STATISTICS IN MEDICINE
Statist. Med. 2009; **28**:3316–3327
Published online 18 August 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.3698

An improved test of latent-variable model misspecification in structural measurement error models for group testing data

Xianzheng Huang*, †

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

SUMMARY

We consider structural measurement error models for group testing data. Likelihood inference based on structural measurement error models requires one to specify a model for the latent true predictors. Inappropriate specification of this model can lead to erroneous inference. We propose a new method tailored to detect latent-variable model misspecification in structural measurement error models for group testing data. Compared with the existing diagnostic methods developed for the same purpose, our method shows vast improvement in the power to detect latent-variable model misspecification in group testing design. We illustrate the implementation and performance of the proposed method via simulation and application to a real data example. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS: group testing; latent variable; remeasurement method; structural measurement error models

1. INTRODUCTION

Group testing, or pooled testing, is best known for its application in infectious disease screening [1]. In group testing design, the test of the presence of a disease is conducted on a pool of individuals instead of each individual. Assuming perfect test (without testing error), a positive test can only result if at least one individual in the pool is infected. When the disease is rare, group testing is more cost-effective than individual testing. This, along with many other benefits, such as lowering false positive rate in low prevalence settings [2, 3] and ensuring confidentiality of individuals, makes group testing a compelling alternative to individual testing in infectious disease screening, genetics, systems biology, and drug discovery [4–6].

Historically, research on group testing has focused on prevalence estimation, optimal design of group testing to reduce cost and improve efficiency, and retesting strategies to identify infected individuals. Most of these work assumes homogeneous population. About a decade ago researchers

*Correspondence to: Xianzheng Huang, Department of Statistics, University of South Carolina, Columbia, SC 29208, U.S.A.
†E-mail: huang@stat.sc.edu

Received 29 January 2009
Accepted 13 July 2009

started to consider heterogeneous population. Vansteelandt *et al.* [7, 8] justified the necessity of accounting for relevant individual covariate information in making inference about human immunodeficiency virus (HIV) prevalence based on pooled serum samples. They conducted regression analysis on group testing data with individual covariates as predictors. Xie [9] also tackled a regression model for group testing data with individual covariates and developed an EM algorithm to estimate the regression parameters. As pointed out by these authors, it is common in the practice of group testing that information about individual characteristics is collected besides the pooled response, and it is sensible to use these information to improve efficiency of statistical inference.

However, as in many regression applications, individual covariates may be measured with error in group testing design. In the group testing literature, measurement error is often referred to testing error that causes misclassification of pooled response. Measurement error in the predictors, which is our focus here, has not been studied in the context of group testing. To focus on the issue of measurement error in covariates, we assume perfect test in this article. Relaxing this assumption imposes little technical difficulty on our proposed method. It is well documented that ignoring measurement error and using the contaminated covariates as if they were the true latent predictors can bias inference. A popular strategy to account for measurement error in regression models is to use structural measurement error models [10]. In the framework of structural measurement error models, a measurement error model is used to relate the observed covariates to the latent predictors, and a model for the latent variables is also assumed. A general concern of this framework is the choice of latent-variable model, which often lacks solid support from the observed data. Inappropriate specification of the latent-variable model can lead to erroneous inference [11].

To avoid stringent model assumptions on the latent variables, Carroll *et al.* [12] proposed nonparametric approaches based on the idea of simulation-extrapolation (SIMEX) [13] and regression spline; Schafer [14] and Rabe-Hesketh *et al.* [15] considered nonparametric maximum likelihood estimation where the latent-variable distribution is estimated by a discrete distribution with nonzero probabilities at a finite set of points. The discrete nature the estimated latent-variable distribution is unattractive when the distributional characteristics of true predictors are of interest. In line with parametric modeling, Carroll *et al.* [16] used mixtures of a fixed number of normals as the assumed latent-variable model. They admitted that the resulting inference might not be reliable when the true distribution cannot be approximated well by normal mixtures or when the number of normal components is large. Similar flexible parametric modeling was also adopted by Richardson *et al.* [17] in a Bayesian framework, where the number of normal components in the mixtures is left unknown. They pointed out that their approach is inadvisable in the absence of validation data. These existing nonparametric approaches and flexible parametric approaches usually entail complicated algorithms and heavy computation that limit their popularity. The complexity and computational burden are expected to increase for group testing data as the likelihood function is usually less tractable than that of individual testing data. Hence, diagnostic technique to guard against model misspecification is especially valuable because, with an effective diagnostic tool, one may find a parsimonious parametric model based on which reliable and efficient inference can be achieved without invoking nonparametric methods or unnecessarily complex parametric methods.

