Estimation of Background Serum 2,3,7,8-TCDD Concentrations By Using Quantile Regression in the UMDES and NHANES Populations

Qixuan Chen,[a,b] *David H. Garabrant,*[c,d] *Elizabeth Hedgeman,*[c] *Roderick J. A. Little,*[b] *Michael R. Elliott,*[b,e] *Brenda Gillespie,*[b] *Biling Hong,*[c] *Shih-Yuan Lee,*[b] *James M. Lepkowski,*[e] *Alfred Franzblau,*[c,d] *Peter Adriaens,*[f] *Avery H. Demond,*[f] *and Donald G. Patterson, Jr*[g]

Background: The goal of the present study was to quantify the population-based background serum concentrations of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) by using data from the reference population of the 2005 University of Michigan Dioxin Exposure Study (UMDES) and the 2003–2004 National Health and Nutrition Examination Survey (NHANES).

Methods: Multiple imputation was used to impute the serum TCDD concentrations below the limit of detection by combining the 2 data sources. The background mean, quartiles, and 95th percentile serum TCDD concentrations were estimated by age and sex by using linear and quantile regressions for complex survey data.

Results: Any age- and sex-specific mean, quartiles, and 95th percentiles of background serum TCDD concentrations of study participants between ages 18 and 85 years can be estimated from the regressions for the UMDES reference population and the NHANES non-Hispanic white population. For example, for a 50-year-old man in the reference population of UMDES, the mean, quartiles, and 95th percentile serum TCDD concentrations are estimated to be 1.1, 0.6, 1.1, 1.8, and 3.3 parts per trillion, respectively. The study also shows that the UMDES reference population is a valid reference population for serum TCDD concentrations for other predominantly white populations in Michigan.

Conclusion: The serum TCDD concentrations increased with age and increased more over age in women than in men, and hence estimation of background concentrations must be adjusted for age and sex. The methods and results discussed in this article have wide application in studies of the concentrations of chemicals in human serum and in environmental samples.

(*Epidemiology* 2010;21: S51–S57)

Submitted 12 August 2008; accepted 23 February 2009; posted 11 March 2010.

From the [a]Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY; Departments of [b]Biostatistics and [c]Environmental Health Sciences, University of Michigan School of Public Health, Ann Arbor, MI; [d]Risk Science Center, University of Michigan School of Public Health, Ann Arbor, MI; [e]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI; [f]Department of Civil and Environmental Engineering, University of Michigan College of Engineering, Ann Arbor, MI; and [g]Enviro Solutions Consulting, Inc., Jasper, GA.

Supported by the Dow Chemical Company.

Correspondence: Qixuan Chen, 722 West 168 St., R652, New York, NY 10032. E-mail: qc2138@columbia.edu.

ISSN: 1044-3983/10/2104-0051
DOI: 10.1097/EDE.0b013e3181ce9550

The University of Michigan Dioxin Exposure Study (UMDES) was conducted in response to concern that people's body burdens of dioxins might be elevated in Midland and Saginaw counties, Michigan, because of environmental contamination from the Dow Chemical Company facilities in the City of Midland and sediments in the Tittabawassee River flood plain. To assess whether the concentrations of blood serum dioxins are elevated among residents in Midland/Saginaw, they need to be compared with the background concentrations of serum dioxins in other areas where there are no known, unusual sources of dioxin exposures. The most studied dioxin congener, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD), is formed as an unintentional byproduct of incomplete combustion and has been classified as a probable human carcinogen.[1] Our goal in this paper is to estimate the mean and quantiles of serum TCDD levels in the general population.

Studies have shown that, among the general public, the concentrations of serum dioxins increase with age.[2,3] These increases are most likely the result of higher levels of dioxins in the environment in the 1960s and 1970s than in recent years, the number of years of past exposure, and slower elimination among older people. In addition, the difference in serum dioxin concentrations by sex may be due to differences in elimination between men and women.[4] Therefore, the estimation of background concentrations of serum dioxins must be adjusted for these factors.

In exposure assessment, quantiles are sometimes of more interest than means, from a public health perspective. In the presence of a skewed distribution, quantiles can also catch important information that might be missed by measurements of central tendency and dispersion. Because age and sex are associated with serum TCDD concentrations, an age- and sex-specific quantile estimate among the reference population is of greater interest than a univariate quantile estimate. Quantile regression is used to estimate and allow inferences about conditional quantile functions given covariates,[5] similar to linear regression, which is used to predict conditional means given covariates.

