

On Latent-Variable Model Misspecification in Structural Measurement Error Models for Binary Response

Xianzheng Huang* and Joshua M. Tebbs

Department of Statistics, University of South Carolina, Columbia, South Carolina 29208, U.S.A.

* *email*: huang@stat.sc.edu

SUMMARY. We consider structural measurement error models for a binary response. We show that likelihood-based estimators obtained from fitting structural measurement error models with pooled binary responses can be far more robust to covariate measurement error in the presence of latent-variable model misspecification than the corresponding estimators from individual responses. Furthermore, despite the loss in information, pooling can provide improved parameter estimators in terms of mean-squared error. Based on these and other findings, we create a new diagnostic method to detect latent-variable model misspecification in structural measurement error models with individual binary response. We use simulation and data from the Framingham Heart Study to illustrate our methods.

KEY WORDS: Group testing; Latent variable; Measurement error; Pooled response; Reliability ratio; Robustness; Simulation extrapolation.

1. Introduction

Covariate measurement error is a problem commonly encountered in epidemiological investigations (Carroll, 2005). One popular study, cited by numerous authors in the measurement error literature, is the Framingham Heart Study (Kannel et al., 1986), a study where one of the primary goals is to characterize the relationship between the risk of coronary heart disease and long-term systolic blood pressure. The true predictor in this application, long-term systolic blood pressure, cannot be observed. Instead, systolic blood pressure readings are collected at periodic clinic visits for each subject and are viewed as error-contaminated versions of the true predictor.

In this article, we consider structural measurement error models for a binary response Y , e.g., whether or not evidence of heart disease is detected. In a structural measurement error model, the true predictor can be viewed as a latent variable with its own distribution. Likelihood-based inference subsequently depends on this assumed latent-variable model, and its misspecification can adversely affect inference (Carroll et al., 2006, Section 5.6.3). This problem is exacerbated by the fact that diagnosing latent-variable model misspecification with the observed data directly is not possible. To address this issue, Huang, Stefanski, and Davidian (2006) developed simulation-based remeasurement methods to diagnose latent-variable model misspecification in structural measurement error models. Their methods provide a general framework to assess the adequacy of an assumed latent-variable model and the robustness of a target estimator to measurement error.

This research expands on the study presented in Huang et al. (2006) in the context of binary response. First, we extend the methods proposed in Huang et al. (2006) to investigate the effects of latent-variable model misspecification on likelihood-based estimators in structural measurement

error models with pooled binary response. We show that using pooled binary response in place of individual binary response can result in better regression parameter estimates, and it leads to likelihood-based inference that is much less sensitive to covariate measurement error when the latent-variable model is misspecified. Second, using the pooled responses, we propose a new diagnostic method to detect latent-variable model misspecification with individual response data. This method possesses attractive properties and avoids the time-consuming remeasurement technique proposed by Huang et al. (2006).

2. Structural Measurement Error Models

We first define the structural measurement error model for individual response, as in Huang et al. (2006), and then extend its utility to pooled response.

2.1 Individual Response

For the i th individual, let Y_i , W_i , and X_i denote the binary response, the observed predictor value, and the true predictor value, respectively, for $i = 1, 2, \dots, n$. A structural measurement error model consists of three component models. The first component model is a generalized linear model for Y_i , conditional on X_i , given by $\Pr(Y_i = 1 | X_i) = h(\beta_0 + \beta_1 X_i)$, where $h(\cdot)$ is a known inverse link function. Inference on the regression parameters $\boldsymbol{\theta} = (\beta_0, \beta_1)^T$ is of central interest. The second component model is the classical measurement error model given by $W_i = X_i + U_i$, where U_i is the nondifferential measurement error (Carroll et al., 2006, Section 2.5) and $U_i \sim N(0, \sigma_U^2)$. It follows that $W_i | X_i \sim N(X_i, \sigma_U^2)$. The nondifferentiability of the measurement error implies that given X_i , Y_i is independent of W_i . The third component model is the assumed model for X , with density denoted by $f_X^{(a)}(x; \boldsymbol{\tau})$,

where $\boldsymbol{\tau}$ is a parameter vector of length t . For individual i , the joint density of the observed datum, (Y_i, W_i) , is given by

$$f_{Y,W}(Y_i, W_i; \boldsymbol{\theta}, \boldsymbol{\tau}, \sigma_U^2) = \int f_{Y|X}(Y_i | x; \boldsymbol{\theta}) f_{W|X}(W_i | x; \sigma_U^2) f_X^{(a)}(x; \boldsymbol{\tau}) dx, \quad (1)$$

where $f_{Y|X}(Y_i | x; \boldsymbol{\theta}) = h(\beta_0 + \beta_1 x)^{Y_i} \{1 - h(\beta_0 + \beta_1 x)\}^{1-Y_i}$, $f_{W|X}(W_i | x; \sigma_U^2) = \sigma_U^{-1} \phi\{\sigma_U^{-1}(W_i - x)\}$, and $\phi(\cdot)$ denotes the $N(0, 1)$ probability density function. We assume throughout that the n individuals are independent so that the log likelihood of the observed data is

$$l(\boldsymbol{\theta}, \boldsymbol{\tau}, \sigma_U^2) = \sum_{i=1}^n \log \left\{ \int f_{Y|X}(Y_i | x; \boldsymbol{\theta}) f_{W|X}(W_i | x; \sigma_U^2) f_X^{(a)}(x; \boldsymbol{\tau}) dx \right\}. \quad (2)$$

The appearance of the primary regression model $f_{Y|X}(Y_i | x; \boldsymbol{\theta})$ and the assumed latent-variable model $f_X^{(a)}(x; \boldsymbol{\tau})$ within the integrand in equation (1) suggests that the choice of the assumed model for X can affect inference for $\boldsymbol{\theta}$ based on equation (2).

