HHS Public Access

Author manuscript

Environ Int. Author manuscript; available in PMC 2016 December 01.

Published in final edited form as:

Environ Int. 2015 December ; 85: 40–45. doi:10.1016/j.envint.2015.08.003.

Confidence Interval Estimation for Pooled-Sample Biomonitoring from a Complex Survey Design

Samuel P. Caudill, PhD

Division of Laboratory Sciences, National Center for Environmental Health, Centers for Disease Control & Prevention, Public Health Service, US Department of Health and Human Services, Atlanta, Georgia 30333

Abstract

The National Centers for Disease Control and Prevention (CDC) is using a weighted pooled-sample design to characterize concentrations of persistent organic pollutants (POPs) in the U.S. population. Historically, this characterization has been based on *individual measurements* of these compounds in body fluid or tissue from representative samples of the population using stratified multistage selection. Pooling samples before making analytical measurements reduces the costs of biomonitoring by reducing the number of analyses. Pooling samples also allows for larger sample volumes which can result in fewer left censored results. But because samples are pooled across the sampling design cells of the original survey, direct calculation of the design effects needed for accurate standard error and confidence interval (CI) estimation is not possible. So in this paper I describe a multiple imputation (MI) method for calculating design effects associated with pooled-sample estimates. I also evaluate the method presented, by simulating NHANES individual sample data from which artificial pools are created for use in a comparison of pooled-sample estimates with estimates based on individual samples. To further illustrate and evaluate the method proposed in this paper I present geometric mean and various percentile estimates along with their 95% CIs for two chemical compounds from NHANES 2005-2006 pooled samples and compare them to individual-sample based estimates from NHANES 1999-2004.

1. INTRODUCTION

Recently Caudill (2012) demonstrated how a weighted pooled-sample design can be used with individual samples collected in conjunction with NHANES. He incorporated survey sampling weights into a pooled-sample design by using a different volume of material from each sample contributing to a pool. The volume chosen for each sample in a pool was based on the ratio of its sampling weight to the sum of the sampling weights of all samples in the same pool (Caudill, 2010). Because samples were pooled across the sampling design cells of the original NHANES design, direct calculation of design effects was not possible. So in accordance with previous pooled-sample studies based on NHANES data (Calafat et al., 2006; Kato et al., 2009), Caudill (2012) presented unadjusted confidence intervals (CIs) assuming simple random sampling and then adjusted these CIs using design effects from a

Telephone/fax numbers, email address of the corresponding author: Phone: (770) 488-4622, FAX: (770) 488-4839, SPC1@CDC.GOV, Mailing Address: 4770 Buford Highway, NE MI-F25, Atlanta, GA 30341.

previous NHANES. There is no guarantee, however, that the design effects from a previous survey are applicable to a current survey. So in this paper I present and evaluate a multiple imputation (MI) method for generating individual-sample data from pooled-sample estimates and then use the imputed individual-sample data to obtain point estimates and associated CIs adjusted for design effects. Unlike the pooled-sample estimates, these imputation-based estimates and associated CIs do not have to be limited to the demographic groups specified by the original pooled-sample design. That is, point estimates and their associated CIs can be calculated for demographic groupings such as the total civilian non-institutionalized U.S. population or all males in the civilian non-institutionalized U.S. population.

To evaluate the MI method, I simulate multiple individual sample data sets and use a pooled-sample design similar to the one used with NHANES 2005-2006 to create artificial pools. I then compute the average bias of the pooled-sample point estimates and the average coverage probability of the corresponding 95% CIs. Finally, to illustrate the MI method using actual data and to show how pooled-sample estimates from one survey compare to individual-sample based estimates from previous surveys, I present geometric means, various percentiles, and 95% CIs adjusted for design effects for 2,2’,4,4’,5,5’-hexachlorobiphenyl (or polychlorinated biphenyl, PCB153) and for 1,1’-(2,2-dichloroethenylidene)-bis[4-chlorobenzene] (or *p*,*p*’-DDE) in the U.S. population using pooled-samples from NHANES 2005-2006 for comparison with estimates from NHANES 1999-2004 individual samples.

2. METHODS

2.1 NHANES Survey Design

The sampling scheme for NHANES 2005–2006 is a complex multistage, probability sampling design that selects participants who are representative of the civilian, non-institutionalized U.S. population. Over-sampling of certain population subgroups is done to increase the reliability and precision of health status indicator estimates for those groups. Because each sample person does not have an equal probability of selection, sample weighting is needed to produce correct population estimates of means, percentiles, and other descriptive statistics. Also, because of the use of stratified multistage selection, incorporation of the sampling design is needed to calculate sampling variances (NCHS, 1994). These variances can be related to variances based on simple random sampling via the design effects (NCHS, 1969). For POPs measured as part of CDC's biomonitoring program, instead of using the full NHANES sample, a random one-third subsample of NHANES participants is used along with appropriately adjusted sampling weights (Curtin et al., 2012). After collection, serum specimens are divided into aliquots, transferred to clean cryovials, frozen, shipped on dry ice to CDC's National Center for Environmental Health, and stored at −70 °C.

2.2 Pooled-Sample Design

In order to implement a pooled-sample design for NHANES 2005-2006 each aliquot was identified as belonging to one of 32 demographic groups based on race/ethnicity (non-

Hispanic white: NHW; non-Hispanic black: NHB; Mexican American: MA; Other Hispanic and non-Hispanic multiracial: Other); gender (Male: M, Female: F); and age group (12-19, 20-39, 40-59, and 60+ years of age and older). For this analysis a pooled-sample design consisting of 32 demographic groups and 8 samples per pool was chosen based on the results of simulation experiments presented in Caudill (2010). The number of pools created for each of the 32 demographic groups varied depending on the total number of individual aliquots available. The one-third subset of NHANES 2005-2006 represents 2345 individual samples/aliquots, but because the pooled-sample design calls for the same number of samples in each pool and requires that all samples be of sufficient volume, only 1973 samples were available to create 247 pools with 8 samples per pool (with the exception that one pool for M Other 40-59 consisted of only 7 samples and one pool for F Other 60+ consisted of only 6 samples).

