFMPIR), pseudo-stratum (SDMVSTRA), pseudo-PSU (SDMVPSU), and the two year MEC weight (WTMEC2YR). The laboratory file l06_b.xpt contains SEQN and the blood lead (LBXBPB). The two files are merged using the common variable SEQN.

- NHANES 2003-2004: Demographic file demo_c.xpt. Laboratory file l06bmt_c.xpt. The demographic file demo_c.xpt is a SAS transport file that contains the subject identifier (SEQN), age (RIDAGEYR), sex (RIAGENDR), race/ethnicity (RIDRETH1), poverty income ratio (INDFMPIR), pseudo-stratum (SDMVSTRA), pseudo-PSU (SDMVPSU), and the two year MEC weight (WTMEC2YR). The laboratory file l06bmt_c.xpt contains SEQN and the blood lead (LBXBPB). The two files are merged using the common variable SEQN.

- NHANES 2005-2006: Demographic file demo_d.xpt. Laboratory file pbcd_d.xpt. The demographic file demo_d.xpt is a SAS transport file that contains the subject identifier (SEQN), age (RIDAGEYR), sex (RIAGENDR), race/ethnicity (RIDRETH1), poverty income ratio (INDFMPIR), pseudo-stratum (SDMVSTRA), pseudo-PSU (SDMVPSU), and the two year MEC weight (WTMEC2YR). The laboratory file pbcd_d.xpt contains SEQN and the blood lead (LBXBPB). The two files are merged using the common

variable SEQN.

- NHANES 2007-2008: Demographic file demo_e.xpt. Laboratory file pbcd_e.xpt. The demographic file demo_e.xpt is a SAS transport file that contains the subject identifier (SEQN), age (RIDAGEYR), sex (RIAGENDR), race/ethnicity (RIDRETH1), poverty income ratio (INDFMPIR), pseudo-stratum (SDMVSTRA), pseudo-PSU (SDMVPSU), and the two year MEC weight (WTMEC2YR). The laboratory file pbcd_e.xpt contains SEQN and the blood lead (LBXBPB). The two files are merged using the common variable SEQN.

- NHANES 2009-2010: Demographic file demo_f.xpt. Laboratory file pbcd_f.xpt. The demographic file demo_f.xpt is a SAS transport file that contains the subject identifier (SEQN), age (RIDAGEYR), sex (RIAGENDR), race/ethnicity (RIDRETH1), poverty income ratio (INDFMPIR), pseudo-stratum (SDMVSTRA), pseudo-PSU (SDMVPSU), and the two year MEC weight (WTMEC2YR). The laboratory file pbcd_f.xpt contains SEQN and the blood lead (LBXBPB). The two files are merged using the common variable SEQN.

- NHANES 2011-2012: Demographic file demo_g.xpt. Laboratory file pbcd_g.xpt. The demographic file demo_g.xpt is a SAS transport file that contains the subject identifier (SEQN), age (RIDAGEYR), sex (RIAGENDR), race/ethnicity (RIDRETH1), poverty income ratio (INDFMPIR), pseudo-stratum (SDMVSTRA), pseudo-PSU (SDMVPSU), and the two year MEC weight (WTMEC2YR). The laboratory file pbcd_g.xpt contains SEQN and the blood lead (LBXBPB). The two files are merged using the common variable SEQN.

National Health and Nutrition Examination Surveys (NHANES)

Since the 1970s, the National Center for Health Statistics, a division of the Centers for Disease Control and Prevention, has conducted the National Health and Nutrition Examination Surveys (NHANES), a series of U.S. national surveys of the health and nutrition status of the noninstitutionalized civilian population. The National Center for Environmental Health at CDC measures environmental chemicals in blood and urine samples collected from NHANES participants. Indicators B1 and B2 use blood lead measurements in children ages 5 and under. NHANES II (1976-1980) included blood lead data for children from six months to 5 years. NHANES III (1988-1994) and the NHANES 1999-2000, 2001-2002, 2003-2004, 2005-2006, 2007-2008, 2009-2010, and 2011-2012 surveys included blood lead data for children ages 1 to 5 years. The NHANES data were obtained from the NHANES website: http://www.cdc.gov/nchs/nhanes.htm Following the CDC recommended approach, values below the analytical limit of detection (LOD) were replaced by LOD/√2.[ii]

