Published in final edited form as:

Stat Med. 2010 July 20; 29(16): 1661–1672. doi:10.1002/sim.3905.

Likelihood-based methods for estimating the association between a health outcome and left- or interval-censored longitudinal exposure data[‡]

Kathleen A. Wannemuehler^{a,*}, Robert H. Lyles^b, Amita K. Manatunga^b, Metrecia L. Terrell^c, and Michele Marcus^{c,d}

^aDivision of Foodborne, Bacterial and Mycotic Diseases, National Center for Zoonotic, Vectorborne and Enteric Diseases, Centers for Disease Control and Prevention, Atlanta, GA, U.S.A.

^bDepartment of Biostatistics and Bioinformatics, The Rollins School of Public Health of Emory University, Atlanta, GA, U.S.A.

^cDepartment of Epidemiology, The Rollins School of Public Health of Emory University, Atlanta, GA, U.S.A.

^dDepartment of Environmental and Occupational Health, The Rollins School of Public Health of Emory University, Atlanta, GA, U.S.A.

Abstract

The Michigan Female Health Study (MFHS) conducted research focusing on reproductive health outcomes among women exposed to polybrominated biphenyls (PBBs). In the work presented here, the available longitudinal serum PBB exposure measurements are used to obtain predictions of PBB exposure for specific time points of interest via random effects models. In a two-stage approach, a prediction of the PBB exposure is obtained and then used in a second-stage health outcome model. This paper illustrates how a unified approach, which links the exposure and outcome in a joint model, provides an efficient adjustment for covariate measurement error. We compare the use of empirical Bayes predictions in the two-stage approach with results from a joint modeling approach, with and without an adjustment for left- and interval-censored data. The unified approach with the adjustment for left- and interval-censored data resulted in little bias and near-nominal confidence interval coverage in both the logistic and linear model setting.

Keywords

detection limits; censoring; environmental epidemiology; interval; measurement error; prediction; random effects

1. Introduction and background

This research is motivated by a reproductive health study of women exposed to polybrominated biphenyls (PBBs). The availability of the longitudinal serum PBB exposure measurements lends itself to the use of random effects models to obtain subject-specific

[‡]This article is a U.S. Government work and is in the public domain in the U.S.A.

*Correspondence to: Kathleen A. Wannemuehler, Division of Foodborne, Bacterial and Mycotic Diseases, National Center for Zoonotic, Vectorborne and Enteric Diseases, Centers for Disease Control and Prevention, Atlanta, GA, U.S.A..

[†]kpw9@cdc.gov

measures of exposure that are potentially related to health outcomes. In this study, the goal is to associate unmeasured PBB serum levels at a specific time point, t_i^* , with menstrual function outcomes such as abnormal/normal bleed length in exposed women. Because the level of PBBs in the body decays over time, t_i^* is defined as the exposure time that is most relevant to the health outcome of interest. In a two-stage approach, a prediction of the exposure at time t_i^* is obtained and then used as the independent variable in a second-stage health outcome model. A unified approach, which links the exposure and outcome in a joint model, arguably provides an inherent and efficient adjustment for covariate measurement error. The complexity of either approach increases when the exposure data are subject to detection limits, are coarse due to rounding, or are otherwise interval-censored. We propose a likelihood-based joint model to tie together the exposure and outcome data, while adjusting for the interval-censored exposure data.

Assessment of the effects of occupational and environmental exposures on specific health outcomes depends on the accurate measurement of the exposure of interest. Often true exposure concentration measurements are not available, either due to an inability to directly or accurately measure the exposure of interest or because direct measuring is too costly. Books by Fuller [1] and Carroll *et al.* [2], and in review-type publications such as those by Carroll [3], Thomas *et al.* [4] and Thurigen *et al.* [5] provide extensive coverage of conceptual and analytic issues related to measurement error. Studies dealing with longitudinal data to monitor exposure trends over time often seek to estimate an unmeasured exposure at critical time points [6]. These studies often rely on estimating exposure trends based on separate ordinary least-squares (OLS) fits to the longitudinal exposure measurements for each subject. These exposure estimates are then used in subsequent analyses linking exposure to various health outcomes of interest [7]. Wannemuehler and Lyles [8] refer to this as an *associated* measurement error problem.

A second type of predictor measurement error, common in environmental and virology studies, can occur when measurements fall above or below a limit of detection (LOD) of the measuring device. There are various proposed methods for handling unobserved exposure measurements due to a lower LOD. *Ad hoc* approaches include setting the unobserved measurements to zero or to some fraction of the LOD [9]. Hughes [10], Jacqmin-Gadda *et al.* [11], and Lyles *et al.* [12] propose the methods to address the issue of limits of detection in repeated HIV ribonucleic acid data using linear mixed effect models. Albert [13] explores jointly modeling measurements from two different assays with two different, but known LODs. Papers by Wu and Carroll [14], DeGruttola and Tu [15], and more recently Tsonaka *et al.* [16], propose methods that address informative censoring by jointly modeling the censoring process. These approaches fall within the scope of shared parameter models. Wannemuehler and Lyles [8] proposed a unified model in a full maximum likelihood framework to address both *associated* measurement error and non-detectable exposures in an occupational health setting. Measurement error due to detection issues may not be limited to LOD concerns, as data may be coarse due to rounding or otherwise effectively interval-censored. Such data are often treated as if measured without error, i.e. the rounded values or the midpoints of the known intervals are used at face value. In addition to the LOD references cited earlier, research by Heitjan and Rubin [17] suggests that these practices could introduce biases in estimation and/or compromise statistical inference.

2. Motivating example

In 1973, Michigan residents unknowingly consumed contaminated meat and dairy products after an inadvertent substitution of a fire retardant substance containing PBBs for a nutritional supplement in livestock feed. From 1976 through 1978, 3500+ individuals were enrolled into a registry, in an effort to assess the possible adverse health effects of the

exposure. Initial enrollment included a health questionnaire and a serum sample. Additional serum samples were drawn at various time points on some members of the registry cohort for use in a subsequent follow-up study. Further details of the incident and cohort are given elsewhere [18, 19]. In 1997–1998, the Michigan Female Health Study (MFHS) was conducted within the cohort of exposed women with the goal of associating reproductive health outcomes with exposure. Details of this follow-up study are provided by Blanck *et al.* [6, 20].

