



Published in final edited form as:

*J Agric Biol Environ Stat.* 2013 March 1; 18(1): 22–38. doi:10.1007/s13253-012-0115-9.

## Approximate and Pseudo-Likelihood Analysis for Logistic Regression Using External Validation Data to Model Log Exposure

**Robert H. LYLES [Associate Professor]** and

Department of Biostatistics and Bioinformatics, The Rollins School of Public Health of Emory University, 1518 Clifton Rd. N.E., Mailstop 1518-002-3AA, Atlanta, GA 30322, USA

**Lawrence L. KUPPER [Professor Emeritus]**

Department of Biostatistics, Gillings School of Global Public Health, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420, USA

### Abstract

A common goal in environmental epidemiologic studies is to undertake logistic regression modeling to associate a continuous measure of exposure with binary disease status, adjusting for covariates. A frequent complication is that exposure may only be measurable indirectly, through a collection of subject-specific variables assumed associated with it. Motivated by a specific study to investigate the association between lung function and exposure to metal working fluids, we focus on a multiplicative-lognormal structural measurement error scenario and approaches to address it when external validation data are available. Conceptually, we emphasize the case in which true untransformed exposure is of interest in modeling disease status, but measurement error is additive on the log scale and thus multiplicative on the raw scale. Methodologically, we favor a pseudo-likelihood (PL) approach that exhibits fewer computational problems than direct full maximum likelihood (ML) yet maintains consistency under the assumed models without necessitating small exposure effects and/or small measurement error assumptions. Such assumptions are required by computationally convenient alternative methods like regression calibration (RC) and ML based on probit approximations. We summarize simulations demonstrating considerable potential for bias in the latter two approaches, while supporting the use of PL across a variety of scenarios. We also provide accessible strategies for obtaining adjusted standard errors to accompany RC and PL estimates.

### Keywords

Consistency; Likelihood; Multiplicative measurement error; Probit; Validation

## 1. MOTIVATION AND BACKGROUND

### 1.1. Motivating Example

Our motivating example arises from a study of the association between exposure to metal working fluids (MWF) and lung function as measured by the presence or absence of wheeze, leading to methodological challenges as discussed by Weller et al. (2007). Data were collected on 1040 subjects from a prior epidemiological study (Greaves et al. 1997; Eisen et

al. 1997, 2001), including the binary response variable wheeze ( $Y$ ), vectors of covariates  $\mathbf{C}$  capturing information on workplace, age, race, and smoking status, and vectors  $\mathbf{D}$  capturing machine types used by workers and whether they were exposed to straight or synthetic MWF. The main objective is to relate  $Y$  to a quantitative risk factor ( $X$ , thoracic aerosol fraction in  $\text{mg}/\text{m}^3$ ) induced by MWF exposure, while controlling for the vector  $\mathbf{C}$ .

The primary challenge in this analysis is the fact that data on the true exposure  $X$  was not collected on any of the 1040 main study subjects. Fortunately, an independent sample of 83 subjects was available as an external validation study (Woskie et al. 1994), consisting of data on  $X$  and  $\mathbf{Z} = (\mathbf{C}, \mathbf{D})$  collected on each subject. Thus, the problem posed by the motivating example is readily viewed as one of exposure measurement error.

The approach taken by Weller et al. (2007) was to first fit a measurement error model (MEM) to the external validation data that relates  $X$  to the variables in  $\mathbf{Z}$ . They then applied a new variant on the popular regression calibration (RC) method that utilized the parameter estimates from the first model to estimate the coefficients of a logistic regression model relating  $Y$  to  $(X, \mathbf{C})$  among the main study subjects. The following section provides some relevant background on RC and other measurement error correction approaches, along with introductory material and specifics about the objectives of the current paper.

## 1.2. Background and Objectives

The handling of one or more mismeasured continuous predictors in logistic regression figures is prominent in the broad literature on exposure measurement error (e.g., Schafer 1987; Rosner, Willett, and Spiegelman 1989; Rosner, Spiegelman, and Willett 1990; Stefanski and Carroll 1990; Liang and Liu 1991; Kuha 1994; Thoresen and Laake 2000; Spiegelman, Carroll, and Kipnis 2001; Thurston, Spiegelman, and Ruppert 2003). As reflected in these and other references, the nature of this problem has led researchers to approach it from several different angles. In part, these various treatments arise from options related to philosophical views on measurement error modeling. For example, measurement error can be viewed as structural or functional (Carroll et al. 2006). The former term implies that the true unknown predictor  $X$  is viewed as a random variable following some specified distribution  $f(X)$ ; the modern view of the latter suggests that  $X$  may be fixed or random, but if random, then few or no assumptions are made about  $f(X)$ .

Variety in approaches to address measurement error in logistic regression also arises due to practical issues inherent in epidemiological research. These include study design considerations critical to the estimation of measurement error parameters, for which the two primary strategies involve collection of repeated or “replication” data on mismeasured exposure, or the use of validation data (e.g., Thomas, Stram, and Dwyer 1993). The validation approach has the drawback of requiring a gold standard method for measuring exposure, but arguably permits the most flexible and robust strategy to reveal the nature of an underlying MEM.

Epidemiologists are adept at designing effective validation studies, while tending to favor intuitively and computationally accessible statistical methods for practical use. This in part underlies the popularity of the RC approach, which arose from two concurrent and conceptually appealing schools of thought (Rosner, Willett, and Spiegelman 1989, 1990; Carroll and Stefanski 1990; Stefanski and Carroll 1990). The Rosner et al. strategy essentially applies correction factors to estimated coefficients from fitting the “naïve” logistic model, in which an error-prone surrogate exposure variable is used in place of the unknown  $X$ . The Carroll and Stefanski strategy instead “calibrates” by replacing  $X$  by its conditional expectation given one or more surrogates. Under main/validation study designs

with additive measurement error, the two strategies tend to yield equivalent estimators under a broad class of generalized linear models (Thurston, Spiegelman, and Ruppert 2003).

For logistic regression, it is well known that the RC estimator is only approximately unbiased asymptotically under certain conditions. Specifically, the approximation has been justified when the outcome is rare and measurement error is normally distributed, and/or the magnitudes of the exposure effect and measurement error are both “small” (e.g., Kuha 1994). Alternative structural approaches to develop a fully consistent estimator are generally maximum likelihood (ML)-, pseudo-likelihood (PL)-, or quasi-likelihood (QL)-based (Carroll et al. 2006). A full ML approach uses distributional assumptions about the MEM to specify an overall joint likelihood in main/validation study settings, while PL (Gong and Samaniego 1981) offers a scaled back alternative by inserting estimates of MEM parameters into a main study-only likelihood. QL (e.g., Liang and Liu 1991) permits a similar approach that engenders a set of estimating equations that rely in theory only on the first two conditional moments of  $X$  given its surrogate(s) and other covariates. Practical use of the ML, PL, and QL methods has remained rather limited to date, however, due to real and/or perceived computational complexities.