The NHANES use a complex multi-stage, stratified, clustered sampling design. Certain demographic groups were deliberately over-sampled, including Mexican-Americans, Blacks,

[ii] See Hornung RW, Reed LD. 1990. Estimation of average concentration in the presence of nondetectable values. *Appl Occup Environ Hyg* 5:46–51.

and, from 2007 onwards, All Hispanics, then, from 2011 onwards, Asians, to increase the reliability and precision of estimates of health status indicators for these population subgroups. The publicly released data includes survey weights to adjust for the over-sampling, non-response, and non-coverage. The statistical analyses used the applicable MEC survey weights (N2LB0300 for NHANES II, WTPFEX1 for Phase I of NHANES III, WTPFEX2 for Phase 2 of NHANES III and WTMEC2YR for 1999 and later) to re-adjust the blood lead data to represent the national population.

Race/Ethnicity and Family Income

For Indicator B2, the percentiles were calculated for demographic strata defined by the combination of race/ethnicity and family income.

The family income was characterized based on the INDFMPIR variable, which is the ratio of the family income to the poverty level. The National Center for Health Statistics used the U.S. Census Bureau Current Population Survey definition of a "family" as "a group of two people or more (one of whom is the householder) related by birth, marriage, or adoption and residing together" to group household members into family units, and the corresponding family income for the respondent was obtained during the 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. Family income was stratified into the following groups:

- Below Poverty Level: PIR < 1
- Above Poverty Level: PIR ≥ 1
- Unknown Income: PIR is missing

For the four year period 2009-2012, the weighted percentage of children ages 1 to 5 years with unknown income was 6%.

Race/ethnicity was characterized using the RIDRETH1 variable. The possible values of this variable are:

- 1. Mexican American
- 2. Other Hispanic
- 3. Non-Hispanic White
- 4. Non-Hispanic Black
- 5. Other Race – Including Multi-racial
- "." Missing

Category 5 includes: all Non-Hispanic single race responses other than White or Black; and multi-racial responses.

For indicator B2, the RIDRETH1 categories 2, 5, and missing were combined into a single "All Other Races/Ethnicities" category. This produced the following categories:

- White non-Hispanic: RIDRETH1 = 3
- Black non-Hispanic: RIDRETH1 = 4
- Mexican-American: RIDRETH1 = 1
- All Other Races/Ethnicities: RIDRETH1 = 2 or 5 or missing

The "All Other Races/Ethnicities" category includes multiracial persons and individuals whose racial or ethnic identity is not White non-Hispanic, Black non-Hispanic, or Mexican-American. Except for Asian non-Hispanics in 2011-2012, persons of "All Other Races/Ethnicities" are selected into the survey with a probability that is very much lower than White non-Hispanic, Black non-Hispanic and Mexican-American individuals, and as a group they are not representative of all other race and ethnicities in the United States.

Calculation of Indicator

Indicator B1 is the median and 95th percentile for blood lead in children of ages 1 to 5 years. Table B1a presents the median and 95th percentile for blood lead in children of ages 1 to 17 years, stratified by age group. Indicator B2 is the median of blood lead in children of ages 1 to 5 years stratified by race/ethnicity and family income. Table B2a presents the 95th percentile of blood lead in children of ages 1 to 5 years stratified by race/ethnicity and family income. Table B2b presents the median of blood lead in children of ages 1 to 5 years, stratified by race/ethnicity and family income for 1991-1994. The median is the estimated concentration such that 50% of all noninstitutionalized civilian children ages 1 to 5 years during the survey period have blood lead concentrations below this level. The 95th percentile is the estimated concentration such that 95% of all noninstitutionalized civilian children ages 1 to 5 years during the survey period have blood lead concentrations below this level.

