*Submitted to the Annals of Applied Statistics*

# STATISTICAL METHODS FOR ANALYSIS OF COMBINED CATEGORICAL BIOMARKER DATA FROM MULTIPLE STUDIES

By Chao Cheng[†], Bernard Rosner[‡] and Molin Wang[‡,*]

*Yale University*[†] *and Harvard University*[‡]

In the analysis of pooled data from multiple studies involving a biomarker exposure, the biomarker measurements can vary across laboratories and usually require calibration to a reference assay prior to pooling. Previous researches consider the measurements from a reference laboratory as the gold standard, even though measurements in the reference laboratory is not necessarily closer to the underlying truth in reality. In this paper, we do not treat any laboratory measurements as the gold standard, and we develop two statistical methods, the exact calibration and cut-off calibration methods, for the analysis of aggregated categorical biomarker data. We compare the performance of both methods for estimating the biomarker-disease relationship under a random sample or controls-only calibration design. Our finding includes: (1) the exact calibration method provides significantly less biased estimates and more accurate confidence intervals than the other method. (2) the cut-off calibration method could yield estimates with minimal bias and valid confidence intervals under small measurement errors and/or small exposure effects. (3) Controls-only calibration design can result in additional bias but the bias is minimal if the exposure effects and/or disease prevalences are small. Finally, we illustrate the methods in an application evaluating the relationship between circulating vitamin D levels and colorectal cancer risk in a pooling project.

**1. Introduction.** It has become increasingly common to pool biomarker data from different studies together to investigate biomarker-disease relationships. Examples of pooling projects examining biomarker-disease associations include Cohort Consortium Vitamin D Pooling Project of Rarer Cancers [Gallicchio et al. (2010)], Vitamin D Pooling Project of Breast and Colorectal Cancer [McCullough et al. (2018)], the Breast and Prostate Cancer Cohort Consortium [Tsilidis et al. (2013)] and the Endogenous Hormones, Nutritional Biomarkers, and Prostate Cancer Collaborative Group [Crowe et al. (2014); Key et al. (2015)].

There is between-laboratory variability in biomarker measurements. For instance, circulating vitamin D (25(OH)D) have highly variable measurements between laboratories and assays, which may vary up to 40% [Lai et al. (2012); Snellman et al. (2010)]. Many

*Corresponding author

*Keywords and phrases:* between-study variability, calibration, multiple studies, categorical biomarker data, measurement error

hormones, such as testosterone and estrone, also have highly variable measurements between assays and laboratories [Key et al. (2010); Tworoger and Hankinson (2006)]. Under such circumstances, calibration is implemented by choosing a reference laboratory and re-assaying a subset of biospecimens from each contributing study at the reference laboratory, which can be used to estimate between-study measurement variability [Sloan et al. (2019)]. Owing to potential concerns about biomarker measurements in cases, investigators typically use only non-cases in the calibration study subsets.

This paper focuses on analyses of biomarker data pooled from multiple studies for evaluating the odds ratio (OR) representing the association of a categorized version of the biomarker and a binary disease outcome. In previous researches, the measurements from the reference laboratory were treated as the "gold standard" measurements [Sloan et al. (2019); Gail et al. (2016)]. For example, the method in Sloan et al. (2019) assumes $E(H_r|H_s) = \alpha_s + \beta_s H_s$, where $H_r$ and $H_s$ are the measurements from the reference and study-specific laboratory respectively, and uses the estimated reference laboratory measurement, $\hat{\alpha}_s + \hat{\beta}_s H_s$, in the place of the biomarker exposure in the logistic regression model. In reality, however, measurements in reference laboratory is not necessarily closer to the underlying truth than the measurements from other study-specific laboratories. For example, the US Institute of Medicine (IOM) recommendation for vitamin D concluded that serum 25(OH)D>50 nmol/l is 'sufficient'. This recommendation was based on data from many laboratories. Categorizing 25(OH)D to 'sufficient' vs. 'insufficient' based on the observed or calibrated measurements from only the reference laboratory could result in misclassification and biased odds ratio (OR) estimates, and may not the best practice given that data from multiple laboratories are indeed available in a pooling project.

In this paper, we take advantage of the existence of measurements from multiple laboratories, and the reference laboratory or any study-specific laboratory measurements are no longer treated as the gold standard. We develop two approaches, the exact calibration method and cut-off calibration method, for the aggregated biomarker data. The framework of this paper is as follows: We present the statistical models and methods in Section 2. In Section 3, we evaluate the methods in a simulation study. We illustrate the methods in a circulating vitamin D and colorectal cancer example in Section 4, and we offer a theoretical adjustment when the biomarker does not follow a normal distribution in Section 5. Finally, a concluding discussion is presented in Section 6.

## 2. Methods.

2.1. *Notations and Models.* Suppose that there are $M$ studies, each associated with a study-specific local laboratory $j$, where $j = 1, 2, \ldots, M$. Let $X_{jk}$ be the unobserved true value of the continuous biomarker for the $k^{th}$ individual in the $j^{th}$ study, and $Y_{jk}$ the binary disease outcome. We assume the following logistic regression model:

$$\text{logit}(P(Y_{jk} = 1|X_{jk}, \boldsymbol{Z}_{jk})) = \beta_{0j} + \sum_{l=2}^{G} \mathbb{I}(g_{l-1} \leq X_{jk} < g_l)\beta_{x,l} + \boldsymbol{\beta}_z^T \boldsymbol{Z}_{jk}, \tag{2.1}$$

where $\mathbb{I}(.)$ is the indicator function, and $\boldsymbol{Z}_{jk}$ represents potential confounders for the X–Y relationship. Without further specifications, all vectors are column vectors. Note that the true biomarker measurement, $X_{jk}$, is ranged $(g_0, g_G)$, $g_0$ and $g_G$ can be 0 and $\infty$, respectively, and categorized at cut-off points $g_1 <, \ldots, < g_{G-1}$. The parameter of interest is $\boldsymbol{\beta}_x = [\beta_{x,2}, \ldots, \beta_{x,G}]^T$, and $\boldsymbol{\beta_0} = [\beta_{01}, \ldots, \beta_{0M}]^T$ contains study-specific intercepts. As $X_{jk}$ is unavailable, we can not use the standard logistic regression to estimate $\boldsymbol{\beta}_x$.

Suppose there are a total of $N_c$ individuals, with $N_j$ individuals in the $j^{th}$ study. Without loss of generality, we assume that the first $n_j$ individuals in study $j$ are included in the calibration subset. The calibration subset consists of a random selection of controls or a random selection of both cases and controls. We use the terminology controls only calibration study (COCS) and random sample calibration study (RSCS) respectively to refer to these designs. Biospecimens from individuals in the calibration subset are re-assayed at a reference laboratory. For individuals selected into the calibration subset, let $H_{jk,d}$ be the biomarker measurement from laboratory $d$, where $d = 0$ refers to the reference laboratory and $d = j > 0$ refers to the $j^{th}$ study's local laboratory. Individuals who are not selected into the calibration subset only have the local laboratory measurements $H_{jk,j}$. For brevity, we use $\boldsymbol{H}_{jk}$ to denote all the biomarker measurements; i.e, for the individuals in the calibration subset, $\boldsymbol{H}_{jk} = [H_{jk,0}, H_{jk,j}]^T$, and for the individuals who only have local laboratory measurements, $\boldsymbol{H}_{jk} = H_{jk,j}$.

We assume that the measurement $H_{jk,d}$ and the underlying truth $X_{jk}$ follow the classic additive measurement error model

$$H_{jk,d} = X_{jk} + \epsilon_{jk,d}, \tag{2.2}$$

where $\epsilon_{jk,d}$, the measurement error, is independent from $X_{jk}$, and follows a mean-zero normal distribution with a laboratory-specific variance; i.e, $\epsilon_{jk,d} \sim N(0, \sigma_d^2)$, $d = 0, j$ and $j = 1, \ldots, M$.

We assume that, given variables $\boldsymbol{W}_{jk}$, $X_{jk}$ can be modeled using the following regression

$$X_{jk} = \mu(\boldsymbol{W}_{jk}; \alpha_{0j}, \boldsymbol{\tau}) + \epsilon_{x_{jk}}, \tag{2.3}$$

where $\mu(\boldsymbol{W}_{jk}; \alpha_{0j}, \boldsymbol{\tau})$ denotes the conditional mean of $X_{jk}|\boldsymbol{W}_{jk}$, $\mu(.)$ is a given function, $\alpha_{0j}$ is a unknown study-specific parameter, $\boldsymbol{\tau}$ represents unknown parameters common for all the studies, and $\epsilon_{x_{jk}}$ is the error term following $N(0, \sigma_x^2)$. Typically, we can use the linear regression model $X_{jk} = \alpha_{0j} + \boldsymbol{\tau}^T \boldsymbol{W}_{jk} + \epsilon_{x_{jk}}$, where $\alpha_{0j}$ is a study-specific intercept. If $\boldsymbol{W}_{jk}$ is null, the conditional mean, $\mu(\boldsymbol{W}_{jk}; \alpha_{0j}, \boldsymbol{\tau})$, degenerates to $\alpha_{0j}$.

