Published in final edited form as:

J Agric Biol Environ Stat. 2015 March ; 20(1): 65–82. doi:10.1007/s13253-015-0196-3.

Assessing Assay Variability of Pesticide Metabolites in the Presence of Heavy Left-Censoring

Haiying Chen,

Division of Public Health Sciences, Department of Biostatistical Sciences, Wake Forest School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157, USA

Sara A. Quandt,

Division of Public Health Sciences, Department of Epidemiology and Prevention, Wake Forest School of Medicine, Winston-Salem, NC, USA

Dana Boyd Barr, and

Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA, USA

Thomas A. Arcury

Department of Family and Community Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA

Haiying Chen: hchen@wakehealth.edu

Abstract

Assessing assay variability for field samples in environmental research is challenging, since a quantitative assay is typically constrained by a lower limit of detection. The purpose of this paper is to compare three parametric models for assessing assay variability using duplicate data subject to heavy left-censoring. Efron information criterion (EIC) and Bayesian information criterion (BIC) are used to aid in model selections. Distributional parameter estimates are obtained using maximum likelihood estimation for bivariate lognormal, bivariate zero-inflated lognormal, and bivariate 3-component mixture models. We illustrate a practical application using duplicate pesticide data from the Community Participatory Approach to Measuring Farmworker Pesticide Exposure (PACE3) study. Furthermore, a simulation study is conducted to empirically evaluate the performance of the three models. The results from PACE3 indicate that the bivariate zero-inflated lognormal model is fairly competitive based on EIC or BIC. Further, total variability for the lognormal component can be decomposed into between-subject and within-subject variance based on this model. Assay variability estimates such as within-subject coefficient variation, minimum detectable change, and probability of k -fold difference can be easily derived under the bivariate zero-inflated lognormal model. Additionally, the assay variability is rather large for the PACE3 data. Therefore, apparent longitudinal change in pesticide exposure should be examined cautiously in the context of substantial assay variability.

Keywords

Coefficient of variation; Limit of detection; Minimum detectable change; Mixture distribution; Probability of k -fold difference

1. INTRODUCTION

In quantitative bioassays, knowledge of assay variability can assist researchers to infer how likely an observed difference represents a real change versus a random variation in assay results. In the context of replicate data, within-subject coefficient of variation (WCV) is a commonly accepted parameter to reflect assay variability (Chinn 1990; Chow and Tse 1990; Quan and Shih 1996; Rosner 2006). Alternative parameters that can help evaluate assay variability include repeatability (Bland and Altman 1986, 1996), also known as smallest detectable change or minimum detectable change (MDC) in the biomedical fields (de Vet et al. 2006; Euser et al. 2008; Donoghue 2009), and probability of k -fold difference (Reed et al. 2002). The estimation of these parameters for environmental data is often challenging, since a quantitative assay is typically constrained by a lower limit of detection (LOD). This is referred to as left-censoring in the literature. For instance, the duplicate pesticide metabolites data collected in our Community Participatory Approach to Measuring Farmworker Pesticide Exposure (PACE3) study have a large number of pairs with both values $\leq$ LOD (see more details in Sect. 2). Commonly adopted method of simple substitution of a single constant (e.g., $\text{LOD}/2$, $\text{LOD}/\sqrt{2}$, LOD) for values $\leq$ LOD is apparently unsatisfactory in this situation. The individual WCVs for these pairs will be zero, which in turn will lead to biased estimation for the overall WCV. Therefore, maximum likelihood estimation (MLE) which has been increasingly used to account for the left-censoring present in concentration data (Jacqmin-Gadda et al. 2000; Lyles et al. 2001a,b; Thiébaud and Jacqmin-Gadda 2004; Barnhart et al. 2005) is a compelling alternative for estimating assay variability.

The performance of MLE relies heavily on the correct assumption of underlying parametric distributions. The most commonly used parametric distribution for fitting environmental concentration data is lognormal. For values below the LOD, one reasonable assumption is that they are the left-censored values from the same lognormal distribution as the values above the LOD. For example, acephate (APE) is a pesticide widely used to treat tobacco. Farmworkers involved in activities such as topping, harvesting, or curing pesticide-treated tobacco are almost definitely exposed to APE (Arcury et al. 2010). In such situations, assuming the left-censored data coming exclusively from a lognormal distribution is appropriate. However, sometimes this assumption may not be realistic. For instance, in pesticide research, a farmworker may not be exposed to a certain pesticide, since he/she does not work on the crops on which the particular pesticide is applied; or even if he/she is exposed, it is possible that the pesticide has been metabolized and cleared from the body and fails to be detected in biological samples. In these situations, the values below the LOD can be either zeroes or left-censored values from the lognormal component. The zero component, together with the lognormal component, leads to a zero-inflated lognormal distribution (Moulton and Halsey 1995; Taylor et al. 2001; Berk and Lachenbruch 2002; Chu et al. 2005, 2008).

Limited research has been conducted to examine the appropriate parametric forms for use of MLE in analyzing heavily left-censored replicate assay data. The purpose of this paper is to compare three parametric models for assessing assay variability using duplicate data subject to heavy left-censoring. The first is a bivariate lognormal model where all left-censored values are considered to be strictly positive and come from the lower end of the lognormal. The second is a bivariate zero-inflated lognormal model where the left-censored values can be either zeroes or come from the lognormal component. Since we observed a large number of pairs with one value $> \text{LOD}$ and one $\leq \text{LOD}$ in the PACE3 study (more details in Sect. 2), we also investigate a third model that consists of three components. We refer to it as a bivariate 3-component mixture model hereafter. It not only contains a zero component but also assumes that there are two lognormal components with different variances for the positive values. Efron information criterion (EIC) and Bayesian information criterion (BIC) are used to aid in model selections. In the rest of the paper, we first describe the duplicate pesticide data from the PACE3 study that motivate the consideration of the three models. Next we present the likelihood function for the three different models. Then we illustrate the application of the three models to the duplicate pesticide data. Finally, a simulation study is conducted to evaluate the performance of the three models.