Another feature that distinguishes approaches to exposure measurement error problems is the nature and complexity of the MEM itself. While the typical additive-normal model is by far the most common (e.g., Fuller 1987; Thoresen and Laake 2000), epidemiological studies often motivate incorporation of covariates into the MEM. Further, environmental epidemiology sometimes invites consideration of multiplicative measurement error (e.g., Hwang 1986; Carroll 1989; Pierce et al. 1992; Iturria, Carroll, and Firth 1999). Specifically, several references (Rappaport 1991; Lyles and Kupper 1997; Rappaport and Kupper 2008) suggest that environmental exposure on an untransformed scale is often appropriate as a predictor of health outcomes, while additive-normal MEMs for exposure tend to be more reasonable on the log scale. While the methods we discuss apply readily under additive error models, this view directly motivates a multiplicative-lognormal MEM. Treatments of multiplicative error have been relatively rare, particularly under logistic, as opposed to linear, disease models.

The remainder of this paper proceeds as follows. In Section 2, we introduce the assumed logistic regression model and the MEM stemming from the motivating example. Section 3 outlines the exposure measurement error adjustment methods to be studied. We focus on RC, PL and an approximate full ML approach, with a view toward making the latter two methods more accessible via sharable programs that invoke optimization routines from standard statistical software. We also emphasize accessible proposals for computing appropriate standard errors to accompany point estimates based on these methods. In Section 4, we present results from the analysis of the motivating data. Section 5 summarizes simulation studies of the relative performances of the approaches considered. In particular, we focus on the sensitivity of the RC and approximate full ML approaches to the magnitudes of the exposure effect and the measurement error variance. We conclude with a discussion in Section 6.

## 2. LOGISTIC MODEL AND MEM FOR MOTIVATING EXAMPLE

As discussed in Section 1.1, we revisit an example (Weller et al. 2007) based on a study of the association between MWF exposure and lung function. The example uses  $n_m = 1040$  main study observations, together with  $n_v = 83$  observations from an external validation sample. The objective is to relate the binary outcome  $Y$  to a true exposure  $X$ , while controlling for a vector of covariates ( $C$ ). Only  $Y$  and  $Z$  are measured on main study subjects, where  $C$  is a subset of  $Z$  and  $Z = (C, D)$  is a vector of variables assumed associated

with  $X$ . In contrast, validation study subjects only contribute data on  $X$  and  $\mathbf{Z}$ , characteristics of which are assumed “transportable” to the main study (Carroll et al. 2006). In this example, all variables in  $\mathbf{Z}$  happen to be binary.

Of primary interest is a logistic regression model, i.e.,

$$\text{logit} [\Pr (Y_i=1|L_i, C_i)] = \eta_i = \alpha + \beta X_i + \sum_{j=1}^p \delta_j C_{ij}, \quad (2.1)$$

where  $i$  indexes subject. Model (2.1) is the “true disease model” (TDM) in the terminology of Clayton (1992). We assume the following linear measurement error model (MEM):

$$L_i = \ln (X_i) = \theta_i + \varepsilon_i, \quad (2.2)$$

where

$$\theta_i = \gamma_0 + \sum_{k=1}^q \gamma_k Z_{ik} \quad \text{and} \quad \varepsilon_i \sim N(0, \sigma_e^2).$$

Further details regarding the covariates in  $\mathbf{Z} = (\mathbf{C}, \mathbf{D})$  are reserved for Section 4. Note that the homoscedasticity of errors associated with model (2.2) reflects an assumption of non-differential (with respect to  $Y$  and  $\mathbf{Z}$ ) measurement error.

Weller et al. (2007) fitted model (2.2) to the untransformed exposure ( $X$ ) data and presented a new variant of RC modeled after the version due to Rosner, Willett, and Spiegelman (1989, 1990), in which a subset of the variables in  $\mathbf{Z}$  are designated as surrogates and corrected estimates based on each surrogate are weighted to produce a final estimate of  $\beta$ . As they acknowledge, this variation on RC is likewise only approximately unbiased asymptotically for logistic regression and is expected to perform well under “small  $\beta$ ” and “small measurement error” conditions.

Upon fitting the MEM in (2.2) to the external validation sample both with and without the log transformation, standard residual analysis reveals that the normality assumption is better justified on the log scale (Shapiro-Wilk p-value = 0.25, vs. <0.001). This finding is consistent with frequent evidence in the environmental sciences literature suggesting the appropriateness of lognormal models for exposure. Thus, our treatment of this example differs from Weller et al.’s in that we fit the MEM on the log scale while retaining the untransformed  $X$  as the predictor of interest in model (2.1). This effectively induces a multiplicative measurement error model, as seen in the following section. A more important difference is that in what follows, we primarily focus on likelihood-based methods that provide asymptotically unbiased estimators of logistic regression coefficients under models (2.1) and (2.2).

### 3. METHODS

#### 3.1. RC and Likelihood-Based Methods

As a regression calibration alternative for comparison, we utilize the version of RC proposed by Carroll and Stefanski (1990). In this approach, we first fit the MEM in (2.2) to the validation data to produce the estimates  $(\widehat{\gamma}, \widehat{\sigma}_e^2)$  via ordinary least squares (OLS). Then, model (2.1) is fit to the main study subjects after replacing  $X_i$  by its estimated conditional expectation, i.e.,

$$\widehat{\mu}_{.vi} = \widehat{E}(X_i|Z_i) = \exp\left(\widehat{\theta}_i + \widehat{\sigma}_e^2/2\right), \quad (3.1)$$

where the form of  $\widehat{\mu}_{.vi}$  stems from the underlying multiplicative-lognormal MEM that relates  $X$  to  $\mathbf{Z}$  upon exponentiating both sides of Equation (2.2). Note that prior literature (Lyles and Kupper 1997; Carroll et al. 2006) suggests that Carroll and Stefanski's version of RC is preferred to that of Rosner, Willett, and Spiegelman (1989, 1990) and its variants when measurement error is multiplicative, in contrast to the similarity of the two versions under additive error models (Thurston, Spiegelman, and Ruppert 2003).

As opposed to RC, a full ML approach must be based upon a likelihood proportional to the product ( $L_m \times L_v$ ) of the main and external validation likelihoods, where we write

$$L_m = \prod_{i=1}^{n_m} \Pr(Y_i = y_i | Z_i) = \prod_{i=1}^{n_m} \int_{-\infty}^{\infty} \Pr(Y_i = y_i | L_i, C_i) f(L_i | Z_i) dL_i \quad (3.2)$$

and

$$L_v = \prod_{i=1}^{n_v} f(L_i | Z_i),$$

with  $\Pr(Y_i = y_i | L_i, C_i) = e^{\eta y_i} (1 + e^{\eta})^{-1}$  based on (2.1), and the normal density  $f(L_i | Z_i)$  following from (2.2). While full ML is arguably an ideal solution to this structural measurement error problem, application has typically been hampered by numerical challenges posed by the integral in (3.2) despite attempts to facilitate its computation (e.g., Crouch and Spiegelman 1990). Messer and Natarajan (2008) offered further hope in this regard in the case of additive-normal measurement error, by showing how modern statistical software can be adapted to the problem.