We used 2 referent populations for estimation of the quantiles of serum TCDD levels: the population of Jackson and Calhoun counties in Michigan and the 2003–2004 National Health and Nutrition Examination Survey (NHANES).[6] The Jackson/Calhoun sample was advantageous because it was representative of local Michigan residents. The NHANES sample was advantageous because it included large numbers of participants and was representative of the US general population. However, for the serum TCDD, about half the NHANES data were below the limit of detection (LOD). The LOD is defined as the concentration of analyte that gives a signal equal to a laboratory blank (obtained when no analyte is present) plus 3 times the standard deviation (SD) of the blank.[7] The LOD represents the level below which we cannot be confident whether the analyte is actually present. The high proportion of TCDD samples below LOD in the NHANES data has resulted in difficulty in estimating age- and sex-specific mean, median, and lower quantiles of serum TCDD concentration in general US population based only on the NHANES data.

The present study applies linear and quantile regression methods in the setting of complex survey data to quantify the age- and sex-specific mean and quantiles estimates of the background serum TCDD concentrations in the general Michigan population and separate estimates in the general US population. It also illustrates a multiple imputation approach for imputing values below the LOD. The LOD issue has posed formidable limitations to the estimation of serum TCDD levels (and levels of other environmental contaminants that are commonly measured near the LOD) in the general population. Conventional approaches of imputing the values below the LOD as 0, LOD, LOD/2, or LOD/$\sqrt{2}$ depend on the blood sample volume and the LOD levels of the measurement methods and may lead to biased estimates of serum TCDD concentration, especially in the scenario of high proportion of data above the LOD and high LOD levels.[8]

METHODS

Study Population

Jackson and Calhoun counties, Michigan, are more than 100 miles away from Midland, Michigan. The population of these counties was chosen as the reference population in the UMDES because it was similar to Midland and Saginaw counties in terms of demographics, urban/rural distribution, and percentage of employment in industry—except that there is no known, unusual source of dioxins (such as the Dow Chemical Company). To be eligible for participation in this study, the Jackson/Calhoun residents were required to be 18 years or older and to have lived in their current residence for at least 5 years. The sampling used a 2-stage area probability selection of housing units in Jackson and Calhoun counties and a third stage of selection of an eligible person within each sample housing unit.[9] Participants provided written, informed consent that had been approved by the University of Michigan Health Sciences Institutional Review Board. Participants who met Red Cross criteria for blood donation (no clotting disorders or blood thinner medications, no recent chemotherapy, weight of at least 110 lb, etc.) were invited to provide an 80-mL sample of blood. In Jackson/Calhoun counties, the study cooperation rates (proportion of known eligible persons who provided data) were 82.2% in the interviewing stage and 78.4% in the blood collection stage.[9] A total of 359 persons completed the UMDES study questionnaire and among these 251 gave blood samples in the summer of 2005.[10]

Vista Analytical Laboratory of El Dorado Hills, CA performed all serum TCDD analyses by using high-resolution gas chromatography–mass spectrometry. To ensure the precision and accuracy of the serum results, Vista Analytical Laboratory first synchronized its serum analysis methods with the methods of the National Center for Environmental Health (NCEH) laboratories at the Centers for Disease Control and Prevention before the start of UMDES fieldwork. Additionally, 20 serum quality assurance and quality control samples were supplied by NCEH, which were blind analyzed by Vista Analytical Laboratory during fieldwork, and the results were verified by NCEH laboratories. The mean lipid content of these samples was measured at 586 mg/dL (SD = 20) by Vista, compared with 603 mg/dL (SD = 21) by NCEH. Vista's analytical results for TCDD concentration were within 2 SDs of the sample means determined by NCEH after repeated testing of these samples over time. Serum standard reference materials, supplied by the National Institute of Standards and Technology, and pooled serum samples were analyzed periodically (1 each per 40 samples) to verify the method performance. The serum TCDD concentration was divided by the total lipids and was reported in parts per trillion (ppt) or picograms per gram of lipids. The lipids were determined by measurements of triglycerides and total cholesterol and then the total lipids were calculated using the Phillips method, as was done in the NCEH laboratory.[11]