2.2. *Likelihoods.* The likelihood function corresponding to the logistic regression model in Model (2.1) is

$$L = \prod_j \prod_k \frac{\exp\{Y_{jk}(\beta_{0j} + \sum_{l=2}^{G} \mathbb{I}(g_{l-1} \leq X_{jk} < g_l)\beta_{x,l} + \boldsymbol{\beta}_z^T \boldsymbol{Z}_{jk})\}}{1 + \exp\{\beta_{0j} + \sum_{l=2}^{G} \mathbb{I}(g_{l-1} \leq X_{jk} < g_l)\beta_{x,l} + \boldsymbol{\beta}_z^T \boldsymbol{Z}_{jk}\}}.$$

This likelihood, however, cannot be computed because $X_{jk}$ is not available. In fact, the contribution of a conditional likelihood from an individual can be shown as

$$\begin{aligned} L_{jk} &= P(Y_{jk}|\boldsymbol{H}_{jk}, \boldsymbol{Z}_{jk}) \\ &= \int P(Y_{jk}|X_{jk}, \boldsymbol{H}_{jk}, \boldsymbol{Z}_{jk}) f(X_{jk}|\boldsymbol{H}_{jk}, \boldsymbol{Z}_{jk}) \mathrm{d}X_{jk}. \end{aligned} \tag{2.4}$$

We make the surrogacy assumption that, conditional on $\boldsymbol{Z}_{jk}$, the laboratory measurement $\boldsymbol{H}_{jk}$ does not contain additional information about $Y_{jk}$ if the true biomarker value is known; i.e., $P(Y_{jk}|X_{jk}, \boldsymbol{H}_{jk}, \boldsymbol{Z}_{jk}) = P(Y_{jk}|X_{jk}, \boldsymbol{Z}_{jk})$. Let $\boldsymbol{W}_{jk}$ contain the variables in $\boldsymbol{Z}_{jk}$ that could be associated with $X_{jk}$; that is, we assume $P(X_{jk}|\boldsymbol{H}_{jk}, \boldsymbol{Z}_{jk}) = P(X_{jk}|\boldsymbol{H}_{jk}, \boldsymbol{W}_{jk})$. It follows that the likelihood contribution $L_{jk}$ can be written as

(2.5)

$$\begin{aligned} L_{jk} &= \int P(Y_{jk}|X_{jk}, \boldsymbol{Z}_{jk}) f(X_{jk}|\boldsymbol{H}_{jk}, \boldsymbol{W}_{jk}) \mathrm{d}X_{jk} \\ &= \int \frac{\exp\{Y_{jk}(\beta_{0j} + \sum_{l=2}^{G} \mathbb{I}(g_{l-1} \le X_{jk} < g_l)\beta_{x,l} + \boldsymbol{\beta}_z^T \boldsymbol{Z}_{jk})\}}{1 + \exp\{\beta_{0j} + \sum_{l=2}^{G} \mathbb{I}(g_{l-1} \le X_{jk} < g_l)\beta_{x,l} + \boldsymbol{\beta}_z^T \boldsymbol{Z}_{jk}\}} f(X_{jk}|\boldsymbol{H}_{jk}, \boldsymbol{W}_{jk}) \mathrm{d}X_{jk} \\ &= \sum_{l=2}^{G} \frac{\exp\{Y_{jk}(\beta_{0j} + \beta_{x,l} + \boldsymbol{\beta}_z^T \boldsymbol{Z}_{jk})\}}{1 + \exp(\beta_{0j} + \beta_{x,l} + \boldsymbol{\beta}_z^T \boldsymbol{Z}_{jk})} P(g_{l-1} \le X_{jk} < g_l|\boldsymbol{H}_{jk}, \boldsymbol{W}_{jk}), \end{aligned}$$

which is a function of our parameter of interest, $\boldsymbol{\beta}_x$.

In practice, there could be variables, denoted as $\boldsymbol{W}_{jk}^*$, that are informative about the biomarker exposure $X_{jk}$, but not be part of $\boldsymbol{Z}_{jk}$, a variable in the dataset. To take advantage of the availability of these variables, hereafter, we use $\boldsymbol{W}_{jk}$ to denote the collection of available variables that could be informative about $X_{jk}$, possibly including variables not in $\boldsymbol{Z}_{jk}$. With this extended definition of $\boldsymbol{W}_{jk}$, the conditional likelihood contribution is now $L_{jk} = P(Y_{jk}|\boldsymbol{H}_{jk}, \boldsymbol{Z}_{jk}, \boldsymbol{W}_{jk}^*)$, which can still be written as in (2.5). A discussion about the potential benefit of including the additional variables $\boldsymbol{W}_{jk}^*$ is in Section 2.7.

Next, we derive the analytic forms of $P(g_{l-1} \le X_{jk} < g_l|\boldsymbol{H}_{jk}, \boldsymbol{W}_{jk})$.

2.3. *Conditional Distribution of the Unknown True Biomarker Value.* Although $X_{jk}$ is unobservable, we can derive the conditional distribution of $X_{jk}$ given $\boldsymbol{H}_{jk}$ and $\boldsymbol{W}_{jk}$. Specifically, based on models (2.2) and (2.3), $(X_{jk}|\boldsymbol{W}_{jk}, H_{jk,0}|\boldsymbol{W}_{jk}, H_{jk,j}|\boldsymbol{W}_{jk})^T$ follows the multivariate normal distribution below:

$$\begin{pmatrix} X_{jk}|\boldsymbol{W}_{jk} \\ H_{jk,0}|\boldsymbol{W}_{jk} \\ H_{jk,j}|\boldsymbol{W}_{jk} \end{pmatrix} \sim \mathrm{MVN}\left( \begin{pmatrix} \mu_{X_{jk}|\boldsymbol{W}_{jk}} \\ \mu_{X_{jk}|\boldsymbol{W}_{jk}} \\ \mu_{X_{jk}|\boldsymbol{W}_{jk}} \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_x^2 & \sigma_x^2 \\ \cdot & \sigma_x^2 + \sigma_0^2 & \sigma_x^2 \\ \cdot & \cdot & \sigma_x^2 + \sigma_j^2 \end{pmatrix} \right), \tag{2.6}$$

where $\mu_{X_{jk}|\boldsymbol{W}_{jk}}$ is the abbreviation of $\mu(\boldsymbol{W}_{jk}; \alpha_{0j}, \boldsymbol{\tau})$ in (2.3). It follows that, for individuals who only have local laboratory measurements,

$$X_{jk}|\boldsymbol{H}_{jk}, \boldsymbol{W}_{jk} \sim N\Big(\rho_j H_{jk,j} + (1 - \rho_j)\mu_{X_{jk}|\boldsymbol{W}_{jk}}, \rho_j \sigma_j^2\Big), \tag{2.7}$$

where $\rho_j = \frac{\sigma_x^2}{\sigma_x^2+\sigma_j^2}$, and for individuals in the calibration subset

$$X_{jk}|\boldsymbol{H}_{jk}, \boldsymbol{W}_{jk} \sim N\Big(\rho_j^*(w_j H_{jk,j} + (1-w_j)H_{jk,0}) + (1-\rho_j^*)\mu_{X_{jk}|\boldsymbol{W}_{jk}}, \rho_j^* w_j \sigma_j^2\Big), \tag{2.8}$$

where $\rho_j^* = \sigma_x^2/(\sigma_x^2+\sigma_j^2 w_j)$ and $w_j = \sigma_0^2/(\sigma_j^2+\sigma_0^2)$. Hearafter, we use $\mu_{jk}$ and $s_{jk}$ to denote the mean and standard deviation of $X_{jk}|\boldsymbol{H}_{jk}, \boldsymbol{W}_{jk}$. Next, we describe the procedures for estimating the parameters involved in the conditional mean, $\mu_{jk}$.

2.4. *Estimation of Parameters in the Conditional Mean.* Let $\boldsymbol{\theta} = [\alpha_{01}, \alpha_{02}, \ldots, \alpha_{0M}, \boldsymbol{\tau}^T]^T$ and $\boldsymbol{\sigma}^2 = [\sigma_x^2, \sigma_0^2, \sigma_1^2, \ldots, \sigma_M^2]^T$ denote the unknown parameters in the means and variances of (2.7) and (2.8). Combining models (2.2) and (2.3), we have

$$H_{jk,d} = \mu(\boldsymbol{W}_{jk}; \alpha_{0j}, \boldsymbol{\tau}) + \delta_{x_{jk},d},$$

where $\delta_{x_{jk},d} \sim N(0, \sigma_x^2+\sigma_d^2)$ is the laboratory-specific error term. Note that the outcomes $H_{jk,d}$ are not always independent since each individual in the calibration subset has two correlated measurements, $H_{jk,0}$ and $H_{jk,j}$. We propose the following estimating equations for $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}^2$.