Thus far, there has been limited study on assessing latent-variable model specification in structural measurement error models. Without providing implementation details or theoretical justification, Aitkin and Rocci [18] commented on using the bootstrap to compare the deviance resulting from a normal assumed latent-variable model with the deviance resulting from nonparametric maximum likelihood inference. They also outlined a comparison between the K-masspoint model

DOI: 10.1002/sim

with Gaussian quadrature model with K masspoints via the likelihood ratio test. Huang *et al.* [11] proposed a remeasurement method based on the idea of SIMEX. The remeasurement method is applicable to responses from either individual testing or group testing. Despite its general applicability and satisfactory performance in individual testing design, the remeasurement method can suffer very low power to detect model misspecification in group testing design, as evidenced in later sections. This is not unexpected considering the loss in information due to pooling. Huang and Tebbs [19] constructed another test motivated by the finding that the likelihood inference based on individual testing data and that from the induced group testing data can distinctly disagree in the presence of model misspecification. Their test relies on individual testing data thus is not directly applicable in the current setting where only group testing data are observed.

In this article, we combine the idea of the remeasurement method and the idea behind the diagnostic test in [19] to construct a more powerful diagnostic tool that does not require individual testing data to be observed. The proposed method is elaborated in Section 3. Before presenting the new diagnostic method, we define the structural measurement error models for the individual testing data and group testing data in Section 2. The remeasurement method is also reviewed in Section 2, which is used as the benchmark method that the new method is compared with. In Section 4, we present simulation studies to illustrate the implementation and performance of the new method. In Section 5, we apply the new method and the remeasurement method to a real data. We conclude the article with discussions on the proposed method and future research in Section 6.

2. STRUCTURAL MEASUREMENT ERROR MODELS AND REMEASUREMENT METHOD

2.1. Structural measurement error models

Denote by Y_{ij} the binary response of the jth individual in the ith pool, by X_{ij} the values of the associated unobservable predictors, and W_{ij} the corresponding observed covariate values, for $i=1,\ldots,m$, and $j=1,\ldots,n_i$. Define by $N=\sum_{i=1}^{m} n_i$ the total number of individuals in the study. Define $Y_i=(Y_{i1},\ldots,Y_{in_i})^{\mathrm{T}}$, and similarly define X_i and W_i. In individual testing, the observed data include $\{Y_{ij}, W_{ij}, i=1,\ldots,m, j=1,\ldots,n_i\}$. In group testing, the observed data are $\{Y_i^*, W_i, i=1,\ldots,m\}$, where $Y_i^*=\max_{1\leqslant j\leqslant n_i} Y_{ij}$, $i=1,\ldots,m$. In the regression model, there may exist observable predictors measured without error, say, V. In that case, our concern becomes the assumed model of X conditional on V. For notational simplicity, we do not include V in our example models. Because generalization from univariate covariate to multivariate covariate is straightforward in our study, we assume univariate latent variable X, and thus univariate W. Finally, random pooling, as opposed to homogeneous pooling [8], is assumed when forming the pools, as often the case in practice.

Formulating a structural measurement error model entails specification of the following three component models: (i) the primary regression model that relates the response and the true predictor; (ii) the measurement error model that links the observed covariate to the true predictor; and (iii) the assumed model for the true predictor. Denote by $f_X(x;\tau)$ the assumed model in (iii) indexed by parameters τ. For the second component model, we consider the classical additive measurement error model ([10], Section 1.2),

$$W_{ij}=X_{ij}+U_{ij}, \quad i=1,2,\ldots,m, \quad j=1,\ldots,n_i \tag{1}$$

DOI: 10.1002/sim

where $U_{ij} \sim N(0, \sigma^2)$ are the independent nondifferential measurement error ([10], Section 2.5). An implication of nondifferential measurement error is $U_{ij} \perp X_{ij}$, and thus $Y_{ij} \perp W_{ij}$ conditioning on X_{ij}, $i=1,\ldots,m,\ j=1,\ldots,n_i$. The structural measurement error model for the individual testing data and that for the group testing data have the same second and third component models. But they differ in the first component model. Suppose that the primary regression model for the individual testing response is a generalized linear model given by

$$P(Y_{ij}=1|X_{ij};\beta)=h(\beta_0+\beta_1 X_{ij}), \quad i=1,2,\ldots,m,\ j=1,\ldots,n_i \tag{2}$$

where $h(.)$ is a known inverse link function and $\beta=(\beta_0,\beta_1)^{\mathrm{T}}$ are the regression parameters. By the definition of Y_i^*, the primary regression model for the group testing response can be induced from (2) and is given by

$$P(Y_i^*=1|X_i;\beta)=1-\prod_{j=1}^{n_i}\{1-h(\beta_0+\beta_1 X_{ij})\} \tag{3}$$

We assume throughout the article that the first two component models are correctly specified.