The following research utilizes data on the reproductive health of women enrolled in MFHS who were born prior to the estimated exposure date of July 1, 1973. See Davis *et al.* [21] for study details. The analysis focuses on the association between the reported menstrual bleed length in the past year and the unmeasured level of serum PBB. Because bleed length is considered a nonspecific marker for the reproductive status of women [22], we focus the current work on the association of PBB levels at a specific time point (t_i^*) with a dichotomous health outcome indicating abnormal length of menstrual periods among the exposed women. Abnormal bleed length was defined using the 75th percentile (≥ 7 days versus < 7 days). See Small *et al.* [23] for details regarding menstrual cycle characteristics. t_i^* is defined as the time from exposure to 6 months prior to the interview.

To estimate the decay of PBB in the serum for each woman, previous research used linear regression methods. Measurements of serum PBB levels were used at face-value if they were ≥ 1.0 parts per billion (ppb), the known LOD. For values falling below the LOD, a value of 0.5 was assigned. OLS estimates of the slopes were used in subsequent analyses looking at the age of menarche in daughters of exposed women [20], and daughters' height and weight [7]. These studies ignored the measurement error inherent in the serum PBB levels.

We explore the impact of measurement error on a dichotomous health outcome via simulations and provide a real world example using the MFHS data. We also explore the impact of measurement error on a continuous, normally distributed outcome via simulations. We propose a unified model to estimate the association while addressing both the issues of measurement error and interval censoring of exposure data, via a full-likelihood framework.

3. Methods

3.1. Exposure model

The research of Blanck *et al.* [6] assumed a linear one-compartment model for PBB elimination such that the initial PBB serum level (C_0) and subsequent concentrations ($C(t)$) at time t are related by the following formula:

$$C(t) = C_0 e^{-\lambda t},$$

where λ is the rate of decay. Taking the natural log of the above equation gives a linear relationship: $\ln(C(t)) = \ln(C_0) - \lambda t$, that allows one to estimate the rate of decay from the slope of a linear regression model given the repeated PBB serum levels taken over time. Blanck *et al.* then used linear regression models to estimate the overall decay rate (λ_i) for each woman.

Using the Blanck *et al.* model as a framework, Terrell *et al.* [24] utilized linear mixed model methods [25] to develop a decay model for the serum PBB measurements over time. They demonstrated improvements in the estimates by basing them instead on the empirical Bayes-like predictions of random effects using standard software (e.g. SAS [26]). The following

research builds upon their work by exploring efficient methods for linking health outcomes and latent time-specific exposure via a joint model. We begin by defining the following linear growth curve model [27, 28]:

$$Y_{ij} = \ln(X_{ij}) = \left(\alpha_0 + \sum_{s=1}^S \alpha_s C_{is} + a_i \right) + \left(\beta_0 + \sum_{v=1}^V \beta_v C_{iv}^* + b_i \right) t_{ij} + e_{ij} \quad (1)$$

where Y_{ij} is the j th PBB serum level on the natural log-scale for the i th individual taken at time t_{ij} , for $i = 1, \dots, N$ and $j = 1, \dots, n_i$. Equation (1) defines the linear change over time. The C_{is} represent subject-specific covariates that explain a portion of the variability of the initial serum levels. The C_{iv}^* , which may or may not overlap the set of covariates C_{is} , are covariates that interact with time and thus explain a portion of the variability in the slopes.

We make the typical normality assumptions associated with the vector of random effects (e.g. Laird and Ware [25]), i.e.

$$\begin{pmatrix} a_i \\ b_i \\ e_{ij} \end{pmatrix} \stackrel{iid}{\sim} \mathbf{N} \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_a^2 & \sigma_{ab} & 0 \\ \sigma_{ab} & \sigma_b^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix} \right].$$

Given that the time point of interest, t_i^* , is specific to each subject and typically different from any of the exposure measurement times (t_{ij}), our interest focuses on the true, unobserved model-based exposure level at time t_i^* (see equation (2)). Our approach is based on our belief that Y_i^* , located along the true trajectory for subject i , is the appropriate exposure to link with specific health effects.

$$Y_i^* = E[Y_{ij} | a_i, b_i, t_{ij} = t_i^*] = \left(\alpha_0 + \sum_{s=1}^S \alpha_s C_{is} + a_i \right) + \left(\beta_0 + \sum_{v=1}^V \beta_v C_{iv}^* + b_i \right) t_i^* \quad (2)$$

We note here that in theory both C_{is} and C_{iv}^* can be either time-independent or time-dependent. However, incorporating time-dependent covariates arguably requires that the values of these covariates be known at time t_i^* .

The minimal mean-squared error predictor of the quantity in equation (2) is obtained by replacing the random effects a_i and b_i by their posterior means, given the data $\mathbf{Y}_i$ (see equation (1); Searle *et al.* [29]). In the absence of censoring, one uses in practice the empirical best linear unbiased predictor (EBLUP).

3.2. Measurement error paradigm

The goal is to model the association between PBB levels at a specific time point, t_i^* , and potential reproductive health outcomes. We begin by considering abnormal bleed length as the health outcome of interest. Using the 75th percentile of the reported average bleed length as the cut point, a dichotomous outcome variable was created such that $R_i = 1$ if a woman reported an average length of her menstrual period to be ≥ 7 days, 0 otherwise. In the generalized linear model framework [30] we write the model using the (canonical) logit link. Given that the relevant exposure of interest is Y_i^* , we define a three-model measurement error paradigm [31] as follows:

$$\text{TDM: } \log \left(\frac{Pr[R_i=1]}{1 - Pr[R_i=1]} \right) = \theta_0 + \theta_1 Y_i^* + \sum_{h=2}^H \theta_h B_{ih} \quad i=1, \dots, N_k \quad \text{where } N_k \leq N \quad (3)$$

$$\text{MEM*}: Y_i^* = E[Y_{ij} | a_i, b_i, t_{ij}=t_i^*; \tau] \quad i=1, \dots, N_k \quad \text{where } N_k \leq N \text{ and } t_i^* \text{ is the time point of interest for subject } i, \quad (4)$$

$$\text{EDM: } Y_{ij} | (a_i, b_i) \sim N \left[\left(\alpha_0 + \sum_{s=1}^S \alpha_s C_{is} + a_i + \left(\beta_0 + \sum_{v=1}^V \beta_v C_{iv}^* + b_i \right) t_{ij} \right), \sigma^2 \right] \quad i=1, \dots, N. \quad (5)$$

The observed outcome, R_i , is observed on a subset of N_k women, where $N_k \leq N$. The B_{ih} 's are subject-specific covariates related to the outcome of interest.