I used the procedure described in Caudill (2012) to incorporate sample weighting into the pooled-sample design. While implementing this procedure for the current study, I maintained the unique NHANES specimen identification number (SEQN) of each individual sample so that pool measurements could later be linked to the individual samples used to form the corresponding pool.

The number of subjects in the one-third subset, the number of samples available, the number of these samples that were usable, and the number of pools formed in each demographic group are presented in Table 1. Once the pools were created, summed sampling weights were further adjusted to account for the unused samples.

2.3 Calculation of Point Estimates using Pooled-Samples

Measurements of POPs in samples from individuals tend to be skewed to higher values and are often log-normal or can be approximated by log-normal distributions, so special methods are required when computing means, variances, and percentiles from pooled-sample measurements. As per Caudill (2012) I assume there are d pooled-sample demographic groups, p_i pools in the ith demographic group and that each pool consists of s samples. To simplify the discussion, I assume that the original individual "unmeasured" sample results (x_{ijk};i = 1,2,...d; j = 1,2,...p_i;k = 1,2,...,s) are log-normal with mean and variance of the natural logarithm of individual "unmeasured" results (i.e., $y_{ijk} = \ln(_{ijk})$) equal to μ_{y_i} and $\sigma^2_{y_i}$, respectively. That is, the individual y_{ijk} values are normal with mean $\mu_{y_i} = E(y_{ijk})$ and with variance $\sigma^2_{y_i} = Var(y_{ijk})$. Based on the properties of the log-normal distribution (Aitchison and Brown 1963), the geometric mean of the i^{th} pooled-sample demographic group can be estimated by:

$$G\hat{\mu}_{y_i} = exp\left\{ ln\left[\sum_{j=1}^{p_i} w_{ij.}\left(\bar{x}_{ij.}\right)/w_{i..}\right] - \frac{\hat{\sigma}^2_{y_i}}{2}\right\} \quad (1)$$

where the single measured value of a pool ($\bar{x}_{ij.}$) is comparable to a weighted average of log-normal values $\left[\bar{x}_{ij.} = \sum_{k=1}^{s} w_{ijk}x_{ijk}/w_{ij.}\right]$, w_{ijk} is the sampling weight of the k^{th} sample in the j^{th} pool in the i^{th} pooled-sample demographic group, $w_{ij.}$ is the sum of the s sampling

weights in the j^{th} pool in the i^{th} pooled-sample demographic group, and $\hat{\sigma}^2_{y_i}$ is an estimate of the variance of y_{ijk}[= ln(x_{ijk})] and is calculated as follows:

$$\hat{\sigma}^2_{y_i} = \left\{ \sum_{j=1}^{p_i} w_{ij.} \left[ln \left(w^2_{ij.} C_{V \bar{x}_{ij.}}{}^2 / \sum_{k=1}^{s} w^2_{ijk} + 1 \right) \right] \right\} / \sum_{j=1}^{p_i} w_{ij.} \quad (2)$$

Note that equations 1 and 2 above differ slightly from equations 3 and 1, respectively, in Caudill (2012), based on later work by Li et al (2014) who demonstrated that unbiased estimation at the demographic group level requires averaging across pools prior to subtraction of the bias correction ($\hat{\sigma}^2_{y_i}/2$). Also, note that equation 1 of Caudill (2012) has a typographical error in that the sum should have been from *k*=1 to *s*.

2.4 Design Effect Estimation using Multiple Imputation (MI)

Because individual samples were pooled across the sampling design cells of the original NHANES sampling design to accommodate physical limitations associated with weighted pooling (Caudill 2012), it is not possible to directly estimate the design effects associated with pooled-sample estimates. So to create data for which design effects could be estimated, I used the pooled-sample estimates from the various demographic groups to impute individual sample measurements for every subject in the original one-third subset of NHANES 2005-2006. That way, with every subject in the one-third subset having an estimate, all of the strata and primary sampling units (PSUs) were involved in calculating the design effects. I used the estimated mean $\left(\hat{\mu}_{y_i}\right)$ and variance $\left(\hat{\sigma}^2_{y_i}\right)$ of the natural logarithm of the individual samples in the i^{th} demographic group to impute individual sample measurements for each sample within each pool in the i^{th} demographic group. Imputed measurements were also assigned to *samples in the one-third subset that were not used* to form pools.

The imputed sample measurements were assigned to individual samples in the appropriate demographic group based on their unique specimen identification number (SEQN). To incorporate the uncertainty associated with $\hat{\mu}_{y_i}$ and $\hat{\sigma}^2_{y_i}$, I drew a random variance $\breve{\sigma}^2_{y_i}$ such that $\breve{\sigma}^2_{y_i} \sim \left(df_{\hat{\sigma}^2_{y_i}}\right) \hat{\sigma}^2_{y_i} / \chi^2_{\hat{\sigma}^2_{y_i}}$ to obtain a random sample measurement defined by

$$\breve{y}_{ijkl} = \hat{\mu}_{y_i} + R_{N(0,1)} \sqrt{\breve{\sigma}^2_{y_i}} \quad (3)$$

where $df_{\hat{\sigma}^2_{y_i}}$ is the degrees-of-freedom associated with $\hat{\sigma}^2_{y_i}$, $\chi^2_{\hat{\sigma}^2_{y_i}}$ is a random chi-square deviate with degrees-of-freedom ($df_{\hat{\sigma}^2_{y_i}}$), $R_{N(0,1)}$ is a random normal deviate with zero mean and unit variance, *i*, *j*, and *k* are defined as before, and *l* (= 1, 2, . . ., *n*) represents the number of imputations. The variance $\breve{\sigma}^2_{y_i}$ was drawn in this way so that new parameter estimates are used for each imputed data set (Rubin, 1987).