2. THE MOTIVATING EXAMPLE

This research is motivated by the pesticide metabolites data collected in PACE3. In this study, a total of 287 farmworkers from 11 counties in eastern North Carolina were recruited in 2007 (Arcury et al. 2009). Urinary concentrations for the six dialkylphosphate (DAP) urinary metabolites of organophosphorus (OP) pesticides (dimethylphosphate (DMP), dimethylthiophosphate (DMTP), dimethyldithiophosphate (DMDTP), diethylphosphate (DEP), diethylthiophosphate (DETP), and diethyldithiophosphate (DEDTP)) were measured longitudinally across four periods in the agricultural season (period 1, May 1 to June 8; period 2, June 9 to July 7; period 3, July 8 to August 5; and period 4, August 6 to September 4). For the purpose of assay quality control (QC), approximately 20% of the participants were selected to provide duplicate samples, that is, two subsamples taken from and representative of the same sample from a participant. Specifically, the data collectors followed a pre-specified protocol and randomly selected urine samples from 57 farmworkers over the course of the study. All samples were thoroughly mixed, split into two separate aliquots, and sent for quantifications with unique identifiers unknown to the lab. Of the 57 samples, 14 (24.6%), 16 (28.0%), 13 (22.8%), and 14 (24.6%) were from periods 1, 2, 3, and 4, respectively. The pesticide metabolites data were analyzed at the Pesticide Laboratory at National Center for Environmental Health, Centers for Disease Control and Prevention. The LODs for DMP, DMTP, DMDTP, DEP, DETP, and DEDTP are 0.6, 0.2, 0.1, 0.2, 0.1, and 0.1 $\mu\text{g/L}$, respectively.

Figure 1 summarizes the degree of left-censoring for the six different analytes quantified in the 57 pairs of duplicates. Research has shown that the estimation process is highly unreliable with above 80% censoring (Helsel 2005). Given that both measures of DEDTP have more than 80% values $\leq \text{LOD}$, it is not included in subsequent analyses. For the remaining five analytes, the proportion of values $\leq \text{LOD}$ ranges from about 20% for DMTP to about 65% for DETP for a given measure. Therefore, it is important to examine the

presence of potential zeroes to account for the large amount of left-censoring. We use Quantile–Quantile (Q–Q) plots to examine the univariate distribution for each measure of each analyte to help determine the appropriate parametric form for further analyses.

Let X follow a univariate zero-inflated lognormal distribution. That is, with probability γ , X takes on a value of zero; with probability $1 - \gamma$, X is distributed as a lognormal (θ, τ^2) . Then the expected p th ($p > \gamma$) quantile on the log scale is calculated as $\theta + \tau \times \Phi^{-1}((p - \gamma)/(1 - \gamma))$, where Φ^{-1} is the inverse cumulative distribution function (CDF) of a standard normal. If γ is equal to 0, this is reduced to a univariate lognormal distribution without a zero component and the expected quantiles can be calculated accordingly. Based on these calculations, we created Q–Q plots for values $> \text{LOD}$ on log scale for the each measure of each analyte under both lognormal and zero-inflated lognormal distributions using estimated γ , θ , and τ^2 from the sample. It is apparent that there is a lack of fit for the lognormal distributions, which assume all values $> \text{LOD}$ come from a lognormal component. The expected quantiles were underestimated for the lower end of the distribution and overestimated for the higher end of the distribution. The zero-inflated lognormal improved the fit of the model substantially for DMP (Fig. 2), DMDTP, DEP, and DETP (details not shown), supporting the appropriateness of including a zero component for all these analytes. For DMTP that has the least amount of left-censoring among the five analytes, the inclusion of a zero component also appears beneficial (Fig. 2).

Next we examine the bivariate distribution of the duplicate measures for each analyte. Figure 1 displays the scatter plot of two measurements for each analyte on the log scale. In this plot, we substituted all values $< \text{LOD}$ with LOD to graphically depict the spread of the data. The single dot on the lower leftmost corner represents all pairs of observations with both values $< \text{LOD}$. The dots that fall on either the horizontal or vertical reference lines represent the pairs of observations with one value $> \text{LOD}$ and one $< \text{LOD}$. For example, one pair has the measure 1 of DMP $< \text{LOD}$, i.e., $0.6\mu\text{g/L}$ (-0.5 on the log scale), while the measure 2 of DMP is as high as $13.8\mu\text{g/L}$ (2.6 on the log scale). The rest of the dots that scatter around the unity reference line represent the pairs of observations with both values $> \text{LOD}$. Overall, there appears to be a considerable amount of variability in the duplicate data for all analytes. Since the pairs were indeed collected as duplicate samples, two models are natural candidates for assessing assay variability. The first is a bivariate lognormal model where all left-censored values are considered to be strictly positive and come from the lower end of the lognormal. The second is a bivariate zero-inflated lognormal model where the left-censored values can be either zeroes or come from the lognormal component. In both models, we assume that the bivariate lognormal distribution has a common mean and variance with a non-negative correlation coefficient, similar to the setup of a linear mixed effects model when censoring is not of a concern.