2.2 Pooled Response

Suppose that G pools are formed from the n individuals, and let n_g denote the pool size for pool g , $g = 1, 2, \dots, G$, so that $\sum_{g=1}^G n_g = n$. Denote the individual binary responses in pool g by $(Y_{g1}, Y_{g2}, \dots, Y_{gn_g})^T$ and the corresponding pooled binary response by $Y_g^* = \max\{Y_{g1}, Y_{g2}, \dots, Y_{gn_g}\}$. Denote the observed and true predictor values in pool g by $\mathbf{W}_g^* = (W_{g1}, W_{g2}, \dots, W_{gn_g})^T$ and $\mathbf{X}_g^* = (X_{g1}, X_{g2}, \dots, X_{gn_g})^T$, respectively, and assume, as before, that $W_{gj} = X_{gj} + U_{gj}$, where $U_{gj} \sim N(0, \sigma_U^2)$, for all g and j . The major difference in the structural measurement error model for pooled response is in the first component model; i.e., the primary regression model. Here, the primary regression model relates the pooled response to the associated true predictors and is given by

$$\Pr(Y_g^* = 1 | \mathbf{X}_g^*) = 1 - \prod_{j=1}^{n_g} \{1 - h(\beta_0 + \beta_1 X_{gj})\}. \quad (3)$$

Our derivation shows that the joint density of the observed data for pool g , (Y_g^*, \mathbf{W}_g^*) , can be written as

$$\begin{aligned} f_{Y^*, \mathbf{W}^*}(Y_g^*, \mathbf{W}_g^*; \boldsymbol{\theta}, \boldsymbol{\tau}, \sigma_U^2) &= Y_g^* \prod_{j=1}^{n_g} \int \frac{1}{\sigma_U} \phi\left(\frac{W_{gj} - x}{\sigma_U}\right) f_X^{(a)}(x; \boldsymbol{\tau}) dx + (1 - 2Y_g^*) \\ &\times \prod_{j=1}^{n_g} \int \{1 - h(\beta_0 + \beta_1 x)\} \frac{1}{\sigma_U} \phi\left(\frac{W_{gj} - x}{\sigma_U}\right) f_X^{(a)}(x; \boldsymbol{\tau}) dx. \end{aligned}$$

Assuming that the G pools are independent, the log likelihood of the observed data based on the pooled responses is

$$l^*(\boldsymbol{\theta}, \boldsymbol{\tau}, \sigma_U^2) = \sum_{g=1}^G \log \{f_{Y^*, \mathbf{W}^*}(Y_g^*, \mathbf{W}_g^*; \boldsymbol{\theta}, \boldsymbol{\tau}, \sigma_U^2)\}. \quad (4)$$

Unless otherwise stated, we assume that σ_U^2 is known. In practice, σ_U^2 can be estimated when there are replicate surrogate measurements for each X , and it is straightforward to

revise our approach (see Section 6). For either data structure, individual or pooled, it is worth emphasizing that if the first two component models in the structural measurement error model are correct, likelihood-based inference is consistent if the latent-variable model is correctly specified or if $\sigma_U^2 = 0$. In the absence of measurement error, Vansteelandt, Goetghebuer, and Verstraeten (2000) considered regression models of the form in equation (3) for modeling the prevalence of HIV using pooled responses from group testing. In such applications, forming pools is a natural and commonly used technique to reduce testing costs. Our use of pooling in this article is more general and is not restricted to applications in group testing.

3. Empirical Methods for Assessing Robustness

Huang et al. (2006) use remeasurement-based methods to assess the robustness of target estimators to measurement error in structural measurement error models. We summarize the salient aspects of this approach with models for individual binary response and then extend this technique to models for pooled response.

3.1 Individual Response

The remeasurement method involves further contaminating the observed covariate W . We call the further contaminated data the λ -remeasured data, where λ is a prespecified positive constant that controls the degree of further contamination. Specifically, for a chosen $\lambda > 0$, we first generate B sets of n independent random errors from a standard normal distribution, $\{Z_{bi}, i = 1, 2, \dots, n\}_{b=1}^B$, and form the λ -remeasured data defined by

$$W_{bi}(\lambda) = W_i + \sqrt{\lambda} \sigma_U Z_{bi}, \quad (5)$$

for $b = 1, 2, \dots, B$ and $i = 1, 2, \dots, n$. Note that, by this construction, the measurement error variance associated with $W_{bi}(\lambda)$ is equal to $(1 + \lambda)\sigma_U^2$ for all b and i .

Let $\boldsymbol{\Omega} = (\boldsymbol{\theta}^T, \boldsymbol{\tau}^T)^T$ denote the $r \times 1$ vector of unknown parameters, where $r = 2 + t$, and suppose that $\boldsymbol{\Omega}$ is to be estimated by solving the vector-valued estimating equation

$$\sum_{i=1}^n \boldsymbol{\psi}(Y_i, W_i; \boldsymbol{\theta}, \boldsymbol{\tau}, \sigma_U^2) = \mathbf{0}, \quad (6)$$

in the absence of further contamination. For example, if the maximum likelihood estimator (MLE) of $\boldsymbol{\Omega}$ is desired, then $\sum_{i=1}^n \boldsymbol{\psi}(Y_i, W_i; \boldsymbol{\theta}, \boldsymbol{\tau}, \sigma_U^2) = (\partial/\partial \boldsymbol{\Omega}) l(\boldsymbol{\theta}, \boldsymbol{\tau}, \sigma_U^2)$. In general, denote the estimator that solves equation (6) by $\hat{\boldsymbol{\Omega}}(0) = \{\hat{\boldsymbol{\theta}}(0)^T, \hat{\boldsymbol{\tau}}(0)^T\}^T$. Based on equation (6), we construct the vector-valued estimating equation evaluated at the λ -remeasured data as

$$\sum_{i=1}^n \boldsymbol{\psi}_B \{Y_i, \tilde{W}_i(\lambda); \boldsymbol{\theta}, \boldsymbol{\tau}, (1 + \lambda)\sigma_U^2\} = \mathbf{0}, \quad (7)$$

where $\boldsymbol{\psi}_B \{Y_i, \tilde{W}_i(\lambda); \boldsymbol{\theta}, \boldsymbol{\tau}, (1 + \lambda)\sigma_U^2\} = B^{-1} \sum_{b=1}^B \boldsymbol{\psi} \{Y_i, W_{bi}(\lambda); \boldsymbol{\theta}, \boldsymbol{\tau}, (1 + \lambda)\sigma_U^2\}$ and $\tilde{W}_i(\lambda) = \{W_{bi}(\lambda)\}_{b=1}^B$. Solving equation (7) yields the same type of estimator as $\hat{\boldsymbol{\Omega}}(0)$ based on the λ -remeasured data; we denote this estimator by $\hat{\boldsymbol{\Omega}}(\lambda) = \{\hat{\boldsymbol{\theta}}(\lambda)^T, \hat{\boldsymbol{\tau}}(\lambda)^T\}^T$. Plotting one component of $\hat{\boldsymbol{\Omega}}(\lambda)$ versus λ , for $\lambda \geq 0$, produces the simulation extrapolation, SIMEX, plot for that estimate (Cook and Stefanski, 1994; Stefanski and

Cook, 1995). A constant or nearly constant SIMEX plot suggests that the considered estimator is robust to measurement error under the assumed model for X . A substantial deviation from constancy indicates that the estimator is sensitive to measurement error under the presumed model for X and that the estimator is biased, with the magnitude of bias depending on the size of measurement error variance.