To simply demonstrate the calculations, we will use the NHANES 2009-2010 blood lead values for children ages 1 to 5 years of all race/ethnicities and all incomes as an example. We have rounded all the numbers to make the calculations easier:

We begin with all the non-missing NHANES 2009-2010 blood lead values for children ages 1 to 5 years. Assume for the sake of simplicity that valid data on blood lead were available for every sampled child. Each sampled child has an associated survey weight, WTMEC2YR, that estimates the annual number of U.S. children represented by that sampled child. For example, the lowest blood lead measurement for a child between 1 and 5 years of age is 0.18 μg/dL with a survey weight of 30,000, and so represents 30,000 children between 1 and 5 years of age. The total of the survey weights for the sampled children equals 20 million, the total U.S. population of children between 1 and 5 years of age. The second lowest measurement is 0.27 μg/dL with a survey weight of 40,000, and so represents another 40,000 U.S. children between 1 and 5 years of age. The highest measurement was 17 μg/dL, with a survey weight of 13,000, and so represents another 13,000 U.S. children between 1 and 5 years of age.

To calculate the median, we can use the survey weights to expand the data to the entire U.S. population of 20 million children ages 1 to 5 years. We have 30,000 values of 0.18 μg/dL from the lowest measurement, 40,000 values of 0.27 μg/dL from the second lowest measurement, and so on, up to 13,000 values of 17 μg/dL from the highest measurement. Arranging these 20

million values in increasing order, the 10 millionth value is 1.15 μg/dL. Since half of the values are below 1.15 and half of the values are above 1.15, the median equals 1.15 μg/dL. To calculate the 95th percentile, note that 95% of 20 million equals 19 million. The 19 millionth value is 3.4 μg/dL. Since 95% of the values are below 3.4, the 95th percentile equals 3.4 μg/dL.

In reality, the calculations need to take into account that blood lead measurements were not available for every respondent, and to use exact rather than rounded numbers. There were blood lead measurements for only 836 of the 1,295 sampled children ages 1 to 5 years. The survey weights for all 1,295 sampled children add up to 21.1 million, the U.S. population of children ages 1 to 5 years. The survey weights for the 836 sampled children with blood lead data add up to 13.5 million. Thus the available data represent 13.5 million values and so represent only 64% of the U.S. population of children ages 1 to 5 years. The median and 95th percentiles are given by the 6.75 millionth (50% of 13.5 million) and 12.83 millionth (95% of 13.5 million) U.S. child's value. These calculations assume that the sampled children with valid blood lead data are representative of the children without valid blood lead data.

Equations

These percentile calculations can also be given as the following mathematical equations, which are based on the default percentile calculation formulas from Statistical Analysis System (SAS) software. Exclude all missing blood lead values. Suppose there are n children of ages 1 to 5 years with valid blood lead values. Arrange the blood lead concentrations in increasing order (including tied values) so that the lowest concentration is x(1) with a survey weight of w(1), the second lowest concentration is x(2) with a survey weight of w(2), …, and the highest concentration is x(n) with a survey weight of w(n).

1. Sum all the survey weights to get the total weight W:

$$W = \Sigma[1 \le i \le n]\ w(i)$$

2. Find the largest number i so that the total of the weights for the i lowest values is less than or equal to W/2.

$$\Sigma[j \le i]\ w(j) \le W/2 < \Sigma[j \le i + 1]\ w(j)$$

3. Calculate the median using the results of the second step. We either have

$$\Sigma[j \le i]\ w(j) = W/2 < \Sigma[j \le i + 1]\ w(j)$$

or

$$\Sigma[j \le i]\ w(j) < W/2 < \Sigma[j \le i + 1]\ w(j)$$

In the first case we define the median as the average of the i'th and i + 1'th values:

$$\text{Median} = [x(i) + x(i + 1)]/2 \text{ if } \Sigma[j \le i]\ w(j) = W/2$$

In the second case we define the median as the i + 1'th value:

$$\text{Median} = x(i + 1) \text{ if } \Sigma[j \leq i]\, w(j) < W/2$$

(The estimated median does not depend upon how the tied values of x(j) are ordered).

