Environment International 105 (2017) 112–125

Contents lists available at ScienceDirect

Environment International

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

Estimating serum concentrations of dioxin-like compounds in the U.S. population effective 2005–2006 and 2007–2008: A multiple imputation and trending approach incorporating NHANES pooled sample data

Anne Bichteler[a], Daniele S. Wikoff[b,*], Francis Loko[a], Mark A. Harris[c]

[a] ToxStrategies, Inc., Austin, TX, United States
[b] ToxStrategies, Inc., Asheville, NC, United States
[c] ToxStrategies, Inc., Houston, TX, United States

ARTICLE INFO

Keywords:
Dioxin
Biomonitoring
Pooled samples
NHANES
Referent
TEQ
Multiple imputation
Uncertainty

ABSTRACT

Dioxin-like compounds (DLCs) are monitored in the U.S. population using data collected with the National Health and Nutrition Examination Survey (NHANES). Until recently, participants' serum samples have been analyzed individually, and summary statistics defining reference ranges by age, gender, and race/ethnicity have served as the background by which other biomonitoring data can be evaluated. In the most recent NHANES DLC data, 2005–2006 and 2007–2008, participants' sera have been physically pooled prior to laboratory analysis, introducing major challenges to their utility as a reference population: variability among individuals and relations with covariates are lost, and individual design effects cannot be applied. Further, the substantial drop in limits of detection (LODs) in pooled sample biennials prevents reliable comparisons to individual data, and has complicated estimates of change over time. In this study, we address the drawbacks introduced by pooled samples by generating U.S. population reference ranges based on individual-level data adjusted to 2005–2006 and 2007–2008 levels. Using publicly available data, multiple imputation (MI) generated four NHANES biennials (2001–2008) of individual DLC data; we then trended the change over time in each DLC by demographic stratum. NHANES 2003–2004 individuals were adjusted by the trended change over time. Population estimates of toxic equivalency (TEQ) concentrations were calculated using traditional MI survey analysis methods and reference tables provided for 2005–2006 and 2007–2008 by age, race, and gender. Demographic differences in TEQ concentrations and trended change are reported, e.g. TEQ continues to drop in young adults aged 20–39, but distributions appear stable in older adults 60+; Mexican Americans have consistently lowest dioxins, furans, and PCBs, with non-Hispanic Blacks dropping to the same levels as non-Hispanic Whites in dioxins and PCBs and significantly below non-Hispanic Whites in furans by 2007–2008. Additionally, the ratio of 95th percentile to mean in DLC distributions was found to vary by age, between dioxins, furans, and PCBs, and across mean, making a simple ratio approach impractical for describing population concentrations using pooled samples. We discuss the practical implications of the pooled sample method, the performance of this trending solution in the context of other methods, and expected effects of distribution assumptions on variability and TEQ estimates, particularly in largely undetected congeners. These updated reference populations of individuals, along with information on trending, provide a common and valid basis for interpreting other individually sampled biomonitoring data.

1. Introduction

Polychlorinated dibenzodioxins (PCDDs), polychlorinated dibenzofurans (PCDFs), and coplanar PCBs, collectively referred to as dioxin-like compounds (DLCs), are a subset of persistent organic pollutants (van den Berg et al., 2006) that continue to be studied in the environment and monitored in populations around the world. Established population reference ranges of DLC concentrations based toxic equivalence (TEQ) by age, gender, and race/ethnicity serve as the background by which sample biomonitoring data can be evaluated (e.g., Ferriby et al., 2006; LaKind et al., 2009; Pavuk et al., 2014; Wong et al., 2008). In the United States, data collected as part of the biannual National Health and Nutrition Examination Survey (NHANES), a large-scale biomonitoring effort conducted by National Centers for Disease

* Corresponding author at: ToxStrategies, Inc., 825C Merrimon Ave, #103, Asheville, NC, 28804, United States.
E-mail address: dwikoff@toxstrategies.com (D.S. Wikoff).

http://dx.doi.org/10.1016/j.envint.2017.05.003
Received 21 October 2016; Received in revised form 3 May 2017; Accepted 3 May 2017

Control and Prevention (CDC), are the basis for U.S. population reference ranges, commonly referred to as background levels, of measured serum analytes.

In NHANES's most recent release of DLC data, for sera collected in 2005–2006 and 2007–2008, CDC and its National Center for Health Statistics (NCHS) collaborated on a new serum analysis method that pooled the sera of eight individuals within the same demographic stratum—age group, gender, and race/ethnicity—into one sample, prior to measurement (Curtin et al., 2012). Until that time, sera collected from participants had been analyzed and reported on an individual basis. Their recent change from an individual to a pooled sample method for analyzing DLCs has precluded the use of traditional analytic methods for generating reference ranges and estimating change over time.