In addition to the graphical SIMEX diagnostic, Huang et al. (2006) provide a quantitative assessment of robustness in the form of the test statistic

$$t_1^*(\lambda) = \hat{\nu}_1^{-1}\{\hat{\gamma}(\lambda) - \hat{\gamma}(0)\},$$

where γ is an element in Ω , $\hat{\gamma}(\cdot)$ is the target estimator for γ , and $\hat{\nu}_1^2$ is an estimator for $\text{var}\{\hat{\gamma}(\lambda) - \hat{\gamma}(0)\}$. The derivation of an estimator for $\text{var}\{\hat{\gamma}(\lambda) - \hat{\gamma}(0)\}$ is provided in Web Appendix A. The statistic $t_1^*(\lambda)$ evaluates the amount of change in the estimate as the measurement error variance increases from σ_U^2 to $(1 + \lambda)\sigma_U^2$ after adjusting for background noise. If $t_1^*(\lambda)$ deviates significantly from zero, one may conclude that the estimator is not robust to measurement error under the assumed latent-variable model. The operating characteristics of $t_1^*(\lambda)$ are studied in Huang et al. (2006).

There is a delicate difference between the remeasurement method described above and the method described in Huang et al. (2006) in computing $\hat{\Omega}(\lambda)$ at $\lambda > 0$. In Huang et al. (2006), when $\lambda > 0$, Ω is estimated B times, each time based on one set of λ -remeasured data of size n , and the final $\hat{\Omega}(\lambda)$ is defined as the average of these B estimates. The modification here is to compute $\hat{\Omega}(\lambda)$ only once based on all B sets of λ -remeasured data, which are combined in one vector-valued estimating equation (7). The solution to equation (7) is asymptotically equivalent to the older version of $\hat{\Omega}(\lambda)$, but is far less cumbersome to compute.

3.2 Pooled Response

Generating the remeasured data in Section 3.1 does not involve the binary response Y . Therefore, the same data-generation step applies when pooled responses are considered. Suppose that the target estimator for Ω , based on the data with pooled response, solves the vector-valued estimating equation

$$\sum_{g=1}^G \psi^*(Y_g^*, \mathbf{W}_g^*; \theta, \tau, \sigma_U^2) = \mathbf{0}. \tag{8}$$

Analogously to the construction of equation (7), the estimator for Ω of the same type, based on the λ -remeasured data with pooled response, is obtained by solving

$$\sum_{g=1}^G \psi_B^*\{Y_g^*, \tilde{\mathbf{W}}_g^*(\lambda); \theta, \tau, (1 + \lambda)\sigma_U^2\} = \mathbf{0}, \tag{9}$$

where $\psi_B^*\{Y_g^*, \tilde{\mathbf{W}}_g^*(\lambda); \theta, \tau, (1 + \lambda)\sigma_U^2\} = B^{-1} \sum_{b=1}^B \psi^*\{Y_g^*, \mathbf{W}_{bg}^*(\lambda); \theta, \tau, (1 + \lambda)\sigma_U^2\}$, $\mathbf{W}_{bg}^*(\lambda) = \{W_{bgj}^*(\lambda)\}_{j=1}^{n_g}$, for $b = 1, 2, \dots, B$, and $\tilde{\mathbf{W}}_g^*(\lambda) = \{\mathbf{W}_{bg}^*(\lambda)\}_{b=1}^B$, for $g = 1, 2, \dots, G$. Denote by $\hat{\Omega}^*(0) = \{\hat{\theta}^*(0)^T, \hat{\tau}^*(0)^T\}^T$ and $\hat{\Omega}^*(\lambda) = \{\hat{\theta}^*(\lambda)^T, \hat{\tau}^*(\lambda)^T\}^T$ the solutions to equations (8) and (9), respectively. As with individual response, the SIMEX plot of one component of $\hat{\Omega}^*(\lambda)$ can provide evidence of robustness or lack thereof. Furthermore, assuming independent pools, the construction of $t_1^*(\lambda)$ based on the solutions to equations (8) and (9) is

identical to that based on individual response. As described in Web Appendix A, the variance estimator $\hat{\nu}_1^2$ in this case is changed to reflect the fact that pools are the experimental units; not the individuals.

4. Robustness Findings

We now apply the previously outlined diagnostic methods to assess the robustness of target estimators resulting from data with different types of binary response.

4.1 Random Versus Homogeneous Pooling

In studying the structural measurement error models with pooled responses, we consider two pooling strategies. The first strategy is to form the pools randomly, so that the composition of pools is independent of the individual covariate information. We henceforth refer to this as random pooling. The second pooling strategy we consider is motivated by Vansteelandt et al. (2000), who showed that, in the absence of covariate measurement error, a more precise estimator of $\text{Pr}(Y_g^* = 1 | \mathbf{X}_g^*)$ results when one pools individuals with similarly observed covariates, that is, when $X_{g1}, X_{g2}, \dots, X_{gn_g}$ are as homogeneous as possible. We refer to this as homogeneous pooling. In the presence of measurement error, the best homogeneous composition one can hope for is to pool individuals with similar W values. To implement this strategy, one can sort and partition the ordered W 's into G pools, so that the datum for the first pool, say, is $(Y_1^*, W_{11}, W_{12}, \dots, W_{1n_1}) = (Y_1^*, W_{(1)}, W_{(2)}, \dots, W_{(n_1)})$, where $W_{(1)}, W_{(2)}, \dots, W_{(n_1)}$ are the first n_1 order statistics of W_1, W_2, \dots, W_n . When generating λ -remeasured data, regardless of pool composition, we always contaminate the unsorted observed data W_1, W_2, \dots, W_n with the unsorted random noise $\{Z_{bi}, i = 1, 2, \dots, n\}_{b=1}^B$ to obtain $\{W_{bi}(\lambda), i = 1, 2, \dots, n\}_{b=1}^B$.