(2.9)

$$\begin{aligned}
\Psi_{\boldsymbol{\theta}} = &\sum_{j=1}^{M}\sum_{k=1}^{n_j} \frac{d}{d\boldsymbol{\theta}} \begin{bmatrix}\mu(\boldsymbol{W}_{jk};\alpha_{0j},\boldsymbol{\tau})\\ \mu(\boldsymbol{W}_{jk};\alpha_{0j},\boldsymbol{\tau})\end{bmatrix}^T \begin{bmatrix}\sigma_x^2+\sigma_0^2 & \sigma_x^2\\ \sigma_x^2 & \sigma_x^2+\sigma_j^2\end{bmatrix}^{-1} \begin{bmatrix}e_{jk,0}\\ e_{jk,j}\end{bmatrix}\\
&+\sum_{j=1}^{M}\sum_{k=n_j+1}^{N_j} \frac{d\mu(\boldsymbol{W}_{jk};\alpha_{0j},\boldsymbol{\tau})}{d\boldsymbol{\theta}}(\sigma_x^2+\sigma_j^2)^{-1}e_{jk,j} = \mathbf{0},\\
\Psi_{\boldsymbol{\sigma}^2} = &\sum_{j=1}^{M}\sum_{k=1}^{n_j} \frac{d}{d\boldsymbol{\sigma}^2} \begin{bmatrix}\sigma_x^2+\sigma_0^2\\ \sigma_x^2\\ \sigma_x^2+\sigma_j^2\end{bmatrix}^T \begin{bmatrix}e_{jk,0}^2-(\sigma_x^2+\sigma_0^2)\\ e_{jk,0}e_{jk,j}-\sigma_x^2\\ e_{jk,j}^2-(\sigma_x^2+\sigma_j^2)\end{bmatrix} + \sum_{j=1}^{M}\sum_{k=n_j+1}^{N_j} \frac{d(\sigma_x^2+\sigma_j^2)}{d\boldsymbol{\sigma}^2}\big(e_{jk,j}^2-(\sigma_x^2+\sigma_j^2)\big) = \mathbf{0},
\end{aligned}$$

where $e_{jk,d} = H_{jk,d} - \mu(\boldsymbol{W}_{jk};\alpha_{0j},\boldsymbol{\tau})$. For example, if $\mu(\boldsymbol{W}_{jk};\alpha_{0j},\boldsymbol{\tau}) = \alpha_{0j} + \boldsymbol{\tau}^T\boldsymbol{W}_{jk}$, $\mu(\boldsymbol{W}_{jk};\alpha_{0j},\boldsymbol{\tau})$ can be written as $\boldsymbol{C}_{jk}^T\boldsymbol{\theta}$, where $\boldsymbol{C}_{jk} = [\boldsymbol{D}_{jk}^T, \boldsymbol{W}_{jk}^T]^T$, $\boldsymbol{D}_{jk}$ is a $M\times 1$ vector with one on $j^{th}$ element and zeros elsewhere. Note that $\Psi_{\boldsymbol{\theta}} = \mathbf{0}$ is a first-order estimating equation for $\boldsymbol{\theta}$ with $\boldsymbol{\sigma}^2$ treated as known, and $\Psi_{\boldsymbol{\sigma}^2} = \mathbf{0}$ is a second-order estimating equation for $\boldsymbol{\sigma}^2$ with $\boldsymbol{\theta}$ considered known.

We adopt a two-stage iteration method [Lanzkron, Rose and Szyld (1990)] to estimate $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}^2$. We start the procedure by setting $\hat{\boldsymbol{\sigma}}^{2(0)} = [\hat{\sigma}_x^{2(0)}, \hat{\sigma}_0^{2(0)}, \hat{\sigma}_1^{2(0)}, \ldots, \hat{\sigma}_M^{2(0)}]^T$ with $\hat{\sigma}_x^{2(0)} = 0$, $\hat{\sigma}_0^{2(0)} = \hat{\sigma}_j^{2(0)} = 1$ for $j = 1, \ldots, M$; that is, the working variance-covariance matrix in the first stage estimating equation is set to the identity matrix as the starting value. Specifically, the two stages in the $m^{th}$ iteration is

- First stage: fix $\boldsymbol{\sigma}^2 = \hat{\boldsymbol{\sigma}}^{2(m)}$ and obtain $\hat{\boldsymbol{\theta}}^{(m)}$ by solving $\Psi_{\boldsymbol{\theta}} = \mathbf{0}$;
- Second stage: fix $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(m)}$ and obtain $\hat{\boldsymbol{\sigma}}^{2(m+1)}$ by solving $\Psi_{\boldsymbol{\sigma}^2} = \mathbf{0}$.

The iteration continues until convergence. The convergence criteria can be based on both relative differences $\frac{||\hat{\boldsymbol{\theta}}^{(m+1)} - \hat{\boldsymbol{\theta}}^{(m)}||}{||\hat{\boldsymbol{\theta}}^{(m)}||}$ and $\frac{||\hat{\boldsymbol{\sigma}}^{2(m+1)} - \hat{\boldsymbol{\sigma}}^{2(m)}||}{||\hat{\boldsymbol{\sigma}}^{2(m)}||}$, where $||\bullet||$ denotes the Euclidean norm.

One or more elements in the variance estimates ($\hat{\boldsymbol{\sigma}}^2$) above could be negative, which is unlikely to happen if the calibration sample size is sufficient. However, when the calibration subset sample size is small and some laboratories' measurement errors are close to zero, the negative variance problem is more likely to happen. The negative-estimates problem of variance components has been discussed in previous literatures [El Leithy, Wahed and Abdallah (2016); Fletcher and Underwood (2002); Thompson et al. (1962)]. Here, according to the suggestion in Rao (1972), we use a restriction method to solve this problem. Specifically, in the second stage above, instead of solving $\Psi_{\boldsymbol{\sigma}^2} = \mathbf{0}$, we minimize the following sum of squares

(2.10)
$$L_{\boldsymbol{\sigma}^2} = \sum_{j=1}^{M}\sum_{k=1}^{n_j} \begin{bmatrix} e_{jk,0}^2 - (\sigma_x^2 + \sigma_0^2) \\ e_{jk,0}e_{jk,j} - \sigma_x^2 \\ e_{jk,j}^2 - (\sigma_x^2 + \sigma_j^2) \end{bmatrix}^T \begin{bmatrix} e_{jk,0}^2 - (\sigma_x^2 + \sigma_0^2) \\ e_{jk,0}e_{jk,j} - \sigma_x^2 \\ e_{jk,j}^2 - (\sigma_x^2 + \sigma_j^2) \end{bmatrix} + \sum_{j=1}^{M}\sum_{k=n_j+1}^{N_j} \left(e_{jk,j}^2 - (\sigma_x^2 + \sigma_j^2)\right)^2,$$

subject to the condition that $\boldsymbol{\sigma}^2 \geq \mathbf{0}$. If the roots of $\Psi_{\boldsymbol{\sigma}^2} = \mathbf{0}$ are all nonnegative, this restriction method leads to the same calibration parameter estimates as the unrestricted one. We conducted a simulation study to investigate the impact on $\boldsymbol{\beta}_x$ when using this restriction method in the presence of negative variance; see Section 3, Appendix E and Tables S7–S9 in the Supplementary Material.

2.5. *Exact Calibration Method.* This section focuses on a likelihood-based method for the estimation of exposure effects using aggregated data. Denoting $P(g_{l-1} \leq X_{jk} < g_l|\boldsymbol{H}_{jk}, \boldsymbol{W}_{jk})$ as $p_{jk,l}$ $(l = 2, 3, \ldots, G)$, and following the distribution of $X_{jk}|\boldsymbol{H}_{jk}, \boldsymbol{W}_{jk}$ in (2.7) and (2.8), we have

$$p_{jk,l} = \Phi\left(\frac{g_l - \mu_{jk}}{s_{jk}}\right) - \Phi\left(\frac{g_{l-1} - \mu_{jk}}{s_{jk}}\right), \tag{2.11}$$

where $\Phi(.)$ is the standard normal cumulative density function. We plug the estimators $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\sigma}}^2$ in Section 2.4 into $p_{jk,l}$ in (2.11), leading to $\hat{p}_{jk,l}$. The likelihood contribution in (2.5) with $p_{jk,l}$ replaced by $\hat{p}_{jk,l}$ becomes

$$\tilde{L}_{jk} = \sum_{l=2}^{G} \frac{\exp\{Y_{jk}(\beta_{0j} + \beta_{x,l} + \boldsymbol{\beta}_z^T \boldsymbol{Z}_{jk})\}}{1 + \exp(\beta_{0j} + \beta_{x,l} + \boldsymbol{\beta}_z^T \boldsymbol{Z}_{jk})} \hat{p}_{jk,l}. \tag{2.12}$$

Let $\boldsymbol{\beta} = [\boldsymbol{\beta}_{\mathbf{0}}^T, \boldsymbol{\beta}_x^T, \boldsymbol{\beta}_z^T]^T$, which contains the coefficients in model (2.1), and $\boldsymbol{\pi} = [\boldsymbol{\theta}^T, {\boldsymbol{\sigma}^2}^T, \boldsymbol{\beta}^T]^T$, which includes all the unknown parameters. Estimates of $\boldsymbol{\beta}$ can be obtained by maximizing the pseudo-likelihood $\tilde{L} = \Pi_{j,k}\tilde{L}_{jk}$. We name this method as *Exact Calibration Method* (ECM) and denote the $\boldsymbol{\beta}$-estimator from ECM as $\hat{\boldsymbol{\beta}}^{(E)}$.

