Methods

Indicator

B7. PCBs in women ages 16 to 49 years: Median concentrations in blood serum, by race/ethnicity and family income, 2001-2004.

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] This indicator uses lipid-adjusted serum PCB measurements of PCBs 118, 138, 153 and 180 in women ages 16 to 49 years, summed together. The NHANES 2001-2002 and 2003-2004 surveys included serum PCB data for children and adults ages 12 years and over.

Indicator B7 is the median concentrations of the sum of the four PCBs for women ages 16 to 49 years for 2001-2004, stratified both by race/ethnicity and family income. The median is the estimated concentration such that 50% of all noninstitutionalized civilian women ages 16 to 49 years during the survey period have a sum of the four PCB concentrations below this level; a birth rate-adjusted distribution of women's PCB levels is used in calculating this indicator, meaning that the data are weighted using the age-specific probability of a woman giving birth. Table B7a presents the 95th percentile concentrations of the sum of the four PCBs for women ages 16 to 49 years for 2001-2004, stratified both by race/ethnicity and family income. The 95th percentile is the estimated concentration such that 95% of all noninstitutionalized civilian women ages 16 to 49 years during the survey period have a sum of the four PCB concentrations below this level; the population distribution was weighted to account for over-sampling, non-response, and non-coverage.

Data Summary

Indicator	B7. PCBs in women ages 16 to 49 years: Median concentrations in blood serum, by race/ethnicity and family income, 2001-2004.							
Time Period	2001-2004							
Data	Serum PCB (lipid adjusted) for four PCBs							
Years/PCB	2001-2002/ 118	2001-2002/ 138	2001-2002/ 153	2001-2002/ 180	2003-2004/ 118	2003-2004/ 138	2003-2004/ 153	2003-2004/ 180
Limits of Detection (ng/g lipid)*	10.5	10.5	10.5	10.5	0.6	0.4	1.1	0.4

[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.

Indicator	B7. PCBs in women ages 16 to 49 years: Median concentrations in blood serum, by race/ethnicity and family income, 2001-2004.							
Number of Values	700	700	700	700	627	627	627	627
Number of Non-missing Values**	644 (92%)	641 (92%)	644 (92%)	643 (92%)	519 (83%)	520 (83%)	519 (83%)	520 (83%)
Number of Missing Values**	56 (8%)	59 (8%)	56 (8%)	57 (8%)	108 (17%)	107 (17%)	108 (17%)	107 (17%)
Percentage Below Limit of Detection***	50	18	11	32	0	0	0	0

* The Limit of Detection (LOD) is defined as the level at which the measurement has a 95% probability of being greater than zero. LODs vary among samples. The highest LOD among all the samples is shown.
**Non-missing values include those below the analytical limit of detection (LOD), which are reported as LOD/$\sqrt{2}$. 1,158 sampled women 16 to 49 years had non-missing values for all four PCB congeners. 6 sampled women 16 to 49 years had missing values for some but not all of the four PCB congeners. Missing values are the number of sampled women ages 16 to 49 years in the Mobile Examination Center (MEC) sub-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 and is weighted by age-specific birth rates.

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 NHANES website: http://www.cdc.gov/nchs/nhanes.htm.

- NHANES 2001-2002: Demographic file demo_b.xpt. Laboratory 28POC file l28poc_b. 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) and the pseudo-PSU (SDMVPSU). The laboratory file l28poc_b.xpt contains SEQN, the lipid-adjusted PCB 118, 138, 153, and 180 (LBX118LA, LBX138LA, LBX153LA, LBX180LA), the PCB 118, 138, 153, and 180 non-detect comment codes (LBD118LC, LBD138LC, LBD153LC, LBD180LC), and the PCB sub-sample two year Mobile Examination Center (MEC) weight (WTSPO2YR). The two files are merged using the common variable SEQN.

