Environment International 85 (2015) 40–45

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

Full length article

# Confidence interval estimation for pooled-sample biomonitoring from a complex survey design

Samuel P. Caudill *

*Division of Laboratory Sciences, National Center for Environmental Health, Centers for Disease Control & Prevention, Public Health Service, US Department of Health and Human Services, United States*

ARTICLE INFO

*Article history:*
Received 1 October 2014
Received in revised form 20 July 2015
Accepted 1 August 2015
Available online xxxx

*Keywords:*
Design effects
Persistent organic pollutants
Pooled-samples
Biomonitoring
Complex survey design

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.

Published by Elsevier Ltd.

## 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 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.

* 4770 Buford Highway, NE MS-F25, Atlanta, GA 30341, United States.
*E-mail address:* SPC1@CDC.GOV.

http://dx.doi.org/10.1016/j.envint.2015.08.003
0160-4120/Published by Elsevier Ltd.

## 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 $(x_{ijk}; i = 1,2,\ldots d; j = 1,2,\ldots p_i; k = 1,2,\ldots,s)$ are log-normal with mean and variance of the natural logarithm of individual "unmeasured" results (i.e., $y_{ijk} = \ln(x_{ijk})$) equal to $\mu_{y_i}$ and ${\sigma_{y_i}}^2$, respectively. That is, the individual $y_{ijk}$ values are normal with mean $\mu_{y_i} = E(y_{ijk})$ and with variance ${\sigma_{y_i}}^2 = Var(y_{ijk})$. 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:

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

where the single measured value of a pool $(\overline{x}_{ij.})$ is comparable to a weighted average of log-normal values $\left[\overline{x}_{ij.} = \sum_{k=1}^{s} w_{ijk}x_{ijk}/w_{ij.}\right]$, $w_{ijk}$ 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:

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

Note that Eqs. (1) and (2) above differ slightly from Eqs. (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 Eq. (1) of Caudill (2012) has a typographical error in that the sum should have been from $k = 1$ to $s$.

**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.