We note that the large number of pairs with one value $> \text{LOD}$ and one $< \text{LOD}$ is especially striking, ranging approximately from 20% for DMDTP to 35% for DMP. We attempt to use a third model, a bivariate 3-component mixture model, to better accommodate this portion of the data. Specifically, we use a zero component to account for the heavy left-censoring. Then we use two different lognormal components to separately model the pairs with one value $> \text{LOD}$ and one $< \text{LOD}$ and the pairs with both values $> \text{LOD}$. Since correlations

between pairs with one value $> \text{LOD}$ and one $\leq \text{LOD}$ cannot be estimated directly because of the left-censored values in one of the pairs, we allow the two different lognormal components to have the same correlation but different variances. Given the data pattern observed in Fig. 1, we anticipate that one lognormal component will have low variability, while the other will have high variability. Additionally, the range of values $> \text{LOD}$ among the pairs with one value $> \text{LOD}$ and one $\leq \text{LOD}$ is similar to that of the pairs with both values $> \text{LOD}$. Therefore, we further assume that the two different lognormal components have the same location parameter.

3. METHOD

In this section, we describe the likelihood functions for the three different bivariate models. We start with the most complex model, the 3-component mixture model. Let Y_{i1} and Y_{i2} represent two individual assay measurements made on the duplicate samples from subject i . We assume that (Y_{i1}, Y_{i2}) follows a bivariate mixture distribution with three components: a zero component with probability ω_0 , a low variability lognormal component with probability ω_1 , and a high variability lognormal component with probability ω_2 , subject to the constraint that $\omega_0 + \omega_1 + \omega_2 = 1$. Additionally, we assume that even though the two lognormal components have different variances, they share a common location parameter μ and a common correlation coefficient of ρ . If both values $> \text{LOD}$, then we assume that the pair comes from the low variability lognormal component. Since Y_{i1} and Y_{i2} are duplicates, it is reasonable to assume that if one is $> \text{LOD}$ and one is $\leq \text{LOD}$, then the pair of observations comes from the high variability lognormal component. Similarly, it is sensible to assume that if both Y_{i1} and Y_{i2} are $\leq \text{LOD}$, then the pair will both take on the value 0 or both be left-censored from the two lognormal components. Specifically, we define $\omega_0 = P(Y_{i1} = 0 \cap Y_{i2} = 0)$. That is, ω_0 is the probability that the sample on the i th subject from which the two duplicate measures are taken is indeed a 0. Let (Z_{i1}, Z_{i2}) and L denote the log-transformed concentration values and $\log(\text{LOD})$ of the assay, respectively. The likelihood function can be derived as follows.

When both (Z_{i1}, Z_{i2}) are $> L$, their contribution to the likelihood function (l_{1i}) is simply the joint density function of a bivariate normal distribution conditioning on the pair coming from the low variability lognormal component. That is,

$$l_{1i} = \omega_1 \times \left(2\pi\sigma_1^2 \sqrt{1-\rho^2} \right)^{-1} \times \exp \left(-\frac{1}{2(1-\rho^2)} \left(\left(\frac{Z_{i1}-\mu}{\sigma_1} \right)^2 - 2\rho \left(\frac{Z_{i1}-\mu}{\sigma_1} \right) \left(\frac{Z_{i2}-\mu}{\sigma_1} \right) + \left(\frac{Z_{i2}-\mu}{\sigma_1} \right)^2 \right) \right).$$

When Z_{i1} is $> L$ and Z_{i2} is $\leq L$, we assume that the pair comes from the high variability lognormal component. Their contribution to the likelihood function (l_{2i}) can then be expressed as the product of the marginal distribution of Z_{i1} and the conditional probability of $Z_{i2} \leq L$ given the observed Z_{i1} . That is,

$$l_{2i} = \omega_2 \times (2\pi\sigma_2^2)^{-1/2} \exp\left(-\frac{1}{2}\left(\frac{Z_{i1}-\mu}{\sigma_2}\right)^2\right) \times \Phi\left(\frac{L-(\mu+\rho(Z_{i1}-\mu))}{\sigma_2\sqrt{(1-\rho^2)}}\right),$$

where Φ is the CDF for a standard normal distribution. When Z_{i2} is $> L$ and Z_{i1} is $\leq L$, their contribution to the likelihood function (l_{3i}) can be defined similarly as l_{2i} . That is,

$$l_{3i} = \omega_2 \times (2\pi\sigma_2^2)^{-1/2} \exp\left(-\frac{1}{2}\left(\frac{Z_{i2}-\mu}{\sigma_2}\right)^2\right) \times \Phi\left(\frac{L-(\mu+\rho(Z_{i2}-\mu))}{\sigma_2\sqrt{(1-\rho^2)}}\right).$$

When both (Z_{i1}, Z_{i2}) are $\leq L$, the pair can either come from the zero component or the two lognormal components. Therefore, their contribution to the likelihood function (l_4) is

$$l_4 = \omega_0 + \omega_1 \times P(Z_{i1} \leq L \cap Z_{i2} \leq L | \mu, \sigma_1^2, \rho) + \omega_2 \times P(Z_{i1} \leq L \cap Z_{i2} \leq L | \mu, \sigma_2^2, \rho).$$

The final log-likelihood function is then

$$\begin{aligned} & \sum_{i=1}^n I(Z_{i1} > L \cap Z_{i2} > L) \log(l_{1i}) \\ & + I((Z_{i1} > L \cap Z_{i2} \leq L) \log(l_{2i}) \\ & + I(Z_{i1} \leq L \cap Z_{i2} > L) \log(l_{3i}) \\ & + I(Z_{i1} \leq L \cap Z_{i2} \leq L) \log(l_4), \end{aligned}$$

where n is the total number of sample pairs and $I(\cdot)$ is an indicator function taking on value 1 if the argument is true and 0 otherwise. The likelihood function indicates that the estimated proportion of zeroes depends on both the observed proportion of data $\leq LOD$ and the distribution of the lognormal components. Intuitively, if the lognormal components have low means and/or large variances, then it is more likely to contribute more values $\leq LOD$, which will lead to a smaller estimated proportion of zeroes. Under this 3-component mixture model, the overall mean (μ) and variance (σ^2) for the two lognormal components are μ and $\omega_1/(\omega_1+\omega_2) \times \sigma_1^2 + \omega_2/(\omega_1+\omega_2) \times \sigma_2^2$, respectively, on the log scale.