- NHANES 2003-2004: Demographic file demo_c.xpt. Dioxins, Furans, and Coplanar PCBs Laboratory file l28dfp_c.xpt. The Non-dioxin-like PCB Laboratory file l28npb_c. 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) and the pseudo-PSU (SDMVPSU). The Dioxins, Furans and Co-planar PCB laboratory file l28dfp_c.xpt contains SEQN, the lipid-adjusted PCB 118 (LBX118LA), the PCB 118 non-detect comment code (LBD118LC), and the sub-sample C two year MEC weight (WTSC2YR). The Non-dioxin-like PCB laboratory file l28npb_c.xpt contains SEQN, the lipid-adjusted PCB 138, 153, and 180 (LBX138LA, LBX153LA, LBX180LA), the PCB 138, 153, and 180 non-detect comment codes (LBD138LC, LBD153LC, LBD180LC), and the PCB sub-sample C two year MEC weight (WTSC2YR). The three 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. This indicator uses serum PCB measurements of four PCBs from NHANES 2001-2002 and 2003-2004 in women ages 16 to 49. 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] This analysis uses the sum of the four PCB congeners 118, 138, 153 and 180 listed in the following table. If some but not all of the four PCB congeners are missing, then the sum is over the non-missing PCB congeners. In the rest of this section, we will refer to this sum as the total serum PCB.

PCB Congener	Full name	SAS name (lipid-adjusted)	SAS name for non-detect comment code*
118	2,3',4,4',5-pentachlorophenyl	LBX118LA	LBD118LC
138	2,2',3,4,4',5- and 2,3,3',4,4',6-hexachlorophenyl	LBX138LA	LBD138LC
153	2,2',4,4',5,5'-hexachlorophenyl	LBX153LA	LBD153LC
180	2,2',3,4,4',5,5'-heptachlorophenyl	LBX180LA	LBD180LC

*The non-detect comment code equals 1 if the measurement is below the analytical limit of detection, and equals 0 if the measurement is at or above the analytical limit of detection.

The NHANES use a complex multi-stage, stratified, clustered sampling design. Certain demographic groups were deliberately over-sampled, including Mexican-Americans and Blacks, 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 PCB sub-sample MEC survey weights (WTSPO2YR for 2001-2002 and WTSC2YR for 2003-2004) to re-adjust the serum PCB data to represent the national population.

[ii] See Hornung RW, Reed LD. 1990. Estimation of average concentration in the presence of nondetectable values. *Applied Occupational and Environmental Hygiene* 5:46–51.

Age-Specific Birth Rates

In addition to the NHANES survey weights, the data for women of child-bearing age (ages 16 to 49) were also weighted by the birth rate for women of the given age and race/ethnicity to estimate prenatal exposures. Thus the overall weight in each two year period is the product of the NHANES MEC survey weight and the total number of births in the two calendar years for the given age and race/ethnicity, divided by twice the corresponding population of women at the midpoint of the two year period:[iii]

Adjusted Survey Weight =
MEC survey weight × U.S. Births (NHANES cycle, age, race/ethnicity) /
{Number of years in NHANES cycle × U.S. Women (NHANES cycle midpoint, age, race/ethnicity)}.

Race/Ethnicity and Family Income

For this indicator, 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 2001-2004, the weighted percentage of women ages 16 to 49 years with unknown income was 5%.

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

[iii] Axelrad, D.A., Cohen, J. 2011. Calculating summary statistics for population chemical biomonitoring in women of child-bearing age with adjustment for age-specific natality. *Environmental Research* 111 (1): 149-155..

- 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 this indicator, 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. 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 B7 is the median for total serum PCB in women of ages 16 to 49 years, stratified by race/ethnicity and family income. The median is the estimated concentration such that 50% of all noninstitutionalized civilian women ages 16 to 49 years during the survey period have total serum PCB concentrations below this level. Table B7a presents the 95th percentile for total serum PCB in women of ages 16 to 49 years, stratified by race/ethnicity and family income. The 95th percentile is the estimated concentration such that 95% of all noninstitutionalized civilian women ages 16 to 49 years during the survey period have total serum PCB concentrations below this level. To adjust the NHANES data to represent prenatal exposures, the data for each woman surveyed was multiplied by the estimated number of births per woman of the given age and race/ethnicity.