A complication with homogeneous pooling is that it renders a small amount of dependence among the pools due to the ordering of the observed covariates. Therefore, the MLE of Ω based on the homogeneous-pooling responses is not obtained by maximizing equation (4) as equation (4) is derived under the assumption of independent pools. In fact, we have found that maximizing the true log likelihood from homogeneous pooling is practically infeasible. To estimate Ω in this situation, a sensible solution is to first estimate τ by maximizing the log likelihood of $\mathbf{W} = (W_1, W_2, \dots, W_n)^T$ given by

$$l_W(\tau, \sigma_U^2) = \sum_{i=1}^n \log \left\{ \int f_{W|X}(W_i | x; \sigma_U^2) f_X^{(a)}(x; \tau) dx \right\}.$$

This provides a consistent estimator for τ because the ordering does not enter at this level. We then estimate θ by maximizing equation (4) with τ fixed at its estimate. This simplifies the objective function to be optimized by assuming pools to be independent and yields a pseudo-MLE for θ , which is biased, but much less so when compared to the estimator for θ resulting from maximizing equation (4) with respect to τ and θ simultaneously. The vector-valued estimating equation evaluated at the observed data corresponding to this two-stage approach is

$$\sum_{g=1}^G \psi^*(Y_g^*, \mathbf{W}_g^*; \theta, \tau, \sigma_U^2) = \begin{pmatrix} (\partial/\partial\tau)l_W(\tau, \sigma_U^2) \\ (\partial/\partial\theta)l^*(\theta, \tau, \sigma_U^2) \end{pmatrix} = \mathbf{0}.$$

Another small complication with homogeneous pooling relates to the variance estimator for the pseudo-MLE of Ω and the variance estimator $\hat{\nu}_1^2$ in $t_1^*(\lambda)$. Both estimators are initially derived under the assumption of independent experimental units, which, in the case of pooled response, corresponds to independent pools. For the reasons previously outlined, these estimators are not appropriate for use with homogeneous pooling. To obviate this difficulty, we use the bootstrap to estimate these variances when homogeneous pooling is used.

4.2 Simulation Evidence

We conducted several simulations to evaluate the merit of pooling in structural measurement error models. For illustration, we take $h(\beta_0 + \beta_1 X) = \Phi(\beta_0 + \beta_1 X)$, where $\Phi(\cdot)$ is the $N(0, 1)$ distribution function, $\theta = (\beta_0, \beta_1)^T = (-2, 1)^T$, and the true model for X to be the two-component mixture $f_X^{(a)}(x) = 0.1f_1(x) + 0.9f_2(x)$, where $f_1(x)$ and $f_2(x)$ denote the $N(2.35, 0.41)$ and $N(-0.26, 0.38)$ density functions, respectively. The two normal components are chosen to produce a bimodal distribution with mean 0, variance 1, skewness 1.3, and kurtosis 2. A random sample of size $n = 2000$ is generated from the structural measurement error model for individual data defined in Section 2.1 with $\sigma_U^2 = 0.25$, so that the reliability ratio $\sigma_X^2 / (\sigma_X^2 + \sigma_U^2) = 0.8$ (Carroll et al. 2006,

Section 3.2.1). For both pooling protocols, we set $n_g = 10$, for $g = 1, 2, \dots, G$, yielding $G = 200$ pools. Thus, in the results that follow, the MLEs from the data with individual responses are based on 2000 responses whereas the estimates from the data with pooled response are based on only 200. The assumed models for X we consider are (a) a normal distribution, representing the situation wherein one misspecifies the latent-variable model and (b) a two-component mixture normal distribution, representing the situation in which one serendipitously chooses the correct parametric model. The estimators resulting from (b) serve as gold standards with which the estimators from (a) are compared. In the remeasurement method, we set $B = 50$ and take $\lambda = 1$.

For $\lambda \geq 0$, denote by $\hat{\theta}_I^{(n)}(\lambda)$, $\hat{\theta}_R^{(n)}(\lambda)$, and $\hat{\theta}_H^{(n)}(\lambda)$ the MLEs for θ from the data with individual response, the data with random-pooling response, and the pseudo-MLE for θ from data with homogeneous-pooling response, respectively, when the assumed model for X is normal, and by $\hat{\theta}_I^{(m)}(\lambda)$, $\hat{\theta}_R^{(m)}(\lambda)$, and $\hat{\theta}_H^{(m)}(\lambda)$ the corresponding estimates when the assumed model for X is mixture normal. From the SIMEX plots in Figure 1, it is evident that $\hat{\theta}_H^{(n)}(\lambda)$ is more robust to measurement error than either $\hat{\theta}_I^{(n)}(\lambda)$ or $\hat{\theta}_R^{(n)}(\lambda)$, despite the fact