We note that if the pairs with one value $> LOD$ and one $\leq LOD$ are assumed to come from the same distribution as the pairs with both values $> LOD$, then we have a bivariate zero-inflated lognormal. Namely, with probability $1 - \omega_0$, (Z_{i1}, Z_{i2}) follows a bivariate normal with a common mean μ , a common variance σ^2 , and a correlation coefficient of ρ . Further, if ω_0 equals zero, then all pairs of (Z_{i1}, Z_{i2}) come from a bivariate normal with parameters (μ, σ^2, ρ) . The likelihood functions described above can then be modified accordingly for these two models (Chu et al. 2008). These log-likelihood functions can be maximized using various optimization routines available in many commercial software packages. In this

paper, we used a nonlinear optimization routine by Newton-Raphson ridge method in SAS IML(SAS Institute, Cary, NC). The SAS macro is available from the authors upon request.

To help determine which parametric form provides the best fit for the data, we evaluate both EIC and BIC. EIC is a bootstrap-based information criterion developed for small samples (Chu et al. 2005). It is calculated as

$$-2\log l(\mathbf{Z}; \hat{\beta}) + 2/B \sum_{b=1}^B \left(\log l(\mathbf{Z}_b^*; \hat{\beta}_b^*) - \log l(\mathbf{Z}; \hat{\beta}_b^*) \right),$$

where $l(\cdot)$ represents the likelihood function, $\hat{\beta}$ is the MLE estimated from the original sample, B is the number of bootstrap samples, and $\mathbf{Z}_b^*$ and $\hat{\beta}_b^*$ are the b th bootstrap sample and the MLE estimated from the b th bootstrap sample, respectively. BIC is calculated as $-2 \log$ -likelihood plus $\log(n)$ times the number of parameters in the model. A smaller EIC or BIC is indicative of a model with a better fit.

Under the bivariate lognormal or the zero-inflated lognormal models, we can estimate assay variability for the lognormal component via the following parameters. The within-subject (σ_e^2) and between-subject (σ_b^2) variances on the log scale for the lognormal component are calculated as $\sigma_e^2 = \sigma^2(1-\rho)$ and $\sigma_b^2 = \sigma^2(\rho)$, respectively. The WCV for the bivariate

lognormal data is then $\exp(\sigma_b^2/2) \sqrt{\exp(\sigma_e^2) - 1} \times 100\%$ (Quan and Shih 1996). The MDC is associated with a confidence level, usually 95% (MDC95). For lognormal distribution, the

MDC95 on the natural scale is expressed as $\exp\left(1.96 \sqrt{2\sigma_e^2}\right)$, representing the ratio between Y_{i1} and Y_{i2} . It means that the change (the ratio) between two duplicate measurements for the same subject is expected to be less than MDC95 for 95% pairs of observations. A probability of k -fold difference is defined as the probability that two duplicate measurements differ by a factor of $k(k > 1)$ or more. Namely,

$p(k) = P(Y_{i2}/Y_{i1} \geq k \text{ or } Y_{i1}/Y_{i2} \geq k) = 2\Phi\left(-\log(k)/\sqrt{2\sigma_e^2}\right)$. Under the 3-component mixture model, the WCV will have to be computed separately for each lognormal component. However, the overall MDC95 and $p(k)$ for the two combined lognormal components can be derived based on the distribution of $Z_{i1} - Z_{i2}$. Specifically, $Z_{i1} - Z_{i2}$ follows a mixture normal distribution. With probability $\omega_1/(\omega_1 + \omega_2)$, it comes from a normal distribution with mean 0 and variance $2\sigma_1^2(1-\rho)$. With probability $\omega_2/(\omega_1 + \omega_2)$, it comes from a normal distribution with mean 0 and variance $2\sigma_2^2(1-\rho)$. Since the distribution is symmetric, then the overall MDC95 on the natural scale can be calculated as $\exp(\mathbf{F}^{-1}(97.5))$, where $\mathbf{F}^{-1}$ is the inverse CDF of $Z_{i1} - Z_{i2}$. Similarly, the overall $p(k)$ is

$$2\omega_1/(\omega_1 + \omega_2)\Phi\left(-\log(k)/\sqrt{2\sigma_1^2(1-\rho)}\right) + 2\omega_2/(\omega_1 + \omega_2)\Phi\left(-\log(k)/\sqrt{2\sigma_2^2(1-\rho)}\right).$$

4. RESULTS FROM THE PACE3 STUDY

Table 1 presents the parameter estimates obtained using MLE based on the three different models described in Sect. 3. We also obtained standard deviations for the proportions of zeroes under the bivariate zero-inflated lognormal using 5,000 bootstrap samples. Under the bivariate zero-inflated lognormal models, the estimated proportions of zeroes are about 8% ($\pm 4\%$) for DMTP, close to a quarter ($\pm 8\%$) for DMP, 35% ($\pm 7\%$) for DEP, and a little more than a half ($\pm 7\%$) for DMDTP and DETP, respectively. These estimates of ω_0 are similar to those obtained from the 3-component mixture models. The correlation between the two duplicate measurements from the lognormal component (including the pairs with only one value $> \text{LOD}$ and the pairs with both values $> \text{LOD}$) is weak under the zero-inflated lognormal models. The MLEs for DEP and DETP lie on the lower bound of the parameter space and are, therefore, zero. The observed largest correlation is about 0.4 for DMTP. The estimated correlations are stronger under the bivariate lognormal models, ranging from 0.53 for DMP to 0.80 for DMDTP. The 3-component mixture models yield correlations of similar magnitude to those under the bivariate lognormal models without a zero component. Finally, the estimated variances under the bivariate lognormal models are markedly larger than those under the zero-inflated lognormal models. For the 3-component mixture models, the variances for the pairs with both values $> \text{LOD}$ are small, while the variances for the pairs with one value $> \text{LOD}$ and one $\leq \text{LOD}$ are extremely large as expected.