With the three component models specified, the joint density of the observed datum in individual testing, (Y_{ij}, W_{ij}), is given by, for $i=1,\ldots,m,\ j=1,\ldots,n_i$,

$$f_{Y,W}(Y_{ij},W_{ij};\beta,\tau,\sigma^2)=\int f_{Y|X}(Y_{ij}|X_{ij};\beta)f_{W|X}(W_{ij}|X_{ij};\sigma^2)f_X(x;\tau)\,\mathrm{d}x \tag{4}$$

where $f_{Y|X}(Y_{ij}|X_{ij};\beta)=h(\beta_0+\beta_1 X_{ij})^{Y_{ij}}\{1-h(\beta_0+\beta_1 X_{ij})\}^{1-Y_{ij}}$, $f_{W|X}(W_{ij}|X_{ij};\sigma^2)=\sigma^{-1}\phi\{\sigma^{-1}(W_{ij}-X_{ij})\}$, and $\phi(.)$ denotes the standard normal probability density function. The joint density of the observed datum in group testing, (Y_i^*, W_i), can be shown to be, for $i=1,\ldots,m$,

$$\begin{aligned} f_{Y^*,W}(Y_i^*,W_i;\beta,\tau,\sigma^2) = {} & Y_i^*\prod_{j=1}^{n_i}\int\frac{1}{\sigma}\phi\left(\frac{W_{ij}-x}{\sigma}\right)f_X(x;\tau)\,\mathrm{d}x+(1-2Y_i^*) \\ & \times\prod_{j=1}^{n_i}\int\{1-h(\beta_0+\beta_1 x)\}\frac{1}{\sigma}\phi\left(\frac{W_{ij}-x}{\sigma}\right)f_X(x;\tau)\,\mathrm{d}x \end{aligned} \tag{5}$$

The dependence between the primary regression model and the assumed latent-variable model within the integrands in (4) and (5) suggests that the choice of the assumed model for X can affect the likelihood-based inference for β.

For simplicity, we assume herein that σ^2 is known. In practice, the measurement error variance σ^2 can be easily estimated when there are replicate measures for each value of the true predictor ([10], Section 4.4.2). The diagnostic methods discussed in this article still apply with minor modification when σ^2 is estimated. Let $\Omega=(\beta^{\mathrm{T}},\tau^{\mathrm{T}})^{\mathrm{T}}$ denote the vector of all unknown parameters. The central interest of inference lies in β, and in particular, the maximum likelihood estimator (MLE), $\hat{\beta}$. It is worth pointing out that the likelihood-based inference is consistent if the latent-variable model is correctly specified or if $\sigma^2=0$.

2.2. The remeasurement method

The remeasurement method [11, 20] exploits the potential dependence of $\hat{\beta}$ on σ^2 implicitly indicated by (4) and (5). More specifically, when the model for the observed data is correct, $\hat{\beta}$ is

DOI: 10.1002/sim

consistent for β regardless of the magnitude of σ^2, whereas in the presence of model misspecification, $\hat{\beta}$ can be inconsistent with bias depending on σ^2. To reveal this potential dependence, one generates the so-called λ-remeasured data defined by, for a chosen $\lambda>0$,

$$W_{b,ij} = W_{ij} + \sqrt{\lambda}\sigma Z_{b,ij} \tag{6}$$

where $Z_{b,ij}$ are independent standard normal random errors, for $b=1,\ldots,B$, $i=1,\ldots,m$, $j=1,\ldots,n_i$. The definition in (6) is in accordance with the assumption of classical additive measurement error model, which suggests that, conditional on X_i, $W_{b,i}(\lambda)=\{W_{b,ij}(\lambda), j=1,\ldots,n_i\}$ follows the same distribution as $W_i=\{W_{ij}, j=1,\ldots,n_i\}$, except that the measurement error variance of the former is equal to $(1+\lambda)\sigma^2$. Using the remeasured data along with the original binary responses, Ω is estimated by maximizing the likelihood function, in which σ^2 is replaced by $(1+\lambda)\sigma^2$. Denote the resulting MLE for Ω by $\hat{\Omega}(\lambda)$ and similarly denote by $\hat{\beta}(\lambda)$ the MLE for β based on the λ-remeasured data. Varying the value of λ over a range, $[0, \lambda_{\max}]$, and repeatedly estimating β yields a realization of a random function $\hat{\beta}(\lambda)$ for $\lambda\in[0, \lambda_{\max}]$. When the model is correct, $\hat{\beta}(\lambda)$ is expected to be nearly a constant function, but it can exhibit a nonconstant pattern in the presence of model misspecification.

Following the above idea, Huang *et al.* [11, 20] defined a statistic to examine the change of $\hat{\beta}(\lambda)$ as λ changes. Let γ be one of the parameters in Ω. Huang *et al.* [11, 20] proposed a t test to detect model misspecification via the following t statistic, $t=\{\hat{\gamma}(\lambda)-\hat{\gamma}(0)\}\hat{v}^{-1}$, where $\hat{\gamma}(0)$ denotes the MLE computed from the observed data and $\hat{v}$ is an estimator for the standard error of $\hat{\gamma}(\lambda)-\hat{\gamma}(0)$. A value of t that significantly differs from zero provides evidence of model misspecification. The formula and derivation of $\hat{v}$ are provided in [20].