In the 'true disease model' (TDM, equation (3)), which links the exposure of interest (Y_i^*) to the health outcome (R_i), the primary focus is on the parameter θ_1 . Since Y_i^* is unobserved in practice, the insertion of an estimate or surrogate introduces a covariate measurement error problem. Typically the 'measurement error model' (MEM) is a model that relates the surrogate to the true Y_i^* . Here we write the *MEM*, denoted with a '*' in equation (4), not as a model but rather a statement that defines the true, but unobserved, Y_i^* as the conditional expectation of Y_{ij} at time t_i^* , given the random effects (a_i, b_i) and τ . Here,

$\tau = (\alpha_0, \dots, \alpha_s, \beta_0, \dots, \beta_v, \sigma_a^2, \sigma_b^2, \sigma_{ab}, \sigma^2)'$ is the vector of nuisance parameters from the exposure model given in equation (1). The 'exposure density model' (EDM, equation (5)) specifies the distribution of Y_{ij} .

For a continuous outcome such as the age at onset of menarche for daughters of exposed women, we would replace the TDM in equation (3) with the following linear model:

$$R_i = \theta_0 + \theta_1 Y_i^* + \sum_{h=2}^H \theta_h B_{ih} + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma_r^2). \quad (6)$$

3.3. Likelihood for the unified model

For completely observed exposure data (i.e. exposures not subject to rounding or detection issues), the likelihood for estimating the primary parameter vector $\theta = (\theta_0, \dots, \theta_H)'$, can be written as:

$$\begin{aligned} \mathcal{L}(\theta, \tau, \mathbf{R}, \mathbf{Y}) &= \prod_{i=1}^N \int \int f(R_i | Y_i, a_i, b_i, \mathbf{B}_i, \theta)^{J_i} f(Y_i | a_i, b_i, \tau) f(a_i | b_i) f(b_i) da_i db_i \\ &= \prod_{i=1}^N \int \int f(R_i | Y_i^*, \mathbf{B}_i, \theta)^{J_i} f(Y_i | a_i, b_i, \tau) f(a_i | b_i) f(b_i) da_i db_i \\ J_i &= \begin{cases} 1 & \text{if } i \leq N_k \\ 0 & \text{if } i > N_k. \end{cases} \end{aligned} \quad (7)$$

Note that once a_i and b_i are known, the exposure of interest, Y_i^* , is also known and thus $f(R_i | \mathbf{Y}_i, a_i, b_i, \mathbf{B}_i, \theta)^{J_i}$ reduces to $f(R_i | Y_i^*, \mathbf{B}_i, \theta)^{J_i}$.

3.3.1. Interval censoring adjustment—In longitudinal studies, it is not uncommon to encounter study data that are a combination of left-censored, coarse (i.e. rounded) and completely observed measurements. In the case of the MFHS, the censored and rounded measurements are primarily a result of early laboratory methods that lacked precision, while the fully observed data resulted from more advanced lab techniques that were introduced later. We propose a proper interval-censoring adjustment designed to improve upon the common approach of inserting fixed values for censored or coarse exposure data, by taking advantage of the actual information known about each measurement.

For both left- and interval-censored data we do not observe the true measurement, but rather know that it falls within a given interval, say $(c_p, c_{p+1}]$. For simplicity, we specify a finite set of intervals as: $(c_1, c_2], (c_2, c_3], \dots, (c_p, c_{p+1}], \dots, (c_P, c_{P+1}]$, and we let

$$I_{ijp} = \begin{cases} 1 & \text{if } y_{ij} \in \text{interval } p \ (p=1, \dots, P) \\ 0 & \text{otherwise} \end{cases}$$

Note that on the natural log-scale we can specify $c_1 = -\infty$ and $c_{P+1} = +\infty$, allowing for left- and right-censoring, respectively. Also, for a given lower LOD, $c_2 = \log(\text{LOD})$.

We propose to incorporate the appropriate likelihood-based adjustment, similar to that of Lyles *et al.* [12] for strictly left-censored data, by substituting $f^*(\mathbf{I}_i, \mathbf{K}_i)$ for $f(\mathbf{Y}_i | a_i, b_i, \tau)$. Here, f^* is defined as:

$$f^*(\mathbf{I}_i, \mathbf{K}_i) = \prod_{j=1}^{n_i} \left\{ f(Y_{ij} | a_i, b_i, \tau)^{K_{ij}} \prod_{p=1}^P [F(c_{p+1}) - F(c_p) | a_i, b_i]^{(1-K_{ij})I_{ijp}} \right\}, \quad (8)$$

where

$$K_{ij} = \begin{cases} 1 & \text{if } y_{ij} \text{ is completely observed} \\ 0 & \text{if } y_{ij} \text{ is interval-censored,} \end{cases}$$

and $F(\cdot)$ is the cumulative distribution function corresponding to the density $f(Y_{ij} | a_i, b_i)$.

The integrated likelihood proposed in equation (7), with or without the adjustment given in equation (8), can be maximized by a variety of optimization techniques. We utilize the SAS *NLMIXED* (SAS version 9.13; SAS [32]) procedure's ability to accommodate a user-specified likelihood. We favor the trust-region method that requires the objective function to support continuous first- and second-order derivatives, and is reported to perform well for a wide variety of small- to medium-sized problems. The SAS syntax used to fit the joint model with interval-censored adjustment is available upon request.