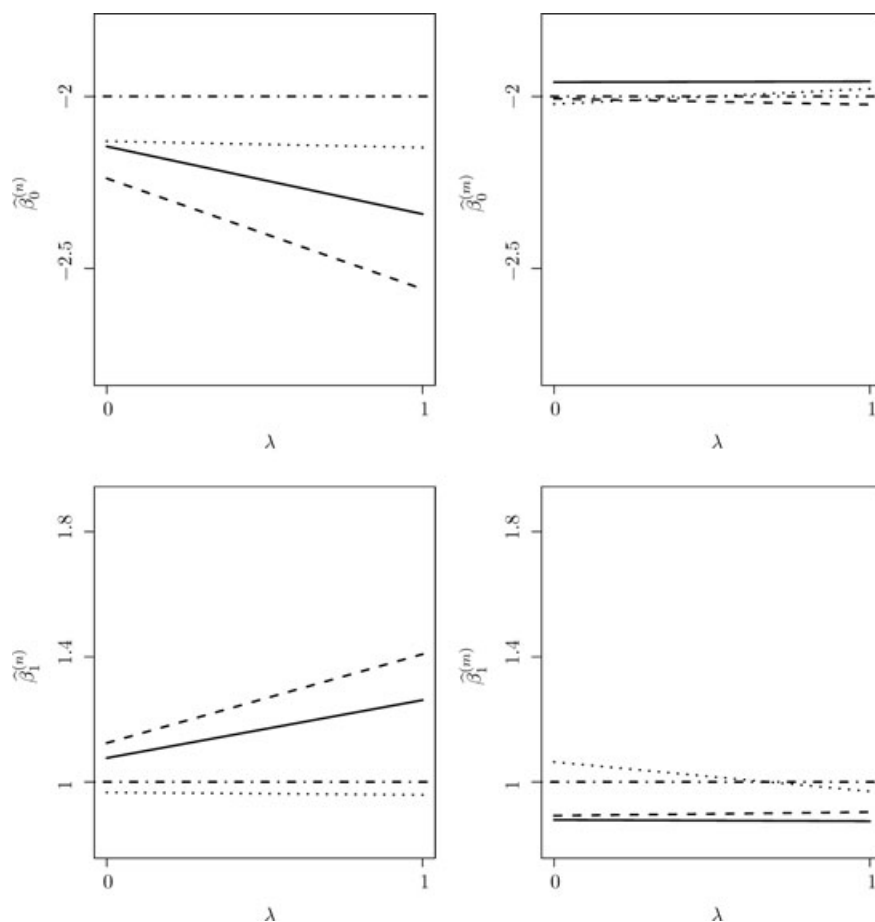


Figure 1. SIMEX plots for regression parameter estimates. The horizontal dot-dashed line is a reference line placed at the true parameter values. The left (right) panel depicts the estimates under a normal (mixture normal) assumption for X . The solid line, dashed line, and dotted line correspond to $\hat{\theta}_I$, $\hat{\theta}_R$, and $\hat{\theta}_H$, respectively.

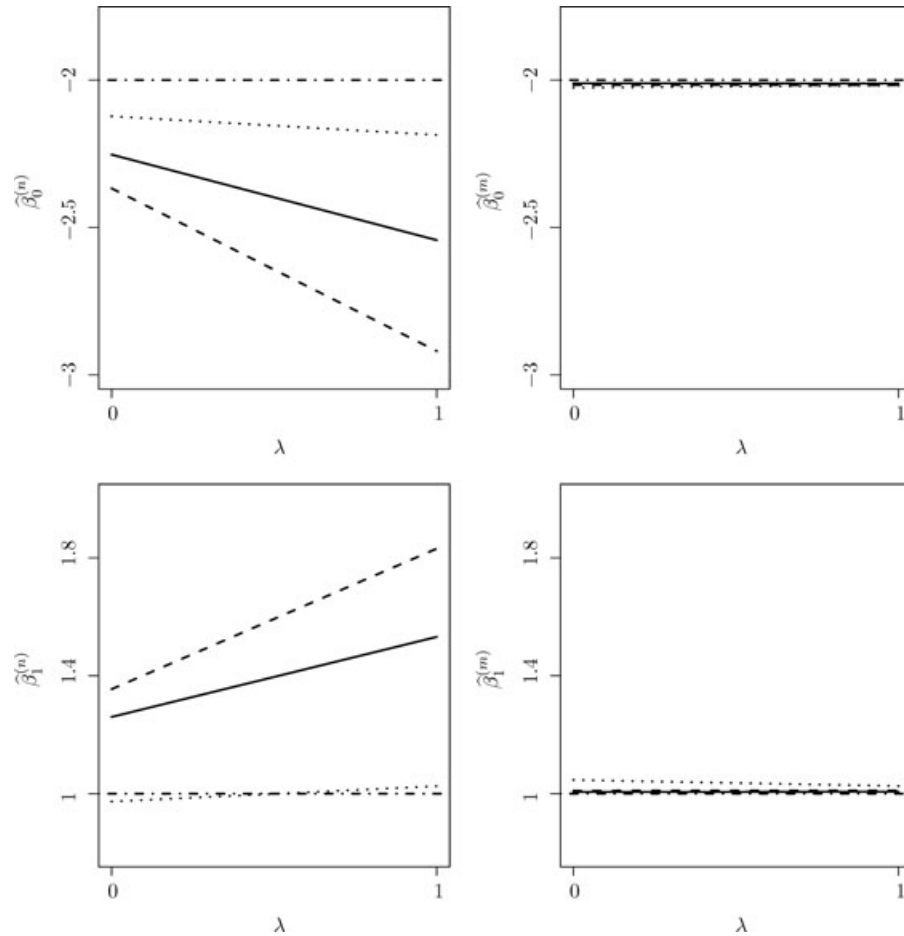


Figure 2. SIMEX plots averaged over 300 Monte Carlo replications. The horizontal dot-dashed line is a reference line placed at the true parameter values. The left (right) panel depicts the estimates under a normal (mixture normal) assumption for X . The solid line, dashed line, and dotted line correspond to $\hat{\theta}_I$, $\hat{\theta}_R$, and $\hat{\theta}_H$, respectively.

that the latent-variable model is misspecified and $\hat{\theta}_H^{(n)}(\lambda)$ is a pseudo-MLE. For this simulation, all three estimates for θ under correct modeling do not appear to be sensitive to measurement error. To provide an overall assessment, we repeated the same simulation 300 times to observe the average pattern exhibited by the SIMEX plots in the setting described above. The averages of the 300 sets of $\hat{\theta}_I^{(n)}(\lambda)$, $\hat{\theta}_R^{(n)}(\lambda)$, and $\hat{\theta}_H^{(n)}(\lambda)$, as well as $\hat{\theta}_I^{(m)}(\lambda)$, $\hat{\theta}_R^{(m)}(\lambda)$, and $\hat{\theta}_H^{(m)}(\lambda)$, are plotted in Figure 2, and the values of these averages for $\lambda = 0, 1$ are given in Table 1. The observations from Figure 2 and Table 1 reconcile the patterns observed from the one Monte Carlo replicate depicted in Figure 1.

4.3 Interpretation

Our overall simulation results indicate that, in the presence of latent-model misspecification, estimates computed from data with homogeneous-pooling response can be more robust to measurement error than those from data with individual response or with random-pooling response. An interesting related observation is given in Weinberg and Umbach (1999), who used pooled exposure assessments to improve the efficiency in case-control studies. Unlike our use of pooling,

Table 1
Mean regression parameter estimates from 300 Monte Carlo replications under the normal and mixture normal latent-variable assumption with $\beta_0 = -2$, $\beta_1 = 1$, and $\lambda = 0, 1$. Monte Carlo mean-squared errors are in parentheses. IND, RP, and HP represent individual response, random-pooling response, and homogeneous-pooling response, respectively.

