Appendix C. Confidence Interval Estimation for Percentiles

A common practice to calculate confidence intervals from survey data is to use large-sample normal approximations. Ninety-five percent confidence intervals on point estimates of percentiles are often computed by adding and subtracting from the point estimate a quantity equal to twice its standard error. This normal approximation method may not be adequate, however, when estimating the proportion of subjects above or below a selected value (especially when the proportion is near 0.0 or 1.0 or when the effective sample size is small).

In addition, confidence intervals on proportions deviating from 0.5 are not theoretically expected to be symmetric around the point estimate. Further, adding and subtracting a multiple of the standard error to an estimate near 0.0 or 1.0 can lead to impossible confidence limits (i.e., proportion estimates below 0.0 or above 1.0).

We used the method of Korn and Graubard (1998) to compute Clopper-Pearson 95% confidence intervals about percentile estimates. We describe the method below, using SAS Proc Univariate and SUDAAN. SAS code for calculating these confidence intervals can be downloaded from http://www.cdc.gov/exposurereport.

Procedure to calculate confidence intervals about percentiles

Step 1: Use SAS (SAS Institute Inc., 1999) Proc Univariate to obtain a point estimate of the percentile of a chemical's results for the demographic group of interest (e.g., the 90th percentile of blood lead results for children aged 1-5 years). Use the Freq option to assign the correct sample weight for each chemical result.

Step 2: Use SUDAAN (SUDAAN Users Manual, 2001) 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 below the percentile estimate 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

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

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 (e.g., males).

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:

$$P_L(x,n_{df}) = v_1F_{v1,v2}(0.025)/(v_2 + v_1F_{v1,v2}(0.025)) \quad \& \quad P_U(x,n_{df}) = v_3F_{v3,v4}(0.975)/(v_4 + v_3F_{v3,v4}(0.975)) \qquad (2)$$

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 d_1 and d_2 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 Freq option to assign weights) to determine the chemical values that correspond to the proportion obtained in Step 2 and the lower and upper limits on this proportion obtained in Step 3.

Example:

To estimate the 75th percentile, use SAS Proc Univariate with the Freq option to get a weighted point estimate of the chemical value that corresponds to the 75th percentile. Then use SUDAAN to estimate the weighted proportion of subjects with results below the 75th percentile (which should be very near 0.75). Next, obtain a confidence interval on this proportion by computing the weighted Clopper-Pearson 95% confidence limits using the degrees-of-freedom adjusted effective sample size. Suppose these confidence limits are 0.67 and 0.81, then use SAS Proc Univariate with the Freq option to determine the chemical values corresponding to the weighted 67th and 81st percentiles. These point estimates are the lower and upper confidence limits on the 75th percentile.