3.4. Posterior predicted subject-specific mean exposures

We also explore the impact of left- and interval-censored data on the posterior predicted subject- and time-specific exposure levels. The 'Bayes' prediction of the quantity of interest

in equation (2) is given by the posterior mean of Y_i^* given the observed data. Omitting covariates for simplicity, we may write,

$$\tilde{Y}_{i(\text{Bayes})}^* = (\alpha_0 + \tilde{a}_i) + (\beta_0 + \tilde{b}_i)t_i^*, \quad (9)$$

where

$$\tilde{a}_i = E(a_i | \mathbf{I}_i, \mathbf{K}_i) = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a_i f(a_i | b_i) f(b_i) f^*(\mathbf{I}_i, \mathbf{K}_i) da_i db_i}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(a_i | b_i) f(b_i) f^*(\mathbf{I}_i, \mathbf{K}_i) da_i db_i},$$

and

$$\tilde{b}_i = E(b_i | \mathbf{I}_i, \mathbf{K}_i) = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} b_i f(b_i | a_i) f(a_i) f^*(\mathbf{I}_i, \mathbf{K}_i) da_i db_i}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(b_i | a_i) f(a_i) f^*(\mathbf{I}_i, \mathbf{K}_i) da_i db_i}.$$

Equation (9) theoretically provides the minimal mean-squared error of prediction (MSEP) estimate of Y_i^* , sometimes known as the ‘best predictor’ (BP) [29]. In practice, to obtain empirical Bayes-like predictions of exposure at the time point of interest for each subject via e.g. equation (9), we insert the maximum likelihood estimates of the variance components based on the fit of model (7) with the adjustment defined in equation (8). We have verified that the predicted values for a_i and b_i produced by *SAS NLMIXED* [32] matched those using direct numerical integration techniques to evaluate equation (9).

4. Simulations: assessment of six models

In the following section, a simulation-based approach is used to compare the results from various modeling strategies focusing on the estimation of the association parameter θ_1 (equation (3)). We compare the mean and standard deviation of 1000 estimated θ_1 's obtained from the proposed joint model (*Joint IC*, equation (7)) using the interval-censored adjustment (equation (8)) with the results based on five alternative strategies. The *Joint IC* model accommodates exposure data that are measured with and without error (i.e. either fully observed or known only within an interval due to assay limitations). The *Joint ALL IC* model applies the same approach except that it treats all exposure observations as interval-censored.

The *two-stage IC* approach first models the observed exposure data to obtain estimates of Y_i^* accounting for interval censoring. See Section 3.4 for details on empirical Bayes estimates. These estimates are then used in the second-stage health outcome model. This regression calibration-like method incorporates the interval-censoring adjustment in the first-stage exposure model, accommodating those measured with and without error, and then treats the estimated Y_i^* as if measured without error in the second stage.

In the *Joint Midpoint* model, each exposure observation, whether measured without error,

set to a fraction ($\frac{1}{2}$) of the LOD, or rounded to the nearest integer, is treated as if measured without error. This model addresses the *associated* measurement error using the joint model (equation (7)), but ignores potential issues inherent in relying upon rounded values.

The *two-stage Midpoint* approach first models the observed exposure data, treated expediently as in the joint midpoint approach, to obtain standard empirical Bayes estimates of Y_i^* . These estimates, analogous to those proposed by Terrell *et al.* [24], are then inserted into the second-stage health outcome model.

The *two-stage OLS* method obtains OLS estimates of the slope for each subject. These slopes are then regressed against potential covariates (i.e. body mass index (BMI)) to obtain adjusted slopes. In the second stage it treats the estimated $\hat{Y}_i^* = y_{i1} + (\text{adj slope}) t_i^*$, as if measured without error. This method parallels that by Blanck *et al.* [6].

The *True* method, though not available in practice, takes advantage of all (unobserved) true y_{ij} 's available in the simulated data. This model addresses the *associated* measurement error using the joint model (equation (7)) approach, and suffers from no other source of error.

4.1. Simulation I: binary outcome

Four exposure measurements were simulated for each of 500 subjects. Time between measurements was generated via a uniform distribution with bounds of 0.5–2 years. BMI at the initial time of exposure was generated from a transformed Beta(2,3) distribution with bounds of 15 and 38. BMI was then centered around the sample mean.

Exposure measurements were generated under the model described in equation (1) with BMI included as a covariate interacting with time. Covariance parameters were set to the following values: $\sigma_a^2 = 0.7$, $\sigma_b^2 = 0.03$, $\sigma_{ab} = -0.04$, and $\sigma^2 = 0.7$. Coefficient parameters were: $\alpha_0 = 6$, $\beta_0 = -0.3$, and $\beta_1 = 0.002$.

The lower LOD was set at 3, resulting in ≈ 10 per cent of measurements falling below the LOD. The remaining measures were randomly assigned via a Bernoulli distribution with probability = 0.5 to be fully observed or to be rounded to the nearest integer in order to mimic interval censoring. Simulated data ranged from 0–12. Interval-censored observations fell into one of the following categories: (0, 3), [3,3.5), [3.5,4.5), [4.5,5.5), ..., [11.5,12.5), ..., with 95 per cent of the data falling within the first six intervals.

A binary outcome was simulated for each subject under the model described in equation (3) using the simulated Y_i^* and age as a covariate. The event time for each subject, t_i^* , was generated from a uniform distribution with bounds of 1 and 3 years after the final exposure measurement. Age at the time of the health outcome was generated via a transformed Beta(1,3) with bounds 16–60 and was centered around the sample mean, and associated with the outcome via the coefficient $\theta_2 = -0.05$. The overall probability of a subject having the health outcome of interest, i.e. $Pr[R_i = 1]$, was chosen to be ≈ 0.25 . Three parallel simulations were run, setting the parameter of interest, θ_1 , to be 0.2, 0.4, and 0.8, respectively.