Normal	$\hat{\beta}_0^{(n)}(0)$	$\hat{\beta}_0^{(n)}(1)$	$\hat{\beta}_1^{(n)}(0)$	$\hat{\beta}_1^{(n)}(1)$
IND	-2.25 (0.064)	-2.54 (0.295)	1.26 (0.068)	1.53 (0.283)
RP	-2.37 (0.135)	-2.92 (0.846)	1.35 (0.126)	1.83 (0.691)
HP	-2.12 (0.015)	-2.19 (0.035)	0.97 (0.001)	1.03 (0.001)
Mixture normal	$\hat{\beta}_0^{(m)}(0)$	$\hat{\beta}_0^{(m)}(1)$	$\hat{\beta}_1^{(m)}(0)$	$\hat{\beta}_1^{(m)}(1)$
IND	-2.01 (<0.001)	-2.01 (<0.001)	1.01 (<0.001)	1.01 (<0.001)
RP	-2.02 (<0.001)	-2.02 (<0.001)	1.01 (<0.001)	1.01 (<0.001)
HP	-2.03 (0.001)	-2.02 (<0.001)	1.05 (0.002)	1.03 (0.001)

these authors pool the exposure covariates instead. When the exposure assessment is obtained from an error-prone assay, the authors show that the attenuation effect of measurement error on the naive MLEs (that is, estimators which ignore

measurement error) can be magnified when the pooled exposure assessments are used in place of the individual exposure assessments. The authors reached this conclusion by showing that the reliability ratio of the pooled exposure data is lower than that of the individual exposure data.

Our use of pooling, which can lead to more robust estimators, involves the responses based on the sorted observed predictor values and is thus structurally different from the approach taken by Weinberg and Umbach (1999). However, the gains in robustness from homogeneous pooling can also be explained by the change in the reliability ratio. When we form pooled responses based on homogeneous covariates, the variation in the true predictor among the pools is usually larger than the variation in the true predictor based on the individuals. This is true so long as the reliability ratio of the individual data is not too small, that is, as long as the process of sorting the W data also approximately sorts the X s. Higher between-pool variance of the true predictor yields a higher reliability ratio for the pooled data. In this light, homogeneous pooling can alleviate the effect of measurement error.

Our findings are especially encouraging when group testing is used for cost concerns, because one may be able to realize large gains in robustness to measurement error in the presence of latent-model misspecification at a small fraction of the testing cost. One might naturally conjecture that using pooled responses in place of individual responses would cause a notable loss in precision when estimating θ . However, we have found that this may not be true. Inference based on the pooled responses can actually be more efficient than inference based on the individual responses when the between-pool variance of X is large and the pool sizes n_g are not so large that the pools conceal too much information. In fact, in our presented simulation study, with $n_g = 10$, we found that the mean-squared errors associated with $\hat{\theta}_I^{(n)}$ are generally much higher than the mean-squared errors associated with $\hat{\theta}_H^{(n)}$ (see Table 1). The reduction in bias due to measurement error in $\hat{\theta}_H^{(n)}$ can again be explained by the increase in the variance of the true predictor among the experimental units after forming homogeneous pools, which in turn increases the reliability ratio.

5. New Diagnostic Method

The statistic $t_1^*(\lambda)$ proposed by Huang et al. (2006) is designed to test the hypothesis $H_{01} : \gamma(0) = \gamma(\lambda)$, where γ is one of the elements in Ω , and where $\gamma(0)$ and $\gamma(\lambda)$ are the probability limits of the target estimators for γ based on the observed data and the λ -remeasured data, respectively. In the context herein, it is understood to mean that asymptotic results apply when the number of experimental units grows large without bound; that is, n for individual-response data and G for pooled-response data. Rejecting H_{01} implies that the considered estimator for γ is not robust to measurement error. As nonrobustness can be a consequence of misspecifying the latent-variable model, rejection of H_{01} can suggest latent-variable model misspecification. Because the model parameters are re-estimated based on the noisier data sets in the remeasurement method, computing $t_1^*(\lambda)$ can be time consuming. We propose a new test statistic for diagnosing model

misspecification that does not require generating remeasured data or re-estimation based on noisier data.

The creation of our new statistic is motivated by the fact that when the latent-variable model is correctly specified, $\hat{\theta}_I$ and $\hat{\theta}_R$ are in close agreement, but that they can differ largely when the latent-variable model is misspecified (see Figures 1 and 2). This occurs because if the latent-variable model is correctly specified, so that the likelihood of the data with individual response and with random-pooling response are both correct, $\hat{\theta}_I$ and $\hat{\theta}_R$ are both consistent estimators of θ . However, close agreement between $\hat{\theta}_I$ and $\hat{\theta}_R$ is not guaranteed when the latent-variable model is misspecified.

5.1 Test Statistic

For $G > r$, the new test statistic is defined by

$$t_2^* = \frac{G - r}{r(G - 1)} \times \{\hat{\Omega}_R(0) - \hat{\Omega}_I(0)\}^T \hat{\Sigma}^{-1} \{\hat{\Omega}_R(0) - \hat{\Omega}_I(0)\},$$

where $\hat{\Omega}_{(\cdot)}(0)$ is the MLE of $\Omega = (\theta^T, \tau^T)^T$, subscripts I and R denote individual response and random-pooling response, respectively, and $\hat{\Sigma}$ is an estimator of the variance-covariance matrix of $\hat{\Omega}_R(0) - \hat{\Omega}_I(0)$, derived in Web Appendix B. Also shown in Web Appendix B, $(G - r)^{-1}r(G - 1)t_2^*$ is a Hotelling's T^2 -type statistic; thus, t_2^* follows an $F(r, G - r)$ distribution under H_{02} . Structurally different from $t_1^*(\lambda)$, the statistic t_2^* tests the null hypothesis $H_{02} : \Omega_R(0) = \Omega_I(0)$, where $\Omega_R(0)$ and $\Omega_I(0)$ are the probability limits of $\hat{\Omega}_R(0)$ and $\hat{\Omega}_I(0)$, respectively. It is worth noting that t_2^* does not depend on the remeasured data, so it is far easier to compute than $t_1^*(\lambda)$.