Thus for the l^{th} imputed data set the natural logarithm of the individual sample measurement ($\breve{y}_{ijkl}\left[=ln\left(\breve{x}_{ijkl}\right)\right]$) was imputed for each sample using the estimated mean and variance of samples in the corresponding pooled-sample demographic group. I repeated this process *n* times, where *n* represents the number of imputations. With these imputed measurements it was then possible to calculate point estimates with standard errors adjusted for estimated design effects for any demographic groupings consistent with the original NHANES 2005-2006 individual-sample sampling design. At this point in the estimation process, I chose the standard demographic groupings included in CDC's National Reports on Human Exposure to Environmental Chemicals [2009, 2012]. These demographic groupings are as follows: Total (civilian non-institutionalized U.S. population 12 years and older); Age group 12-19 years, Age group 20+ years; Males, Females; non-Hispanic Whites (NHW), non-Hispanic Blacks (NHB), Mexican Americans (MA). Then for each of the *n* sets of 2345 imputed results ($\breve{y}_{ijkl}$), I used Proc SurveyReg from SAS software version 9.3 (SAS Institute, Cary, NC) to calculate point estimates ($\hat{P_{i..l}}$) and their corresponding standard errors ($\hat{\sigma}_{\hat{P}_{i..l}}$) adjusted for design effects, resulting in *n* sets of estimates for each demographic grouping. Note that the point estimate $\hat{P_{i..l}}$ can be obtained directly from the Proc SurveyReg output, but the standard error of $\hat{\sigma}_{\hat{P}_{i..l}}$ of $\hat{P_{i..l}}$ is not readily available when $\hat{P_{i..l}}$ represents a percentile. It can be obtained indirectly, however, using the Woodruff (1952) method combined with the method proposed by Korn and Graubard (1999) for obtaining Clopper and Pearson 1-α confidence limits from complex surveys. This method uses the standard error of the empirical distribution function at the selected percentile and constructs a 1-α confidence interval, followed by back transformation using the inverse of the empirical distribution. So to estimate the standard error, I first calculated the difference between the upper and lower Clopper-Pearson 1-α confidence limits for the log transformed parameter estimate (i.e. $U_{(1-\alpha/2)}{}_{log\left(\hat{P}_{i..l}\right)}-L_{(\alpha/2)}{}_{log\left(\hat{P}_{i..l}\right)}$). I then divided this difference by twice the (1-α/2) critical value of a t-distribution with degrees of freedom ($df_{\hat{\sigma}_{\hat{P}_{i..l}}}$) associated with the simulated point estimate in the i^{th} demographic group and l^{th} simulation (l = 1, 2, ..., *n*). After this variance (square of the standard error) was estimated for each of the *d* demographic groups and *n* simulations, I used the method of Rubin (1987) to calculate the averages ($\hat{P}_i=\sum_l^n \hat{P}_{i..l}/n$) and ($\hat{\sigma}^2_{AW_i}=\sum_l^n \hat{\sigma}^2_{\hat{P}_{i..l}}/n$), respectively, acroos the *n* imputations to obtain a point estimate ($\hat{P_i}$) and within-imputation variance estimate ($\hat{\sigma}^2_{AW_i}$) for each demographic group *i* (=1, 2, . . . , *d*). I estimated the between-imputation variance ($\hat{\sigma}^2_{AB_i}$) for the i^{th} demographic group by computing the sample variances of the *n* imputed point estimates of interest (Rubin 1987). These adjusted within- and between-imputation variance estimates reflect the within- and between-imputation variability due to the use of estimated design effects. I estimated the adjusted total variance ($\hat{\sigma}^2_{AT_i}$) as the sum of the adjusted within-imputation variance ($\hat{\sigma}^2_{AW_i}$) and (*n*+1)/*n* times the adjusted between-imputation variance ($\hat{\sigma}^2_{AB_i}$).

The overall estimated standard error (adjusted for design effects) associated with a point estimate for the i^{th} demographic group is the square root of this adjusted variance estimate. I calculated approximate 100(1-α)% confidence intervals for $(\hat{P_i})$ as follows:

$$LL_{\hat{P}_i} = exp\left\{\hat{P}_i - t_{1-\alpha/2,df_{AT_i}}\,\hat{\sigma}_{ATi}\right\} \quad \text{and}$$
$$UL_{\hat{P}_i} = exp\left\{\hat{P}_i + t_{1-\alpha/2,df_{AT_i}}\,\hat{\sigma}_{ATi}\right\},$$

based on a Student's t-distribution, with estimated degrees of freedom as given by (Rubin 1987):

$$df_{ATi} = (n-1)\left[1+\hat{\sigma}^2_{AW_i}/\left((1+1/n)\,\hat{\sigma}^2_{AB_i}\right)\right]^2 \quad (4)$$

where $\hat{\sigma}^2_{AWi}$ and $\hat{\sigma}^2_{ABi}$, are the within- and between-imputation variances, respectively, and n is the number of imputed data sets.