Table 2 presents EIC calculated based on 5,000 bootstrap samples. For all five analytes, EIC favors the bivariate zero-inflated lognormal models over the bivariate lognormal models, indicating the presence of some zeroes in the left-censored data. In addition, the zero-inflated lognormal models yield substantially smaller EIC for DMP and DEP, slightly larger EIC for DMDTP and DETP, and substantially larger EIC for DMTP when compared to the 3-component mixture models. Results based on BIC are similar. Overall, the bivariate zero-inflated lognormal is a fairly competitive model for all analytes except for DMTP.

Next, we proceed to estimate assay variability under three different models. The estimated within-subject SDs are strikingly similar under the bivariate lognormal and zero-inflated lognormal models (Table 2). This indicates that the observed higher variability ($\hat{\sigma}^2$) under the bivariate lognormal models in Table 2 compared to the zero-inflated lognormal models is mostly due to the larger between-subject variances. The estimated WCVs are very large under both models. Specifically, the WCVs are greater than 100% for all pesticide metabolites, even greater than 600% under the bivariate lognormal model. Both MDC95 and $p(2)$ depend solely on the within-subject SD under the bivariate lognormal and zero-inflated lognormal models, and therefore, are similar under these two models. The MDC can be interpreted as the minimum amount of change in a person's pesticide concentrations that ensures the change is not likely the result of measurement error. For example, DMTP, which has the smallest estimated measurement error among the five analytes, has a MDC95 of 42 under both the bivariate lognormal and zero-inflated lognormal models, indicating that one is 95% confident that an at least 42-fold of increase or decrease in the concentrations can be considered a real change. Compared to the bivariate lognormal and zero-inflated lognormal models, the MDC95 under the bivariate 3-component mixture models is markedly larger, while $p(2)$ is slightly smaller. The probability that two duplicates differ by 2-fold or more is

at least about 50% for all analytes. It shows that a twofold difference between paired concentrations on the same person measured at different time points in a longitudinal study can be frequently expected and that such difference is likely due to assay variability instead of an actual change.

5. A SIMULATION STUDY

We conducted a simulation study to evaluate the sampling properties of the distributional parameter estimates under the three different models. The bivariate zero-inflated lognormal is used as the true underlying distribution for simulating bivariate data. Without loss of generality, we assume that for the lognormal component, $\mu = 0$, $\sigma^2 = 1$, and $\rho = 0.5$. The proportion of zeroes (ω_0) varies from 0.1 to 0.4. The LODs are selected so that the proportion of values $\geq$ LOD for the lognormal component is 0.2. The sample size is fixed at 50 pairs, and MLE is performed under the correct distributional assumption (bivariate zero-inflated lognormal) as well as the incorrect distributional assumptions (bivariate lognormal and 3-component mixture models). A total of 5,000 replicates are generated for each scenario to approximate the sampling distribution of the parameter estimates.

Figure 3 presents the means of MLEs from 5,000 simulated samples for ω_0 , μ , σ^2 , and ρ . The error bars represent the SDs of the 5,000 MLEs. The results indicate that $\hat{\omega}_0$ is similar under the bivariate zero-inflated and 3-component mixture models. As expected, as true ω_0 increases, the sample size for the zeroes increases, and the bias decreases. It is evident that the incorrect assumption of bivariate lognormal and 3-component mixture models lead to substantially biased estimates for μ , σ^2 , and ρ . Especially, the performance of the bivariate lognormal model becomes worse when ω_0 increases to greater than 20%. Figure 4 presents the assay variability estimates under the three different models. Interestingly, the MLEs for the within-subject SD, MDC95, and $p(k)$ have minimal bias under both the bivariate lognormal and zero-inflated lognormal models regardless of the presence of zeroes. However, the MLEs for WCV under the bivariate lognormal models and the MLEs for MDC95 under the bivariate 3-component mixture models have considerable bias and variability even with relatively small proportion of zeroes. The performance of MLEs for $p(k)$ is fairly robust under the three different models, with minimal bias even in the presence of large amount of zeroes.

Finally, as a sensitivity analysis, we repeat the above analyses using the bivariate lognormal ($\omega_0 = 0$) as the true underlying distribution. The proportion of values $\geq$ LOD varies from 0.1 to 0.5. The results indicate that the estimates for μ , σ^2 , and ρ under the incorrect assumption of bivariate zero-inflated lognormal models have minimal bias when the left-censoring is not severe (Supplement figure 1). When the proportion of values $\geq$ LOD increases to 40–50%, the bias for these parameters becomes larger, especially for the mean. However, the estimates for within-subject SD, WCV, MDC95, and $p(k)$ remained having minimal bias even in the presence of heavy left-censoring (Supplement figure 2). In comparison, the estimates under the incorrect assumption of bivariate 3-component mixture model in general have large bias except for $p(k)$. We note that the bias for σ^2 and MDC95 tends to become smaller with the increase of left-censoring, suggesting that this model may become competitive in practice when the true underlying distribution is unknown.