As suggested by prior authors, one option to lessen computational difficulties associated with full ML is to consider a pseudo-likelihood (PL) alternative. To apply PL here, we insert the OLS-based estimates ( $\widehat{\gamma}, \widehat{\sigma}_e^2$ ) from the fit of the MEM (2.2) into the expression for  $L_m$  in (3.2). We then maximize  $L_m$  only with respect to the model (2.1) primary parameters ( $\alpha, \beta, \delta$ ). Note that the integral in (3.2) remains to be dealt with, however. Also, relatively complex sandwich estimators (e.g., Gong and Samaniego 1981) have historically been required for calculating standard errors to accompany PL estimates in order to account for uncertainty in ( $\widehat{\gamma}, \widehat{\sigma}_e^2$ ). In Section 3.2, we suggest a computational shortcut to circumvent the need for the sandwich estimator.

An alternative approach designed to get around the numerical integration problem is based on the idea of approximating the logistic function via the probit, i.e.,  $H(t) \cong \Phi(t/k)$ , where  $k$

is a constant,  $H(t) = \frac{\exp(t)}{1 + \exp(t)}$ , and  $\Phi(\cdot)$  denotes the standard normal c.d.f. Such a single-probit approximation tends to hold closely for values of  $k$  in the neighborhood of 1.6–1.8 (e.g., Carroll et al. 1984; Savalei 2006), and implies the following:

$$\Pr(Y_i = 1 | Z_i) \cong H\left(\frac{\alpha + \beta E(X_i | Z_i) + \sum_{j=1}^p \delta_j C_{ij}}{\sqrt{1 + \beta^2 \text{var}(X_i | Z_i) / k^2}}\right). \quad (3.3)$$

The derivation of (3.3) is explicitly based on the assumption that the  $X|Z$  distribution is normal. However, (3.3) has been shown to remain accurate, at least in some cases, when  $X|Z$  is far from normal (Carroll et al. 2006, pp. 91–92). While Monahan and Stefanski (1991) and Demidenko (2004) proposed potentially more accurate approximations to  $H(t)$  using linear combinations of probits, in the current study we evaluate the use of (3.3) with  $k = 1.7$  and  $k = 1.749$ . These two values of  $k$  stem from invoking minimax and Kullback–Leibler criteria, respectively (Savalei 2006).

We use the probit approximation to facilitate optimizing approximate versions of the full likelihood ( $L_m \times L_v$ ) by inserting (3.3) into the expression for  $L_m$  in (3.2), thus eliminating the integral. This approach jointly estimates all parameters ( $\boldsymbol{\gamma}, \sigma_e^2, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\delta}$ ), as opposed to inserting external estimates ( $\widehat{\boldsymbol{\gamma}}, \widehat{\sigma_e^2}$ ) into an approximate main study likelihood as in a PL strategy. The form of  $E(X_i|Z_i)$  is given in (3.1), while the multiplicative-lognormal MEM also implies that

$$\text{var}(X_i|Z_i) = \mu_{xi}^2 [\exp(\sigma_e^2) - 1]. \quad (3.4)$$

While one could undertake a quasi-likelihood (QL) approach that would hinge upon these same two conditional moments (Liang and Liu 1991; Lyles and Kupper 1997), we focus here upon investigating the relative performances of RC, PL, full ML, and approximate full ML methods.

### 3.2. Computational Issues and Standard Error Estimation

With respect to point estimation, RC is clearly the simplest of the methods to apply. For the ML and PL methods, we favor the use of numerical optimization, integration, and Hessian approximation functions directly available in modern software, with a view toward assessing their reliability for the current problem. Specifically, we utilize the dual quasi-Newton function NLPQN for optimization, the finite differences second derivative approximation function NLPFDD, and the QUAD function for adaptive numerical integration available in SAS IML (SAS Institute, Inc. 2004). Similar functions are featured in other software (e.g., the R package). Given convergence of such algorithms, the full and approximate ML methods enjoy the advantage of direct applicability of the numerically-derived Hessian to estimate standard errors.

In contrast, RC and PL pose additional challenges for standard error estimation. Options to account for uncertainty in the  $\widehat{\mu}_{xi} = \widehat{E}(X_i|Z_i)$  values used in RC include sandwich variance adaptations (e.g., Liang and Liu 1991; Lyles and Kupper 1997; Carroll et al. 2006), or bootstrapping (Thoresen and Laake 2000). Sandwich estimators (Gong and Samaniego 1981; Spall 1989) are an option for PL as well, but bootstrapping is less attractive due to computing time and risk of instabilities with the required numerical integration and optimization.

In this study, we propose and evaluate practical approaches designed to facilitate estimation of standard errors to accompany the RC and PL methods. For PL, our idea is simple. We apply a second derivative approximation function (NLPFDD in SAS IML) to the full likelihood ( $L_m \times L_v$ ) [see (3.2)], evaluating the full Hessian at the final PL estimates of ( $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\delta}$ ) and the externally-derived OLS estimates of ( $\boldsymbol{\gamma}, \sigma_e^2$ ). The resulting standard error estimates adjust for uncertainty in ( $\widehat{\boldsymbol{\gamma}}, \widehat{\sigma_e^2}$ ), with the advantage that only a single call to the second derivative function is required subsequent to convergence of a PL (rather than full ML) estimation process.

For RC, we borrow ideas originally proposed by Rubin (1987) in conjunction with multiple imputation. Based on the original OLS fit to the external validation sample, we generate  $M$  replicates  $(\widehat{\gamma}, \widehat{\sigma}_e^2)_m$  ( $m = 1, \dots, M$ ) using standard methods involving draws from chi-square and multivariate normal distributions dictated by model (2.2), the OLS estimates, and their variance-covariance matrix (see Schafer 1999). For each of these, we calculate a corresponding replicate of the RC estimate by inserting the  $m$ -th set of  $\widehat{\mu}_{xi}$ 's [see (3.1)] in place of the  $X_i$ 's in the logistic regression model fit to the main study data. We then use Rubin's well-known approach to combine the latter replicates and estimate the variance of each component of the vector of RC parameter estimates, accounting for within- and between-replication sources of variability. While no actual imputation of data is involved, we use the acronym 'MI' in subsequent tables to refer to the resulting standard error estimates due to connections with the multiple imputation paradigm. An advantage over other computational approaches (e.g., bootstrapping) is that we require only a few replicates; we take  $M = 5$  in the example and simulations to follow.

#### 4. RESULTS

We return to the motivating example drawn from Weller et al. (2007), for which the covariates ( $\mathbf{C}$ ) in model (2.1) include the following binary variables measured on automobile industry workers: workplace (i.e., plant number, 2 vs. 1), three indicators for age group (30-39, 40-49, and 50+ years, with <30 as reference), race (Caucasian vs. non-Caucasian), and current smoker status (yes vs. no). The vector  $\mathbf{D}$  consists of three other indicators: machine type (grinding vs. non-grinding), exposure to straight metal working fluid (MWF; yes vs. no), and exposure to synthetic MWF (yes vs. no). For reasons explained by Weller et al. (2007), workers exposed to soluble MWF were excluded. The vectors  $\mathbf{Z} = (\mathbf{C}, \mathbf{D})$  for each subject are utilized in the MEM (2.2).