We define joint estimating equation $\boldsymbol{\Psi}(\boldsymbol{\pi}) = [\Psi_{\boldsymbol{\theta}}^T, \Psi_{\boldsymbol{\sigma}^2}^T, \Psi_{\boldsymbol{\beta}}^T]^T = \mathbf{0}$, where $\Psi_{\boldsymbol{\theta}}$ and $\Psi_{\boldsymbol{\sigma}^2}$ are defined in (2.9) and $\Psi_{\boldsymbol{\beta}}$ is the score function based on $\tilde{L}$ (see Appendix A in the Supplementary Material for the derivation of $\Psi_{\boldsymbol{\beta}}$). The $\boldsymbol{\beta}$ estimates from solving this joint estimating equation is identical to $\hat{\boldsymbol{\beta}}^{(E)}$ obtained by maximizing $\tilde{L}$. We consider the following two approaches to obtain variance estimates of $\hat{\boldsymbol{\beta}}$:

- Approach 1, Standard sandwich method: $\text{var}(\hat{\boldsymbol{\pi}}) \approx \hat{\boldsymbol{Q}}^{-1}\hat{\boldsymbol{U}}(\hat{\boldsymbol{Q}}^{-1})^T$, where $\hat{\boldsymbol{Q}} = \sum_{j=1}^{M}\sum_{k=1}^{N_j}\left[\frac{d\boldsymbol{\psi}_{\boldsymbol{\pi},jk}}{d\boldsymbol{\pi}}|_{\boldsymbol{\pi}=\hat{\boldsymbol{\pi}}}\right]$, $\hat{\boldsymbol{U}} = \sum_{j=1}^{M}\sum_{k=1}^{N_j}\left[\boldsymbol{\psi}_{\hat{\boldsymbol{\pi}},jk}\boldsymbol{\psi}_{\hat{\boldsymbol{\pi}},jk}{}^T\right]$, and $\boldsymbol{\psi}_{\boldsymbol{\pi},jk}$ is the piece in $\boldsymbol{\Psi}_{\boldsymbol{\pi}}$ corresponding to the $k^{th}$ individual of the $j^{th}$ study.
- Approach 2, Pseudo-likelihood hessian matrix method: estimate the variance of $\hat{\boldsymbol{\beta}}$ through calculating the hessian matrix for $\tilde{L}$; i.e, $\text{var}(\hat{\boldsymbol{\beta}}) \approx \left[\sum_{j=1}^{M}\sum_{k=1}^{N_j} -\frac{d^2 \ln \tilde{L}_{jk}}{d\boldsymbol{\beta}^2}|\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}\right]^{-1}$. For the variance of $\hat{\boldsymbol{\theta}}$, we have $\text{var}(\hat{\boldsymbol{\theta}}) \approx \left[\frac{d\boldsymbol{\Psi}_{\boldsymbol{\theta}}}{d\boldsymbol{\theta}}|\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}, \boldsymbol{\sigma}^2 = \hat{\boldsymbol{\sigma}}^2\right]^{-1}$.

The standard sandwich method takes into account variation due to estimating $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\sigma}}^2$ in the variance estimator of $\hat{\boldsymbol{\beta}}$. By contrast, the pseudo-likelihood hessian matrix method assumes $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\sigma}}^2$ are fixed values in estimating $\text{var}(\hat{\boldsymbol{\beta}})$. Therefore, the estimated variance of $\hat{\boldsymbol{\beta}}$ by standard sandwich method is expected to be a little larger than that from the pseudo-likelihood hessian matrix method.

2.6. *Cut-off Calibration Method.* Alternatively, the regression coefficients in (2.1) can be estimated by maximizing the following approximate likelihood

$$\tilde{L}^{(c)} = \prod_j \prod_k \frac{\exp\{Y_{jk}(\beta_{0j} + \sum_{l=2}^{G}\mathbb{I}(g_{l-1} \le \hat{X}_{jk} < g_l)\beta_{x,l} + \boldsymbol{\beta}_z^T\boldsymbol{Z}_{jk})\}}{1 + \exp(\beta_{0j} + \sum_{l=2}^{G}\mathbb{I}(g_{l-1} \le \hat{X}_{jk} < g_l)\beta_{x,l} + \boldsymbol{\beta}_z^T\boldsymbol{Z}_{jk})}, \tag{2.13}$$

where $\hat{X}_{jk}$ is the estimated value for $X_{jk}$ with $\hat{X}_{jk} = \hat{\mu}_{jk}$. We name it *Cut-off Calibration Method* (CCM) as it categorizes the estimated biomarker values directly, and we denote the estimates from the CCM as $\hat{\boldsymbol{\beta}}_x^{(C)}$.

As shown in Supplementary Material Appendix B, the CCM performs best when $\sigma_d^2$, $d = 0, 1, \ldots, M$, are small and/or the association between $Y_{jk}$ and $X_{jk}$ is not too strong. The approach for estimating $\text{var}(\hat{\boldsymbol{\beta}}_x^{(C)})$ is similar to that in section 2.5. See Supplementary Material Appendix C for more details.

2.7. *Calibration Study Designs.* Under the RSCS design, since $\Psi_\theta$ and $\Psi_{\boldsymbol{\sigma}^2}$ are unbiased estimating functions, the calibration parameter estimates, $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\sigma}}^2)$, in section 2.4 are consistent as $n_j \to \infty$ for $j = 1, \ldots, M$ [Godambe (1991)]. As a result, $\hat{p}_{jk}$ in (2.12), $\hat{\mu}_{jk}$ in (2.13) , and $\hat{\boldsymbol{\beta}}^{(E)}$ in Section 2.5 are also consistent estimates. Note that this consistency property holds whether or not $\boldsymbol{W}_{jk}$ includes the additional variables $\boldsymbol{W}_{jk}^*$. Under the linear regression model $\mu(\boldsymbol{W}_{jk}; \alpha_{0j}, \boldsymbol{\tau}) = \alpha_{0j} + \boldsymbol{\tau}^T\boldsymbol{W}_{jk}$, including $\boldsymbol{W}_{jk}^*$ in $\boldsymbol{W}_{jk}$ typically increases

the precision of $\hat{\boldsymbol{\theta}}$ and reduces $\sigma_x^2$ [Robinson and Jewell (1991)], and thus may increase the precision of the resulting $\boldsymbol{\beta}$-estimators.

Under the COCS design, $\hat{p}_{jk}$ and $\hat{\mu}_{jk}$ are typically asymptotic biased due to the following reasons. First, under the COCS design, the calibration subset data can be used to estimate the parameters in model $E(H_{jk,d}|\boldsymbol{W}_{jk}, Y_{jk} = 0) = \alpha_{0j,co} + \boldsymbol{\tau}_{co}^T \boldsymbol{W}_{jk}$, where the subscript $co$ denotes controls only enrollment, and $\hat{\alpha}_{0j,co}$ and $\hat{\boldsymbol{\tau}}_{co}$ are not consistent estimates for $\alpha_{0j}$ and $\boldsymbol{\tau}$. Second, the conditional distribution of $X_{jk}$ that can be estimated in the calibration subset is actually $P(X_{jk}|\boldsymbol{H}_{jk}, \boldsymbol{W}_{jk}, Y_{jk} = 0)$, which is also not identical to $P(X_{jk}|\boldsymbol{H}_{jk}, \boldsymbol{W}_{jk})$ in (2.5). However, under certain conditions we can show that $\alpha_{0j,co} \approx \alpha_{0j}$, $\boldsymbol{\tau}_{co} \approx \boldsymbol{\tau}$ and $P(X_{jk}|\boldsymbol{H}_{jk}, \boldsymbol{W}_{jk}, Y_{jk} = 0) \approx P(X_{jk}|\boldsymbol{H}_{jk}, \boldsymbol{W}_{jk})$. The conditions include (i) small exposure effect (i.e, $\boldsymbol{\beta}_x \approx \mathbf{0}$) and/or (ii) rare disease prevalence (see Supplementary Material Appendix D for more details). Under these conditions, the calibration parameter estimates under the COCS design are typically less biased.