Results are given in Table I. Fitting the *true* model, where only *associated* measurement error is involved, results in little bias and optimal coverage within the specified logistic model setting. The *OLS* method results in an average $\hat{\theta}_1$ that underestimates the true θ_1 by more than 50 per cent, together with markedly sub-nominal coverage rates. The *midpoint* methods, both joint and two-stage, perform considerably better than the OLS method, although the results also show consistent underestimation of the association and sub-nominal coverage rates. The proposed interval-censored adjusted methods perform quite well, showing very little bias for the two smaller magnitudes of the association. The *two-stage IC* approach exhibits greater bias when $\theta = 0.08$. Also evident is a markedly lower coverage rate, 83 per cent, for the two-stage IC approach when $\theta = 0.08$. The coverage rate of the

Joint IC method is 92 per cent, though below the nominal 95 per cent, is quite acceptable relative to the *True* model.

In this simulation, and the two that follow in the next sections, convergence rates were near 99 per cent or better for all methods (see Table footnotes). The *Joint IC* and *Joint Midpt* models were fit in *SAS NLMIXED* using the parameter estimates from the *two-stage Midpt* model as the starting values.

To explore the impact of sample size, a second simulation study was run simulating data for 1000 subjects for $\theta_1 = 0.2$ and 0.4 independently. The models were fit to a random set of 100, 200, 300, 400, 500, 750, and 1000 subjects for each simulated data set. A plot of the means of $\hat{\theta}_1$ against the sample size (see Figure 1) shows that the joint *IC* model consistently improves the estimation of θ_1 for both $\theta_1 = 0.2$ and 0.4 as compared with the two-stage joint and both midpoint models. The model performs quite well even with a sample size of 100 though we note that convergence is under 90 per cent when $N = N_k = 100$. Model convergence improves to better than 95 per cent when N increases to 200 and to nearly 100 per cent when $N \geq 400$.

4.2. Simulation II: binary outcome—coarser exposure

In the second simulation study, we explore the impact of fewer and wider intervals. We assume the identical exposure model and coefficient parameters as described in Section 4.1, however interval-censored observations were assigned to one of the following intervals: (0, 3), [3,4.5), [4.5,6.5), [6.5,8.5), [8.5,10.5), [10.5,12.5),.... Nearly 100 per cent of the data fell within the first four intervals.

Results are given in Table II, and continue to indicate that the *IC* adjustment methods perform quite well for coarser exposure data. Bias is negligible and coverage is near-optimal for both $\theta_1 = 0.2$ and 0.4 . We note that the *midpoint* methods still exhibit underestimation of the association, and sub-nominal coverage for larger θ_1 .

Here we also emphasize the joint interval-censored adjusted method when all observations are treated as interval-censored (*Joint All IC*) versus a model that distinguishes between those measured without error and those interval-censored (*Joint IC*). While these methods were virtually indistinguishable in Table I, in Table II *Joint IC* yields a smaller average SE, reflecting the potential increase in efficiency when the method is tuned to handle both types of observations. The flexibility and ease with which the tuning is accomplished is an advantage of the proposed implementation via the *NLMIXED* procedure [32]. The MFHS study (Section 5) provides an example in which evolving laboratory methods logically lead to a desire to account for both completely observed and rounded (i.e. interval-censored) exposure data.

4.3. Simulation III—continuous outcome

In the third simulation study, we explore similar methods assuming a linear health outcome model. Exposure measurements were simulated and interval-censored in the same manner as described in Section 4.1. In the MFHS context, a potential outcome in this context is the age of onset of menarche in daughters of exposed mothers. The linear health outcome was simulated under the model described in equation (6). A single covariate (B_{i2}) representing whether the daughter was breastfed or not was included, generated via a Bernoulli distribution with a probability 0.5. We let $N = 400$, $N_k = 219$, and $E[n_i] = 4$. The mean simulated age of onset of menarche (R_i) was ≈ 10.5 years. Three parallel simulations were run with $\theta_1 = -0.05$, -0.1 , and -0.2 . For each simulation, $\theta_0 = 11$, $\theta_2 = -0.3$, and $\sigma_r^2 = 0.3$.

Results are given in Table III. The *IC* adjustment method again shows very little bias and provides good coverage for all three magnitudes of association, in both the joint and two-stage models. Both *midpoint* methods produce an underestimation of the association, with satisfactory coverage when $\theta_1 = -0.05$, but sub-nominal coverage for the larger associations. The OLS method again results in considerable bias and sub-nominal coverage.

5. MFHS: menstrual bleed length

The following analysis utilizes data from the MFHS where women were asked to report their usual length of menstrual bleeding during the past year. The analysis focuses on the association between menstrual bleed length, dichotomized using the 75th percentile (≥ 7 days versus < 7 days), and the unmeasured level of serum PBB 6 months prior to the interview. This time point (t_i^*) of 6 months prior to the interview was chosen because it falls at the midpoint of the reporting period.

PBB exposure data were available on 406 women. The range of age of these women at the time of exposure was 2–61 years (mean = 28). Between 2 and 6 longitudinal serum PBB exposure measurements were available for each woman (median = 3). Approximately 3 per cent of the measures fell below the known LOD of 1 ppb; 63 per cent were rounded to the nearest integer (i.e. interval censored) generally due to the use of a cruder assay method during the early years of the study; and 34 per cent were recorded more precisely and considered measured without error. Each woman's BMI at the time of the first exposure measurement was available, and BMI ranged from 15 to 61 kg/m² (mean = 24). Because decay rates were thought to vary more directly with initial PBB exposure than could be accounted for solely by the covariance parameter σ_{ab} (equation (1)), women were classified as having a high level of exposure if their initial PBB exposure measurement was ≥ 10 ppb. Fifty-three (13 per cent) women were classified as such, with initial exposure measurements ranging from 10 to 560 ppb (mean = 72).

Outcome data were available for 109 women with 30 (28 per cent) reporting menstrual bleed length ≥ 7 days. A logit model (equation (3)) was fit to the dichotomous bleed length variable. Age at the time of the interview was included in this model (range: 26–54 years), centered around the sample mean of 42 years. We assume that the subset of 109 women were a representative sample of the 406.

The exposure model (equation (1)) assumes that elimination of (log-transformed) PBBs occurs linearly as a function of time. To account for individual-level differences in elimination, the model includes BMI and age at exposure, centered around their respective sample means, as covariates interacting with time. Also included was the indicator variable for high initial PBB exposure. Including this indicator variable as a main effect and via an interaction term with time allows for unique population average intercepts and slopes for the low exposure group and high exposure group.