3. AN IMPROVED DIAGNOSTIC TEST

3.1. The combined evidence test

Empirical evidence suggests that, when the sample size (N or m) is small or when the error contamination is substantial, the performance of the remeasurement method can be severely degraded. We now propose a new test that combines two sources of model-misspecification evidence to raise the power of detecting model misspecification when only group testing data are observed. In the sequel, we call the new test the *combined evidence test* (CET). The rationale of CET is as follows.

In the remeasurement method, the t statistic measures the discrepancy between $\hat{\beta}(0)$ and $\hat{\beta}(\lambda)$ that results from the interaction of model misspecification and the increase in measurement error variance (from σ^2 to $(1+\lambda)\sigma^2$). On the other hand, Huang and Tebbs [19] showed that, in the presence of model misspecification, MLE for β based on the individual testing data, denoted by $\hat{\beta}^{(I)}$, can differ significantly from its counterpart estimate based on the induced group testing data, denoted by $\hat{\beta}^{(G)}$. They proposed a test for $H_0:\tilde{\beta}^{(I)}=\tilde{\beta}^{(G)}$, where $\hat{\beta}^{(I)} \xrightarrow{p} \tilde{\beta}^{(I)}$ as $N\to\infty$, $\hat{\beta}^{(G)} \xrightarrow{p} \tilde{\beta}^{(G)}$, as $m\to\infty$ at the rate of $O(N)$, and '$\xrightarrow{p}$' stands for 'converge in probability.' Evidence sufficient enough to reject H_0 can also be evidence of model misspecification. Heuristically, similar to the implicit dependence of $\hat{\beta}$ on σ^2 suggested by (4), (5) indicates dependence of $\tilde{\beta}^{(G)}$ on the group size n_i (unless all models are correct). It is important that, in the asymptotics for group testing design, one assumes $m\to\infty$ while $\max_{1\leqslant i\leqslant m} n_i$ bounded so that the intrinsic dependence

DOI: 10.1002/sim

of $\hat{\beta}^{(G)}$ on n_i is retained in limit. Individual testing data can be viewed as a special case of group testing data with $n_i=1$. Now we have another indicator of model misspecification based on the discrepancy between $\hat{\beta}^{(I)}$ and $\hat{\beta}^{(G)}$, which results from the interaction of model misspecification and the increase in group size (from 1 to n_i). The above two types of discrepancy indicators are associated with different consequences of model misspecification. Interestingly, they are both products of some form of information reduction. With N fixed, increasing the measurement error variance conceals data information, and so does increasing the group size. The new test statistic in CET, defined next, combines these two distinct evidences in the hope of a more powerful test.

3.2. Algorithm for the new test

Define the new test statistic by, for an unknown parameter γ and a prespecified constant $\lambda>0$,

$$C_\gamma(\lambda)=\{\hat{\gamma}^{(G)}(\lambda)-\hat{\gamma}^{(G)}(0)\}^2+\{\hat{\gamma}^{(G)}(0)-\hat{\gamma}^{(I)}(0)\}^2 \tag{7}$$

where $\hat{\gamma}^{(G)}(0)$ denotes the MLE for γ based on the observed group testing data, $\hat{\gamma}^{(G)}(\lambda)$ is the MLE based on the λ-remeasured group testing data, and $\hat{\gamma}^{(I)}(0)$ is the MLE computed from the pseudo-individual testing data simulated based on the observed group testing data. It is clear from Section 2.2 how to obtain the first two estimators in (7). To obtain $\hat{\gamma}^{(I)}(0)$, we simulate the pseudo-individual testing data as follows.

We first estimate X_{ij} by its best linear predictor ([10], Section A.4.1) given by

$$\hat{X}_{ij}=\mu_x+\frac{\sigma_x^2}{\sigma_x^2+\sigma^2}(W_{ij}-\mu_x),\quad i=1,\ldots,m,\ \ j=1,\ldots,n_i \tag{8}$$

where μ_x and σ_x^2 are the mean and variance of the true predictor X, respectively. Note that the derivation of the best linear predictor for X_{ij} does not depend on the specification of $f_X(x;\tau)$. Hence, the estimate $\hat{X}_{ij}$ in (8) is the same for all candidate assumed models for X. Carroll *et al.* [21] justified the use of best linear estimator for X when the disease is rare in structural measurement error models. Because μ_x and σ_x^2 are unknown parameters, we estimate them by the method of moments and use $\hat{\mu}_x=\overline{W}=N^{-1}\sum_{i=1}^m\sum_{j=1}^{n_i}W_{ij}$ and $\hat{\sigma}_x^2=\hat{\sigma}_w^2-\sigma^2$ in place of μ_x and σ_x^2 in (8), respectively, where $\hat{\sigma}_w^2$ is the sample variance of W. Then we generate the pseudo-individual testing data of size N according to (1) and (2), where β is evaluated at $\hat{\beta}^{(G)}(0)$, $\{X_{ij},j=1,\ldots,n_i\}_{i=1}^m$ are replaced by the best linear predictors, and $\{U_{ij},j=1,\ldots,n_i\}_{i=1}^m$ are generated from $N(0,\sigma^2)$. To reduce Monte Carlo variation, we repeat the data generation procedure B^* times, and stack these $N\times B^*$ realizations in one big pseudo-individual testing data set, based on which the MLE of Ω is computed. The resulting estimate is denoted by $\hat{\Omega}^{(I)}(0)$, and $\hat{\gamma}^{(I)}(0)$ is one of the elements.