## 3. Simulation Studies.

3.1. *Simulation Setup and Results.* In this section, we first describe how the unobserved biomarker $X_{jk}$, local and reference laboratory measurements $H_{jk,0}$ and $H_{jk,j}$, and the binary disease outcome $Y_{jk}$ were generated. For simplicity, we set $W_{jk}$ to follow one-dimensional normal distribition with mean 0 and variance 1. We generated $(X_{jk}|W_{jk}, H_{jk,0}|W_{jk}, H_{jk,j}|W_{jk})^T$ according to the multivariate normal distribution in (2.6) with $\mu(\boldsymbol{W}_{jk}; \alpha_{0j}, \boldsymbol{\tau}) = \alpha_{0j} + \tau W_{jk}$, where $\alpha_{0j}$'s were generated from $N(6, 0.05)$ and $\tau$ was set as 2. We assumed there were 5 studies, i.e, $M = 5$, each with 1000 individuals. We first assumed a sample size of 100 for each calibration subset. For each study, we considered disease prevalences of 5%, 25%, and 50%. The calibration subsamples were randomly selected from both cases and controls for the RSCS calibration design and randomly selected from controls for the COCS design. In the variance-covariance matrix, we set $\sigma_x^2 = 3$ and assumed $\sigma_d^2 \sim \text{Unif}(1.5, 2.5)$, where $d = 0, 1, \ldots, 5$. Note that the Intra-laboratory Correlation Coefficient (ICC), which is $\frac{\sigma_x^2}{\sigma_x^2 + \sigma_d^2}$ for laboratory $d$, in the design above is from 55% to 67%. The binary outcome $Y_{jk}$'s were generated based on

$$\text{logit}(P(Y_{jk} = 1|X_{jk})) = \beta_{0j} + \sum_{l=2}^{5} \mathbb{I}(g_{l-1} \leq X_{jk} < g_l)\beta_{x,l}, \tag{3.1}$$

where the regression coefficients across the biomarker categories are linearly increasing; that is, $\beta_{x,2} = \frac{1}{2}\beta_{x,3}$. We considered four values of $\beta_{x,3}$, including log(1.2),log(1.5), log(2) and log(3). The cut-off points used in (3.1) were 33% and 66% percentiles of $X_{jk}$. Note that we assumed there is a single variable $\boldsymbol{W}_{jk}^*(=W_{jk})$ that is correlated with $X_{jk}$, but not the outcome $Y_{jk}$, in our simulation study.

At each $\beta_{x,3}$ and prevalence considered, we completed 1000 simulation replicates and compared the ECM and CCM with regard to the following operating characteristics: mean

percent bias, mean squared error (MSE), empirical and estimated standard errors (SE and $\widehat{\text{SE}}$), and coverage rate of 95% confidence interval. For purpose of comparison, we considered a naive method as a benchmark, which replaced $X_{jk}$ in model (2.1) with the average of $H_{jk,0}$ and $H_{jk,j}$ if $H_{jk,0}$ was available and with $H_{jk,j}$ otherwise, and fitted this regression model to obtain $\boldsymbol{\beta}_x$.

We first considered the COCS design (Table 1). The naive method performed poorly for percent bias at every OR and prevalence considered, and as the prevalence and OR increased, the coverage rates were more different from 95%. Both ECM and CCM reduced the percent bias, and the ECM typically minimized the percent bias. For $\beta_{x,2}$, the CCM typically halved the percent bias of the naive estimates, but for $\beta_{x,3}$, its percent bias still exceeded -15%. The CCM typically minimized the MSEs of $\hat{\beta}_{x,2}$, while MSEs of the ECM estimates were one or two times larger than those of the CCM estimates; however, the ECM minimized the MSEs of $\hat{\beta}_{x,3}$ in most simulation scenarios. The estimated standard errors by the naive method, CCM and ECM were close to their corresponding empirical standard errors. For the coverage rates of the ECM estimates, the sandwich variance and hessian matrix methods performed quite similarly; both were close to 95% and significantly outperformed the naive method and the CCM. For all the disease prevalences considered, the coverage rates of the CCM estimates improved largely from the naive method estimates, especially for larger prevalences and effect sizes, but were still typically less than 90% except in the scenarios with small ORs ($< 1.5$). The simulation results under RSCS were similar (Table 2).

Next, although $X_{jk}$ was generated based on $W_{jk}$, when fitting model (2.3), we assumed $X_{jk} = \alpha_{0j} + \epsilon_{x_{jk}}$ $(j = 1, \ldots, 5)$; that is, $\boldsymbol{W}_{jk}$ was not in the analysis. As presented in Table 3, the results were comparable to those in Table 1, but the MSEs of the CCM and ECM were larger than those in Table 1. These results indicated that incorporating additional variables associated with $\boldsymbol{X}_{jk}$ into the model for $X_{jk}$ could improve the precisions of the OR estimators.

We implemented several additional simulation experiments under the COCS design; the results are summarized as below: (I) When percentage of subjects involved in the calibration study increased (Supplementary Material Tables S1), the percent bias and coverage rates of the naive estimates were worse, those of the CCM improved, and the improvement in the ECM was less noticable than the CCM. (II) When $\sigma_x^2$ decreased (Supplementary Material Table S2), the percent bias and coverage rate of the CCM improved and those of the ECM did not change much. (III) As measurement errors in all laboratories (i.e for all $\sigma_d^2$'s) decreased (Supplementary Material Table S3), all methods improved in terms of percent biases and MSEs. Moreover, the coverage rates of the CCM improved dramatically as the measurement errors decreased, indicating that the CCM can be regarded as appropriate methods for small measurement errors. (IV) In model (3.1), we introduced covariate $\boldsymbol{Z} = W$ with $\beta_z = 1$, which indicated that $W$ is not only associated with $X$, but also associated with $Y$ directly. The simulation result (Supplementary Material Table S4) showed that the percent biases from the naive method were typically larger, and for the

TABLE 1

*Comparison of operating characteristics under a COCS design for the naive ($\hat{\beta}^{(N)}$), cut-off calibration ($\hat{\beta}^{(C)}$) and exact calibration ($\hat{\beta}^{(E)}$) methods.*