3. RESULTS

3.1 Simulation Experiment to Evaluate the Multiple Imputation (MI) Method

To evaluate the MI method, I simulated 1000 NHANES individual data sets and formed 8-sample pools according to the NHANES 2005-2006 pooled-sample design. From these pooled-sample results I calculated MI method estimates of the 50th, 75th, 90th, and 95th percentiles along with their corresponding design effect adjusted 95% CIs based on n = 25 imputations. Twenty-five imputations were chosen in order to achieve imputation relative efficiencies (Rubin 1987) of at least 0.97, where a relative efficiency of 1.0 corresponds to an infinite number of imputations.

Theoretically, equation (1) with $\bar{x_{ij.}}$ and $\hat{\sigma}^2_{y_i}$ replaced by the true pooled-sample means and the true variances among the natural logarithms of the individual samples, respectively, provides an unbiased estimate of the geometric mean of the i^{th} pooled-sample demographic group. Similarly, equation (2) provides an unbiased estimate of the variance of the natural logarithms of the individual samples. Thus, on average, the percentile estimates obtained from the imputed individual results generated by equation (3) should be unbiased. To validate this assertion, I computed the average biases of the pooled-sample based point estimates and the average coverage probabilities of the 95% confidence intervals. These biases and coverage probabilities are displayed in Table 2 for each demographic group.

From Table 2 it can be seen that the percent bias in estimating the 50th percentile ranges from −5.4% to 8.1% and on average is 1.2%. The coverage probability of the 95% CIs around the true 50th percentile ranges from 0.855 to 0.993 and on average is 0.956 indicating that the CIs on average are right on target, but when the bias for a particular demographic group is between −5% and +5% the 95% CIs tend to be conservative. The percent bias in estimating the 95th percentile ranges from −4.1% to 5.9% and on average is 0.0%. The

coverage probability of the 95% CIs around the true 95th percentile ranges from 0.946 to 0.996 and on average is 0.983 indicating that the CIs on average are conservative.

3.2 Application of the Multiple Imputation (MI) Method to Actual Pooled-Sample Data from NHANES 2005-2006

To illustrate the MI estimation method using actual pooled-sample data from NHANES 2005-2006, I present geometric mean and various percentile estimates along with their design effect adjusted CIs for PCB153 and *p,p*'-DDE. Because the pooled-sample estimation method relies on variance estimates to correct biases, I chose for the illustration PCB153 and *p,p*'-DDE, which differ in terms of their between-subject variance structure. The between-subject variance for PCB153 measurements decreases slightly with increasing concentration, whereas the between-subject variance for *p,p*'-DDE increases with increasing concentration. The geometric mean and variance estimates for PCB153 and *p,p*'-DDE from NHANES 2005-2006 were calculated using equations 1 and 2 and then the MI method described in Methods (using $n = 25$ imputed data sets) was used to obtain pooled-sample point estimates and 95% confidence limits adjusted for design effects.

To demonstrate how well the *pooled-sample* estimates agree with *individual-sample* based estimates from previous NHANES surveys, I present *individual-sample* based estimates from NHANES 1999-2000, 2001-2002, and 2003-2004 along with the *pooled-sample* estimates from NHANES 2005-2006. Tables 3 and 4 display for PCB153 and *p,p*'-DDE, respectively, geometric mean and various percentile estimates along with 95% confidence limits for each of the 8 demographic groupings typically reported in the National Exposure Reports of the Centers for Disease Control (2009, 2012). These demographic groupings are as follows: Total (civilian non-institutionalized U.S. population 12 years and older); Age group 12-19 years, Age group 20+ years; Males, Females; and non-Hispanic Whites (NHW), non-Hispanic Blacks (NHB), and Mexican Americans (MA).

The geometric mean and percentile estimates based on *pooled-samples* from NHANES 2005-2006 in Table 3 are consistent with the steady decline of PCB153values that has been observed in previous NHANES (1999-2000, 2001-2002, and 2003-2004) based on *individual samples*. The PCB instrument limit of detection improved after survey years 2001-2002 so geometric mean and percentile estimates based on *individual samples* were reported for all demographic groups for survey years 2003-2004. Interestingly, even though PCB levels continued to decline, the *pooled-sample* method (see survey years 2005-2006) was able to provide estimates of geometric means and the selected percentiles for all demographic groups, which might not have been the case (or might not be the case in future survey years) if estimates were based on *individual samples*.

The geometric mean and percentile estimates based on pooled-samples from NHANES 2005-2006 in Table 4 are also consistent with the rather stable levels of *p,p*'-DDE that have been observed in previous NHANES (1999-2000, 2001-2002, and 2003-2004) based on individual samples. For example, 95th percentile estimates of *p,p*'-DDE in the 12-19 age group for survey years 1999-2000, 2001-2002, and 2003-2004 are 528, 456, and 522 ng/g of lipid, respectively, and the 2005-2006 pooled-sample estimate in Table 4 is 667 ng/g of lipid. Similarly, 95th percentile estimates of *p,p*'-DDE in the 20 years of age and older group

for survey years 1999-2000, 2001-2002, and 2003-2004 are 2020, 2550, and 1990 ng/g of lipid, respectively, and the 2005-2006 pooled-sample estimate is 2157 ng/g of lipid. Interestingly, for Mexican Americans the (2005-2006) *pooled-sample* results are more in line with the *individual-sample* results for 1999-2000 and 2001-2002 than are the *individual-sample* results for 2003-2004, suggesting that the *p,p*'-DDE estimates for Mexican Americans may have been underestimated in the 2003-2004 survey.

4. DISCUSSION

Using pooled samples from NHANES requires pooling across the sampling design cells of the original survey. As a result, direct calculation of design effects is not possible. In this paper I present a method that creates imputed individual-sample data from pooled-sample based estimates. With the imputed individual-sample results it is possible to approximate the complex survey design of the original survey and, thereby, estimate standard errors and CIs adjusted for design effects. Of course, the true design effects associated with the original individual samples will not be captured by any one set of imputed results because the imputed results will not necessarily exhibit the true intra-cluster correlations. So when estimating design effects I use multiple imputation to incorporate the within- and between-imputation variation as a means of capturing the possible range in intra-cluster correlations.