In order to use $C_\gamma(\lambda)$ for testing, the critical points associated with its null distribution are needed. Here, the null hypothesis is $H_0:\tilde{\gamma}^{(I)}(0)=\tilde{\gamma}^{(G)}(0)=\tilde{\gamma}^{(G)}(\lambda)$. Correct model specification or $\sigma^2=0$ necessarily implies H_0. A large value of $C_\gamma(\lambda)$ casts doubt on the veracity of H_0. Noting the sum-of-squares form of $C_\gamma(\lambda)$, one may propose to approximate the null distribution of $C_\gamma(\lambda)$ via a scaled χ^2, with the scale constant and the degrees of freedom estimated by matching the first two moments. But the complex correlation of $\hat{\gamma}^{(G)}(0)$ and $\hat{\gamma}^{(I)}(0)$ hinders explicit derivation of the second moment of $C_\gamma(\lambda)$. Hence we resort to parametric bootstrap to obtain the approximated critical points of the null distribution, or equivalently, to obtain an empirical p-value of the

DOI: 10.1002/sim

test. The algorithm of CET, including computing $C_\gamma(\lambda)$ and obtaining an empirical p-value, is described next.

Step 1: Compute the MLE for Ω based on the observed group testing data, $\{Y_i^*, W_i, i=1,\ldots,m\}$. The resulting MLE is $\hat{\Omega}^{(G)}(0)$.

Step 2: Generate B sets of remeasured data for a chosen $\lambda>0$, $\{Y_i^*, W_{b,i}(\lambda), i=1,\ldots,m\}_{b=1}^{B}$. Compute the MLE for Ω based on them. This gives $\hat{\Omega}^{(G)}(\lambda)$.

Step 3: Compute $\{\hat{X}_{ij}, i=1,\ldots,m, j=1,\ldots,n_i\}$ according to (8).

Step 4: Use $\hat{\beta}^{(G)}(0)$ and $\hat{X}_{ij}$ in place of β and X_{ij}, respectively, to generate B^* sets of pseudo-individual testing data according to (1) and (2), each of size N. Denote this collection of pseudo-individual testing data by $\{\hat{Y}_{b^*,ij}, \hat{W}_{b^*,ij}, i=1,\ldots,m, j=1,\ldots,n_i\}_{b^*=1}^{B^*}$. Compute the MLE for Ω, denoted by $\hat{\Omega}^{(I)}(0)$, using the pseudo-individual testing data of size $N\times B^*$.

Step 5: Compute the CET test statistic $C_\gamma(\lambda)$ in (7) for each element of Ω.

Step 6: For each $q=1,\ldots,Q$,

- Simulate X from the estimated assumed model $f_X\{x;\hat{\tau}^{(G)}(0)\}$. Denote the simulated X by $\{X_{q,ij}, i=1,\ldots,m, j=1,\ldots,n_i\}$.
- Generate W according to (1) with X_{ij} replaced by $X_{q,ij}$ and U_{ij} generated from $N(0,\sigma^2)$. Denote these W by $\{W_{q,ij}, i=1,\ldots,m, j=1,\ldots,n_i\}=\{W_{q,i}, i=1,\ldots,m\}$.
- Generate Y according to (2) with β evaluated at $\hat{\beta}^{(G)}(0)$ and X_{ij} evaluated at $X_{q,ij}$, resulting in $\{Y_{q,ij}, i=1,\ldots,m, j=1,\ldots,n_i\}$.
- Finally, form the group testing responses after random pooling, $\{Y_{q,i}^*, i=1,\ldots,m\}$.

The above data generation procedure produces the qth parametric bootstrap sample of group testing data, $\{Y_{q,i}^*, W_{q,i}, i=1,\ldots,m\}$, for $q=1,\ldots,Q$.

Step 7: For the qth bootstrap sample, $q=1,\ldots,Q$, repeat steps 1 through 5 treating $\{Y_{q,i}^*, W_{q,i}, i=1,\ldots,m\}$ as the 'observed' group testing data. In the end, one collects Q realizations of CET statistic, denoted by $\{C_{\gamma,q}(\lambda)\}_{q=1}^{Q}$, for each element of Ω. The empirical p-value for a test associated with γ is defined by the proportion of $\{C_{\gamma,q}(\lambda)\}_{q=1}^{Q}$ that exceed $C_\gamma(\lambda)$.