To simply demonstrate the calculations, we will use the NHANES 2001-2004 total serum PCB values for women ages 16 to 49 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 2001-2004 total serum PCB values for women ages 16 to 49 years. Assume for the sake of simplicity that valid data on total serum PCB were available for every sampled woman. Each sampled woman has an associated annual survey weight that estimates the annual number of U.S. women represented by that sampled woman. Since two 2-year periods are combined for these analyses, the associated annual survey weight for each woman is defined as WTSPO2YR/2 for 2001-2002 and WTSC2YR/2 for 2003-2004, so that the combined 2001-2004 sample represents the annual population. Each sampled woman also has an associated birth rate giving the numbers of annual births per woman of the given age, race, and ethnicity. The product of the annual survey weight and the birth rate estimates the annual number of U.S. births represented by that sampled woman, which we will refer to as the adjusted survey weight. For example, the lowest total serum PCB measurement for a woman between 16 and 49 years of age is 4.6 ng/g lipid with an annual survey weight of 10,000, a birth rate of 0.03, and thus an adjusted survey weight of 300, and so represents 300 births. The total of the adjusted survey weights for the sampled women equals 4 million, the total number of annual U.S. births to women ages 16 to 49 years. The second lowest measurement is also 4.6 ng/g lipid with an adjusted survey weight of 800, and so represents another 800 U.S. births. The highest measurement was 715.5 ng/g lipid, with an adjusted survey weight of only 4, because it was for a 48 year old woman, and so represents another 4 U.S. births.

To calculate the median, we can use the adjusted survey weights to expand the data to the entire U.S. population of births to women ages 16 to 49. We have 300 values of 4.6 ng/g lipid from the lowest measurement, 800 values of 4.6 ng/g lipid from the second lowest measurement, and so on, up to 4 values of 715.5 ng/g lipid from the highest measurement. Arranging these 4 million values in increasing order, the 2 millionth value is 30.1 ng/g lipid. Since half of the values are below 30.1 and half of the values are above 30.1, the median equals 30.1 ng/g lipid. To calculate the 95th percentile, note that 95% of 4 million equals 3.8 million. The 3.8 millionth value is 106.2 ng/g lipid. Since 95% of the values are below 106.2, the 95th percentile equals 106.2 ng/g lipid.

In reality, the calculations need to take into account that total serum PCB measurements were not available for every respondent, and to use exact rather than rounded numbers. There were total serum PCB measurements for only 1,164 of the 1,327 sampled women ages 16 to 49 years. These 1,164 sampled women included 1,158 women with measured concentrations for all four congeners and 6 women with measured concentrations for only one, two, or three of the four congeners; for those six women the total serum PCB is defined as the sum of the non-missing congeners. The adjusted survey weights for all 1,327 sampled women add up to 4.2 million, the U.S. population of births to women ages 16 to 49. The adjusted survey weights for the 1,164 sampled women with total serum PCB data add up to 3.7 million. Thus the available data represent 3.7 million values and so represent only 90 % of the U.S. population of births. The median and 95th percentiles are given by the 1.85 millionth (50 % of 3.7 million) and 3.52 millionth (95 % of 3.7 million) U.S. birth's value. These calculations assume that the sampled women with valid total serum PCB data are representative of women giving birth without valid total serum PCB data. The calculations also assume that the sampled women are representative of women that actually gave birth in 2001-2004, since NHANES information on pregnancy and births was not incorporated into the analysis.

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 total serum PCB values. Suppose there are n women of ages 16 to 49 years with valid total serum PCB values. Arrange the total serum PCB concentrations in increasing order (including tied values) so that the lowest concentration is x(1) with an adjusted survey weight of w(1), the second lowest concentration is x(2) with an adjusted survey weight of w(2), …, and the highest concentration is x(n) with an adjusted survey weight of w(n).

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

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

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.