One limitation of the Jackson/Calhoun data is that relatively few participants (n = 20) were older than 75 years, especially men (n = 3). As a result, estimating age- and sex-specific upper percentiles was problematic in this group. However, a substantial data set of serum TCDD concentrations in adults aged 18–85 years exists in the 2003–2004 NHANES.[6] The serum dioxin analyses in the NHANES were performed by NCEH. Because the methods for serum dioxin and lipid quantification are comparable between the NCEH laboratories and the Vista Analytical Laboratory and the results were verified via blind sample introduction, the blood serum data of the UMDES and the NHANES can be combined with little or no expectation of bias. Because the population in Jackson/Calhoun counties was predominantly non-Hispanic whites (91%) and pregnant women were excluded, we examined information from the NHANES subsample of 719 non-Hispanic whites (excluding pregnant

women) for whom there were serum TCDD measures. There were 98 participants (40 men) older than 75 years in whom there were serum TCDD concentrations above the LOD in the NHANES data, and who could be useful in improving the age- and sex-specific percentile estimates among older people in Jackson/Calhoun.

Statistical Analyses

The 719 observations from NHANES data were concatenated with the 251 observations from UMDES Jackson/Calhoun data, with an indicator for data source (1 for NHANES, 0 for UMDES). To be consistent with the NHANES data set, participants who were older than 85 years in the Jackson/Calhoun dataset were recorded as being age 85 to preserve the anonymity of people participating in the study. Age and sex were fully observed for both samples. Other covariates potentially associated with serum TCDD concentration were body mass index (BMI), recent BMI change, cigarette smoking, income, education, and breast-feeding history among women.[12] All of these covariates had less than 7.5% of data missing in both samples. They were imputed separately by using a sequential regression imputation method in both samples before the combination.[13] The survey sampling weights were standardized within each data source by dividing by their respective mean sampling weights and then multiplying by 100 to maintain the ratio of the data source sample sizes; this prevents the analysis results from being overwhelmed by the NHANES data due to its much larger sampling weights (each individual observation represents many more people in the population).

A multiple imputation technique was performed to impute the TCDD concentrations for those below the LOD in the combined data of Jackson/Calhoun and NHANES.[14,15] For each imputation, a bootstrap sample of 970 (n = 251 for Jackson/Calhoun, n = 719 for NHANES) observations was generated from the combined Jackson/Calhoun and NHANES data, and a survey-weighted left-censored (Tobit) linear regression model, assuming a lognormal distribution, was fitted on the bootstrap sample with important covariates including data source, age, sex, BMI, BMI change in the past 12 months, pack-years of cigarette smoking, number of children breast-fed (among women), income, education, and the 2-way interaction terms among age, sex, and data source. Then, for those participants having values below the LOD, the natural logarithm–transformed imputed values were drawn from a normal distribution with mean and variance estimated from the left-censored regression model, with left truncation at their corresponding natural logarithm LOD. This procedure was repeated 5 times to generate 5 imputed data sets.

A natural logarithmic transformation was applied to the serum TCDD concentrations because there is an approximately linear association between log (serum TCDD) and age. To estimate the age- and sex-specific serum TCDD measures, survey-weighted mean, quartiles (25th percentile, median, and 75th percentile), and 95th percentile of quantile regression models of serum TCDD concentrations were fitted on age, sex, data source indicator, and their 3 2-way interaction terms. Age was centered at 50 years to facilitate interpretation of the intercept and to remove collinearity of age with its interaction terms with sex and data source. Because neither the interaction term between age and data source indicator nor the interaction term between sex and data source indicator was significant in any of the mean or quantile regression models, they were removed from all of the models. For the mean regression, we used the conventional method for complex surveys: the SURVEYREG procedure in SAS, version 9.1 (SAS Institute Inc., Cary, NC). However, because there is no statistical software package currently available that provides correct standard error estimates for quantile regression in complex surveys, we corrected the estimates of standard errors of the regression coefficients by using 1000 bootstrap samples.[16] The estimates for each parameter from 5 imputed data sets were averaged to get the combined parameter estimate, and the variances were computed using standard multiple imputation-combining rules that account for between and within imputation variances.[14]