The statistic $C_\gamma(\lambda)$ in (7) can be easily revised to define a test based on the entire vector of MLE for Ω as follows,

$$C_\Omega(\lambda)=\{\hat{\Omega}^{(G)}(\lambda)-\hat{\Omega}^{(G)}(0)\}^{\mathrm{T}}\{\hat{\Omega}^{(G)}(\lambda)-\hat{\Omega}^{(G)}(0)\}+\{\hat{\Omega}^{(G)}(0)-\hat{\Omega}^{(I)}(0)\}^{\mathrm{T}}\{\hat{\Omega}^{(G)}(\lambda)-\hat{\Omega}^{(G)}(0)\}$$

which can serve as an omnibus test of model misspecification when one is not interested in the effect on each estimator of certain model assumption.

Implementation of this algorithm is straightforward. With some price in computation time, we show in the next section that, CET can realize great gain in power compared with the remeasurement method.

4. SIMULATION STUDY

We now investigate the performance of CET in comparison with the remeasurement method via simulation. In the simulation, we apply the remeasurement method to the simulated individual

DOI: 10.1002/sim

testing data and the induced group testing data in turn to show the decrease in power when group testing data are used, and also to show that CET can achieve much higher power.

The individual testing data are generated according to (1) and (2), where, in (2), we use the probit link. The integrals in (4) and (5) can be derived explicitly in this case. We choose the true latent-variable model to be a mixture normal, $0.1N(2.35, 0.41)+0.9N(-0.26, 0.38)$, resulting in $\mu_x=0$ and $\sigma_x^2=1$. The true values of the other parameters are $\beta_0=-2$, $\beta_1=1$, and $\sigma^2=0.25$ and 0.33. The current choices of σ_x^2 and σ^2 result in two levels of reliability ratio ([10], Section 3.2.1), $r=\sigma_x^2/(\sigma_x^2+\sigma^2)$, which are 0.8 and 0.75. We set $N=500$ and 800, with $n_i=5$, for $i=1,\ldots,m$, leading to $m=100$ and 160, respectively. In the remeasurement method and CET algorithm, we let $\lambda=1$, $B=B^*=50$, $Q=200$, and use 200 Monte Carlo replicates for each scenario. It is recommended in the SIMEX literature that $\lambda\in[1,\lambda_{\max}]$ with $\lambda_{\max}\leqslant 3$. Because large λ gives very noisy data that can make numerical procedure unstable, we start with small λ such as $\lambda=1$. The operating characteristics of the remeasurement method and CET are fairly stable over a range of λ where the numerical optimization in obtaining MLE is not too problematic.

For each of the four scenarios with different levels of N and r, we specify two assumed latent-variable models, a normal and a two-component mixture normal. The first assumed model leads to a case with model misspecification, allowing one to study the power of the tests; the second assumed model produces a scenario of correct model specification, allowing one to study the size of the tests. The MLEs of β when $r=0.8$ are presented in Table I. The rejection rates of the t statistic, $C_\gamma(1)$, and $C_\Omega(1)$ are given in Table II.

Results in Table I reinforce that MLEs for the primary regression parameters can be severely biased when the model for X is misspecified. And the nonrobustness of MLEs is evident as more noise is added to W. Also the estimates from the individual testing data tend to disagree with their counterpart estimates based on the group testing data. The bias does not diminish when group testing data are used as opposed to the individual testing data. However, when the group testing data are used, the rejection rates of t are much lower than the same test applied to the

Table I. Monte Carlo averages of MLEs for the regression parameters computed from the individual testing data and the group testing data across 200 Monte Carlo replicates.

	$N=500$		$N=800$	
	Normal	Mixture normal	Normal	Mixture normal
$\hat{\beta}_0^{(I)}(0)$	−2.29 (0.021)	−2.02 (0.013)	−2.29 (0.016)	−2.03 (0.010)
$\hat{\beta}_1^{(I)}(0)$	1.30 (0.018)	1.02 (0.010)	1.30 (0.013)	1.02 (0.007)
$\hat{\beta}_0^{(I)}(1)$	−2.65 (0.042)	−2.03 (0.014)	−2.63 (0.026)	−2.03 (0.010)
$\hat{\beta}_1^{(I)}(1)$	1.63 (0.034)	1.03 (0.011)	1.60 (0.021)	1.02 (0.007)
$\hat{\beta}_0^{(G)}(0)$	−2.43 (0.032)	−2.04 (0.018)	−2.44 (0.025)	−2.06 (0.015)
$\hat{\beta}_1^{(G)}(0)$	1.42 (0.026)	1.04 (0.013)	1.41 (0.019)	1.04 (0.010)
$\hat{\beta}_0^{(G)}(1)$	−3.10 (0.082)	−2.03 (0.014)	−3.14 (0.089)	−2.06 (0.015)
$\hat{\beta}_1^{(G)}(1)$	1.99 (0.064)	1.03 (0.011)	1.99 (0.062)	1.04 (0.010)

The reliability ratio is 0.8. Numbers in parentheses are the Monte Carlo standard error of the averages. The true parameter values are $\beta_0=-2$ and $\beta_1=1$.

DOI: 10.1002/sim

Table II. Rejection rates of the tests (with $\lambda = 1$) associated with remeasurement method and CET across 200 Monte Carlo replicates.