Results are given in Table IV. Use of the OLS surrogate results in a considerably smaller estimate of the association between exposure and the dichotomous response variable, bleed length, as compared with both the *IC* and *midpoint* methods. Differences between the *IC* and *midpoint* methods are small, likely due to the narrowness of the intervals on the log-scale.

The full-likelihood model proposed here provides an attractive approach to modeling the PBB data; however it does come with distributional assumptions of the random effects, and thus other models might be preferred. Weighted normal-plots (Figures 2(a) and (b)) proposed by Lange and Ryan [33] were examined to assess the normality of the estimated random effects. These plots appear reasonable under our assumed model, however they have

limitations with respect to differentiating incorrect distributional assumptions [34]. Further model assessments are recommended in practice.

5.1. Posterior predicted subject-specific mean exposures

We also explored the impact of adjusting for left- and interval-censored data on the prediction of Y_i^* for the 109 women in the bleed-length study. Table V gives the estimates for the parameter vector τ , based on fitting the linear growth curve model to the exposure data (see equation (1) and Section 5 for details). $\tilde{Y}_{i(\text{Bayes})}^*$ was obtained for each of the 109 women in the bleed-length study via the *IC* and the *midpoint* methods, with calculations made via the *NLMIXED* procedure [32]. Figure 3 suggests that the greatest impact of the *IC* adjustment upon prediction occurs in those women with the lower exposures at time t_i^* . On the log-scale, the intervals are wider in the left-tail than in the right; this may explain the smaller discrepancies as the predicted values get larger.

6. Discussion

In this article we have explored the impact of both *associated* measurement error that is inherent when exposure measurements are not available at critical time points with respect to health outcomes of interest, and measurement error due to left- and interval-censored exposure measurements. We proposed a likelihood-based approach that links the exposure and health outcome through a unified model that includes an interval-censoring adjustment. We explored this method in both logistic and linear health outcome model scenarios.

Results of the MFHS analysis showed a modest increase in the estimated θ_1 from the *joint IC* as compared with the *joint midpoint* method, however there was negligible impact on the odds ratio and subsequent inference about the association between high bleed length and PBB serum levels 6 months prior to interview. Because many subjects had only two or three PBB measurements, the exposure model was restricted to a linear model; however incorporating higher order terms into equation (1) is easily accomplished if necessary.

Simulation results indicated that relying upon estimated exposures ($\hat{Y}_i^*$) via the ordinary least-squares (OLS) method (e.g. Blanck *et al.* [6]) resulted in significant underestimation of the association of interest (θ_1) as well as poor coverage. The *two-stage midpoint* method, with Y_i^* estimated as proposed by Terrell *et al.* [24], leads to improved estimation of θ_1 , but still resulted in underestimation and sub-nominal coverage, especially for larger associations. Only slightly better coverage was obtained using the *joint midpoint* method. This clearly indicates that ignoring the measurement error inherent in left- and interval-censored exposure data can lead to bias and poor coverage properties when estimating the association between exposure and a health outcome. Methods that incorporated a proper adjustment for left- and interval-censored data resulted in little bias and excellent coverage in both the logistic and linear model setting. The *joint IC* method yielded slightly better coverage rates than did the *two-stage IC* method.

Research by Liang and Liu [35] indicated that in the logistic regression setting, the use of the conditional expectation as a surrogate can produce inconsistent estimates for the association parameter, θ_1 . They note that this should be less of a problem when θ_1 is small, or the probabilities (p) for $R = 1$ are not extreme. Our research provides empirical evidence that for moderate p and small θ_1 , the *CE* method (referred to here as *two-stage IC*) can perform quite well in the logistic model setting even though we did not incorporate a correction to the variance due to uncertainty in the posterior mean estimates of Y_i^* used in the second stage. Further research is necessary to better understand the boundaries of the *CE*

method within this setting. Nevertheless, because both the *two-stage IC* and *Joint IC* methods require numerical maximization of a non-standard integrated likelihood function, we recommend the unified *Joint IC* strategy for general use given the lack of lingering doubts about the validity of the standard error estimates.

Acknowledgments

A. K. Manatunga and R. H. Lyles were partially supported by an R01 from the National Institute of Environmental Health Sciences (ES012458). Additional funding was provided by the U.S. Environmental Protection Agency (R 825300), the National Institute of Environmental Health Sciences (R01 ES08341 and R01 ES012014), and by the Centers for Disease Control and Prevention U37/CCU500392. Our appreciation goes out to the investigators and participants of the Michigan Female Health Study. We also thank Lorraine Cameron and staff of the Michigan Department of Community Health for access to archived PBB exposure data. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

References

- Fuller, WA. Measurement Error Models. New York: Wiley; 1987.
- Carroll, RJ.; Ruppert, D.; Stefanski, LA. Measurement Error in Nonlinear Models. New York: Chapman & Hall; 1995.
- Carroll RJ. Covariance analysis in generalized linear measurement error models. *Statistics in Medicine*. 1989; 8:1075–1093. [PubMed: 2678349]
- Thomas D, Stram D, Dwyer J. Exposure measurement error: influence on exposure-disease relationships and methods of correction. *Annual Review of Public Health*. 1993; 14:69–93.
- Thurigen D, Spiegelman D, Blettner M, Heuer C, Brenner H. Measurement error correction using validation data: a review of methods and their applicability in case-control studies. *Statistical Methods in Medical Research*. 2000; 9(5):447–474. [PubMed: 11191260]
- Blanck HM, Marcus M, Hertzberg VS, Tolbert PE, Rubin C, Henderson AK, Zhang RH. Determinants of polybrominated biphenyl serum decay among women in the michigan pbb cohort. *Environmental Health Perspectives*. 2000; 108:147–152. [PubMed: 10656855]
- Blanck HM, Marcus M, Rubin C, Tolbert PE, Hertzberg VS, Henderson AK, Zhang RH. Growth in girls exposed *in utero* and postnatally to polybrominated biphenyls and polychlorinated biphenyls. *Epidemiology*. 2002; 13:205–210. [PubMed: 11880762]
- Wannemuehler KA, Lyles RH. A unified model for covariate measurement error adjustment in an occupational health study while accounting for non-detectable exposures. *Journal of the Royal Statistical Society Series C—Applied Statistics*. 2005; 54(1):259–271.
- Hornung RW, Reed LD. Estimation of average concentration in the presence of nondetectable values. *Applied Occupational and Environmental Hygiene*. 1990; 5:46–51.
- Hughes JP. Mixed effects models with censored data with application to HIV RNA levels. *Biometrics*. 1999; 55:625–629. [PubMed: 11318225]
- 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; 1:355–368. [PubMed: 12933561]
- Lyles RH, Lyles CM, Taylor DJ. Random regression models for human immunodeficiency virus ribonucleic acid data subject to left censoring and informative drop-outs. *Applied Statistics*. 2000; 49:485–497.
- Albert PS. Modeling longitudinal biomarker data from multiple assays that have different known detection limits. *Biometrics*. 2008; 64:527–537. [PubMed: 17764483]
- Wu MC, Carroll RJ. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*. 1988; 44:175–188.
- DeGruttola V, Tu XM. Modeling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics*. 1994; 50:1003–1014. [PubMed: 7786983]
- Tsonaka R, Verbeke G, Lesaffre E. A semi-parametric shared parameter model to handle nonmonotone nonignorable missingness. *Biometrics*. 2009; 65:81–87. [PubMed: 18373713]