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

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

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

or

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

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

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

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:

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

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

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

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,[iv] 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,[v] p. 65). The following text is a revised version of the Appendix C. For the birth rate adjusted calculations for women ages 16 to 49, the sample weight is adjusted by multiplying by the age-specific birth rate.

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 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.

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

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, 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 and, for women, for the age-specific birth rate.

Using a weighted linear regression model, the percentile was assumed to be the sum of explanatory terms for age, 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/ income/race/ethnicity combination. The adjusted analyses add age, 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 serum PCB 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 women ages 16 to 49 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, income)" comparisons, the explanatory variables are terms for each race/ethnicity group together with terms for each age, 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 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, so that race/ethnicity/income groups are compared after accounting for any differences due to age. 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 women ages 16 to 49 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 those with unknown income. The adjusted p-value includes adjustment terms for age and race/ethnicity in the model.

For women, the age groups used were 16-19, 20-24, 25-29, 30-39, and 40-49.

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

Table 1. Statistical significance tests comparing the percentiles of PCBs in women ages 16 to 49 years, between pairs of race/ethnicity groups, for 2001-2004.

				P-VALUES					
Variable	**Percentile**	**First race/ethnicity group**	**Second race/ethnicity group***	**All incomes**	**All incomes (adjusted for age, income)**	**Below Poverty Level**	**Below Poverty Level (adjusted for age)**	**At or Above Poverty Level**	**At or Above Poverty Level (adjusted for age)**
PCB	50	White non-Hispanic	Black non-Hispanic	0.622	0.211	0.893	0.033	0.759	0.363
PCB	50	White non-Hispanic	Mexican-American	< 0.001	< 0.001	0.002	0.009	0.001	< 0.001
PCB	50	White non-Hispanic	Other	0.691	< 0.001	0.588	0.477	0.702	< 0.001
PCB	50	Black non-Hispanic	Mexican-American	0.001	< 0.001	< 0.001	< 0.001	0.010	< 0.001
PCB	50	Black non-Hispanic	Other	0.957	< 0.001	0.572	0.307	0.991	< 0.001
PCB	50	Mexican-American	Other	0.002	0.134	0.326	0.020	0.001	0.299
PCB	95	White non-Hispanic	Black non-Hispanic	0.601	< 0.001	0.402	< 0.001	0.947	0.121
PCB	95	White non-Hispanic	Mexican-American	< 0.001	0.002	0.041	< 0.001	0.007	< 0.001
PCB	95	White non-Hispanic	Other	0.003	0.034	< 0.001	< 0.001	0.096	< 0.001
PCB	95	Black non-Hispanic	Mexican-American	< 0.001	< 0.001	0.085	< 0.001	0.020	< 0.001

[vi] 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	Percentile	First race/ethnicity group	Second race/ethnicity group*	All incomes	All incomes (adjusted for age, income)	Below Poverty Level	Below Poverty Level (adjusted for age)	At or Above Poverty Level	At or Above Poverty Level (adjusted for age)
PCB	95	Black non-Hispanic	Other	0.001	< 0.001	< 0.001	0.550	0.125	< 0.001
PCB	95	Mexican-American	Other	< 0.001	< 0.001	< 0.001	< 0.001	0.002	0.009

* "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 PCBs in women ages 16 to 49 years, between those below poverty level and those at or above poverty level, for 2001-2004.

			P-Values for difference between income levels	
Variable	Percentile	Population*	Unadjusted	Adjusted (for age)**
PCB	50	All	0.047	0.274
PCB	50	White non-Hispanic	0.307	0.004
PCB	50	Black non-Hispanic	0.214	0.684
PCB	50	Mexican-American	0.272	0.488
PCB	50	Other	0.798	0.103
PCB	95	All	0.193	< 0.001
PCB	95	White non-Hispanic	0.216	< 0.001
PCB	95	Black non-Hispanic	0.044	0.035
PCB	95	Mexican-American	0.432	0.002
PCB	95	Other	0.089	< 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.
** Comparison for "All" is adjusted for age and race/ethnicity; comparisons for race/ethnicity categories are adjusted for age.