5.2 Simulation Evidence

To examine the performance of the test statistics, we first computed $t_1^*(1)$ and t_2^* using the simulated data from Figure 1 with $n = 2000$, $G = 200$, and $n_g = 10$. Table 2 displays the results. The variance estimate $\hat{\nu}_1^2$, in the case of homogeneous pooling, is computed using 100 bootstrap resamples. Using 1.96 as a large-sample critical value, the values of $t_1^*(1)$ in Table 2 reinforce the visual findings revealed in Figure 1. That

Table 2

Values of $t_1^*(1)$ and t_2^* associated with the regression parameter estimates from Figure 1 under the normal and mixture normal assumption. IND, RP, and HP represent individual response, random-pooling response, and homogeneous-pooling response, respectively. The numbers in parentheses in the rows for t_2^* are the p -values associated with t_2^* .

	$\hat{\beta}_0^{(n)}$			$\hat{\beta}_1^{(n)}$		
	IND	RP	HP	IND	RP	HP
Normal						
$t_1^*(1)$	-4.18	-1.92	-0.34	4.91	2.27	-0.08
t_2^*	4.49 (0.002)	—	—	—	—	—
	$\hat{\beta}_0^{(m)}$			$\hat{\beta}_1^{(m)}$		
	IND	RP	HP	IND	RP	HP
Mixture normal						
$t_1^*(1)$	0.35	-0.72	1.11	-0.83	0.71	-0.68
t_2^*	0.37 (0.92)	—	—	—	—	—

Table 3

Rejection rates of $t_1^*(1)$ and t_2^* based on 300 Monte Carlo replications. Monte Carlo standard errors are in parentheses. IND, RP, and HP represent individual response, random-pooling response, and homogeneous-pooling response, respectively.

Normal	$\hat{\beta}_0^{(n)}$			$\hat{\beta}_1^{(n)}$		
	IND	RP	HP	IND	RP	HP
$t_1^*(1)$	1 (0)	0.45 (0.03)	0.16 (0.02)	1 (0)	0.74 (0.03)	0.12 (0.02)
t_2^*	0.87 (0.02)	—	—	—	—	—
Mixture normal	$\hat{\beta}_0^{(m)}$			$\hat{\beta}_1^{(m)}$		
	IND	RP	HP	IND	RP	HP
$t_1^*(1)$	0.04 (0.01)	0.01 (0.01)	0.02 (0.01)	0.04 (0.01)	0.02 (0.01)	0.07 (0.01)
t_2^*	0.03 (0.01)	—	—	—	—	—

is, when the model for X is misspecified as normal, $\hat{\theta}_H$ is more robust to measurement error than $\hat{\theta}_I$ and $\hat{\theta}_R$. Furthermore, when the model for X is correctly specified, none of the three values of $t_1^*(1)$ provide sufficient evidence to conclude that the estimates are sensitive to measurement error or that the latent-variable model has been misspecified. Similarly, when the latent-variable model is misspecified as normal, the small probability values associated with t_2^* , computed with respect to the $F(4, 196)$ reference distribution, provide strong evidence of model misspecification. When the latent-variable model is correct, t_2^* does not indicate significant disagreement between $\hat{\Omega}_R(0)$ and $\hat{\Omega}_I(0)$, as seen through the large probability values, computed with respect to the $F(7, 193)$ distribution.

To explore the power and size properties of the two statistics, we repeated the same simulation using 300 Monte Carlo replicates and recorded the proportion of times that H_{01} and H_{02} are rejected. Table 3 summarizes the results. Under correct modeling, our findings suggest that both test statistics have nearly nominal size characteristics. In addition, the quantile–quantile plot of the 300 values of t_2^* (not shown) strongly supports the limiting F distribution when the latent-variable model is correct. When the assumed latent-variable model is incorrect, the power of $t_1^*(1)$ under homogeneous pooling is low. This finding reinforces our robustness discovery.

Under incorrect modeling, although high, the power of t_2^* is slightly lower than that of $t_1^*(1)$ based on the individual responses. This is not totally unexpected, given the fact that $t_1^*(\lambda)$ and t_2^* detect latent-model misspecification in very different ways. When the latent-variable model is misspecified in a way that compromises likelihood-based inference, $t_1^*(\lambda)$ reveals nonrobustness of a single target estimator $\hat{\gamma}$ to measurement error as the noise level increases for the same data type, whereas t_2^* detects the discrepancy between estimators of Ω resulting from different data types at a fixed level of noise. When the analyst has misspecified the true latent-variable model, $\hat{\Omega}_R(0)$ and $\hat{\Omega}_I(0)$ may be affected similarly, making misspecification difficult to detect (see Section 6). However, we have found that when n and G are large enough to produce reasonably precise estimates, the power of t_2^* is often above 80% when the pool size $n_g \geq 4$, making its use attractive

because of computational ease. In practice, we recommend choosing G so that $\hat{\text{se}}(\hat{\beta}_0)/\hat{\beta}_0$ and $\hat{\text{se}}(\hat{\beta}_1)/\hat{\beta}_1$ are both small. Here, $\hat{\text{se}}(\cdot)$ denotes the estimated standard error, which we compute using the sandwich formula.

6. Application to Framingham Heart Study

We now apply the remeasurement method and the new testing procedure with data from the Framingham Heart Study. Our data set consists of 1615 male subjects who are followed for the development of coronary heart disease over six examination periods. At each of the second and third examination periods, each subject's systolic blood pressure is measured twice during two clinic visits. Additionally, the first evidence of coronary heart disease within the 8-year follow-up period from the second examination period is recorded for each subject. Define Y as the binary indicator of the first evidence of coronary heart disease within this follow-up period and the explanatory variable X as the long-term systolic blood pressure. The true predictor X is unobserved, and the two systolic blood pressure readings collected during the clinic visits can be viewed as error-contaminated versions of X .

In our analysis, we do not view X as time dependent, and we take the average of the two systolic blood pressure readings measured in the second examination period as the value of the observed predictor, denoted by W . We use the average of the two systolic blood pressure readings measured in the third examination period as a replicate measurement of X so that we can estimate the measurement error variance σ_U^2 , as in Carroll et al. (2006, Section 5.4.2.1). Because σ_U^2 is estimated, we stack an additional estimating equation associated with σ_U^2 on top of the estimating equations in (6)–(9); in addition, σ_U in equation (5) is replaced with its estimate $\hat{\sigma}_U$. To relate Y to X , we posit the probit model, $\Pr(Y = 1 | X) = \Phi(\beta_0 + \beta_1 X)$. For the assumed latent model for X , we choose a normal distribution and a two-component mixture normal distribution, assumptions under which we compute the statistics $t_1^*(1)$ and t_2^* . For illustration, we set $n_g = 5$, for $g = 1, 2, \dots, 323$. For the remeasurement method, we take $B = 50$ and $\lambda = 1$.