| Disease | | Percent Bias | | | MSE | | | SE×100($\widehat{\text{SE}}$ × 100) | | | Coverage Rate | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Prevalence | $\beta_{x,2}$ | $\hat{\beta}^{(N)}_{x,2}$ | $\hat{\beta}^{(C)}_{x,2}$ | $\hat{\beta}^{(E)}_{x,2}$ | $\hat{\beta}^{(N)}_{x,2}$ | $\hat{\beta}^{(C)}_{x,2}$ | $\hat{\beta}^{(E)}_{x,2}$ | $\hat{\beta}^{(N)}_{x,2}$ | $\hat{\beta}^{(C)}_{x,2}$ | $\hat{\beta}^{(E)}_{x,2}(\hat{\beta}^{(E)}_{x,2},\hat{\beta}^{(E*)}_{x,2})$ | $\hat{\beta}^{(N)}_{x,2}$ | $\hat{\beta}^{(C)}_{x,2}$ | $\hat{\beta}^{(E)}_{x,2}$ | $\hat{\beta}^{(E*)}_{x,2}$ |
| 5% | $\frac{1}{2}\log(1.2)$ | -33.4 | -5.6 | -3.1 | 0.028 | 0.026 | 0.079 | 16.4(16.5) | 16.2(16.2) | 28.0(28.0,27.8) | 95.2 | 94.8 | 96.3 | 96.3 |
| | $\frac{1}{2}\log(1.5)$ | -25.2 | -13.1 | -2.5 | 0.033 | 0.029 | 0.085 | 17.3(16.8) | 16.9(16.5) | 29.2(29.0,28.8) | 95.1 | 93.9 | 96.3 | 96.2 |
| | $\frac{1}{2}\log(2)$ | -25.5 | -11.4 | -1.1 | 0.037 | 0.032 | 0.090 | 17.2(17.1) | 17.3(16.8) | 30.1(30.0,29.8) | 92.5 | 94.9 | 95.6 | 95.6 |
| | $\frac{1}{2}\log(3)$ | -22.2 | -13.6 | -3.8 | 0.049 | 0.038 | 0.104 | 18.6(17.6) | 17.9(17.2) | 32.1(31.8,31.6) | 88.1 | 91.9 | 96.4 | 96.3 |
| 25% | $\frac{1}{2}\log(1.2)$ | -36.1 | -5.6 | 4.9 | 0.008 | 0.007 | 0.019 | 8.1(8.2) | 8.1(8.0) | 13.7(13.7,13.6) | 92.7 | 95.4 | 96.0 | 95.9 |
| | $\frac{1}{2}\log(1.5)$ | -30.3 | -13.2 | -2.6 | 0.011 | 0.007 | 0.019 | 8.3(8.3) | 8.0(8.0) | 13.8(14.0,13.8) | 88.0 | 93.4 | 95.0 | 94.9 |
| | $\frac{1}{2}\log(2)$ | -26.1 | -13.8 | -0.4 | 0.015 | 0.009 | 0.019 | 8.5(8.4) | 8.4(8.2) | 13.8(14.3,14.1) | 80.6 | 89.8 | 96.2 | 95.8 |
| | $\frac{1}{2}\log(3)$ | -24.6 | -13.1 | -0.2 | 0.026 | 0.012 | 0.022 | 8.6(8.6) | 8.5(8.4) | 14.8(14.8,14.6) | 66.1 | 86.3 | 94.9 | 94.4 |
| 50% | $\frac{1}{2}\log(1.2)$ | -42.3 | -4.8 | 2.9 | 0.006 | 0.005 | 0.014 | 6.9(7.1) | 6.9(6.9) | 11.6(11.7,11.5) | 92.7 | 95.3 | 95.4 | 94.9 |
| | $\frac{1}{2}\log(1.5)$ | -33.6 | -12.1 | -1.4 | 0.009 | 0.005 | 0.012 | 6.7(7.1) | 6.7(6.9) | 11.0(11.7,11.5) | 85.5 | 93.9 | 97.3 | 97.2 |
| | $\frac{1}{2}\log(2)$ | -28.3 | -13.7 | 0.1 | 0.015 | 0.007 | 0.013 | 7.1(7.1) | 6.9(6.9) | 11.5(11.8,11.6) | 71.1 | 89.3 | 95.5 | 95.3 |
| | $\frac{1}{2}\log(3)$ | -26.8 | -15.5 | 0.5 | 0.027 | 0.012 | 0.014 | 7.0(7.1) | 6.7(6.9) | 11.8(12.0,11.7) | 45.9 | 77.3 | 95.6 | 95.2 |
| Disease | | Percent Bias | | | MSE | | | SE×100($\widehat{\text{SE}}$ × 100) | | | Coverage Rate | | | |
| Prevalence | $\beta_{x,3}$ | $\hat{\beta}^{(N)}_{x,3}$ | $\hat{\beta}^{(C)}_{x,3}$ | $\hat{\beta}^{(E)}_{x,3}$ | $\hat{\beta}^{(N)}_{x,3}$ | $\hat{\beta}^{(C)}_{x,3}$ | $\hat{\beta}^{(E)}_{x,3}$ | $\hat{\beta}^{(N)}_{x,3}$ | $\hat{\beta}^{(C)}_{x,3}$ | $\hat{\beta}^{(E)}_{x,3}(\hat{\beta}^{(E)}_{x,3},\hat{\beta}^{(E*)}_{x,3})$ | $\hat{\beta}^{(N)}_{x,3}$ | $\hat{\beta}^{(C)}_{x,3}$ | $\hat{\beta}^{(E)}_{x,3}$ | $\hat{\beta}^{(E*)}_{x,3}$ |
| 5% | log(1.2) | -23.3 | -15.7 | 2.1 | 0.026 | 0.027 | 0.035 | 15.7(15.6) | 16.3(16.3) | 18.7(19.1,19.0) | 94.3 | 95.1 | 96.2 | 96.1 |
| | log(1.5) | -23.8 | -17.1 | 0.9 | 0.035 | 0.033 | 0.038 | 15.9(15.7) | 16.7(16.4) | 194(19.3,19.3) | 90.0 | 93.3 | 95.1 | 95.0 |
| | log(2) | -24.3 | -17.3 | 0.7 | 0.053 | 0.044 | 0.041 | 15.6(15.7) | 17.2(16.4) | 20.2(19.9,19.8) | 81.1 | 87.0 | 94.9 | 94.8 |
| | log(3) | -24.7 | -18.1 | -0.2 | 0.100 | 0.068 | 0.044 | 16.1(16.0) | 16.9(16.6) | 21.0(21.3,21,2) | 57.7 | 75.7 | 95.5 | 95.3 |
| 25% | log(1.2) | -20.8 | -14.3 | 3.8 | 0.007 | 0.007 | 0.009 | 7.7(7.8) | 8.1(8.2) | 9.4(9.5,9.5) | 93.0 | 93.5 | 95.3 | 95.3 |
| | log(1.5) | -24.1 | -16.9 | 0.1 | 0.016 | 0.012 | 0.009 | 7.9(7.8) | 8.4(8.2) | 9.7(9.6,9.5) | 77.8 | 86.2 | 93.7 | 93.6 |
| | log(2) | -24.7 | -17.9 | -0.6 | 0.037 | 0.023 | 0.011 | 8.6(7.9) | 9.0(8.3) | 10.5(9.9,9.8) | 42.4 | 67.2 | 93.4 | 93.1 |
| | log(3) | -24.7 | -17.5 | 0.1 | 0.080 | 0.044 | 0.011 | 8.3(8.0) | 8.6(8.3) | 10.4(10.3,10.1) | 8.1 | 37.1 | 94.7 | 94.0 |
| 50% | log(1.2) | -25.7 | -19.1 | -2.9 | 0.007 | 0.006 | 0.007 | 6.6(6.8) | 7.0(7.1) | 8.1(8.2,8.2) | 89.9 | 92.8 | 95.5 | 95.5 |
| | log(1.5) | -24.2 | -17.2 | -0.6 | 0.014 | 0.010 | 0.007 | 6.6(6.8) | 7.0(7.1) | 8.1(8.2,8.2) | 70.8 | 83.5 | 95.5 | 95.3 |
| | log(2) | -23.7 | -16.9 | 0.3 | 0.032 | 0.019 | 0.008 | 7.0(6.8) | 7.4(7.2) | 8.7(8.4,8.3) | 33.5 | 62.8 | 93.6 | 93.3 |
| | log(3) | -24.3 | -17.5 | -0.0 | 0.076 | 0.042 | 0.008 | 6.8(6.9) | 7.3(7.2) | 8.7(8.7,8.6) | 3.1 | 24.6 | 95.4 | 95.0 |

NOTE: Percent bias and MSE were computed by averaging $(\hat{\beta}-\beta)/\beta$ and $(\beta-\hat{\beta})^2$ over 1000 simulations. Empirical standard error (SE) is the square root of the empirical variance over all replicates. Coverage rate represents the coverage of a 95% confidence interval. We used the sandwich variances for the $\widehat{\text{SE}}$'s and confidence intervals of the naive and cut-off calibration methods. For the exact calibration method, we applied both the sandwich and hessian matrix approaches for estimating the variances in the coverage rate and estimated standard error evaluation, denoting by $\hat{\beta}^{(E)}$ and $\hat{\beta}^{(E*)}$ in the "Coverage Rate" columns respectively. In the SE×100($\widehat{\text{SE}}$ × 100) section, the numbers in the brackets denote the average estimated standard error ($\widehat{\text{SE}}$) of $\hat{\beta}$ over all replicates; for the exact calibration method, the first and second number denote the $\widehat{\text{SE}}$ by standard variance and pseudo-likelihood hessian matrix methods respectively.

CCM, the impacts varied with $\beta_{x,2}$ and $\beta_{x,3}$. The result from the ECM was only slightly affected by this change, which still provided excellent point and interval estimates. (V) In the above simulation scenarios, we assumed $\alpha_{0j} \sim N(6, 0.05)$ $(j = 1, \ldots, M)$, where $\alpha_{0j}$ among all the studies were quite similar. In this simulation study, we considered the situations where $\alpha_{0j}$ were quite different from each other. Specifically, we set $\alpha_{0j} = 2j$ for $j = 1, \ldots, 5$. As shown in the Supplementary Material Table S5, the performance of the methods were similar to the scenarios where $\alpha_{0j}$'s were similar. (VI) We applied the Full Calibration method in Sloan et al. (2019). Specifically, we fitted $H_{jk,0} = \eta_j + \gamma_j H_{jk,j} + \tilde{\epsilon}_j$ $(j = 1, \ldots, M)$ for the calibration subset in each study using the ordinary least square method, and obtained the estimated reference lab value, $\hat{H}_{jk,0}$, for all individuals. Next, we fitted the logistic regression model (2.1) by substituting $X_{jk}$ with $\hat{H}_{jk,0}$, and denotes the corresponding estimates as $\hat{\boldsymbol{\beta}}^{(F)}_x$. The variance of $\hat{\boldsymbol{\beta}}^{(F)}_x$ was estimated using a sandwich variance method. As presented in Supplementary Material Table S6, this method could reduce the percent bias and improve the coverage rate in comparison with the naive method, but still performed worse than the CCM and ECM. (VII) The ECM showed good robustness

TABLE 2

*Comparison of operating characteristics under a RSCS design for the naive ($\hat{\beta}^{(N)}$), cut-off calibration ($\hat{\beta}^{(C)}$) and exact calibration ($\hat{\beta}^{(E)}$) methods.*