We used the bootstrap method for stratified multistage samples in both the multiple imputation and the quantile regression.[17] For a single replicate of bootstrap, for each stratum h, draw, from the n_h primary sampling units (PSUs) in the sample, a simple random sample with replacement of $m_h = (n_h - 1)$ PSUs. Let $r_{hi}^{(t)}$ denote the number of times that PSU i from stratum h is included in replicate t and let w_{hij} denote the sample weight for unit j in the PSU i and stratum h; the bootstrap weights were calculated as $w_{hij}^{(t)} = w_{hij} \cdot \frac{n_h}{n_h - 1} \cdot r_{hi}^{(t)}$. The bootstrap weights were then used for statistical analysis in the bootstrap samples.

Predictions of conditional mean, quartiles, and 95th percentile of the serum TCDD concentrations were plotted in raw scale versus age. Each value below the LOD was plotted using the average of its imputed values in 5 imputed data. All P values are based on 2-sided hypothesis tests. The statistical analyses were performed using SAS, version 9.1, and the Figure was created by R version 2.6.1 (R Development Core Team, Vienna, Austria).

RESULTS

Table 1 presents characteristics of the 251 Jackson/Calhoun UMDES participants with the 719 NHANES non-Hispanic white participants shown for comparison. The proportion of serum TCDD values below the LOD was 48% in the NHANES data compared with 21% in the Jackson/Calhoun data. The median LOD levels among the samples below the LOD were 1.1 ppt in the NHANES data, but 0.5 ppt in the Jackson/Calhoun data. The differences in the

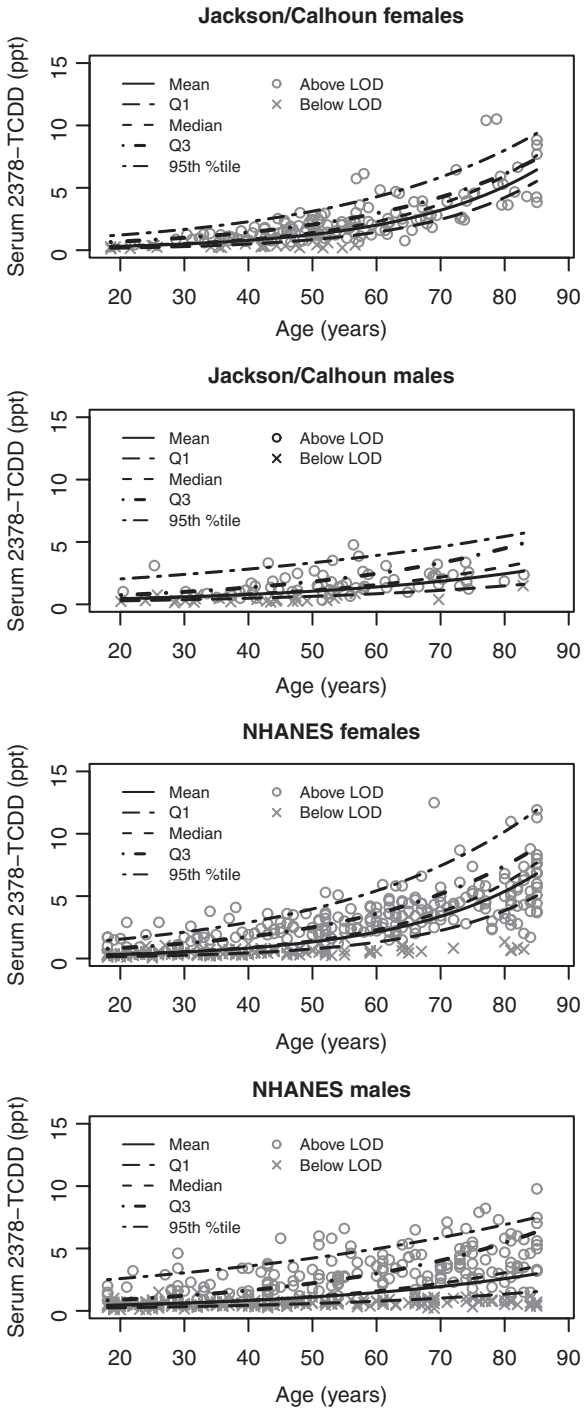

FIGURE. Comparisons of predicted mean, quartiles, and 95th percentile of serum TCDD levels over age by sex between the Jackson/Calhoun, Michigan, 2005, and the NHANES 2003–2004 populations.