Because there was no NHANES data for which both individual-sample results and pooled-sample results were available, I simulated individual-sample data and created artificial pooled-samples using a pooled-sample design similar to the one used with the 2005-2006 NHANES data. I then compared pooled-sample percentile estimates and their corresponding CIs for various demographic groups with estimates based on individual samples. The average bias of the percentile estimates ranged from 0.0% to 1.2% and the average coverage probability ranged from 0.956 to 0.987.

To demonstrate the output that can be obtained using the MI method discussed in this paper I present geometric mean and various percentile estimates along with their associated CIs for PCB153 and for *p,p*'-DDE for the standard demographic groupings in the National Reports on Human Exposure to Environmental Chemicals [CDC, 2009, 2012]: Total (civilian non-institutionalized U.S. population 12 years and older); Age group 12-19 years, Age group 20+ years; Males, Females; and non-Hispanic Whites (NHW), non-Hispanic Blacks (NHB), and Mexican Americans (MA).

In this paper I have shown how pooled-sample estimates from a complex multistage, probability sampling design can be used to impute individual sample data and thereby obtain estimates of summary statistics and their corresponding standard errors or confidence limits. Such imputed data sets may not be appropriate for use in association studies that model exposure variables to determine relationships among variables at the individual level (Greenland and Robins 1994), due to the fact that actual values randomly assigned to individual samples would not reflect the true association with other variables based on individual- or pooled-sample measurements. But those problems may be circumvented to some extent if pools are formed in the right way (e.g., forming pooled-samples by using

very discrete categories of multiple demographic variables). More work would need to be done to determine the possibilities and limitations of such an approach.

Future areas of research into the use of pooled-samples from surveys, such as NHANES, will likely include consideration of alternative variance estimation and modeling methods, exploration of the effects of left censoring (i.e., left censoring in individual samples may or may not lead to left censoring in pooled-samples), and determination of the extent to which association studies and longitudinal studies based on pooled-samples might be possible. Weinberg and Umbach (1999) have shown how additional independent categorical and continuous variables, whether based on pooling or measured individually, might be included in a logistic model containing a pooled exposure (independent) variable. They also show how interactions and transformations might be handled in these models. Saha-Chaudhuri et al (2011) have extended the work of Weinberg and Umbach and have shown that unbiased estimation of the individual-level odds ratio parameter can be based on pooled exposure measurements for a fine-matched case-control study. More work would need to be done to determine whether these modeling techniques would extend to models in which the dependent variable is based on pooled-samples from log-normally distributed populations.

5. CONCLUSIONS

Because pooling samples reduces the costs of biomonitoring and can lower limits of detection by allowing for larger sample volumes, use of pooled biomarkers is becoming more common in an increasing number of exposure assessments and environmental epidemiologic investigations. When biomonitoring is based on pooled-samples from a complex survey design, direct calculation of the design effects needed for accurate standard error estimation is not possible. The multiple imputation method presented here can be used to obtain pooled-sample estimates with standard errors adjusted for design effects and those estimates do not have to be limited to the demographic groups used to form the pools. The methodological advance presented in this paper is significant because it offers a way to obtain standard errors (and hence confidence limits and statistical tests) for pooled-sample estimates that correspond to the original individual-sample based sampling design of the survey.

Acknowledgments

I thank Te-Ching Chen, Hua Di, Brenda Lewis, Lester R. Curtin, and Jane Zhang, at NCHS for assistance with NHANES 2005-2006 samples. I also I thank Chevine Anderson, Sarah Anderson, Troy Cash, Kenroy Crawford, Yolanda Dalton, Autumn Decker, Emily DiPietro, Jerry Dublin, Darlinda Harry, Carolyn Hodge, Richard Jones, Cheryl McClure, Andreas Sjodin, Wayman Turner, Wanda Whitfield, and Yalin Zhang in the for pooling and analyzing samples. Finally, I would like to thank Yi Pan in the Quantitative Sciences and Data Management Branch of DHAP/NCHHSTP/CDC and Rey DeCastro in the Emergency Response Branch of DLS/NCEH/CDC for reviewing the manuscript.

REFERENCES

Aitchison, J.; Brown, JAC. The Lognormal Distribution. Cambridge University Press; London: 1963. p. 7-8.

Calafat AM, Kuklenyik Z, Caudill SP, Reidy JA, Needham LL. Perfluorochemicals in pooled serum samples from United States residents in 2001 and 2002. Environ Sci Technol. 2006; 40:2128–2134. [PubMed: 16646443]

Caudill SP. Characterizing populations of individuals using pooled samples. J. Expo. Sci. Environ. Epidemiol. 2010; 20:29–37. [PubMed: 19002216]

Caudill SP. Use of pooled samples from the national health and nutrition examination survey. Stat. Med. 2012; 31(27):3269–3277. [PubMed: 22492247]

Centers for Disease Control and Prevention, 2009. Fourth National Report on Human Exposure to Environmental Chemicals. Centers for Disease Control and Prevention, National Center for Environmental Health. Division of Laboratory Sciences; Atlanta, GA: 2012. Available from http://www.cdc.gov/ExposureReport/pdf/FourthReport.pdf [13 August 2015]

Curtin LR, Mohadjer LK, Dohrmann SM, Montaquila JM, Kruszan-Moran D, Mirel LB, et al. National Health and Nutrition Examination Survey: sample design, 1999-2006. Vital Health Stat. 2012; 2(155):1–39.