A similar calculation applies to the 95th percentile. The first step to calculate the sum of the weights, W, is the same. In the second step, find the largest number i so that the total of the weights for the i lowest values is less than or equal to 0.95W.

$$\Sigma[j \leq i]\, w(j) \leq 0.95W < \Sigma[j \leq i + 1]\, w(j)$$

In the third step we calculate the 95th percentile using the results of the second step. We either have

$$\Sigma[j \leq i]\, w(j) = 0.95W < \Sigma[j \leq i + 1]\, w(j)$$

or

$$\Sigma[j \leq i]\, w(j) < 0.95W < \Sigma[j \leq i + 1]\, w(j)$$

In the first case we define the 95th percentile as the average of the i'th and i + 1'th values:

$$95^{th} \text{ Percentile} = [x(i) + x(i + 1)]/2 \text{ if } \Sigma[j \leq i]\, w(j) = 0.95W$$

In the second case we define the 95th percentile as the i + 1'th value:

$$95^{th} \text{ Percentile} = x(i + 1) \text{ if } \Sigma[j \leq i]\, w(j) < 0.95W$$

Relative Standard Error

The uncertainties of the median and 95th percentile values were calculated using a revised version of the CDC method given in CDC 2005,[iii] Appendix C, and the SAS® program provided by CDC. The method uses the Clopper-Pearson binomial confidence intervals adapted for complex surveys by Korn and Graubard (see Korn and Graubard, 1999,[iv] p. 65). The following text is a revised version of the Appendix C.

Step 1: Use SAS® Proc Univariate to obtain a point estimate P_{SAS} of the percentile value. Use the Weight option to assign the exact correct sample weight for each chemical result.

Step 2: Use SUDAAN® Proc Descript with Taylor Linearization DESIGN = WR (i.e., sampling with replacement) and the proper sampling weight to estimate the proportion (p) of subjects with results less than and not equal to the percentile estimate P_{SAS} obtained in Step 1 and to obtain the standard

[iii] CDC Third National Report on Human Exposure to Environmental Chemicals. 2005
[iv] Korn E. L., Graubard B. I. 1999. *Analysis of Health Surveys.* Wiley.

error (se_p) associated with this proportion estimate. Compute the degrees-of-freedom adjusted effective sample size

$$n_{df} = (t_{num}/t_{denom})^2 \, p(1 - p) / (se_p^2)$$

where t_{num} and t_{denom} are 0.975 critical values of the Student's t distribution with degrees of freedom equal to the sample size minus 1 and the number of PSUs minus the number of strata, respectively. Note: the degrees of freedom for t_{denom} can vary with the demographic sub-group of interest.

Step 3: After obtaining an estimate of p (i.e., the proportion obtained in Step 2), compute the Clopper-Pearson 95% confidence interval ($P_L(x,n_{df})$, $P_U(x,n_{df})$) as follows:

$$P_L(x,n_{df}) = v_1F_{v1,v2}(0.025)/(v_2 + v_1F_{v1,v2}(0.025))$$
$$P_U(x,n_{df}) = v_3F_{v3,v4}(0.975)/(v_4 + v_3F_{v3,v4}(0.975))$$

where x is equal to p times n_{df}, $v_1 = 2x$, $v_2 = 2(n_{df} - x + 1)$, $v_3 = 2(x + 1)$, $v_4 = 2(n_{df} - x)$, and $F_{d1,d2}(\beta)$ is the β quantile of an F distribution with d1 and d2 degrees of freedom. (Note: If n_{df} is greater than the actual sample size or if p is equal to zero, then the actual sample size should be used.) This step will produce a lower and an upper limit for the estimated proportion obtained in Step 2.