Values of $t_1^*(1)$ and t_2^* for the Framingham data are listed in Table 4. Using these values, one cannot find sufficient evidence of misspecification when the assumed latent-variable model is mixture normal. Under the normal model assumption, the

Table 4

Framingham data. Values of $t_1^*(1)$ and t_2^* computed under the normal and mixture normal assumption. IND, RP, and HP represent individual response, random-pooling response, and homogeneous-pooling response, respectively. The numbers in parentheses in the rows for t_2^* are the p -values associated with t_2^* .

Normal	$\hat{\beta}_0^{(n)}$			$\hat{\beta}_1^{(n)}$		
	IND	RP	HP	IND	RP	HP
$t_1^*(1)$	-2.87	-1.88	-1.67	2.92	1.93	1.51
t_2^*	0.21 (0.93)	—	—	—	—	—
Mixture normal	$\hat{\beta}_0^{(m)}$			$\hat{\beta}_1^{(m)}$		
	IND	RP	HP	IND	RP	HP
$t_1^*(1)$	0.46	0.80	-1.44	-0.45	-0.79	1.31
t_2^*	1.51 (0.16)	—	—	—	—	—

statistic $t_1^*(1)$ based on individual response and t_2^* suggest different conclusions, but this is not necessarily contradictory for the reasons noted in Section 5.2. At the current contamination level of the raw data ($\lambda = 0$), $\hat{\theta}_R$ and $\hat{\theta}_I$ are affected similarly by measurement error under the normal assumption. More importantly, our methods reveal a novel finding for these data; namely, if one prefers a simpler assumed latent-variable model, such as the normal, the pseudo-MLE based on homogeneous-pooling responses is less sensitive to measurement error when compared to the MLE based on individual responses.

7. Discussion

When compared to individual binary response, we have shown that likelihood-based inference in structural measurement error models for pooled binary response can be more robust to covariate measurement error when the latent-variable model is misspecified. Moreover, pooling can improve efficiency in estimation. Pooled binary responses do arise naturally in group testing (Gastwirth and Johnson, 1994; Vansteelandt et al., 2000), and our findings suggest that latent-model misspecification may be less of a concern in this setting, so long as pools are composed homogeneously. However, as we have illustrated herein, our use of pooling is not restricted to group testing. Furthermore, including multiple covariates would not pose methodological challenges beyond those seen here. If, in addition to the mismeasured X , the analyst also uses V , measured without error, then the issue of model misspecification is only relevant to the model of X given V , and all of our techniques apply with $f_X^{(a)}(x; \tau)$ being replaced by an assumed conditional model. If the primary regression model depends on a multivariate error-prone covariate \mathbf{X} , homogeneous pools can be formed by sorting the observed \mathbf{W} data by the least noisiest covariate first. The rationale for this is to achieve the largest possible between-pool variance of \mathbf{X} , noting that sorting the observed covariate associated with a heavily error-

prone predictor may not change the between-pool variance of this predictor significantly.

We have also proposed a new test statistic t_2^* for diagnosing model misspecification. The rationale for using $t_1^*(\lambda)$ and t_2^* as indicators of model misspecification are different, yet, they are united by a common, almost-counterintuitive theme. The construction of both statistics involves some form of information reduction, that is, one either adds more noise to the observed W to compute $t_1^*(\lambda)$, or one conceals the individual responses to compute t_2^* . In the context of structural measurement error models for binary response, the fact that one can learn more from the data by reducing information is intriguing. In fact, we feel that this general idea could prove to be useful when examining model selection in a broader context.

It is important to remember that robustness does not guarantee consistency, yet, robustness is still a desirable property for an estimator to possess when measurement error exists. We do not suggest that estimates from pooled analyses should blindly replace those from the individual data. In fact, because the pseudo-MLE is biased, we recommend using the MLE based on the original individual responses when there is insufficient evidence of latent-variable model misspecification. However, if there is a genuine uncertainty about the form of the true latent-variable model and one desires an estimator that is less sensitive to measurement error, the pseudo-MLE based on homogeneous pooling is preferred.

8. Supplementary Materials

The Web Appendices referenced in Sections 3.1, 3.2, and 5.1 are available under the Paper Information link at the *Biometrics* website <http://www.biometrics.tibs.org>.

ACKNOWLEDGEMENT

We are grateful to the associate editor and an anonymous referee for their helpful comments. We also thank Dr Leonard A. Stefanski for very stimulating discussions. The second author's research was supported by NIH grant R01 AI067373.

REFERENCES

- Carroll, R. (2005). Measurement error in epidemiologic studies. *Encyclopedia of Biostatistics*, 2nd edition. Chichester, U.K.
- Carroll, R., Ruppert, D., Stefanski, L., and Crainiceanu, C. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. Boca Raton, Florida: Chapman & Hall/CRC.
- Cook, J. and Stefanski, L. (1994). Simulation extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association* **89**, 1314–1328.
- Gastwirth, J. and Johnson, W. (1994). Screening with cost-effective quality control: Potential applications to HIV and drug testing. *Journal of the American Statistical Association* **89**, 972–981.
- Huang, X., Stefanski, L., and Davidian, M. (2006). Latent-model robustness in structural measurement error models. *Biometrika* **93**, 53–64.
- Kannel, W., Neaton, J., Wentworth, D., Thomas, H., Stamler, J., Hulley, S., and Kjelsberg, M. (1986). Overall and coronary heart disease mortality rates in relation to major risk factors in 325,348 men screened for MRFIT. *American Heart Journal* **112**, 825–836.

- Stefanski, L. and Cook, J. (1995). Simulation extrapolation: The measurement error jackknife. *Journal of the American Statistical Association* **90**, 1247–1256.
- Vansteelandt, S., Goetghebeur, E., and Verstraeten, T. (2000). Regression models for disease prevalence with diagnostic tests on pools of serum samples. *Biometrics* **56**, 1126–1133.
- Weinberg, C. and Umbach, D. (1999). Using pooled exposure assessment to improve efficiency in case-control studies. *Biometrics* **55**, 718–726.

*Received January 2008. Revised May 2008.
Accepted June 2008.*