Greenland S, Robins J. Invited commentary: ecologic studies—biases, misconceptions, and counterexamples. Am. J. Epidemiol. 1994; 139:747–60. [PubMed: 8178788]

Kato K, Calafat AM, Wong LY, Wanigatunga AA, Caudill SP, Needham LL. Polyfluoroalkyl compounds in pooled sera from children participating in the National Health and Nutrition examination Survey 2001-2002. Environ. Sci. Technol. 2009; 43:2641–2647. [PubMed: 19452929]

Korn, EL.; Graubard, BI. Analysis of Health Surveys. Wiley; New York: 1999.

Xiang, Li; Kuk, AYC.; Xu, J. Empirical Bayes Gaussian likelihood estimation of exposure distributions from pooled samples in human biomonitoring. Stat. Med. 2014; 33(28):4999–5014. [PubMed: 25213192]

National Center for Health Statistics. Plan and operation of a health examination survey of U.S. youths 12-17 years of age. Vital Health Stat. 1969; 1(8)

National Center for Health Statistics. Plan and operation of the Third National Health and Nutrition Examination Survey, 1988-94. Vol. 1. National Center for Health Statistics. Vital Health Stat; 1994.

Rubin, DB. Multiple Imputation for Nonresponse in Surveys. Wiley & Sons; New York: 1987.

Saha-Chaudhuri P, Umbach DM, Weinberg CR. Pooled exposure analysis for matched case-control studies. Epidemiology. 2011; 22(5):704–712. [PubMed: 21747285]

SAS Institute Inc.. SAS/STAT 13.1 User's Guide. SAS Institute Inc., Cary; NC: 2013.

Weinberg CR, Umbach DM. Using pooled exposure assessment to improve efficiency in case-control studies. Biometrics. 1999; 55:718–726. [PubMed: 11314998]

Woodruff RS. Confidence intervals for medians and other position measures. JASA. 1952; 57:622–627.

Table 1

Number of subjects in the one-third subsample, number of individual serum samples available, number of usable samples, and number of pools formed from NHANES 2005–2006 participants per demographic group.

Race or Ethnicity	Gender	Number of subjects in the one-third subsample/Number of Samples Available/Number of Usable Samples (Number of Pools)			
		12–19 years	20–39 years	40–59 years	60+ years
Non-Hispanic White	Male	81/76/ 72 (9)	110/108/96 (12)	114/110/96 (12)	141/137/120 (15)
	Female	94/85/80 (10)	143/136/128 (16)	116/111/104 (13)	149/146/136 (17)
Non-Hispanic Black	Male	129/114/104 (13)	60/57/48 (6)	56/51/40 (5)	55/52/40 (5)
	Female	132/117/112 (14)	74/66/56 (7)	65/62/56 (7)	55/48/40 (5)
Mexican American	Male	106/96/88 (11)	87/84/72 (9)	44/43/32 (4)	38/38/32 (4)
	Female	143/133/128 (16)	88/84/72 (9)	50/50/48 (6)	38/37/24 (3)
Other Hispanic and non-Hispanic multiracial	Male	20/19/16 (2)	27/26/24 (3)	24/23/23[1] (3)	9/8/8 (1)
	Female	31/26/24 (3)	38/34/32 (4)	18/18/16 (2)	10/6/6[2] (1)

[1] With only 23 usable samples, two 8 sample pools and one 7 sample pool were created.

[2] With only 6 usable samples, one 6 sample pool was created.

Table 2

Average bias (%Bias) of pooled-sample estimates and coverage of 95% confidence intervals for 8 demographic groups based on 1000 simulations using the Multiple Imputation method with 25 imputed data sets.

Demographic Group	50th Percentile		75th Percentile		90th Percentile		95th Percentile	
	%Bias	Coverage	%Bias	Coverage	%Bias	Coverage	% Bias	Coverage
All	1.2	0.978	0.7	0.982	−0.7	0.994	−1.0	0.994
12-19 Years	−1.0	0.993	−0.9	0.994	−0.1	0.993	0.5	0.993
20 + Years	0.5	0.975	0.6	0.985	−0.8	0.994	−1.1	0.994
Males	8.1	0.855	1.5	0.982	−3.3	0.971	−4.1	0.983
Females	−5.4	0.910	0.0	0.993	2.1	0.989	2.1	0.996
NHWs	0.0	0.974	0.5	0.983	−1.0	0.997	−1.8	0.995
NHBs	2.7	0.987	2.5	0.982	5.6	0.901	−0.6	0.946
MAs	3.8	0.974	−1.2	0.995	5.3	0.966	5.9	0.963
Average	1.2	0.956	0.5	0.987	0.9	0.976	0.0	0.983

Environ Int. Author manuscript; available in PMC 2016 December 01.

Table 3

Geometric Mean and Various Percentile Estimates for Lipid Adjusted PCB153 by NHANES Survey (1999-2004 based on Individual Samples & 2005-2006 based on Pooled-Samples).