6. DISCUSSION

Assay variability can be easily assessed using linear mixed effects models, if all concentration values are fully observed. In this paper, we report for the first time estimated assay variability using duplicate field samples (samples that are collected in our own study with true concentrations unknown) subject to heavy left-censoring. For our duplicate pesticide metabolites data, there are two distinct features that pose analytical challenges when assessing assay variability. One feature is that for most metabolites, there are substantial amounts of values $\leq$ LOD that cannot be fully explained by a usual lognormal model. Our results indicate that the inclusion of a zero component in the model helps account for the heavy left-censoring of the duplicate data. As corroborated by the simulation study, models without a zero component when there are actual zeroes in the data result in substantial underestimation of the mean, overestimation of the correlation, the total variance, and the between-subject variance, which in turn lead to severe overestimation of the WCV.

Surprisingly, the estimation of within-subject variance is not influenced much by the omission of a zero component, suggesting that the pairs with both values $>$ LOD are the major contributors for this estimation. Therefore, there is minimal bias in MDC and probability of k -fold difference even when the underlying model is inappropriate. This has important practical implications, since the bivariate lognormal models are equivalent to mixed effects models and the correlation structure can be imposed with a random subject effect. Therefore, these models can be fit with many standard software packages. For example, SAS procedure PROC NLMIXED can be used to obtain an estimate for within-subject SD, and subsequently, MDC and $p(k)$. This significantly reduces the programming effort needed compared to fitting a bivariate zero-inflated lognormal model while retaining valid estimates for MDC and $p(k)$. However, we stress that it is important to examine the presence of a potential zero component so that an appropriate parametric form can be used for estimating the parameters such as variance and WCV. The incorporation of zeroes for values $\leq$ LOD can be justified on the basis of both biological grounds and formal statistical tests (Self and Liang 1987).

The other feature observed in our duplicate metabolites data is that there are, unexpectedly, a large number of pairs with one value $\leq$ LOD and one $>$ LOD. Furthermore, the magnitude of disagreement within these pairs is rather striking. The values $>$ LOD in these pairs do not necessarily cluster around the LOD as one would predict in the case of small assay variability. On the contrary, some of the values $>$ LOD are among the highest measured concentrations. These pairs are fairly influential for the parameter estimates. For DEP and DETP, the number and magnitude of disagreement drive the estimated correlation to be zero in the bivariate zero-inflated lognormal models. We attempt to use two separate lognormal components to model pairs with one value $\leq$ LOD and one $>$ LOD along with pairs with both values $>$ LOD. This increases the complexity of the model substantially without a substantial improvement in model fit for most metabolites based on EIC or BIC. Moreover, even though the 3-component mixture model may appear to be competitive in terms of EIC or BIC, it does not provide added benefit in terms of the assay variability indicators that we are mostly interested in. The overall WCV is not easily available under this model. MDC is highly sensitive to model misspecifications based on simulation studies. Estimated $p(k)$ is

similar to those under the more parsimonious models. Therefore, we do not recommend the use of the 3-component mixture model for assessing assay variability.

The presence of pairs with one value $\leq$ LOD and one $>$ LOD reminds the researchers of a fact that is often overlooked in analyzing concentration data. For each analytical sample, the unknown true concentration is either $>$ LOD or $\leq$ LOD. This implies that in these pairs, one of the values is a false positive if the true value is $\leq$ LOD or a false negative if the true value is $>$ LOD due to measurement errors. In fact, if the true concentration is within the close vicinity of LOD, then a measurement will have approximately a 50% chance of being reported as $>$ LOD and a 50% chance of reported being $\leq$ LOD (Helsel 2005). When modeling our duplicate data, we implicitly assume that if one value is $>$ LOD then the underlying true value is $>$ LOD and comes from some lognormal component. This is justified since the definition and derivation of LOD inherently guarantee that the false positive rate is very small, usually 1% (USEPA 1982). Thus, a measurement $>$ LOD will be declared to have a non-zero concentration at a 1% false positive rate. More detailed information on false positive and false negative rates from the laboratory may help with the modeling of the duplicate data, but it is beyond the scope of this paper.

We note that the assay variability estimated from unknown field samples should not be interpreted as the standard assessment of precision in routine assay QC procedures. For standard assessment of assay precision, positive controls with known concentrations are usually used in a well-controlled lab environment. While analyzing the PACE3 pesticide data, the lab prepared three QC pools by fortifying blank urine with the target analytes at nominal concentrations of 5, 20, and 50 $\mu\text{g/L}$. These pooled samples, along with a blank, were analyzed in triplicate over a course of 21 independent days after which the mean concentration, standard deviation, and 95th and 99th confidence limits were calculated. In each analytic run, two of each pool were included to ensure proper method application over the course of the analytic process. These QC pool values were plotted on Shewhart Plots to examine both the relative deviation from the mean and the actual deviation from the mean throughout the course of the study. The QC measurements were evaluated per Westgard Rules (Westgard et al. 1994a,b). The QC data indicate that the total relative SD is less than 15%, representing a well-controlled analytical process. As opposed to a few fixed concentrations in QC pools, the range of data from different subjects can spread from zero, to trace amount, to high level of concentrations for field samples. For PACE3, since the pesticide exposure levels vary greatly from farmworker to farmworker in practice, the variability estimated using field samples is more relevant to the current research compared to reported assay variability using lab controls. The variability here is affected by not only instrumentation/calibration measurement errors but also subsampling for the duplicates, differences in equipment, different operators/analysts, and day to day variations in the testing environment. We reported three indicators of assay variability (WCV, MDC, and probability of k -fold difference) for DMP, DMTP, DMDTP, DEP, and DETP using random duplicate samples from our PACE3 study. In summary, MLE for $p(k)$ is the most robust to model misspecifications. All assay variability indicators suggest that the within-subject assay variability is rather large when the farmworkers were exposed to from practically none to high level of pesticides. This alerts the researchers that inferences about apparent

longitudinal change in pesticide exposure should be examined cautiously in the context of substantial assay variability. We, therefore, encourage the collection of duplicate field samples in future research.