Step 4: Use SAS Proc Univariate (again using the Weight option to assign weights) to determine the chemical percentile values P_{CDC}, L_{CDC} and U_{CDC} that correspond to the proportion p obtained in Step 2 and its lower and upper limits obtained in Step 3. Do not round the values of p and the lower and upper limits. For example, if p = 0.4832, then P_{CDC} is the 48.32'th percentile value of the chemical. The alternative percentile estimates P_{CDC} and P_{SAS} are not necessarily equal.

Step 5: Use the confidence interval from Step 4 to estimate the standard error of the estimated percentile P_{CDC}:

$$\text{Standard Error } (P_{CDC}) = (U_{CDC} - L_{CDC}) / (2t_{denom})$$

Step 6: Use the estimated percentile P_{CDC} and the standard error from Step 4 to estimate the relative standard error of the estimated percentile P_{CDC}:

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

The tabulated estimated percentile is the value of P_{SAS} given in Step 1. The relative standard error is given in Step 6, using P_{CDC} and its standard error.

The relative standard error depends upon the survey design. For this purpose, the public release version of NHANES includes the variables SDMVSTRA and SDMVPSU, 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. If the relative standard error is too high, then the estimated percentile will not be accurately estimated. Furthermore, if the degrees of freedom (from Step 2) is too low, then the relative standard error will be less accurately estimated and thus may be underestimated. For these reasons, percentiles with high relative standard errors or with low degrees of freedom are unstable or unreliable.

Percentiles with a relative standard error less than 30% and with 12 or more degrees of freedom were treated as being reliable and were tabulated. Percentiles with a relative standard error that is

30% or greater but less than 40% and with 12 or more degrees of freedom were treated as being unstable; these values were tabulated but were flagged to be interpreted with caution. Percentiles with a relative standard error less than 40% and with between 7 and 11 degrees of freedom were also treated as being unstable; these values were tabulated but were flagged to be interpreted with caution. Percentiles with a relative standard error that is 40% or greater, or without an estimated relative standard error, or with 6 or less degrees of freedom, 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

Statistical analyses of the percentiles were used to determine whether the differences between percentiles for different demographic groups were statistically significant. For these analyses, the percentiles and their standard errors were calculated for each combination of age group, sex, income group (below poverty, at or above poverty, unknown income), and race/ethnicity group using the method described in the "Relative Standard Error" section. In the notation of that section, the percentile and standard error are the values of P_{CDC} and Standard Error (P_{CDC}), respectively. These calculated standard errors account for the survey weighting and design.

Using a weighted linear regression model, the percentile was assumed to be the sum of explanatory terms for age, sex, income and/or race/ethnicity and a random error term; the error terms were assumed to be approximately independent and normally distributed with a mean of zero and a variance equal to the square of the standard error. In this model, the weight is the inverse of the variance, so that percentiles with larger standard errors are given less of a statistical weight in the fitted regression model. Using this model, the difference in the value of a percentile 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 directly compare a percentile 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 percentiles between different race/ethnicity pairs. The adjusted race/ethnicity comparisons use the percentiles for each age/sex/income/race/ethnicity combination. The adjusted analyses add age, sex, and income terms to the statistical model and compare the percentiles 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 blood lead level strongly depends on family income only, then the unadjusted differences between these two race/ethnicity groups would be significant but the adjusted difference (taking into account income) would not be significant.

Comparisons between pairs of race/ethnicity groups and between income groups are shown in Tables 1 and 2, respectively, for children ages 1 to 5 years. 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 percentiles for each pair of race/ethnicity groups. For the adjusted "All incomes (adjusted for age, sex, income)" comparisons, the explanatory variables are terms for each race/ethnicity group together with terms for each age, 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 twelve race/ethnicity/income combinations (combinations of four race/ethnicity groups and three 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 age and sex, so that race/ethnicity/income groups are compared after accounting for any differences due to age or sex. Although these comparisons only involve the two income groups with known incomes, these statistical models were fitted to all three income groups (including those with unknown income) to make a more general, better fitting model; this approach has no impact on the unadjusted p-values but has a small impact on the adjusted p-values. Also in Table 2, the unadjusted p-value for the population "All" compares the percentiles for children ages 1 to 5 years below poverty level with those at or above poverty level, using the explanatory variables for the two income groups (below poverty, at or above poverty), excluding children with unknown income. The adjusted p-value includes adjustment terms for age, sex, and race/ethnicity in the model.