Table 1 summarizes the fit of the MEM in model (2.2) to the external validation study data ( $n_v = 83$ ). For comparison, we present results based on OLS, in addition to those based on the single-probit-based approximate full ML approach [see (3.3)]. The point estimates are quite similar in each case, while estimated standard errors are slightly smaller based on the latter method. This is somewhat expected due to the reduced estimate for the residual variance ( $\sigma_e^2$ ) obtained by approximate ML, but may also reflect that a full ML approach captures limited information about the MEM parameters in the main study data. Certain variables (plant, age 30-39, smoking status) became significant or marginally so when fitting model (2.2) to the logged exposure data, but were non-significant in the identical MEM without transforming  $X$  (Weller et al. 2007).

Table 2 presents measurement error adjusted analyses for model (2.1) using RC, PL, and approximate full ML based on (3.3). These three analyses agree well with respect to all coefficient and standard error estimates, and all point to the conclusion that thoracic aerosol fraction ( $X$ ) is positively associated with the prevalence of wheeze ( $Y$ ) after adjusting for the covariates ( $\mathbf{C}$ ). It is important to appreciate that the similarity in the point estimates across the three methods is predictable here given the very small OLS-based estimate of the measurement error variance ( $\sigma_e^2 = 0.139$ ) obtained when fitting the MEM. We were also able to adapt SAS NLMIXED code provided for the additive-normal error case by Messer and Natarajan (2008), to obtain a full ML solution for the example data set. The results were identical to those obtained via the approximate ML approach in the rightmost column of Table 2.

For RC and PL, note that we report two standard error estimates in Table 2. The first (top) estimate ignores uncertainty due to MEM parameter estimation, while the second (bottom)

adjusts for this uncertainty using the approaches proposed for each method in Section 3.2. In this example, there is relatively little underestimation in the former case. For comparison, we also note that the estimated exposure effects and standard errors obtained via RC, PL, and approximate full ML agree quite well with the values [1.06 (0.38)] reported in Weller et al. (2007), despite the difference in scales used when fitting the MEM to the exposure ( $X$ ) data. Again, this similarity in results stems in large part from the small measurement error ( $\sigma_e^2$ ) operating in this example, regardless of the scale for  $X$  (log or untransformed) that one chooses. In the following section, we explore the effects of variations in this parameter on the log scale.

## 5. SIMULATION STUDIES

### 5.1. Multiplicative-Lognormal MEM

Theory suggests that the RC estimator could deteriorate statistically with increases in the magnitude of the exposure effect ( $\beta$ ) and/or the measurement error variance ( $\sigma_e^2$ ) under models (2.1) and (2.2). Likewise, it stands to reason that the performance of approximate ML based on (3.3) may vary with  $\sigma_e^2$  as a result of its influence upon the extent to which the lognormal  $X|Z$  distribution departs from normality. On the other hand, practical experience suggests that PL- and other ML-based methods may pose greater numerical challenges with increases in these parameters. To explore these possibilities, we conducted simulations under 9 separate conditions distinguished by the assumed ( $\beta, \sigma_e^2$ ) combinations.

Specifically, we considered three values for  $\beta$  (0.50, 1.10, 1.50; designated as “small,” “moderate,” and “large”), and three values for  $\sigma_e^2$  (0.139, 0.50, 1.50; “small,” “moderate,” “large”). The smallest  $\sigma_e^2$  value equates to the estimate obtained in the motivating example. Otherwise, the simulations closely mimicked the example conditions. Each simulated data set included  $n_m = 1040$  “main” and  $n_v = 83$  external “validation” observations. A total of 9 variables were generated independently for each simulation to comprise  $Z$ , each as Bernoulli variates with prevalences identical to those observed for the 9 covariates in the actual main study data. The true parameters generating the main model (2.1) of interest and the MEM (2.2) for each simulation were identical to the OLS and RC estimates presented in Tables 1 and 2, respectively. Because we observed little difference in results by taking  $k = 1.7$  as opposed to  $k = 1.749$  in approximation (3.3) when evaluating approximate full ML, we report results based on  $k = 1.7$  throughout.

In Table 3, we summarize simulations evaluating point and interval estimates of  $\beta$  under 8 of the 9 conditions; the ninth scenario will be considered in detail in Table 4. The table summarizes mean  $\beta$  coefficient estimates across simulations and compares mean standard errors with the empirical (across-simulations) standard deviations of the coefficient estimates. A total of 250 runs were generated under each of the 8 conditions. Note that Scenario 2 ( $\beta = 1.10, \sigma_e^2 = 0.139$ ) mimics the actual conditions of the motivating example extremely closely. It is also important to note that we were unable to comprehensively evaluate direct full ML via our own SAS IML-based programming, or by adapting the SAS NLMIXED implementation of Messer and Natarajan (2008). This was due to a high prevalence of numerical problems in obtaining convergence for all of the “large measurement error” scenarios depicted in Table 3.

A key indication in Table 3 is the fact that the RC method displayed bias toward the null and poor confidence interval coverage in all but the “small beta, small measurement error” case (Scenario 1), and especially as  $\beta$  and  $\sigma_e^2$  increase. The bias in standard RC is noticeable, albeit small, under the conditions of the motivating example (Scenario 2). On the other hand,



the approximate full ML approach performed acceptably in most cases, with the notable exception of the two “large measurement error” settings (Scenarios 7 and 8). Most importantly, although confidence interval (CI) coverages are somewhat below nominal in the higher measurement error cases given the small validation study sample size, only the PL method maintained an acceptable point estimate of  $\beta$  under all conditions. However, PL did exhibit occasional numerical problems when obtaining the adjusted standard error estimate (see footnote to Table 3).

Table 3 also points out the strong tendency for the unadjusted standard errors accompanying RC and PL to underestimate the true standard errors, as the former fail to correct for uncertainty in the estimation of the MEM parameters. At the same time, it highlights the utility of the two practical approaches discussed in Section 3.2 for making such corrections. In particular, the proposed adjusted standard errors match the empirical standard deviation of the estimates relatively well on average. The corresponding adjusted standard errors were used to compute all CIs studied in conjunction with RC and PL.

Table 4 provides a more detailed summary based on 500 simulations for Scenario 9 (“large  $\beta$ , large measurement error”), including information regarding all parameters in model (2.1). In this simulation, we set  $n_m = 1000$  and increased  $n_v$  to 250, to produce comparisons less susceptible to fluctuations due to the validation sample size. Under these conditions, the bias in the RC estimate for  $\beta$  is extreme and its corresponding CI coverage is close to 0 %. We also see attenuation in the RC estimates and poor CI coverage for some other model (2.1) coefficients (in particular, those reflecting effects of the variables generated to mimic “plant” and “smoking status”). The approximate full ML method also produces a highly attenuated estimate for  $\beta$  (though less so than for RC), but performs well with respect to all other regression parameters. As in Table 3, we find that PL performs best overall, with little bias in any of the estimates.