| Disease | | Percent Bias | | | MSE | | | SE×100($\widehat{\text{SE}}$ × 100) | | | Coverage Rate | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Prevalence | $\beta_{x,2}$ | $\hat{\beta}^{(N)}_{x,2}$ | $\hat{\beta}^{(C)}_{x,2}$ | $\hat{\beta}^{(E)}_{x,2}$ | $\hat{\beta}^{(N)}_{x,2}$ | $\hat{\beta}^{(C)}_{x,2}$ | $\hat{\beta}^{(E)}_{x,2}$ | $\hat{\beta}^{(N)}_{x,2}$ | $\hat{\beta}^{(C)}_{x,2}$ | $\hat{\beta}^{(E)}_{x,2}(\hat{\beta}^{(E)}_{x,2},\hat{\beta}^{(E*)}_{x,2})$ | $\hat{\beta}^{(N)}_{x,2}$ | $\hat{\beta}^{(C)}_{x,2}$ | $\hat{\beta}^{(E)}_{x,2}$ | $\hat{\beta}^{(E*)}_{x,2}$ |
| 5% | $\frac{1}{2}\log(1.2)$ | -24.8 | -18.8 | -3.1 | 0.027 | 0.027 | 0.074 | 16.3(16.5) | 16.3(16.2) | 27.2(27.7,27.5) | 95.5 | 94.8 | 96.2 | 95.9 |
| | $\frac{1}{2}\log(1.5)$ | -21.6 | -15.1 | -4.2 | 0.031 | 0.027 | 0.086 | 17.0(16.8) | 16.1(16.4) | 29.3(28.3,28.2) | 94.0 | 95.7 | 94.9 | 94.8 |
| | $\frac{1}{2}\log(2)$ | -18.3 | -13.3 | -0.3 | 0.035 | 0.033 | 0.095 | 17.7(17.1) | 17.7(16.8) | 30.8(29.4,29.2) | 92.5 | 93.0 | 95.3 | 95.2 |
| | $\frac{1}{2}\log(3)$ | -18.6 | -11.2 | 1.9 | 0.042 | 0.035 | 0.097 | 17.8(17.6) | 17.7(17.2) | 31.1(31.0,30.8) | 89.1 | 93.7 | 96.1 | 95.8 |
| 25% | $\frac{1}{2}\log(1.2)$ | -20.2 | -14.7 | 1.9 | 0.007 | 0.007 | 0.018 | 8.0(8.2) | 8.2(8.0) | 13.4(13.5,13.5) | 94.5 | 94.3 | 94.9 | 94.8 |
| | $\frac{1}{2}\log(1.5)$ | -24.2 | -16.1 | 0.6 | 0.010 | 0.007 | 0.017 | 8.5(8.3) | 7.9(8.1) | 13.2(13.7,13.6) | 90.6 | 94.1 | 95.7 | 95.5 |
| | $\frac{1}{2}\log(2)$ | -22.0 | -15.5 | 0.6 | 0.013 | 0.010 | 0.021 | 8.4(8.6) | 8.5(8.4) | 14.5(14.6,14.5) | 84.4 | 88.4 | 94.5 | 94.4 |
| | $\frac{1}{2}\log(3)$ | -22.1 | -15.5 | -1.9 | 0.022 | 0.014 | 0.022 | 8.6(8.4) | 8.5(8.4) | 14.7(14.5,14.5) | 70.5 | 83.8 | 94.3 | 93.9 |
| 50% | $\frac{1}{2}\log(1.2)$ | -25.0 | -16.0 | 1.8 | 0.005 | 0.005 | 0.014 | 7.0(7.1) | 6.9(6.9) | 11.6(11.6,11.5) | 94.2 | 94.4 | 94.9 | 94.7 |
| | $\frac{1}{2}\log(1.5)$ | -22.9 | -16.0 | 3.1 | 0.007 | 0.005 | 0.012 | 6.8(7.1) | 6.5(6.3) | 11.1(11.6,11.5) | 91.7 | 93.1 | 95.3 | 95.3 |
| | $\frac{1}{2}\log(2)$ | -23.9 | -17.3 | -0.6 | 0.012 | 0.008 | 0.014 | 7.2(7.1) | 6.9(6.9) | 11.7(11.6,11.5) | 78.0 | 86.2 | 95.1 | 94.9 |
| | $\frac{1}{2}\log(3)$ | -24.6 | -17.8 | -0.4 | 0.023 | 0.014 | 0.013 | 7.0(7.1) | 6.8(7.0) | 11.3(11.8,11.7) | 54.2 | 71.5 | 95.4 | 95.2 |
| Disease | | Percent Bias | | | MSE | | | SE×100($\widehat{\text{SE}}$ × 100) | | | Coverage Rate | | | |
| Prevalence | $\beta_{x,3}$ | $\hat{\beta}^{(N)}_{x,3}$ | $\hat{\beta}^{(C)}_{x,3}$ | $\hat{\beta}^{(E)}_{x,3}$ | $\hat{\beta}^{(N)}_{x,3}$ | $\hat{\beta}^{(C)}_{x,3}$ | $\hat{\beta}^{(E)}_{x,3}$ | $\hat{\beta}^{(N)}_{x,3}$ | $\hat{\beta}^{(C)}_{x,3}$ | $\hat{\beta}^{(E)}_{x,3}(\hat{\beta}^{(E)}_{x,3},\hat{\beta}^{(E*)}_{x,3})$ | $\hat{\beta}^{(N)}_{x,3}$ | $\hat{\beta}^{(C)}_{x,3}$ | $\hat{\beta}^{(E)}_{x,3}$ | $\hat{\beta}^{(E*)}_{x,3}$ |
| 5% | log(1.2) | -26.3 | -18.5 | -1.1 | 0.026 | 0.027 | 0.034 | 15.4(15.6) | 16.0(16.4) | 18.5(19.0,19.0) | 94.4 | 94.6 | 95.5 | 95.5 |
| | log(1.5) | -23.1 | -16.9 | 0.8 | 0.036 | 0.034 | 0.041 | 16.4(15.7) | 17.1(16.5) | 20.2(19.2,19.2) | 89.1 | 92.9 | 94.8 | 94.9 |
| | log(2) | -24.2 | -17.9 | -0.5 | 0.055 | 0.044 | 0.041 | 16.4(15.8) | 17.0(16.5) | 20.3(19.7,19.7) | 77.8 | 87.3 | 94.8 | 94.6 |
| | log(3) | -23.3 | -16.8 | 1.2 | 0.092 | 0.063 | 0.045 | 16.2(16.0) | 17.1(16.6) | 21.2(21.0,20.9) | 61.9 | 78.2 | 95.5 | 95.4 |
| 25% | log(1.2) | -23.2 | -17.3 | -0.8 | 0.008 | 0.008 | 0.009 | 7.8(7.8) | 8.2(8.2) | 9.4(9.5,9.4) | 92.0 | 92.8 | 95.2 | 95.2 |
| | log(1.5) | -23.9 | -17.1 | -0.4 | 0.016 | 0.012 | 0.010 | 8.0(7.8) | 8.4(8.2) | 9.9(9.5,9.5) | 76.1 | 86.5 | 94.3 | 94.3 |
| | log(2) | -23.9 | -17.1 | -0.1 | 0.034 | 0.021 | 0.010 | 8.2(7.9) | 8.5(8.3) | 9.9(9.7,9.7) | 44.7 | 69.6 | 94.9 | 94.8 |
| | log(3) | -24.4 | -17.7 | -0.4 | 0.078 | 0.045 | 0.010 | 8.2(8.0) | 8.4(8.4) | 10.2(10.2,10.0) | 9.4 | 36.3 | 94.1 | 94.0 |
| 50% | log(1.2) | -22.8 | -15.7 | 0.5 | 0.006 | 0.006 | 0.007 | 6.8(6.8) | 7.2(7.1) | 8.3(8.2,8.2) | 90.4 | 92.6 | 94.0 | 94.0 |
| | log(1.5) | -23.8 | -17.3 | -0.5 | 0.014 | 0.010 | 0.007 | 6.6(6.8) | 7.0(6.8) | 8.1(8.2,8.2) | 71.5 | 82.9 | 94.8 | 94.7 |
| | log(2) | -24.1 | -17.4 | -0.5 | 0.033 | 0.020 | 0.007 | 6.8(6.8) | 7.2(7.2) | 8.2(8.4,8.3) | 31.5 | 60.6 | 96.1 | 95.8 |
| | log(3) | -24.0 | -17.3 | 0.1 | 0.074 | 0.041 | 0.007 | 6.9(6.9) | 7.2(7.2) | 8.6(8.7,8.5) | 3.1 | 25.0 | 94.8 | 94.4 |

NOTE: Percent bias and MSE were computed by averaging $(\hat{\beta}-\beta)/\beta$ and $(\beta-\hat{\beta})^2$ over 1000 simulations. Empirical standard error (SE) is the square root of the empirical variance over all replicates. Coverage rate represents the coverage of a 95% confidence interval. We used the sandwich variances for the $\widehat{\text{SE}}$'s and confidence intervals of the naive and cut-off calibration methods. For the exact calibration method, we applied both the sandwich and hessian matrix approaches for estimating the variances in the coverage rate and estimated standard error evaluation, denoting by $\hat{\beta}^{(E)}$ and $\hat{\beta}^{(E*)}$ in the "Coverage Rate" columns respectively. In the SE×100($\widehat{\text{SE}}$ × 100) section, the numbers in the brackets denote the average estimated standard error ($\widehat{\text{SE}}$) of $\hat{\beta}$ over all replicates; for the exact calibration method, the first and second number denote the $\widehat{\text{SE}}$ by standard variance and pseudo-likelihood hessian matrix methods respectively.

in presence of negative variances in term of the percent bias (see Supplementary Material Appendix E and Tables S7–S9); however, we need to apply the hessian matrix method on estimation of the confidence intervals of $\boldsymbol{\beta}_x$ since sandwich method could significantly overestimate the standard error of $\boldsymbol{\beta}_x$. (VIII) The ECM and CCM showed robustness with regard to relatively small sample sizes in both the contributed studies and the calibration subsets. With 100, 200 or 500 sample size in each of the 5 studies and 10% included in the calibration subset, the 95% coverage rates based on the pseudo-likelihood method ranged from 94.6% to 96.9%, and those based on the sandwich method ranged from 95.1% to 98.1%.(Supplementary Material Table S10).

To summarize, in consideration of the percent bias and confidence interval, the ECM had significant advantages over the CCM and naive method. Besides, for rare disease prevalence and small effect size, the CCM could be an alternative method due to its relative accurate confidence interval coverage rate, easy implementation, and acceptable percent bias. In contrast, the naive method was undesirable because it was heavily biased in most simulation scenarios.