proportion of serum TCDD values below the LOD were due in part to larger serum specimens analyzed in Jackson/Calhoun population (20 mL) compared with the NHANES (5–10 mL). The 2 populations were similar in BMI, BMI change in the last 12 months, pack-years smoking, income, education, and number of children breast-fed (among women). However, the Jackson/Calhoun population was slightly older ($P = 0.05$) and had a smaller proportion of men ($P = 0.04$) than the NHANES population.

Table 2 shows results of the 5 regression models with parameter and standard error estimates. Age was a strong positive predictor in all 5 regression models ($P < 0.01$), and the age and sex interaction term was also significant in the mean, 25th percentile, median, and 95th percentile regressions. For example, for each 10-year increase in age, the mean serum TCDD concentrations were estimated to be increased by 60% ($e^{0.047 \times 10 \text{ years}} = 1.60$) among women and by 34% ($10^{(0.047 - 0.018) \times 10 \text{ years}} = 1.34$) among men. The data source variable was not significant in the mean, 25th percentile, median, or 95th percentile regressions but had marginally positive significant effects in the 75th percentile ($P = 0.07$). This indicates that the Jackson/Calhoun population is similar to the NHANES non-Hispanic white population in the age- and sex-specific serum TCDD concentration.

Any age- (between ages 18 and 85 years) and sex-specific predicted mean, quartiles, and 95th percentile of background serum TCDD concentrations can be obtained for Jackson/Calhoun and for the NHANES from the regression results in Table 2. For example, the predicted 95th percentile of serum TCDD concentrations measured in parts per trillion equals $\exp^{(1.139 + 0.031 \times (\text{age-}50) + 0.063 \times \text{sex} - 0.015 \times \text{sex} \times (\text{age-}50) + 0.239 \times \text{source})}$. The predicted mean and 3 quartiles can be obtained similarly. Table 3 displays these estimates for 50-year-old men and 50-year-old women from Jackson/Calhoun and the NHANES as examples. For a 50-year-old man (woman) in Jackson/Calhoun, the mean, 25th percentile, median, 75th percentile, and 95th percentile serum TCDD concentrations are estimated to be 1.1 (1.3), 0.6 (0.8), 1.1 (1.4), 1.8 (2.1), and 3.3 (3.1) ppt, respectively; and for a 50-year-old man (woman) in the NHANES, the mean, 25th percentile, median, 75th percentile, and 95th percentile serum TCDD concentrations are estimated to be 1.1 (1.3), 0.6 (0.8), 1.1 (1.4), 2.2 (2.5), and 4.2 (4.0) ppt, respectively.

The Figure compares the predicted mean, 3 quartiles, and 95th percentile serum TCDD values over age by sex between the NHANES and the UMDES reference populations. A circle represents an observed serum TCDD concentration above the LOD, and an "x" is the average of the 5 imputations for those below the LOD. The plots show that for people older than 75, the NHANES data improved the estimates in the Jackson/Calhoun population, especially among men (age- and sex-specific upper percentiles among people older than 75 could not be fitted using only the Jackson/Calhoun data). In addition, the background serum TCDD concentrations increased with age and increased more steeply with age in women than in men in both data sources. Moreover, the plots show that the 75th and 95th percentile regres-

TABLE 1. Comparison of LOD and Population-based Demographics Between Jackson/Calhoun, Michigan, 2005 and NHANES 2003–2004 Populations

	NHANES (n = 719)[a]	Jackson/Calhoun (n = 251)	P[b]
Proportion below LOD (amt serum)	48% (5–10 mL)	21% (20 mL)	—
Median LOD levels (range)[c]	1.1 (0.4–3.1) ppt	0.5 (0.3–3.2) ppt	—
Mean age (range)	47.0 (18–85) yrs	49.9 (18–85) yrs	0.051
Mean BMI change in the last 12 months (SE)	0.2 (0.1) kg/m^2	−0.1 (0.2) kg/m^2	0.194
Mean BMI (SE)	27.7 (0.3) kg/m^2	28.7 (0.5) kg/m^2	0.101
Mean pack-years smoking (SE)	11.3 (0.6)	12.5 (1.4)	0.440
Mean no. children breast-fed among women (SE)	0.8 (0.1)	1.0 (0.1)	0.198
Mean income (SE)	$52,000 (2000)	$56,000 (2000)	0.220
Sex (proportion of men)	47.6%	38.1%	0.035
Education (proportion of ≥ high school)	86.6%	86.2%	0.897