## 5.2. Additive-Normal MEM

Despite our primary focus upon the multiplicative-lognormal measurement error problem dictated by models (2.1) and (2.2), it is of interest to assess the RC, PL and approximate full ML methods in the more traditional additive-normal setting. In this direction, we repeated our simulation study for some of the cases summarized in Table 3, but with  $\ln(X)$  rather than  $X$  treated as the exposure of interest in model (2.1). With regard to RC and approximate full ML, note that this change implies simplifications to expressions (3.1) and (3.4). That is, we now utilize  $\mu_{X_i} = E(X_i|Z_i) = \tau_i$  and  $\text{var}(X_i|Z_i) = \sigma_e^2$ , where  $X_i$  here denotes log-transformed exposure.

In Table 5, we summarize the results of 500 simulations under this additive-normal MEM setting, for the “moderate  $\beta$ , large  $\sigma_e^2$ ” scenario. The overall conclusions are similar to those reached in Table 4, with bias toward the null noted with respect to  $\beta$  and certain other coefficients when using the RC approach. While this attenuation is not as severe as it was in the multiplicative MEM case, it is still noteworthy. Approximate ML also displays bias of a similar magnitude (this time away from the null, producing correspondingly larger standard errors) under the assumed additive-normal MEM when estimating  $\beta$  and some other coefficients. This bias may stem from some inaccuracy in the probit approximation to the logistic function under these relatively extreme conditions. As in Table 4, the essential finding is that PL once again performed the best overall with respect to both point and CI estimation.

## 6. DISCUSSION

We have evaluated several approaches for handling a structural measurement error problem involving a logistic regression model with a single continuous exposure variable ( $X$ ) that is imperfectly measured in the main study, by means of a set of variables assumed associated with it via a second regression equation. The true value of  $X$  is observed only for subjects comprising an external validation sample. The motivating example revisits that of Weller et al. (2007), but our primary focus is upon a pseudo-likelihood (PL)-based method geared to maintain consistency for true parameters given the assumptions associated with the two underlying regression models, rather than upon more popular RC methods that lack general consistency but may work well under certain conditions. We address a multiplicative, rather than additive, measurement error problem, by applying a log transformation in model (2.2). Such a lognormal model for exposure is supported by the motivating exposure data, and may often be indicated in environmental epidemiologic studies (e.g., Rappaport 1991; Rappaport and Kupper 2008).

We are unable to recommend standard regression calibration (RC) for general use in the multiplicative error context, due to evidence of substantial bias when estimating  $\beta$  and other model (2.1) regression coefficients except under “small  $\beta$ , small measurement error” conditions. A direct full ML approach offers an attractive solution, and a SAS NLMIXED implementation of Messer and Natarajan (2008) for the additive-normal case was readily adapted to our setting. However, we found numerical stability to be a barrier under simulation conditions in which we assumed a large measurement error variance ( $\sigma_e^2$ ). A full ML approach utilizing probit-based approximations (Carroll et al. 2006; Demidenko 2004) solved this stability problem, performed much better than RC, and matched direct full ML for smaller  $\sigma_e^2$ ; however, we observed substantial bias in the approximate ML estimate of  $\beta$  for large values of  $\sigma_e^2$  (see Tables 3-4). Because both RC and approximate ML are predictably fallible in such cases (Kuha 1994; Carroll et al. 2006), and because direct full ML posed more numerical difficulties, our study favors the PL method overall. It remains fully supported by theory under the assumptions attending models (2.1) and (2.2), and was the only approach to yield an ostensibly valid estimate of  $\beta$  over the complete range of conditions studied.

Our findings differ from those of Thoresen and Laake (2000), who recommended RC over direct and approximate full ML methods based on simulation findings. However, several aspects of the conditions that they assumed differed from ours. They considered the standard additive-normal measurement error model with a single surrogate for  $X$ , used replicates (rather than validation data) to estimate the measurement error variance, and studied smaller magnitudes of the exposure effect. Nevertheless, we also found clear bias in RC estimates under some of our simulation conditions, when  $\ln(X)$  rather than  $X$  was taken to be the true exposure of interest (e.g., Table 5). Thus, as predicted by the known lack of consistency of RC under a logistic link, the deficiencies we observed with its use do not stem only from the fact that our measurement error problem is multiplicative in nature. We expect the approximate ML and/or direct full ML approaches to be generally viable and defensible in the additive-normal case, though we note that PL still outperformed approximate ML under the “large  $\sigma_e^2$ ” case considered in Table 5.

An important objective here was to assess the feasibility of likelihood-based methods for practical use, based on accessible routines for optimization, Hessian approximation, and (in the case of PL) numerical integration that are available in commercial statistical software. With a view toward accessibility, we also proposed computationally straightforward approaches designed to correct standard errors associated with RC and PL for uncertainty in

MEM parameters (Section 3.2). These proposals fared well in our simulation studies in terms of matching mean standard errors with empirical standard deviations of the estimates. The “MI-based” approach suggested for computing standard errors of RC estimates could also be used in conjunction with PL, as well as to avoid the need for sandwich estimators in other contexts (e.g., the quasi-likelihood setting of Liang and Liu 1991).

SAS IML (SAS Institute, Inc. 2004) programs are available from the authors for implementing all methods discussed in this paper. By providing this sharable resource, we hope to encourage broader use of the recommended techniques (particularly PL) for those conducting logistic regression with an error-prone exposure variable under a structural MEM.

## Acknowledgments

This research was supported by an RC4 grant through the National Institute of Nursing Research (1RC4NR012527-01), an R01 grant from the National Institute of Environmental Health Sciences (5R01ES012458-07), and Award number UL1TR000454 via the National Center for Advancing Translational Sciences of the National Institutes of Health. We thank Karen Messer for helpful correspondences, and Ellen Eisen, Edie Weller and Erika Garcia for generously sharing the data for the motivating example.