		$N = 500$		$N = 800$	
		Normal	Mixture normal	Normal	Mixture normal
Reliability Ratio = 0.8					
Remeasurement method	β_0	0.37	0.04	0.92	0.04
(Individual testing data)	β_1	0.72	0.05	0.99	0.03
Remeasurement method	β_0	0.01	0.00	0.15	0.02
(Group testing data)	β_1	0.14	0.01	0.49	0.03
CET	β_0	0.68	0.02	1.00	0.02
(Group testing data)	β_1	0.93	0.03	1.00	0.01
	$\boldsymbol{\Omega}$	0.82	0.06	1.00	0.03
Reliability Ratio = 0.75					
Remeasurement method	β_0	0.07	0.02	0.50	0.03
(Individual testing data)	β_1	0.34	0.03	0.75	0.04
Remeasurement method	β_0	0.01	0.00	0.02	0.01
(Group testing data)	β_1	0.04	0.01	0.10	0.03
CET	β_0	0.55	0.01	0.97	0.02
(Group testing data)	β_1	0.79	0.02	1.00	0.03
	$\boldsymbol{\Omega}$	0.67	0.05	0.98	0.04

individual testing data when the model is misspecified. When N or m is moderate, the power can be unacceptably low even when the adverse effect of misspecification is severe. In contrast, CET attains high power to detect the misspecification, and the power can be even substantially higher than that of the remeasurement method applied to individual testing data. Table II also suggests that CET confers the nominal 0.05 size.

5. APPLICATION TO FRAMINGHAM HEART DATA

We now apply the remeasurement method and CET to a data set from the Framingham Heart Study [22], where one of the interest is the relationship between systolic blood pressure and the risk of coronary heart disease. Even though this is not a group testing application, the availability of individual response data enables us to compare the performance of CET and the remeasurement method, with the latter applied to both individual testing data and group testing data. The group testing data can be simulated from the raw data via random pooling. The data set consists of $N = 1615$ male subjects who were 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 eight-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 true predictor X as the long-term systolic blood pressure, which is unobserved. We use the first systolic blood pressure reading measured in the second examination as a noisy surrogate of X, W, and use the first systolic blood pressure reading from the third examination

DOI: 10.1002/sim

Table III. Values of the test statistics (with $\lambda = 1$) for a data from the Framingham Heart Study when the latent-variable model is assumed to be normal or mixture normal.

		Normal	Mixture normal
Remeasurement method	β_0	−2.81 (0.005)	0.52 (0.603)
(Individual testing data)	β_1	2.87 (0.004)	−0.51 (0.608)
Remeasurement method	β_0	−1.55 (0.122)	0.83 (0.407)
(Group testing data)	β_1	1.59 (0.113)	−0.83 (0.409)
CET	β_0	0.10 (0.046)	0.01 (0.146)
(Group testing data)	β_1	0.05 (0.044)	0.01 (0.168)
	$\boldsymbol{\Omega}$	0.15 (0.046)	0.03 (0.280)

Numbers in parentheses are the associated p-values or empirical p-values.

period as a replicate measurement of X so that we can estimate the measurement error variance σ^2, as in [10] (Section 5.4.2.1). The prevalence rate of heart disease in this sample is $\overline{Y} \approx 0.08$, and $\hat{\sigma}^2 \approx 0.04$, with an estimated reliability ratio of 0.73. The simulated group testing data is formed by randomly pooling five subjects into a group, producing $m = 323$ groups with $n_i = 5$, for $i = 1, \ldots, 323$. The probit link is used in the primary regression model. For the assumed model of X, we experiment on two distributions, a normal distribution and a two-component mixture normal distribution. Then we apply CET to the induced group testing data, and use the remeasurement method with both data sets to assess the appropriateness of each assumed latent-variable model. In the remeasurement method, we use $B = 50$; in CET, we set $B = B^* = 50$ and $Q = 500$. In both methods λ is equal to one. The values of the test statistics resulting from two methods and the associated p-values or empirical p-values are presented in Table III.

Under the assumption of normal latent variable, the tests from the remeasurement method applied to the individual testing data are highly significant, providing strong evidence of model misspecification. This finding agrees with the analyses in [11, 19]. However, when the group testing data is used, the remeasurement method does not find sufficient evidence of model misspecification. In contrast, using the same group testing data, the tests from CET are significant enough (at 0.05 significance level) to suggest that X is not likely to be normally distributed. Under the assumption of mixture normal latent variable, none of the tests strongly suggest presence of model misspecification.

6. DISCUSSION

Regression analysis of group testing data when covariates are measured with error is a natural extension of the work by Vansteelandt *et al.* [7, 8] and Xie [9]. A common concern with structural measurement error models is lack of justification for the assumed latent-variable model. We proposed a new method, CET, to detect latent-variable model misspecification in structure measurement error models for group testing data. Compared with the existing method, CET enjoys a much higher power to detect model misspecification. It is encouraging to observe in simulation that CET is even more powerful than the remeasurement method applied to individual testing data. This leads us to believe that, even for individual testing design, CET can be an attractive alternative to the existing diagnostic methods when N or r is moderate. Of course, in that case, one does not need to generate pseudo-individual testing data and the CET algorithm can be greatly simplified.