Finally, we note that there are some limitations about the bivariate zero-inflated lognormal models we choose for the PACE3 study. For example, the zero-inflated lognormal model is not superior to the 3-component mixture model for DMTP according to EIC or BIC. We also assume that the values $> \text{LOD}$ arise from a lognormal component with a common mean, since the focus is on estimation of within-subject variability. However, as in most longitudinal studies, replicate samples are frequently collected over time as the study continues. Therefore, as exposures (such as crops worked on) change over time, the common mean assumption may not be ideal. In addition, MLE under this model is difficult to perform for more than two repeated measures, since there will be more possible combinations of values $> \text{LOD}$ and/or values $\leq \text{LOD}$. We conjecture that a Bayesian approach might be able to better deal with this complicated situation, and more extensive research is warranted.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research is supported by Grant R01 ES 008739 from the National Institute of Environmental Health Sciences. We thank R. Lyles for providing a sample SAS program. We also thank the associate editor and the three anonymous reviewers for their constructive comments that led to a much improved manuscript.

References

- Arcury TA, Grzywacz JG, Chen H, Vallejos QM, Galván L, Whalley LE, et al. Variation across the agricultural season in organophosphorus pesticide urinary metabolite levels for Latino farmworkers in eastern North Carolina: project design and descriptive results. *American Journal of Industrial Medicine*. 2009; 52:539–550. [PubMed: 19517490]
- Arcury TA, Grzywacz JG, Talton JW, Chen H, Vallejos QM, Galván L, Barr DB, Quandt SA. Repeated pesticide exposure among North Carolina migrant and seasonal farmworkers. *American Journal of Industrial Medicine*. 2010; 53:802–813. [PubMed: 20623661]
- Barnhart HX, Song J, Lyles RH. Assay validation for left-censored data. *Statistics in Medicine*. 2005; 24:3347–3360. [PubMed: 15977288]
- Berk KN, Lachenbruch PA. Repeated measures with zeroes. *Statistical Methods in Medical Research*. 2002; 11:303–316. [PubMed: 12197298]
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986; 5:307–310. [PubMed: 2868172]
- Bland JM, Altman DG. Measurement error. *British Medical Journal*. 1996; 312:1654. [PubMed: 8664723]
- Chinn S. The assessment of methods of measurement. *Statistics in Medicine*. 1990; 9:351–362. [PubMed: 2362975]
- Chow S, Tse S. A related problem in bioavailability/bioequivalence studies – estimation of the intra subject variability with a common CV. *Biometrical Journal*. 1990; 32(5):597–607.
- Chu H, Kensler TW, Muñoz A. Assessing the effect of interventions in the context of mixture distributions with detection limits. *Statistics in Medicine*. 2005; 24:2053–2067. [PubMed: 15803450]

- Chu H, Nie L, Zhu M. On estimation of bivariate biomarkers with known detection limits. *Environmetrics*. 2008; 19:301–317.
- Donoghue D. Physiotherapy Research and Older People (PROP) group, and Stokes EK. How much change is true change? The minimum detectable change of the Berg Balance Scale in elderly people. *Journal of Rehabilitation Medicine*. 2009; 41:343–346. [PubMed: 19363567]
- Euser AM, Dekker FW, Cessie S, et al. A practical approach to Bland-Altman plots and variation of coefficients for log transformed variables. *Journal of Clinical Epidemiology*. 2008; 61:978–982. [PubMed: 18468854]
- Helsel, DR. Non detections and Data Analysis: Statistics for Censored Environmental Data. Hoboken, NJ: John Wiley and Sons; 2005.
- Jacqmin-Gadda H, Thiébaud R, Chêne G, Commenges D. Analysis of left-censored longitudinal data with application to viral load in HIV infection. *Biostatistics*. 2000; 1(4):355–368. [PubMed: 12933561]
- Lyles RH, Fan D, Chuachoowong R. Correlation coefficient estimation involving a left censored laboratory assay variables. *Statistics in Medicine*. 2001a; 20:2921–2933. [PubMed: 11568949]
- Lyles RH, Williams JK, Chuachoowong R. Correlating two viral load assays with known detection limits. *Biometrics*. 2001b; 57:1238–1244. [PubMed: 11764265]
- Moulton LH, Halsey NA. A mixture model with detection limits for regression analyses of antibody response to vaccine. *Biometrics*. 1995; 51:1570–1578. [PubMed: 8589241]
- Reed GF, Lynn F, Meade BD. Use of coefficient of variation in assessing variability of quantitative assays. *Clinical and Diagnostic Laboratory Immunology*. 2002; 9(6):1235–1239. [PubMed: 12414755]
- Rosner, B. Fundamentals of Biostatistics. 6. Belmont CA: Duxbury; 2006.
- Self SG, Liang K-Y. Estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*. 1987; 82:605–10.
- Quan H, Shih WJ. Assessing reproducibility by the within-subject coefficient of variation with random effects models. *Biometrics*. 1996; 52:1195–1203. [PubMed: 8962450]
- Taylor DJ, Kupper LL, Rappaport SM, Lyles RH. A mixture for occupational exposure mean testing with a limit of detection. *Biometrics*. 2001; 57:681–688. [PubMed: 11550915]
- Thiébaud R, Jacqmin-Gadda H. Mixed models for longitudinal left-censored repeated measures. *Computer Methods and Programs in Biomedicine*. 2004; 74:255–260. [PubMed: 15135576]
- U.S. Environmental Protection Agency. Definition and procedure for the determination of the method detection limit. *Code of Federal Regulations* 40. 1982; Part 136:565–567. Appendix B.
- de Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *Journal of Clinical Epidemiology*. 2006; 59:1033–1039. [PubMed: 16980142]
- Westgard JO, Quam EF, Barry PL. Establishing and evaluating QC acceptability criteria. MLO: medical laboratory observer. 1994a; 26(2):22–26. [PubMed: 10171791]
- Westgard JO, Seehafer JJ, Barry PL. Allowable imprecision for laboratory tests based on clinical and analytical test outcome criteria. *Clinical chemistry*. 1994b; 40(10):1909–1914. [PubMed: 7923771]

Figure 1.