## REFERENCES

- Carroll RJ. Covariance Analysis in Generalized Linear Measurement Error Models. *Statistics in Medicine*. 1989; 8:1075–1093. [PubMed: 2678349]
- Carroll RJ, Stefanski LA. Approximate Quasilikelihood Estimation in Models with Surrogate Predictors. *Journal of the American Statistical Association*. 1990; 85:652–663.
- Carroll RJ, Spiegelman CH, Lan KKG, Bailey KT, Abbott RD. On Errors-in-Variables for Binary Regression Models. *Biometrika*. 1984; 71:19–25.
- Carroll, R.J.; Ruppert, D.; Stefanski, L.A.; Crainiceanu, C.M. *Measurement Error in Nonlinear Models: A Modern Perspective*. 2nd ed.. Chapman and Hall; London: 2006.
- 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 University Press; Oxford: 1992. p. 301-325.
- Crouch EA, Spiegelman D. The Evaluation of Integrals of the Form  $\int_{-\infty}^{\infty} f(t)\exp(-t^2)dt$ : Applications to Logistic-Normal Models. *Journal of the American Statistical Association*. 1990; 85:464–467.
- Demidenko, E. *Mixed Models: Theory and Applications*. Wiley; New York: 2004.
- Eisen EA, Holcroft CR, Greaves IA, Wegman DH, Woskie SR, Monson RR. A Strategy to Reduce Healthy Worker Effect in a Cross-Sectional Study of Asthma and Metalworking Fluids. *American Journal of Industrial Medicine*. 1997; 31:671–677. [PubMed: 9131220]
- Eisen EA, Smith TJ, Kreibel D, Woskie SR, Myers DJ, Kennedy SM, et al. Respiratory Effects of Machining Fluids: Pulmonary Function. *American Journal of Industrial Medicine*. 2001; 39:443–453. [PubMed: 11333406]
- Fuller, WA. *Measurement Error Models*. Wiley; New York: 1987.
- Gong G, Samaniego FJ. Pseudo Maximum Likelihood Estimation: Theory and Applications. *Annals of Statistics*. 1981; 9:861–869.
- Greaves IA, Eisen EA, Smith TJ, Pothier LJ, Kreibel D, Woskie SR, et al. Respiratory Health of Automobile Workers Exposed to Metal-Working Fluid Aerosols: Respiratory Symptoms. *American Journal of Industrial Medicine*. 1997; 26:621–634.
- Hwang JT. Multiplicative Errors-in-Variables Models with Applications to Recent Data Released by the U.S. Department of Energy. *Journal of the American Statistical Association*. 1986; 81:680–688.
- Iturria S, Carroll RJ, Firth D. Multiplicative Measurement Error Estimation: Estimating Equations. *Journal of the Royal Statistical Society, Series B*. 1999; 61:547–562.

- Kuha J. Corrections for Exposure Measurement Error in Logistic Regression Models with an Application to Nutritional Data. *Statistics in Medicine*. 1994; 13:1135–1148. [PubMed: 8091040]
- Liang, K-Y.; Liu, X-H. Estimating Equations in Generalized Linear Models with Measurement Error. In: Godambe, VP., editor. *Estimating Functions*. Oxford University Press; Oxford: 1991.
- Lyles RH, Kupper LL. A Detailed Evaluation of Adjustment Methods for Multiplicative Measurement Error in Linear Regression with Applications in Occupational Epidemiology. *Biometrics*. 1997; 53:1008–1025. [PubMed: 9290228]
- Messer K, Natarajan L. Maximum Likelihood, Multiple Imputation and Regression Calibration for Measurement Error Adjustment. *Statistics in Medicine*. 2008; 27:6332–6350. [PubMed: 18937275]
- Monahan, J.; Stefanski, LA. Normal Scale Mixture Approximations to  $F^*(z)$  and Computation of the Logistic-Normal Integral. In: Balakrishnan, N., editor. *Handbook of the Logistic Distribution*. Dekker; New York: 1991. p. 529-540.
- Pierce DA, Stram DO, Vaeth M, Schafer DW. The Errors-in-Variables Problem: Considerations Provided by Radiation Dose-Response Analysis of the A-Bomb Survivor Data. *Journal of the American Statistical Association*. 1992; 87:351–359.
- Rappaport SM. Assessment of Long-Term Exposures to Toxic Substances in Air. *Annals of Occupational Hygiene*. 1991; 35:61–121. [PubMed: 2035954]
- Rappaport, SM.; Kupper, LL. *Quantitative Exposure Assessment*. Lulu Press; Raleigh, NC: 2008.
- Rosner B, Willett WC, Spiegelman D. Correction of Logistic Regression Relative Risk Estimates and Confidence Intervals for Systematic Within-Person Measurement Error. *Statistics in Medicine*. 1989; 8:1051–1069. [PubMed: 2799131]
- Rosner B, Spiegelman D, Willett WC. Correction of Logistic Regression Relative Risk Estimates and Confidence Intervals for Measurement Error: The Case of Multiple Covariates Measured with Error. *American Journal of Epidemiology*. 1990; 132:734–745. [PubMed: 2403114]
- Rubin, DB. *Multiple Imputation for Nonresponse in Surveys*. Wiley; New York: 1987.
- SAS Institute, Inc.. *SAS IML 9.1 User's Guide*. SAS Institute; Cary, NC: 2004.
- Savalei V. Logistic Approximation to the Normal: The KL Rationale. *Psychometrika*. 2006; 71:763–767.
- Schafer DW. Covariate Measurement Error in Generalized Linear Models. *Biometrika*. 1987; 80:889–904.
- Schafer JL. *Multiple Imputation: A Primer*. *Statistical Methods in Medical Research*. 1999; 8:3–15. [PubMed: 10347857]
- Spall JC. Effects of Imprecisely Known Nuisance Parameters on Estimates of Primary Parameters. *Communications in Statistics: Theory and Methods*. 1989; 18:219–237.
- Spiegelman D, Carroll R, Kipnis V. Efficient Regression Calibration for Logistic Regression in Main Study/Internal Validation Study Designs with an Imperfect Reference Instrument. *Statistics in Medicine*. 2001; 10:139–160. [PubMed: 11135353]
- Stefanski, LA.; Carroll, RJ. In: Brown, PJ.; Fuller, WA., editors. *Structural Logistic Regression Measurement Error Models; Proceedings of the Conference of Measurement Error Models*; Providence, RI: American Mathematical Society. 1990;
- 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.
- Thoresen M, Laake P. A Simulation Study of Measurement Error Correction Methods in Logistic Regression. *Biometrics*. 2000; 56:868–872. [PubMed: 10985228]
- Thurston SW, Spiegelman D, Ruppert D. Equivalence of Regression Calibration Methods in Main Study/external Validation Study Designs. *Journal of Statistical Planning and Inference*. 2003; 113:527–539.
- Weller EA, Milton DK, Eisen EA, Spiegelman D. Regression Calibration for Logistic Regression with Multiple Surrogates for One Exposure. *Journal of Statistical Planning and Inference*. 2007; 137:449–461.
- Woskie SR, Smith TJ, Hallock MF, Hammond SK, Rosenthal F, Eisen EA, et al. Size-Selective Pulmonary Dose Indices for Metal-Working Fluid Aerosols in Machining and Grinding

Operations in the Automobile Manufacturing Industry. American Industrial Hygiene Association Journal. 1994; 55:20–29. [PubMed: 8116525]

**Table 1**MEM in Model (2.2) fit to External Validation Data.<sup>a,b</sup>

Variable	OLS $\hat{\gamma}$ (std. error)	Approximate full ML $\hat{\gamma}$ (std. error) <sup>c</sup>	p-value <sup>d</sup>
Intercept	-1.94 (0.17)	-1.96 (0.15)	<0.001
Grinding	0.27 (0.16)	0.22 (0.14)	0.09
Straight MWF	1.53 (0.11)	1.53 (0.10)	<0.001
Synthetic MWF	1.46 (0.14)	1.47 (0.13)	<0.001
Plant 2	-0.47 (0.18)	-0.46 (0.16)	0.01
Age (30–39)	-0.32 (0.14)	-0.30 (0.13)	0.03
Age (40–49)	-0.17 (0.17)	-0.14 (0.16)	0.33
Age (50+)	-0.11 (0.16)	-0.08 (0.15)	0.49
Caucasian Race	0.04 (0.11)	0.04 (0.10)	0.70
Current smoker	0.17 (0.09)	0.16 (0.08)	0.06

<sup>a</sup>Data from Weller et al. (2007);  $n_V = 83$  workers included in validation sample.<sup>b</sup>Residual variance ( $\sigma_e^2$ ) estimates: OLS, 0.139; Approximate ML, 0.122.<sup>c</sup>Using probit approximation for  $\Pr(Y_j = 1|Z_j)$  with  $k = 1.7$ ; see (3.3).<sup>d</sup>Standard  $t$  tests based on OLS.