17. Heitjan DF, Rubin DB. Ignorability and coarse data. *The Annals of Statistics*. 1991; 19:2244–2253.
18. Carter LJ. Michigan's PBB incident: chemical mix-up leads to disaster. *Science*. 1976; 192:240–243. [PubMed: 17831151]
19. Fries GF. The PBB episode in michigan: an overall appraisal. *CRC Critical Review Toxicology*. 1988; 16:105–156.
20. Blanck HM, Marcus M, Tolbert PE, Rubin C, Henderson AK, Hertzberg VS, Zhang RH, Cameron L. Age at menarche and tanner stage in girls exposed *in utero* and postnatally to polybrominated biphenyl. *Epidemiology*. 2000; 11:641–647. [PubMed: 11055623]
21. Davis SI, Blanck HM, Hertzberg VS, Tolbert PE, Rubin C, Cameron LL, Henderson AK, Marcus M. Menstrual function among women exposed to polybrominated biphenyls: a follow-up prevalence study. *Environmental Health: A Global Access Science Source*. 2005; 4:1–14. [PubMed: 15679885]
22. Brenner PF. Differential diagnosis of abnormal uterine bleeding. *American Journal of Obstetrics and Gynecology*. 1996; 175:766–769. [PubMed: 8828559]
23. Small CM, Manatunga AK, Klein M, Feigelson HS, Dominguez CE, McChesney R, Marcus M. Menstrual cycle characteristics. *Epidemiology*. 2006; 17:52–60. [PubMed: 16357595]
24. Terrell ML, Manatunga AK, Small CM, Cameron LL, Wirth J, Blanck HM, Lyles RH, Marcus M. A decay model for assessing polybrominated biphenyl exposure among women in the Michigan longterm PBB study. *Journal of Exposure Science and Environmental Epidemiology*. 2008
25. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics*. 1982; 38:963–974. [PubMed: 7168798]
26. SAS. SAS OnlineDoc 9.1.3. Cary, NC: SAS Institute Inc.; 2004. SAS/STAT: MIXED Procedure.
27. Diggle, PJ.; Liang, KY.; Zeger, SL. *Analysis of Longitudinal Data*. New York: Oxford University Press; 1994.
28. Singer, JD.; Willett, JB. *Applied Longitudinal Data Analysis*. New York: Oxford University Press; 2003.
29. Searle, SR.; Casella, G.; McCulloch, CE. *Variance Components*. New York: Wiley; 1992.
30. McCullagh, P.; Nelder, JA. *Generalized Linear Models*. 2nd edn. Boca Raton: Chapman & Hall, CRC; 1989.
31. Clayton, D. Models for the longitudinal analysis of cohort and case-control studies with inaccurately measured exposures. In: Dwyer, JH., editor. *Statistical Models for Longitudinal Studies of Health*. Oxford: Oxford University Press; 1992. p. 301-325.
32. SAS. SAS OnlineDoc 9.1.3. Cary, NC: SAS Institute Inc.; 2004. SAS/STAT: NLMIXED Procedure.
33. Lange N, Ryan L. Assessing normality in random effects models. *Annals of Statistics*. 1989; 17:624–642.
34. Verbeke, G.; Mohlenberghs, G. *Linear Mixed Models for Longitudinal Data*. New York: Springer; 2000.
35. Liang, KY.; Liu, XH. Estimating equations in generalized linear models with measurement error. In: Godambe, V., editor. *Estimating Functions*. U.S.A.: Oxford University Press; 1991. p. 47-63.

Figure 1.
Plot of means of $\hat{\theta}_1$ against sample size from 500 simulations.

Figure 2.

Weighted p-p plots (a) random intercept deviations (a_i) and (b) random slope deviations (b_i). Vertical axis: expected ($\Phi(\hat{Z}_{(i)})$) Horizontal axis: weighted observed ($\hat{F}(\hat{Z}_{(i)})$), where $\hat{Z}_{(i)}$ refers to the standardized estimates of the random effects.

Figure 3.

Plot of the posterior predicted means, Y_i^* , for 109 women. Estimates of Y_i^* using IC adjusted method (horizontal axis) versus the midpoint method (vertical axis).

Table I

Simulation I: binary outcome.