[a]Non-Hispanic white adults (excluding pregnant women) having serum TCDD measures in 2003–2004 NHANES.
[b]P values using F-tests to compare the population-based demographics between the NHANES and Jackson/Calhoun populations.
[c]The median LOD levels (among the observations below LOD).

TABLE 2. Results of Linear and Quantile Regressions of Log (Serum TCDD Concentration) in the Combined Data of Jackson/Calhoun, Michigan, 2005 and NHANES 2003–2004

Factor	Mean[a]	Q_1[b]	Median[b]	Q_3[b]	95th Percentile[b]
Intercept	0.232 (0.069)*	−0.176 (0.149)	0.346 (0.058)*	0.726 (0.075)*	1.139 (0.119)*
Age[c]	0.047 (0.003)*	0.054 (0.005)*	0.048 (0.003)*	0.036 (0.002)*	0.031 (0.005)*
Sex[d]	−0.183 (0.082)**	−0.273 (0.136)***	−0.223 (0.103)**	−0.126 (0.087)	0.063 (0.108)
Age[c] × sex[d]	−0.018 (0.004)*	−0.025 (0.007)*	−0.015 (0.005)**	−0.006 (0.005)	−0.015 (0.007)**
Source[e]	0.051 (0.084)	−0.094 (0.169)	0.015 (0.100)	0.190 (0.099)***	0.239 (0.151)

[a]The mean model was obtained by fitting a linear regression for complex survey data using SURVEYREG procedure in SAS.
[b]The quantile models were fitted using quantile regressions for complex survey data by using bootstrap method to calculate the standard errors (Q_1: 25th percentile; Q_3: 75th percentile).
[c]Age minus 50 (yrs).
[d]Sex (women = 0, men = 1).
[e]Data source (Jackson/Calhoun = 0, NHANES = 1).
Results are reported as estimate (standard error) P value; *$P \leq 0.01$; **$P \leq 0.05$; ***$P \leq 0.1$.

TABLE 3. Predicted Mean, Quartiles, and 95th Percentile for a 50-year-old Person by Sex

Units = ppt (Lipids)	Mean	Q_1[a]	Median	Q_3[a]	95th Percentile
50-yr-old woman in Jackson/Calhoun	1.3	0.8	1.4	2.1	3.1
50-yr-old man in Jackson/Calhoun	1.1	0.6	1.1	1.8	3.3
50-yr-old woman in NHANES	1.3	0.8	1.4	2.5	4.0
50-yr-old man in NHANES	1.1	0.6	1.1	2.2	4.2

[a]Q_1: 25th percentile, Q_3: 75th percentile.
ppt indicates parts per trillion.

sion models on age and sex were fitted based on serum TCDD measures that were above the LOD for both the NHANES and the Jackson/Calhoun data sets, whereas the 25th percentile, mean, and median among young adults were estimated primarily based on the imputed values in the NHANES and the observed values above the LOD in Jackson/Calhoun study.

DISCUSSION

This study shows that the serum TCDD concentrations in non-Hispanic whites increased with age, and the rates of increase in the mean, 25th percentile, median, and 95th percentile over age were greater among women than men. This difference is probably the results of a longer TCDD half-life among women than men because of higher percentage of body fat in women and the peak level of TCDD in the environment in the 1960s and 1970s.[18] As a result, the overall mean and percentiles of the background concentrations depend on the distribution of age and sex in the reference population, and it is not valid to compare the overall mean or percentiles of serum TCDD concentrations between populations that have different age and sex structures. Therefore, it

is important to quantify the background levels of serum TCDD concentration by age and sex. For example, in comparisons to residents in Midland/Saginaw, we compared the serum TCDD concentrations to the background concentrations of people of the same age and sex to see whether the serum TCDD concentrations were elevated.[19] Quantile regression generalizes a single quantile estimate of serum TCDD concentrations to continuous conditional quantile estimates given age and sex. These age- and sex-adjusted quantile estimates provide better quantification of quantiles than the traditional method of calculating the population quantiles without adjusting for age or of adjusting for a limited number of age groups or strata.