Demographic Group	Survey Years[1]	Geometric Mean	Selected Percentiles[2] (L95, U95)[3]				Sample Size[4]
			50th	75th	90th	95th	
Total	99-00	*[5]	< LOD[6]	< LOD	77.8 (70.2-87.3)	114 (93.0-133)	1926
	01-02	27.2 (24.7-30.1)	30.1 (26.1-34.3)	57.8 (52.1-63.2)	94.7 (86.5-104)	126 (109-142)	2306
	03-04	19.8 (18.8-20.9)	20.8 (18.4-22.2)	43.3 (39.1-46.9)	71.8 (64.4-82.8)	97.1 (88.8-111)	1896
	05-06	14.7 (13.5-15.9)	15.7 (13.5-18.2)	34.9 (30.2-40.3)	65.8 (55.9-77.5)	92.7 (78.8-109)	2345
Age Group 12-19 years	99-00	*	< LOD	< LOD	< LOD	< LOD	668
	01-02	*	< LOD	12.5 (11.1-14.1)	21.2 (17.4-26.7)	31.9 (23.1-64.7)	757
	03-04	5.86 (5.25-6.55)	5.40 (4.70-6.21)	8.50 (7.80-9.85)	15.7 (12.9-18.4)	20.7 (16.9-28.3)	596
	05-06	3.74 (3.32-4.20)	3.76 (3.14-4.50)	6.74 (5.70-7.97)	11.2 (8.66-14.5)	15.1 (11.4-20.0)	736
Age Group 20 years and older	99-00	*	< LOD	< LOD	83.2 (75.9-91.8)	122 (100-139)	1258
	01-02	32.6 (29.5-36.1)	35.1 (31.1-39.0)	62.8 (57.6-68.0)	99.5 (90.7-110)	132 (116-146)	1549
	03-04	23.7 (22.3-25.1)	24.2 (21.8-27.4)	47.1 (43.3-50.5)	77.5 (68.0-87.9)	101 (92.9-119)	1300
	05-06	18.3 (16.8-19.9)	19.3 (17.1-21.9)	39.4 (33.9-45.8)	71.4 (60.6-84.1)	99.0 (83.8-117)	1609
Gender Males	99-00	*	< LOD	< LOD	75.0 (66.7-86.2)	111 (87.7-128)	917
	01-02	28.5 (25.5-32.0)	31.5)26.7-35.2)	57.7 (48.3-66.2)	97.5 (82.1-110)	126 (104-150)	1074
	03-04	20.0 (18.7-21.3)	19.7 (17.7-21.2)	42.9 (37.4-47.6)	72.7 (60.4-88.8)	107 (86.8-122)	947
	05-06	15.9 (14.3-17.7)	17.0 (14.1-20.6)	36.2 (29.4-44.7)	66.2 (53.8-81.6)	92.3 (74.7-114)	1101
Gender Females	99-00	*	< LOD	< LOD	79.0 (70.2-92.0)	119 (91.4-142)	1009
	01-02	26.1 (23.6-28.8)	29.0 (25.1-33.4)	57.9 (52.1-62.9)	94.3 (87.8-98.2)	128 (105-145)	1232
	03-04	19.7 (18.4-21.1)	21.9 (19.0-24.1)	43.8 (39.4-47.7)	70.9 (63.0-81.5)	93.3 (83.8-100)	949
	05-06	13.6 (12.1-15.3)	14.5 (12.4-17.1)	33.6 (27.6-41.0)	65.2 (53.6-79.3)	93.1 (74.4-117)	1244
Race/Ethnicity Mexican Americans	99-00	*	< LOD	< LOD	< LOD	67.5 (59.5-71.8)	634
	01-02	12.5 (10.8-14.4)	11.1 (<LOD-13.3)	24.5 (18.2-33.9)	47.4 (36.2-60.3)	66.7 (55.2-72.3)	567
	03-04	8.75 (7.39-10.4)	7.86 (6.17-9.40)	15.6 (11.4-22.2)	30.3 (25.2-34.9)	37.8 (31.1-45.2)	425

Demographic Group	Survey Years[1]	Geometric Mean	Selected Percentiles[2] (L95, U95)[3]				Sample Size[4]
			50th	75th	90th	95th	
	05-06	5.79 (4.93-6.81)	5.53 (4.67-6.55)	12.9 (9.60-17.2)	26.1 (20.0-34.0)	38.4 (28.8-51.2)	594
Race/Ethnicity Non-Hispanic Black	99-00	*	< LOD	59.4 (<LOD-82.0)	121 (90.3-159)	176 (130-287)	412
	01-02	30.0 (26.2-34.4)	31.0 (25.8-36.4)	65.1 (54.2-82.7)	127 (97.1-152)	170 (126-246)	515
	03-04	22.8 (19.1-27.2)	20.9 (17.0-28.7)	54.1 (37.3-69.2)	126 (92.9-158)	194 (126-294)	464
	05-06	16.2 (13.8-18.9)	16.1 (13.0-19.9)	41.4 (34.2-50.2)	89.6 (70.8-113)	133 (105-170)	626
Race/Ethnicity Non-Hispanic White	99-00	*	< LOD	< LOD	76.4 (69.3-83.9)	102 (87.8-127)	725
	01-02	29.9 (26.8-33.4)	33.0 (28.7-37.1)	61.2 (55.8-66.7)	96.3 (86.5-109)	126 (104-142)	1061
	03-04	21.3 (19.7-23.1)	22.2 (20.4-25.9)	44.9 (39.7-49.5)	70.9 (69.4-82.1)	91.3 (82.1-103)	885
	05-06	16.1 (14.5-17.8)	17.5 (14.7-20.8)	37.7 (31.3-45.3)	68.3 (56.9-81.9)	92.9 (76.7-113)	948

[1] Results for survey years 1999-2004 are based on *individual samples*. Results for survey years 2005-2006 are based on *pooled samples*.

[2] Estimated percentile in ng/g of lipid or parts per billion on a lipid-weight basis.

[3] Lower and upper 95% confidence limits adjusted for design effects (calculated from actual *individual samples* for survey years 1999-2004 and from individual samples imputed from *pooled-samples* for surveys 2005-2006).

[4] Sample Size is the number of individual samples in the random one-third subsample of NHANES 1999-2006 for polychlorinated and polybrominated compounds.

[5] An * indicates that no geometric mean estimate was reported because more than 40% of results were left censored (i.e., below the limit of detection of the instrument).