**Table 2**

Corrected logistic regression results using RC, PL, and approximate full ML methods assuming multiplicative-lognormal measurement error.<sup>a,b</sup>

Variable	RC method <sup>c</sup>		Pseudo-L method <sup>d</sup>		Approximate full ML <sup>e</sup>	
	Estimate	Std. error	Estimate	Std. error	Estimate	Std. error
Intercept	-2.73	0.282 (0.290)	-2.74	0.284 (0.288)	-2.74	0.287
Exposure (X)	1.10	0.383 (0.400)	1.10	0.378 (0.401)	1.14	0.406
Plant 2	0.90	0.204 (0.215)	0.90	0.205 (0.209)	0.90	0.209
Age (30–39)	-0.02	0.193 (0.198)	-0.02	0.194 (0.198)	-0.02	0.198
Age (40–49)	-0.10	0.246 (0.257)	-0.10	0.248 (0.253)	-0.11	0.253
Age (50+)	-0.06	0.260 (0.268)	-0.06	0.262 (0.268)	-0.08	0.269
Race	0.15	0.197 (0.206)	0.15	0.198 (0.201)	0.15	0.200
Current smoker	1.03	0.162 (0.164)	1.04	0.162 (0.164)	1.04	0.164

<sup>a</sup>Data from Weller et al. (2007);  $n_M = 1,040$  main and  $n_V = 83$  validation study subjects included.

<sup>b</sup>Estimated coefficients are for true disease model (TDM) given in Equation (2.1); Measurement error model (MEM) is given in Equation (2.2).

<sup>c</sup>Unadjusted std. error followed by adjusted std. error (in parentheses). Adjusted std. error obtained via adaptation of ML-type methods ( $M = 5$  replications; see Section 3.2).

<sup>d</sup>Unadjusted std. error followed by adjusted std. error (in parentheses). Adjusted std. error obtained by numerically approximating Hessian for full likelihood after inserting OLS estimates of MEM parameters and PL estimates of model (2.1) parameters (see Section 3.2).

<sup>e</sup>Using probit approximation for  $\Pr(Y_j = 1|Z_j)$  with  $k = 1.7$ ; see (3.3). Results for approximate ML agree identically with direct full ML (see text, Section 4).

**Table 3**

Simulations to evaluate corrected estimates for  $\beta$  using RC, PL and approximate full ML methods under multiplicative-lognormal measurement error.<sup>a,b</sup>

RC: Mean estimate (Empirical SD) Mean std. error [unadjusted, adjusted] <sup>c</sup> {95 % CI coverage}	PL: Mean estimate (Empirical SD) Mean std. error [unadjusted, adjusted] <sup>d</sup> {95 % CI coverage}	Approximate full ML <sup>e</sup> : Mean estimate (Empirical SD) [Mean std. error] {95 % CI coverage}
Scenario 1: $\beta = 0.50, \sigma_e^2 = 0.139$ ("SMALL BETA, SMALL MEASUREMENT ERROR")		
0.48 (0.12) [0.11, 0.13] {92.8 %}	0.48 (0.13) [0.11, 0.12] {92.6 %}	0.49 (0.13) [0.12] {92.4 %}
Scenario 2: $\beta = 1.10, \sigma_e^2 = 0.139$ ("MODERATE BETA, SMALL MEASUREMENT ERROR")		
1.01 (0.17) [0.12, 0.17] {84.8 %}	1.10 (0.20) [0.14, 0.19] {93.5 %}	1.10 (0.20) [0.18] {92.4 %}
Scenario 3: $\beta = 1.50, \sigma_e^2 = 0.139$ ("LARGE BETA, SMALL MEASUREMENT ERROR")		
1.32 (0.20) [0.13, 0.21] {80.8 %}	1.51 (0.24) [0.18, 0.24] {96.3 %}	1.52 (0.25) [0.24] {95.2 %}
Scenario 4: $\beta = 0.50, \sigma_e^2 = 0.50$ ("SMALL BETA, MODERATE MEASUREMENT ERROR")		
0.46 (0.14) [0.09, 0.14] {85.2 %}	0.50 (0.16) [0.11, 0.15] {90.5 %}	0.50 (0.16) [0.14] {89.6 %}
Scenario 5: $\beta = 1.10, \sigma_e^2 = 0.50$ ("MODERATE BETA, MODERATE MEASUREMENT ERROR")		
0.78 (0.21) [0.09, 0.20] {52.0 %}	1.07 (0.29) [0.17, 0.27] {90.2 %}	1.08 (0.28) [0.26] {87.6 %}
Scenario 6: $\beta = 1.50, \sigma_e^2 = 0.50$ ("LARGE BETA, MODERATE MEASUREMENT ERROR")		
0.97 (0.25) [0.11, 0.22] {34.0 %}	1.45 (0.36) 1.46 [0.22, 0.34] {90.5 %}	(0.34) [0.31] {88.8 %}
Scenario 7: $\beta = 0.50, \sigma_e^2 = 1.50$ ("SMALL BETA, LARGE MEASUREMENT ERROR")		
0.28 (0.13) [0.05, 0.12] {42.0 %}	0.48 (0.21) [0.13, 0.22] {90.4 %}	0.40 (0.16) [0.16] {80.0 %}
Scenario 8: $\beta = 1.10, \sigma_e^2 = 1.50$ ("MODERATE BETA, LARGE MEASUREMENT ERROR")		
0.38 (0.18) [0.05, 0.16] {11.2 %}	1.04 (0.53) [0.23, 0.45] {83.7 %}	0.80 (0.30) [0.26] {61.2 %}

<sup>a</sup>True disease model (TDM) given in Equation (2.1); Measurement error model (MEM) given in Equation (2.2).

<sup>b</sup>250 replications in each case; Main study sample size = 1040; External validation sample size = 83; True coefficients and individual binary predictor distributions match those observed in real data example except for variations in  $\beta$  and  $\sigma_e^2$ .

<sup>c</sup>Adjusted std. error based on RC obtained using adaptation of MI-type methods ( $M = 5$ ; Section 3.2).



<sup>d</sup>Adjusted std. error based on pseudo-likelihood (PL) obtained by numerical approximation to Hessian for full likelihood after inserting OLS estimates of MEM parameters and PL estimates of model (2.1) parameters (Section 3.2); 2–7 % of runs failed to produce adjusted std. errors due to numerical problems.

<sup>e</sup>Using probit approximation for  $\Pr(Y_j = 1|Z_j)$  with  $k = 1.7$ ; see Equation (3.3).