Some disadvantages accompanied the marked improvement in detection rates and reduction in analytical costs introduced by the pooled sample method (see inset, Summary of analytic challenges). As sera were pooled across cells of the original sampling design, NHANES's complex system of over-sampling in certain sampling units and financial demographics ("design effects"), for example, could no longer be accounted for (Caudill, 2012). As with all survey data, accounting for the stratified survey weighting in NHANES is necessary for generalizing from a selected sample of participants to the general population. In addition, upper percentiles and confidence limits could no longer be assessed directly, as a single measured value of eight pooled sera did not provide information on the variability among individuals.

as well as the limitations associated with survey weighting and variability, the pooled data cannot reliably be used to evaluate individually-collected sample data due to the significant differences in analytical technology.

Though the impact of detection limits has not been explored, recent research has proposed several solutions to the problem of upper percentile and confidence limit estimation on pooled samples (Caudill, 2012; Aylward et al., 2014; Caudill, 2015). One solution developed ratios of the 95th percentile to pooled sample concentration based on individual-level data (Aylward et al., 2014). These ratios could be used for estimating the 95th percentile when only pooled sample concentrations are available. That work was based on congeners measured almost entirely above the LOD, yet the ratio range was ~1.7–~3.0, resulting in a wide range of 95th percentile. DLCs were not included in that study; here we investigate whether those ratios are also found in PCBs, dioxins, and furans, and whether they are consistent across age groups.

NCHS have published two approaches for estimating upper bounds and percentiles: Caudill (2012) modeled variability in retrospectively pooled 2003–2004 individuals and applied it along with 2003–2004 design effects to 2005–2006 pooled samples; Caudill (2015) used the 2003–2004 modeled variability in multiple imputations to the 2005–2006 individuals whose sera had been pooled. Neither of these approaches can be implemented by researchers outside the CDC system, because certain required terms are not made public, e.g., the membership in each pool and the sample weight adjusted for selection into the

Individual Sample Method *Utilized through 2004*	Pooled Sample Method *Utilized 2005–2006 and 2007–2008*
• Each individual serum analyzed; empirical variability among individuals retained	• Combines 8 sera, unknown contributors & volume, unknown variability among individuals
• Direct estimates of upper bounds & percentiles on sample sizes 1100+	• Indirect, model-driven estimates of upper bounds & percentiles
• Survey design/weighting supports generalization to U.S. population	• Survey design/weighting cannot be incorporated
• Supports direct comparisons to other individually biomonitored samples	• Does not support comparisons on upper bounds or percentiles to individually biomonitored samples
• Supports research on relations with participants' exam, diet, location, and other analytes	• Investigating relations among DLCs and participants' other data is no longer supported

Summary of analytic challenges. Challenges to estimating population reference ranges introduced by the pooled versus individual serum sample methodologies.

Image 1

An additional ramification of pooling sera, as yet unexplored in the literature, is how increased sample volume and the resulting drop in limits of detection (LODs) affect estimates of population change over time. Research to date validating pooled sample methods has been exclusively on compounds measured almost entirely above the LOD in both individual and pooled samples. In such compounds, e.g., PCB 153 and PCB 118, a small but steady decline in population means has been observed from the individually analyzed NHANES biennials 1999–2004 to the pooled 2005–2006 biennial, as half-life decay and falling environmental exposures would lead one to expect (Sjödin et al., 2014; Caudill, 2015). However, as part of preliminary scoping efforts associated with this study, it was discovered that some (but not all) compounds with a significant fraction undetected in the individually analyzed sera, the much lower LODs in the pooled samples appear to have caused a precipitous (and biologically implausible) drop in estimates of mean concentration – a phenomenon further explored herein. Thus, given the challenges presented by the large differences in LODs, and considering the importance of proxy estimation (also referred to as substitution) of concentrations for DLC congeners < LOD, pools. The impact of using 2003–2004 variability and design effects to estimate 2005–2006 population statistics is unclear, but it would be preferable to use the variability and design effects estimated directly on the sampling units measured in each NHANES biennial (Caudill, 2015).

The primary objective of the work presented here was to address the drawbacks introduced by pooled samples by generating U.S. population reference ranges for DLCs (based on toxic equivalency, TEQ), including upper percentiles and confidence limits, based on individual-level data. These reference ranges, effective 2005–2006 and 2007–2008, incorporate measured inter-individual variability, account for survey design/weighting, and manage differences in LODs inherent to the pooled sampling methodology. As such, they are appropriate estimates of U.S. background TEQ by age, race, and gender for evaluation and interpretation of other individually-sampled biomonitoring data. Secondary objectives included evaluating the impact of the dramatic decreases in LODs, assessing the validity of the ratio approach (Aylward et al., 2014) for estimating upper percentiles from pooled biomonitoring data for DLCs, and investigating differences in DLCs across demographic strata (age, race/ethnicity, and gender).