Scatter plot for duplicate pesticide metabolite concentration data on log scale. *Note Both horizontal and vertical reference lines represent log (LOD). The oblique reference line represents the unity line. The single dot on the lower leftmost corner represents all pairs with both values $\leq$ LOD. The actual number of pairs is listed aside. The numbers on the upper left and lower right corners represent the number of pairs with one value $\leq$ LOD and one $>$ LOD. The number on the upper right corner represents the number of pairs with both values $>$ LOD. The percentages are included in parentheses.*

Figure 2.

Q-Q plots for values $> \text{LOD}$ on log scale for measure 1 of DMP and DMTP under univariate lognormal and zero-inflated lognormal distributions.

Figure 3. Parameter estimates under three different models in a simulation study. *Note* The true values for mean (μ), correlation (ρ), variance (σ^2), and within-subject variance σ_e^2 are 0, 0.5, 1, and 0.5, respectively. The proportion of values LOD for the lognormal component is 0.2. The error bars represent the estimated standard deviations from the simulated sampling distribution based on 5,000 replicates. The sample size for each replication is 50.

Figure 4.

Assay variability estimates under three different models in a simulation study. *Note* The true values for within-subject SD, within-subject CV, minimum detectable change (MDC95), and $p(2)$ are 0.71, 103.4%, 7.1, and 0.49, respectively. The proportion of values $\geq$ LOD for the lognormal component is 0.2. The *error bars* represent the estimated standard deviations from the simulated sampling distribution based on 5,000 replicates. The sample size for each replication is 50.

Table 1

Parameter estimates under three different models for left-censored duplicate data.

	Bivariate zero-inflated lognormal				Bivariate lognormal				Bivariate 3-component mixture model						
	$\hat{\omega}_0$	$\hat{\mu}$	$\hat{\rho}$	$\hat{\sigma}^2$	$\hat{\mu}$	$\hat{\rho}$	$\hat{\sigma}^2$	$\hat{\omega}_0$	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\mu}$	$\hat{\rho}$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\sigma}^2$
DMP	0.24	0.53	0.27	2.52	-0.21	0.53	3.98	0.25	0.33	0.42	1.47	0.59	0.72	12.53	7.29
DMTP	0.08	1.40	0.40	3.00	1.04	0.58	4.25	0.03	0.72	0.25	1.88	0.77	1.26	63.79	17.43
DMDTP	0.53	-0.16	0.29	3.62	-3.25	0.80	12.89	0.47	0.28	0.25	0.74	0.88	0.98	76.89	36.50
DEP	0.35	0.25	0	3.22	-1.36	0.55	7.79	0.32	0.33	0.35	1.19	0.56	0.57	22.93	11.96
DETP	0.56	-0.19	0	3.28	-3.59	0.70	13.99	0.52	0.24	0.24	0.65	0.63	0.68	36.66	18.26

$\hat{\omega}_0$ is the estimated proportion of zeroes, $\hat{\mu}$, $\hat{\rho}$, and $\hat{\sigma}^2$ are the estimated mean, correlation coefficient, and variance on the log scale for the lognormal component under bivariate lognormal, zero-inflated lognormal, and 3-component mixture distributions. $\hat{\omega}_1$ and $\hat{\omega}_2$ are estimated proportions of low variability and high variability lognormal components for the 3-component mixture distribution, respectively. $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are the estimated variances for the low variability and high variability lognormal components, respectively.

Table 2

Assay variability estimates (standard deviations) under three different models for left-censored duplicate data.

	Bivariate zero-inflated lognormal					Bivariate lognormal				Bivariate 3-component mixture			
	$\hat{\sigma}_e$	WCV(%)	MD95	$p(2)$	EIC	BIC	$\hat{\sigma}_e$	WCV (%)	MDC95	$p(2)$	EIC	BIC	MDC95
DMP	1.36 (0.18)	323 (78)	43 (30)	0.72 (0.04)	322	331	1.37 (0.20)	673 (284)	45 (35)	0.72 (0.04)	324	332	226 (148)
DMTP	1.34 (0.25)	410 (186)	42 (43)	0.72 (0.06)	434	437	1.34 (0.25)	762 (508)	42 (43)	0.72 (0.06)	440	442	1,158 (2,182)
DMDTP	1.60 (0.32)	588 (563)	85 (152)	0.76 (0.06)	280	286	1.62 (0.37)	60,481 (>10,000)	88 (379)	0.76 (0.06)	288	295	1,082 (1,226)
DEP	1.79 (0.23)	490 (291)	145 (138)	0.78 (0.03)	335	345	1.88 (0.30)	4,847 (6,320)	184 (354)	0.79 (0.03)	348	356	3,257 (3,346)
DETP	1.81 (0.29)	508 (542)	152 (257)	0.79 (0.04)	265	274	2.03 (0.41)	10,8046 (>10,000)	275 (1,544)	0.81 (0.04)	282	290	5,226 (5,840)

$\hat{\sigma}_e$ is the estimated within-subject standard deviation on the log scale for the lognormal component. WCV is the within-subject coefficient of variation. MDC95 is the minimum detectable ratio at 95 % confidence level for the concentration data on the natural scale. $p(2)$ is the probability of two duplicate measurements differ by a factor of 2 or more. The estimated standard deviations were presented in parantheses and were obtained from 5,000 bootstrap samples. EIC is the Efron information criterion and computed based on 5,000 bootstrap samples. BIC is the Bayesian information criterion.