**Table 4**

Simulation evaluating corrected estimates using RC, PL and approximate full ML methods under multiplicative-lognormal measurement error.<sup>a,b</sup>

Variable (true coefficient)	RC: Mean estimate (Empirical SD) Mean std. error: [unadjusted, adjusted] <sup>c</sup> {95 % CI coverage}	Pseudo-Likelihood: Mean estimate (Empirical SD) Mean std. error: [unadjusted, adjusted] <sup>d</sup> {95 % CI coverage}	Approximate full ML <sup>e</sup> : Mean estimate (Empirical SD) [Mean std. error] {95 % CI coverage}
	Scenario 9: $\beta = 1.50$ , $\sigma_e^2 = 1.50$ ("LARGE BETA, LARGE MEASUREMENT ERROR")		
Intercept (-2.73)	-1.93 (0.20) [0.19, 0.21] {5.0 %}	-2.75 (0.41) [0.33, 0.39] {91.3 %}	-2.58 (0.32) [0.32] {90.2 %}
Exposure ( <i>X</i> ) (1.50)	0.48 (0.13) [0.06, 0.12] {0.2 %}	1.50 (0.49) [0.31, 0.43] {90.1 %}	0.96 (0.24) [0.23] {37.0 %}
Plant 2 (0.90)	0.64 (0.16) [0.15, 0.17] {66.4 %}	0.92 (0.25) [0.21, 0.25] {94.5 %}	0.89 (0.23) [0.23] {95.2 %}
Age (30–39) (-0.02)	-0.08 (0.19) [0.17, 0.19] {93.8 %}	-0.04 (0.25) [0.20, 0.24] {94.1 %}	-0.06 (0.24) [0.23] {95.4 %}
Age (40–49) (-0.10)	-0.10 (0.18) [0.17, 0.19] {95.2 %}	-0.10 (0.24) [0.20, 0.24] {95.3 %}	-0.10 (0.23) [0.23] {96.0 %}
Age (50+) (-0.06)	-0.07 (0.17) [0.16, 0.18] {96.2 %}	-0.07 (0.23) [0.20, 0.23] {93.6 %}	-0.07 (0.22) [0.23] {94.8 %}
Race (0.15)	0.12 (0.16) [0.15, 0.17] {95.4 %}	0.15 (0.21) [0.18, 0.21] {94.5 %}	0.15 (0.21) [0.21] {94.8 %}
Current smoker (1.04)	0.83 (0.17) [0.15, 0.17] {77.8 %}	1.03 (0.24) [0.19, 0.22] {93.8 %}	1.03 (0.22) [0.21] {94.4 %}

<sup>a</sup>True disease model (TDM) given in Equation (2.1); Measurement error model (MEM) given in Equation (2.2).

<sup>b</sup>500 simulations; Main study sample size = 1000; External validation sample size = 250; True coefficients and individual binary predictor distributions match those observed in real data example.

<sup>c</sup>Adjusted std. error for  $\hat{\beta}$  based on RC obtained using adaptation of MI-type methods ( $M = 5$  replications; see Section 3.2).

<sup>d</sup>Adjusted std. error for  $\hat{\beta}$  based on pseudo-likelihood (PL) obtained by numerically approximating Hessian for full likelihood after inserting OLS estimates of MEM parameters and PL estimates of model (2.1) parameters (see Section 3.2); 23 of 500 runs (4.6 %) failed to produce adjusted std. errors due to numerical problems.

<sup>e</sup>Using probit approximation for  $\Pr(Y_j = 1|Z_j)$  with  $k = 1.7$ ; see (3.3).

**Table 5**

Simulation to evaluate corrected estimates using RC, PL and approximate full ML methods under additive-normal measurement error.<sup>a,b</sup>

Variable (true coefficient)	RC: Mean estimate (Empirical SD) Mean std. error: [unadjusted, adjusted] <sup>c</sup> {95 % CI coverage}	Pseudo-Likelihood: Mean estimate (Empirical SD) Mean std. error: [unadjusted, adjusted] <sup>d</sup> {95 % CI coverage}	Approximate full ML <sup>e</sup> : Mean estimate (Empirical SD) [Mean std. error] {95 % CI coverage}
	$\beta = 1.10, \sigma_e^2 = 1.50$ (“MODERATE BETA, LARGE MEASUREMENT ERROR”)		
Intercept (-2.73)	-2.19 (0.31) [0.27, 0.33] {63.8 %}	-2.85 (0.46) [0.40, 0.47] {96.8 %}	-3.03 (0.82) [0.78] {98.4 %}
Exposure (1.10) [ln(X)]	0.91 (0.14) [0.11, 0.14] {67.8 %}	1.16 (0.32) [0.24, 0.30] {93.6 %}	1.32 (0.56) [0.52] {97.2 %}
Plant 2 (0.90)	0.74 (0.28) [0.24, 0.29] {90.8 %}	0.93 (0.38) [0.33, 0.39] {95.6 %}	1.05 (0.51) [0.54] {96.0 %}
Age (30–39) (-0.02)	-0.02 (0.31) [0.27, 0.32] {96.2 %}	-0.02 (0.40) [0.34, 0.40] {97.0 %}	-0.01 (0.46) [0.48] {97.4 %}
Age (40–49) (-0.10)	-0.07 (0.33) [0.27, 0.32] {94.6 %}	-0.08 (0.42) [0.34, 0.40] {95.2 %}	-0.08 (0.53) [0.50] {96.2 %}
Age (50+) (-0.06)	-0.05 (0.31) [0.25, 0.31] {95.0 %}	-0.06 (0.39) [0.32, 0.38] {96.2 %}	-0.06 (0.48) [0.46] {96.4 %}
Race (0.15)	0.12 (0.28) [0.23, 0.28] {95.0 %}	0.15 (0.35) [0.29, 0.35] {96.0 %}	0.15 (0.40) [0.40] {96.2 %}
Current smoker (1.04)	0.88 (0.29) [0.23, 0.28] {89.6 %}	1.10 (0.39) [0.31, 0.37] {95.0 %}	1.23 (0.56) [0.53] {96.4 %}

<sup>a</sup>True disease model (TDM) given in Equation (2.1), except with ln(X) replacing X as the exposure variable of interest; Measurement error model (MEM) given in Equation (2.2).

<sup>b</sup>500 simulations; Main study sample size = 1000; External validation sample size = 250; True coefficients and individual binary predictor distributions match those observed in real data example.

<sup>c</sup>Adjusted std. error for  $\hat{\beta}$  based on RC obtained using adaptation of MI-type methods ( $M = 5$  replications; see Section 3.2).

<sup>d</sup>Adjusted std. error for  $\hat{\beta}$  based on pseudo-likelihood (PL) obtained by numerically approximating Hessian for full likelihood after inserting OLS estimates of MEM parameters and PL estimates of model (2.1) parameters (see Section 3.2); 1 of 500 runs (0.2%) failed to produce adjusted std. errors due to numerical problems.

<sup>e</sup>Using probit approximation for  $\Pr(Y_j = 1|Z_j)$  with  $k = 1.7$ ; see (3.3).