[6] The estimate was not reported because it was less than the limit of detection (LOD) of the instrument.

Environ Int. Author manuscript; available in PMC 2016 December 01.

Table 4

Geometric Mean and Various Percentile Estimates for Lipid Adjusted *p,p'*-DDE by NHANES Survey (1999-2004 based on Individual Samples & 2005-2006 based on Pooled-Samples).

Demographic Group	Survey Years[1]	Geometric Mean	Selected Percentiles[2] (L95, U95)[3]				Sample Size[4]
			50th	75th	90th	95th	
Total	99-00	260 (226-298)	226 (184-278)	537 (476-631)	1150 (976-1350)	1830 (1410-2300)	1964
	01-02	295 (267-327)	251 (228-278)	598 (521-699)	1410 (1210-1500)	2320 (1830-2780)	2298
	03-04	238 (195-292)	203 (163-275)	509 (376-655)	1170 (836-1570)	1860 (1400-2380)	1956
	05-06	196 (179-216)	179 (163-196)	446 (392-508)	1111 (962-1283)	2001 (1616-2478)	2345
Age Group 12-19 years	99-00	118 (102-135)	108 (97.7-119)	185 (141-237)	339 (243-479)	528 (339-812)	668
	01-02	124 (106-146)	113 (100-140)	213 (172-253)	319 (282-389)	456 (343-722)	758
	03-04	105 (84.7-129)	93.6 (81.0-114)	167 (123-240)	341 (211-586)	522 (313-1430)	588
	05-06	89.9 (78.1-104)	79.3 (67.4-93.3)	158 (132-190)	336 (250-450)	667 (440- 1011)	736
Age Group 20 years and older	99-00	297 (256-344)	269 (213-323)	608 (530-693)	1260 (1030-1550)	2020 (1520-2730)	1278
	01-02	338 (303-376)	285 (249-337)	695 (595-798)	1480 (1310-1700)	2550 (1980-3080)	1540
	03-04	268 (217-332)	233 (175-314)	557 (420-734)	1270 (877-1800)	1990 (1500-2470)	1368
	05-06	223 (202-246)	206 (185-228)	500 (444-562)	1221 (1022-1458)	2157 (1718-2709)	1609
Gender Males	99-00	249 (220-283)	223 (182-262)	494 (380-578)	1010 (789-1130)	1430 (1080-2160)	937
	01-02	285 (252-323)	248 (222-285)	520 (441-627)	1160 (937-1360)	1900 (1580-2490)	1069
	03-04	235 (193-288)	200 (164-262)	466 (331-653)	1000 (763-1400)	1610 (1210-2320)	955
	05-06	211 (188-237)	196 (169-227)	458 (384-547)	1102 (860-1412)	1936 (1471-2548)	1101
Gender Females	99-00	270 (226-322)	234 (184-302)	601 (492-711)	1350 (1040-1720)	2210 (1570-2810)	1027
	01-02	305 (273-341)	256 (219-297)	708 (567-844)	1480 (1410-1710)	2670 (1940-3300)	1229
	03-04	241 (193-301)	207 (161-281)	539 (386-735)	1250 (813-1900)	2010 (1500-2450)	1001
	05-06	184 (163-209)	162 (136-193)	434 (348-542)	1126 (890-1424)	2078 (1476-2925)	1244
Race/Ethnicity Mexican Americans	99-00	674 (574-792)	624 (545-701)	1350 (1090-1660)	3090 (2040-4950)	4950 (3070-9350)	657
	01-02	652 (569-747)	561 (455-690)	1400 (1050-1950)	4110 (2520-6550)	7080 (3080-15600)	566
	03-04	444 (362-545)	373 (283-522)	875 (608-1170)	2150 (1520-2470)	3290 (2380-9240)	457

Demographic Group	Survey Years[1]	Geometric Mean	Selected Percentiles[2] (L95, U95)[3]				Sample Size[4]
			50th	75th	90th	95th	
	05-06	578 (471-709)	561 (435-725)	1472 (1102-1967)	3564 (2502-5077)	6115 (3667-10197)	594
Race/Ethnicity Non-Hispanic Black	99-00	295 (241-362)	251 (199-313)	668 (492-874)	1850 (1040-2220)	2300 (1560-5680)	416
	01-02	324 (262-400)	248 (223-296)	762 (583-999)	1620 (1180-2980)	3260 (1270-6900)	515
	03-04	262 (233-295)	216 (173-267)	589 (453-747)	1620 (1130-2310)	2860 (1880-3440)	487
	05-06	204 (171-244)	185 (153-225)	459 (348-605)	1140 (801-1622)	2106 (1234-3595)	626
Race/Ethnicity Non-Hispanic White	99-00	217 (189-249)	194 (162-238)	438 (355-507)	825 (647-1010)	1160 (1010-1350)	732
	01-02	253 (226-284)	225 (203-254)	463 (402-558)	1150 (878-1340)	1640 (1410-1940)	1053
	03-04	208 (165-263)	177 (148-238)	417 (302-564)	907 (574-1480)	1490 (909-2300)	888
	05-06	157 (141-176)	147 (128-168)	337 (291-391)	749 (600- 935)	1237 (973-1571)	948

[1] Results for survey years 1999-2004 are based on *individual samples*. Results for survey years 2005-2006 are based on *pooled samples*.

[2] Estimated percentile in ng/g of lipid or parts per billion on a lipid-weight basis.

[3] Lower and upper 95% confidence limits adjusted for design effects (calculated from actual *individual samples* for survey years 1999-2004 and from individual samples imputed from *pooled-samples* for surveys 2005-2006).

[4] Sample Size is the number of individual samples in the random one-third subsample of NHANES 1999-2006 for polychlorinated and polybrominated compounds.