DOI: 10.1002/sim

The proposed test gains high testing power at the expense of computation time. Depending on the size of Q (the number of bootstrap samples), the CET can take much longer to implement than the remeasurement method. For instance, in the simulation example in Section 4 with $N = 800$, it takes a few seconds to implement the remeasurement method once whereas it takes a couple of minutes for CET with $Q = 200$. This is expected with the paucity of information in group testing data and it is a price worth paying for such dramatic improvement in testing power.

The idea behind CET is very different from most existing goodness-of-fit tests. It makes use of two types of model-misspecification evidence as discussed in Section 3.1. Different as they are, these two types of evidence can be unified in the same framework of information reduction, which is against the usual intention of striving for maximum use of data information. This counter-intuitive strategy leads to two indicators of model misspecification, which CET combines to achieve excellent performance as a diagnostic tool even when rich data information is lacking. The idea of information reduction has the potential to solve more general model selection problems and merits further investigation.

ACKNOWLEDGEMENTS

The author thanks two anonymous referees and the associate editor for their constructive comments.

REFERENCES

1. Gastwirth J, Johnson W. Screening with cost-effective quality control: potential applications to HIV and drug testing. *Journal of the American Statistics Association* 1994; **89**:972–981.
2. Tu X, Litvak E, Pagano M. On the informativeness and accuracy of pooled testing in estimating prevalence of a rare disease: application to HIV screening. *Biometrika* 1995; **82**:287–297.
3. Hughes-Oliver J, Rosenberger W. Efficient estimation of the prevalence of multiple rare traits. *Biometrika* 2000; **87**:315–327.
4. Gastwirth J. The efficiency of pooling in the detection of rare mutations. *American Journal of Human Genetics* 2000; **67**:1036–1039.
5. Thierry-Mieg N. Pooling in systems biology becomes smart. *Nature Methods* 2006; **3**:161–162.
6. Remlinger K, Hughes-Oliver J, Young S, Lam R. Statistical design of pools using optimal coverage and minimal collision. *Technometrics* 2006; **48**:133–143.
7. Vansteelandt S, Goetghebeur E, Verstraeten T. Adjusting for confounding when estimating a time trend in HIV prevalence based on pooled serum samples. *Archives of Public Health* 1999; **57**:89–105.
8. Vansteelandt S, Goetghebeur E, Verstraeten T. Regression models for disease prevalence with diagnostic tests on pools of serum samples. *Biometrics* 2000; **56**:1126–1133.
9. Xie M. Regression analysis of group testing samples. *Statistics in Medicine* 2001; **20**:1957–1969.
10. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. *Measurement Error in Nonlinear Models*: *A Modern Perspective*. Chapman & Hall, CRC: Boca Raton, FL, 2006.
11. Huang X, Stefanski LA, Davidian M. Latent-model robustness in structural measurement error models. *Biometrika* 2006; **93**:53–64.
12. Carroll RJ, Maca J, Ruppert D. Nonparametric regression in the presence of measurement error. *Biometrika* 1999; **86**:541–554.
13. Cook J, Stefanski LA. Simulation extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association* 1994; **89**:1314–1328.
14. Schafer D. Semiparametric maximum likelihood for measurement error model regression. *Biometrics* 2001; **57**:53–61.
15. Rabe-Hesketh S, Pickles A, Skrondal A. Correcting for covariate measurement error in logistic regression using nonparametric maximum likelihood estimation. *Statistical Modelling* 2003; **3**:215–232.
16. Carroll RJ, Roeder K, Wasserman L. Flexible parametric measurement error models. *Biometrics* 1999; **55**:44–54.

DOI: 10.1002/sim

17. Richardson S, Leblond L, Jaussent I, Green P. Mixture models in measurement error models, with reference to epidemiological studies. *Journal of the Royal Statistical Society A* 2002; **165**:549–566.
18. Aitkin M, Rocci R. A general maximum likelihood analysis of measurement error in generalized linear models. *Statistics and Computing* 2002; **12**:163–174.
19. Huang X, Tebbs JM. On latent-variable model misspecification in structural measurement error models for binary response. *Biometrics* 2009; DOI: 10.1111/j.1541-0420.2008.01128.x.
20. Huang X, Stefanski LA, Davidian M. Latent-model robustness in joint models of a primary endpoint and a longitudinal process. *Biometrics* 2009; DOI: 10.1111/j.1541-0420.2008.01171.x.
21. Carroll RJ, Gail M, Lubin J. Case-control studies with errors in covariates. *Journal of the American Statistical Association* 1993; **88**:185–199.
22. Kannel W, Neaton J, Wentworth D, Thomas H, Stamler J, Hulley S, Kjelsberg M. Overall and coronary heart disease mortality rates in relation to major risk factors in 325 348 men screened for MRFIT. *American Heart Journal* 1986; **112**:825–836.