We expected to see similar results for the Jackson/Calhoun data and the NHANES data because they both represented general populations who were not exposed to any known, unusual sources of dioxins. The present study shows that the effects of age and sex on the serum TCDD concentrations were not significantly different between the Jackson/Calhoun and the NHANES populations and that the Jackson/Calhoun population was not significantly different from the NHANES population in the age- and sex-specific 25th percentile, mean, median, and 95th percentile, but was slightly lower than the NHANES population in the 75th percentile. This implies that the Jackson/Calhoun population is similar to the NHANES population in age- and sex-specific serum TCDD concentration, and thus is a valid reference population for serum TCDD concentration for other predominantly white populations in Michigan. The marginally higher levels of age- and sex-specific 75th percentile in the NHANES than in the Jackson/Calhoun study can be explained as slightly larger variation of serum TCDD concentrations in the US population than in the 2 counties in Michigan. This could be due to more heterogeneity of TCDD exposures among regional US populations. However, the geographic information is not available in the publicly released NHANES data set, so that the geographic variation in serum TCDD concentration cannot be accounted for in the models. With data source indicator in the model, we allow for the effect of the different data source to be incorporated into the model.

Values below the LOD are common in studies of dioxin-like compounds. Simple ways of handling values below the LOD include imputing them with 0, LOD, LOD/2, and LOD/$\sqrt{2}$.[8] However, these imputation methods do not account for imputation uncertainty, and they depend on the blood sample volume and the LOD levels of the measurement methods. (In other words, the serum sample analyzed by 2 different methods having different LODs would be assigned different values.) For studies with a low proportion of data below the LOD and low LOD levels, the estimation of conditional percentiles is less affected by how the LODs are imputed, especially for upper percentiles. However, for environmental contaminants for which the concentrations are near the LOD, a substantial proportion of the analytic results will be below the LOD. Lower percentiles and sometimes even median estimates in such data are more sensitive to the imputation methods used.

In the present study, multiple imputations based on a left-censored regression model using the observed TCDD measures and the LOD levels of the nondetects were employed to impute the values below the LOD to obtain multiple complete data sets, so that complete-data statistical methods (such as quantile regression) could be implemented. In the multiple imputations, we assumed that the serum TCDD concentrations followed a lognormal distribution because the lognormal assumption appeared reasonable for the Jackson/Calhoun data, with 79% of the data that were observed (above LOD). By concatenating the Jackson/Calhoun data with the NHANES data, we improved the imputation for the values below the LOD in the NHANES data by incorporating the observed serum TCDD measures in the Jackson/Calhoun data. At the same time, inclusion of the NHANES data enhanced the estimates of the upper percentiles of serum TCDD values among older people in the Jackson/Calhoun population. The multiple imputation with the combined data set has improved the percentile estimation in both data sources. This method can be applied in other environmental and public health studies where the LOD is an issue and multiple sources of data are available. This article also provides an important example on how to incorporate the complex survey design information in every detail of statistical analysis in a population-based study.

The potential limitation of the multiple imputation approach is the assumption of lognormality. Although the lognormal assumption can be replaced by other statistical distributions, such as Gamma distribution or Weibull distribution according to some prior information, some distribution-free methods for handling values below the LOD, such as Schisterman's method, are of great interest.[20] In using the bootstrap method for stratified multistage samples, we have modified the sample weights with the bootstrap weights. However, the further weight modifications such as nonresponse and poststratification adjustments are not feasible here because of the limited information from the subsample of the non-Hispanic white population in the NHANES data. We combined the UMDES and the NHANES data by concatenating the 2 data sets directly and normalizing their sample weights. In future work, other methods for combining multiple data sources, such as Bayesian hierarchical methods, will also be considered.[21] Finally, we imputed the small fraction of missing covariates before the multiple imputation for values below the LOD to simplify the imputation procedure. In the future, we plan to work on multiple imputation methods that simultaneously impute the missing covariates and values below the LOD.