	Mean(SD)	$\widehat{SE}(\beta)$	CI coverage for β (%)
$\theta_1 = 0.2$			
True	0.200(0.094)	0.094	95
Joint IC	0.199(0.094)	0.095	96
Joint All IC	0.199(0.094)	0.096	96
Two-stage IC	0.198(0.092)	0.094	95
Joint Midpt	0.175(0.083)	0.084	93
Two-stage Midpt	0.173(0.082)	0.082	92
Two-stage OLS	0.094(0.080)	0.080	74
$\theta_1 = 0.4$			
True	0.391(0.101)	0.102	95
Joint IC	0.389(0.102)	0.103	95
Joint All IC	0.389(0.102)	0.104	96
Two-stage IC	0.382(0.097)	0.097	95
Joint Midpt	0.344(0.090)	0.091	89
Two-stage Midpt	0.337(0.086)	0.086	88
Two-stage OLS	0.179(0.076)	0.080	20
$\theta_1 = 0.8$			
True	0.759(0.130)	0.127	92
Joint IC	0.762(0.136)	0.131	92
Joint All IC	0.760(0.138)	0.132	91
Two-stage IC	0.713(0.111)	0.106	83
Joint Midpt	0.671(0.120)	0.114	74
Two-stage Midpt	0.638(0.102)	0.096	57
Two-stage OLS	0.314(0.079)	0.081	0

1000 simulations; $\approx 100\%$ convergence. $N = N_k = 500$.TRUE: Unified model (equation (7)) using unobserved Y_{ijs} measured without error.Joint IC: Unified model (equation (7)) using observed Y_{ijs} w/ interval-censored adjustment accommodating observations measured without error.Joint All IC: Unified model (equation (7)) using observed Y_{ijs} w/ interval-censored adjustment assuming all observations are measured with error.Two-stage IC: Two-stage approach, substituting $E[Y_{ij}|a_i, b_i, t_{ij}=t_i^*; \tau, \mathbf{I}_i, \mathbf{K}_{ij}]$ for Y_i^* .

Joint Midpoint: Unified model (equation (7)) using midpoint of each interval.

Two-stage Midpoint: Two-stage approach, substituting $E[Y_{ij}|a_i, b_i, t_{ij}=t_i^*; \tau]$ for Y_i^* .Two-stage OLS: Two-stage approach, substituting $\widehat{Y}_{it_i^*}$ for Y_i^* .

Table II

Simulation II: binary outcome—coarser exposure.

	Mean(SD)	$\overline{SE}(\hat{\beta})$	CI coverage for β (%)
$\theta_1 = 0.2$			
True	0.194(0.106)	0.108	95
Joint IC	0.196(0.111)	0.112	95
Joint All IC	0.197(0.115)	0.116	95
Two-stage IC	0.196(0.113)	0.113	95
Joint Midpt	0.171(0.097)	0.097	94
Two-stage Midpt	0.169(0.095)	0.095	94
Two-stage OLS	0.070(0.078)	0.078	63
$\theta_1 = 0.4$			
True	0.382(0.118)	0.115	94
Joint IC	0.385(0.121)	0.119	95
Joint All IC	0.388(0.126)	0.124	95
Two-stage IC	0.379(0.119)	0.115	94
Joint Midpt	0.337(0.107)	0.104	89
Two-stage Midpt	0.329(0.102)	0.098	87
Two-stage OLS	0.135(0.079)	0.078	09

1000 simulations; $\approx 99\%$ convergence. $N = N_k = 500$.

See footnote Table I.

Table III

Simulation III: continuous outcome.

	Mean(SD)	$\overline{SE}(\hat{\beta})$	CI coverage for β (%)
$\theta_1 = -0.05$			
True	-0.052(0.027)	0.027	96
Joint IC	-0.051(0.027)	0.028	96
Two-stage IC	-0.051(0.027)	0.027	96
Joint Midpt	-0.043(0.023)	0.023	94
Two-stage Midpt	-0.043(0.023)	0.023	93
Two-stage OLS	-0.023(0.030)	0.029	83
$\theta_1 = -0.1$			
True	-0.104(0.028)	0.029	96
Joint IC	-0.103(0.028)	0.029	97
Two-stage IC	-0.103(0.029)	0.028	96
Joint Midpt	-0.088(0.024)	0.024	91
Two-stage Midpt	-0.087(0.024)	0.024	90
Two-stage OLS	-0.045(0.031)	0.030	55
$\theta_1 = -0.2$			
True	-0.208(0.032)	0.033	97
Joint IC	-0.206(0.033)	0.034	96
Two-stage IC	-0.207(0.034)	0.031	93
Joint Midpt	-0.175(0.027)	0.028	85
Two-stage Midpt	-0.174(0.027)	0.026	81
Two-stage OLS	-0.091(0.035)	0.035	13

1000 simulations; $\approx 100\%$ convergence.

400 exposed subjects, 219 subjects w/ health outcome information.

See footnote Table I.

Table IV

MFHS: Health Outcome Model.

Method	$\hat{\theta}$ (SE)	OR 95% CL
Joint IC	0.384(0.199)	1.47(0.993, 2.17)
Two-stage IC	0.381(0.196)	1.46(0.996, 2.15)
Joint Midpt	0.372(0.191)	1.45(0.996, 2.11)
Two-stage Midpt	0.367(0.188)	1.44(0.998, 2.09)
Two-stage OLS	0.257(0.156)	1.29(0.952, 1.75)

 $N = 406$, $n_{ij} = 2-6$, $N_k = 109$.

See footnote Table I.

Table V

MFHS: exposure model parameters.

Parameter		Joint Midpoint Estimate(SE)	Joint IC Estimate(SE)
Intercept	α_0	0.757(0.040)	0.738(0.040)
HighExp	α_1	2.992(0.108)	3.007(0.109)
t_{ij}	β_0	-0.011(0.003)	-0.007(0.002)
Highexp* t_{ij}	β_1	-0.020(0.006)	-0.023(0.005)
BMI* t_{ij}	β_2	0.001(0.001)	0.001(4×10^{-4})
Age* t_{ij}	β_3	4×10^{-4} (2×10^{-4})	4×10^{-4} (2×10^{-4})
Var[a_i]	σ_a^2	0.416(0.038)	0.446(0.039)
Covar[a_i, b_i]	σ_{ab}	0.002(0.002)	1×10^{-4} (1×10^{-3})
Var[b_i]	σ_b^2	4×10^{-4} (1×10^{-4})	3×10^{-4} (1×10^{-4})
Var[e_{ij}]	σ^2	0.096(0.006)	0.073(0.005)

 $N = 406$, $n_{ij} = 2-6$.

See Section 5 for model details.