Additional comparisons are shown in Table 3. Comparisons are shown for differences between age groups among children ages 1 to 17 years, between children ages 1 to 5 years below poverty and those at or above poverty, and for changes over time (trends) for children ages 1 to 5 years. The Against = "age" unadjusted p-value compares the percentiles for children between the age groups 1, 2, 3-5, 6-10, 11-15, and 16-17 years, using the explanatory variables for these six age groups. The adjusted p-value includes adjustment terms for sex, race/ethnicity, and income in the model. The Against = "income" unadjusted p-value compares the percentiles for children ages 1 to 5 years below poverty level with those at or above poverty level, using the explanatory variables for the two income groups (below poverty, at or above poverty). The adjusted p-value includes adjustment terms for age, sex, and race/ethnicity in the model. The Against = "year" p-value examines whether the linear trend in the percentiles for children ages 1 to 5 years is statistically significant (using the percentiles for each NHANES period regressed against the midpoint of that period); the adjusted model for trend adjusts for demographic changes in the populations from year to year by including terms for age, sex, income, and race/ethnicity.

Table 4 shows comparisons between blood lead levels in children ages 1 to 5 years in 1991-1994 and the blood lead levels in children ages 1 to 5 years in 2009-2012. The Against = "year" p-value examines whether the change in the percentiles is statistically significant (using the percentiles for the periods 1991-1994 and 2009-2012 regressed against the midpoints of those two periods); the adjusted model adjusts for demographic changes in the populations from year to year by including terms for age, sex, income, and race/ethnicity. The rows where the Subset value is not missing show the p-values for different race/ethnicity groups.

The age groups used for the analyses of children ages 1 to 5 years were 1, 2, 3, 4, and 5.

For more details on these statistical analyses, see the memorandum by Cohen (2010).[v]

Table 1. Statistical significance tests comparing the percentiles of blood lead levels in children ages 1 to 5 years, between pairs of race/ethnicity groups, for 2009-2012.

				P-VALUES					
Variable	**Percentile**	**First race/ethnicity group**	**Second race/ethnicity group***	**All incomes**	**All incomes (adjusted for age, sex, income)**	**Below Poverty Level**	**Below Poverty Level (adjusted for age, sex)**	**At or Above Poverty Level**	**At or Above Poverty Level (adjusted for age, sex)**
lead	50	White non-Hispanic	Black non-Hispanic	< 0.001	< 0.001	0.510	0.002	< 0.001	< 0.001
lead	50	White non-Hispanic	Mexican-American	0.704	0.586	0.191	0.471	0.555	0.885
lead	50	White non-Hispanic	Other	0.690	0.076	0.361	< 0.001	0.221	0.644
lead	50	Black non-Hispanic	Mexican-American	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
lead	50	Black non-Hispanic	Other	< 0.001	0.002	< 0.001	0.211	< 0.001	< 0.001
lead	50	Mexican-American	Other	1.000	0.010	0.350	< 0.001	0.555	0.702
lead	95	White non-Hispanic	Black non-Hispanic	0.188	0.017	0.871	0.747	0.264	< 0.001
lead	95	White non-Hispanic	Mexican-American	0.140	< 0.001	0.610	< 0.001	0.441	< 0.001
lead	95	White non-Hispanic	Other	0.202	< 0.001	0.704	< 0.001	0.154	0.121
lead	95	Black non-Hispanic	Mexican-American	0.001	< 0.001	0.115	< 0.001	0.070	< 0.001
lead	95	Black non-Hispanic	Other	0.005	< 0.001	0.373	< 0.001	0.013	< 0.001
lead	95	Mexican-American	Other	0.982	0.002	0.521	0.068	0.606	< 0.001

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

Table 2. Statistical significance tests comparing the percentiles of blood lead levels in children ages 1 to 5 years, between those below poverty level and those at or above poverty level, for 2009-2012.