ACKNOWLEDGMENTS

We thank Sharyn Vantine and Xiaohui Jiang for their assistance, and Linda Birnbaum, Ronald Hites, Paolo Boffetta, Marie Sweeney, and David Kleinbaum for their guidance as members of the UMDES Scientific Advisory Board. The authors also thank the reviewers and editor whose suggestions greatly improved the manuscript.

REFERENCES

1. Agency for Toxic Substances and Disease Registry. *Toxicological Profile for Chlorinated Dibenzo-p-Dioxins.* Atlanta, GA: Public Health Service, US Department of Health and Human Services; 1998.
2. Patterson DG Jr, Patterson D, Canady R, et al. Age specific dioxin TEQ reference range. *Organohal Comp*. 2004;66:2878–2883.
3. Wittsiepe J, Schrey P, Ewers U, Selenka F, Wilhelm M. Decrease of PCDD/F levels in human blood from Germany over the past ten years (1989–1998). *Chemosphere*. 2000;40:1103–1109.
4. Patterson DG Jr, Turner WE, Caudill SP, Needham LL. Total TEQ reference range (PCDDs, PCDFs, cPCBs, mono-PCBs) for the US population 2001–2002. *Chemosphere*. 2008;73:261–277.
5. Koenker R. *Quantile Regression. Econometric Society Monograph Series*. Cambridge: Cambridge University Press; 2005.
6. National Center for Environmental Health. *Third National Report on Human Exposure to Environmental Chemicals.* Atlanta, GA: Department of Health and Human Services, Centers for Disease Control and Prevention; 2005. NCEH Pub No. 05–0570, 1–475.
7. Keith LH, Crummett W, Deegan J, Libby RA, Taylor JK, Wentler G. Principles of environmental analysis. *Anal Chem*. 1983;55:2210–2218.
8. Hornung RW, Reed LD. Estimation of average concentration in the presence of nondetectable values. *Appl Occup Environ Hyg*. 1990;5:46–51.
9. Lepkowski J, Olson K, Ward B, et al. Survey methodology in an environmental exposure study: methods, missing data, and inference. *Organohal Comp*. 2006;68:209–212.
10. University of Michigan. University of Michigan Dioxin Exposure Study. 2008. Available at: http://www.umdioxin.org. Accessed November 4, 2009.
11. Philips DL, Pirkle JL, Burse VW, Bernert JT Jr, Henderson LO, Needham LL. Chlorinated hydrocarbon levels in human serum: effects of fasting and feeding. *Arch Environ Contam Toxicol*. 1989;18:495–500.
12. Garabrant DH, Franzblau A, Lepkowski J, et al. The University of Michigan Dioxin Exposure Study: predictors of human serum dioxin concentrations in Midland and Saginaw, Michigan. *Environ Health Perspect*. 2009;117:818–824.
13. Raghunathan TE, Lepkowski JM, Van Hoewyk J, Solenberger P. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodol*. 2001;27:85–95.
14. Rubin DB. *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley; 1987.
15. Little RJ, Rubin DB. *Statistical Analyses With Missing Data.* New York: John Wiley; 2002.
16. Efron B, Tibshirani R. *An Introduction to the Bootstrap.* London: Chapman & Hall; 1994.
17. Rust KF, Rao JN. Variance estimation for complex surveys using replication techniques. *Stat Meth Med Res*. 1996;5:283–310.
18. Milbrath MO, Wenger Y, Chang CW, et al. Apparent half-lives of dioxins, furans, and PCBs as a function of age, body fat, smoking status, and breastfeeding. *Environ Health Perspect*. 2009;117:417–425.
19. Garabrant D, Chen Q, Hong B, et al. Logistic regression models for high serum 2,3,7,8-TCDD concentrations in residents of Midland, Michigan, USA. *Organohal Comp*. 2007;69:2203–2206.
20. Schisterman EF, Vexler A, Whitcomb BW, Liu A. The limitations due to exposure limits for regression models. *Am J Epidemiol*. 2006;163:374–383.
21. Raghunathan TE, Xie D, Schenker N, et al. Combining information from two surveys to estimate county-level prevalence rates of cancer risk factors and screening. *J Am Stat Assoc*. 2007;102:474–486.