			P-Values for difference between income levels	
Variable	**Percentile**	**Population***	**Unadjusted**	**Adjusted (for age, sex)****
lead	50	All	< 0.001	< 0.001
lead	50	White non-Hispanic	0.022	0.013
lead	50	Black non-Hispanic	0.010	0.020
lead	50	Mexican-American	0.007	0.005
lead	50	Other	< 0.001	< 0.001
lead	95	All	0.026	< 0.001
lead	95	White non-Hispanic	0.593	< 0.001
lead	95	Black non-Hispanic	0.221	0.240

[v] 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 for difference between income levels	
Variable	**Percentile**	**Population***	**Unadjusted**	**Adjusted (for age, sex)****
lead	95	Mexican-American	0.506	< 0.001
lead	95	Other	0.120	0.095

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

Table 3. Other statistical significance tests comparing the percentiles of blood lead levels in children by age for ages 1 to 17 years for 2011-2012, by income for ages 1 to 5 years for 2009-2012, and trends for children ages 1 to 5 years from 1988-2012 and from 1999-2012.

						P-VALUES	
Variable	**Percentile**	**From**	**To**	**Ages**	**Against**	**Unadjusted**	**Adjusted***
lead	50	2011	2012	1 to 17 years	age	< 0.001	< 0.001
lead	50	2009	2012	1 to 5 years	income	< 0.001	< 0.001
lead	50	1988	2012	1 to 5 years	year	< 0.001	< 0.001
lead	50	1999	2012	1 to 5 years	year	< 0.001	< 0.001
lead	95	2011	2012	1 to 17 years	age	< 0.001	< 0.001
lead	95	2009	2012	1 to 5 years	income	0.026	< 0.001
lead	95	1988	2012	1 to 5 years	year	< 0.001	< 0.001
lead	95	1999	2012	1 to 5 years	year	< 0.001	< 0.001

* For Against = "age" the comparison is between the age groups 1, 2, 3-5, 6-10, 11-15 and 16-17 years, and the p-values are adjusted for sex, 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 age, sex, and race/ethnicity.
For Against = "year" the comparison is the trend over different years, and the p-values are adjusted for age, sex, race/ethnicity, and income.

Table 4. Statistical significance tests comparing the percentiles of blood lead levels in children ages 1 to 5 years between 1991-1994 and 2009-2012.

						P-VALUES	
Variable	**Percentile**	**From**	**To**	**Against**	**Subset***	**Unadjusted**	**Adjusted****
Lead	50	1991-1994	2009-2012	year		< 0.001	< 0.001
Lead	50	1991-1994	2009-2012	year	White non-Hispanic	< 0.001	< 0.001
Lead	50	1991-1994	2009-2012	year	Black non-Hispanic	< 0.001	< 0.001
Lead	50	1991-1994	2009-2012	year	Mexican-American	< 0.001	< 0.001
Lead	50	1991-1994	2009-2012	year	Other	< 0.001	< 0.001
Lead	95	1991-1994	2009-2012	year		< 0.001	< 0.001
Lead	95	1991-1994	2009-2012	year	White non-Hispanic	< 0.001	< 0.001
Lead	95	1991-1994	2009-2012	year	Black non-Hispanic	< 0.001	< 0.001
Lead	95	1991-1994	2009-2012	year	Mexican-American	< 0.001	< 0.001
Lead	95	1991-1994	2009-2012	year	Other	< 0.001	< 0.001

* "Other" represents the "All Other Races/Ethnicities" category, which includes all other races and ethnicities not specified, together with those individuals who report more than one race.
** For Against = "year" where Subset is missing, the comparison is between the different periods, and the p-values are adjusted for age, sex, race/ethnicity, and income.
For Against = "year" where Subset is not missing, the comparison is between different periods for each race/ethnicity group , and the p-values are adjusted for age, sex, and income.