TABLE 3

*Comparison of operating characteristics under a COCS design for naive ($\hat{\beta}^{(N)}$), cut-off calibration ($\hat{\beta}^{(C)}$) and exact calibration ($\hat{\beta}^{(E)}$) methods, where the model for $X_{jk}$ was assumed as $X_{jk} = \alpha_{0j} + \epsilon_{x_{jk}}$, $j = 1, \ldots, 5$.*

| Disease | | Percent Bias | | | MSE | | | SE×100($\widehat{\text{SE}} \times 100$) | | | Coverage Rate | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Prevalence | $\beta_{x,2}$ | $\hat{\beta}^{(N)}_{x,2}$ | $\hat{\beta}^{(C)}_{x,2}$ | $\hat{\beta}^{(E)}_{x,2}$ | $\hat{\beta}^{(N)}_{x,2}$ | $\hat{\beta}^{(C)}_{x,2}$ | $\hat{\beta}^{(E)}_{x,2}$ | $\hat{\beta}^{(N)}_{x,2}$ | $\hat{\beta}^{(C)}_{x,2}$ | $\hat{\beta}^{(E)}_{x,2}(\hat{\beta}^{(E)}_{x,2},\hat{\beta}^{(E*)}_{x,2})$ | $\hat{\beta}^{(N)}_{x,2}$ | $\hat{\beta}^{(C)}_{x,2}$ | $\hat{\beta}^{(E)}_{x,2}$ | $\hat{\beta}^{(E*)}_{x,2}$ |
| 5% | $\frac{1}{2}\log(1.2)$ | -29.3 | -2.2 | 4.6 | 0.029 | 0.025 | 0.100 | 16.8(16.5) | 15.9(16.1) | 31.6(32.0,31.3) | 93.8 | 95.5 | 96.2 | 96.2 |
| | $\frac{1}{2}\log(1.5)$ | -28.9 | -14.2 | -8.6 | 0.032 | 0.029 | 0.105 | 17.0(16.8) | 16.8(16.4) | 32.4(33.0,32.1) | 93.4 | 93.7 | 96.7 | 96.7 |
| | $\frac{1}{2}\log(2)$ | -22.4 | -10.9 | 1.8 | 0.036 | 0.032 | 0.119 | 17.4(17.1) | 17.4(16.8) | 34.4(35.2,33.3) | 92.6 | 93.7 | 96.1 | 96.0 |
| | $\frac{1}{2}\log(3)$ | -23.3 | -12.9 | 0.1 | 0.050 | 0.038 | 0.134 | 18.2(17.6) | 18.2(17.4) | 36.6(36.9,35.7) | 87.8 | 92.0 | 96.3 | 96.2 |
| 25% | $\frac{1}{2}\log(1.2)$ | -36.7 | -5.7 | -0.9 | 0.008 | 0.007 | 0.023 | 8.3(8.2) | 8.1(8.0) | 15.0(15.7,15.1) | 92.1 | 95.9 | 96.5 | 96.5 |
| | $\frac{1}{2}\log(1.5)$ | -27.6 | -12.0 | 4.3 | 0.010 | 0.007 | 0.022 | 8.3(8.3) | 8.0(8.1) | 15.0(16.1,15.3) | 90.0 | 93.5 | 96.3 | 96.2 |
| | $\frac{1}{2}\log(2)$ | -26.7 | -14.8 | -1.5 | 0.015 | 0.009 | 0.024 | 8.2(8.4) | 8.3(8.2) | 15.5(16.6,15.7) | 81.7 | 90.0 | 95.4 | 95.4 |
| | $\frac{1}{2}\log(3)$ | -23.3 | -13.9 | 1.6 | 0.023 | 0.013 | 0.025 | 8.3(8.6) | 8.2(8.4) | 15.8(17.7,16.3) | 69.5 | 85.8 | 95.9 | 95.9 |
| 50% | $\frac{1}{2}\log(1.2)$ | -40.9 | 3.4 | 7.6 | 0.006 | 0.005 | 0.016 | 7.1(7.1) | 6.7(6.9) | 12.5(13.9,12.7) | 92.6 | 96.2 | 96.4 | 96.4 |
| | $\frac{1}{2}\log(1.5)$ | -30.9 | -9.9 | 1.7 | 0.009 | 0.005 | 0.016 | 7.3(7.1) | 7.0(6.9) | 12.6(13.9,12.7) | 85.9 | 93.5 | 95.4 | 95.4 |
| | $\frac{1}{2}\log(2)$ | -27.7 | -13.9 | 0.3 | 0.014 | 0.007 | 0.015 | 7.1(7.1) | 6.8(6.9) | 12.2(14.4,12.7) | 70.9 | 89.5 | 96.0 | 96.0 |
| | $\frac{1}{2}\log(3)$ | -26.4 | -16.1 | 0.9 | 0.026 | 0.013 | 0.017 | 7.4(7.1) | 7.0(7.0) | 12.9(14.0,12.8) | 47.6 | 75.7 | 94.4 | 94.4 |
| Disease | | Percent Bias | | | MSE | | | SE×100($\widehat{\text{SE}} \times 100$) | | | Coverage Rate | | | |
| Prevalence | $\beta_{x,3}$ | $\hat{\beta}^{(N)}_{x,3}$ | $\hat{\beta}^{(C)}_{x,3}$ | $\hat{\beta}^{(E)}_{x,3}$ | $\hat{\beta}^{(N)}_{x,3}$ | $\hat{\beta}^{(C)}_{x,3}$ | $\hat{\beta}^{(E)}_{x,3}$ | $\hat{\beta}^{(N)}_{x,3}$ | $\hat{\beta}^{(C)}_{x,3}$ | $\hat{\beta}^{(E)}_{x,3}(\hat{\beta}^{(E)}_{x,3},\hat{\beta}^{(E*)}_{x,3})$ | $\hat{\beta}^{(N)}_{x,3}$ | $\hat{\beta}^{(C)}_{x,3}$ | $\hat{\beta}^{(E)}_{x,3}$ | $\hat{\beta}^{(E*)}_{x,3}$ |
| 5% | $\log(1.2)$ | -23.8 | -21.0 | 0.8 | 0.027 | 0.031 | 0.043 | 15.9(15.6) | 17.3(16.6) | 20.9(20.1,20.0) | 93.5 | 92.6 | 94.0 | 94.1 |
| | $\log(1.5)$ | -24.9 | -21.0 | -0.9 | 0.034 | 0.034 | 0.040 | 15.4(15.6) | 16.5(16.6) | 19.9(20.4,20.1) | 90.2 | 92.2 | 95.5 | 95.4 |
| | $\log(2)$ | -22.8 | -17.9 | 2.8 | 0.051 | 0.046 | 0.048 | 16.2(15.7) | 17.4(16.7) | 21.8(21.9,20.8) | 82.2 | 87.0 | 94.3 | 94.2 |
| | $\log(3)$ | -25.3 | -20.7 | 0.7 | 0.104 | 0.082 | 0.055 | 16.3(16.0) | 17.5(17.0) | 23.4(23.7,22.6) | 58.7 | 70.5 | 95.2 | 95.3 |
| 25% | $\log(1.2)$ | -25.6 | -21.1 | -2.6 | 0.009 | 0.009 | 0.010 | 8.0(7.8) | 8.5(8.3) | 10.2(9.9,9.9) | 89.1 | 91.0 | 95.2 | 95.2 |
| | $\log(1.5)$ | -23.8 | -19.4 | 0.5 | 0.015 | 0.013 | 0.010 | 7.8(7.8) | 8.3(8.3) | 10.2(10.2,9.9) | 77.4 | 84.1 | 94.4 | 94.3 |
| | $\log(2)$ | -24.0 | -19.8 | 0.3 | 0.034 | 0.026 | 0.012 | 7.8(7.9) | 8.4(8.4) | 10.8(10.8,10.1) | 43.2 | 62.1 | 93.7 | 93.7 |
| | $\log(3)$ | -24.2 | -19.9 | 0.8 | 0.077 | 0.055 | 0.014 | 7.8(7.9) | 8.4(8.4) | 11.8(10.8,10.1) | 9.6 | 27.0 | 92.8 | 92.8 |
| 50% | $\log(1.2)$ | -22.4 | -18.2 | 1.6 | 0.006 | 0.006 | 0.007 | 6.9(6.8) | 7.2(7.2) | 8.6(8.6,8.5) | 90.3 | 92.8 | 94.5 | 94.5 |
| | $\log(1.5)$ | -24.1 | -19.9 | -0.9 | 0.014 | 0.012 | 0.007 | 6.7(6.8) | 7.3(7.2) | 8.7(8.9,8.5) | 70.6 | 79.5 | 94.9 | 94.9 |
| | $\log(2)$ | -23.0 | -18.7 | 1.0 | 0.030 | 0.022 | 0.009 | 6.8(6.8) | 7.2(7.2) | 9.2(9.3,8.6) | 35.1 | 56.1 | 93.3 | 93.3 |
| | $\log(3)$ | -24.1 | -19.8 | 0.3 | 0.075 | 0.053 | 0.010 | 6.7(6.9) | 7.3(7.3) | 10.2(10.6,8.9) | 3.2 | 15.2 | 91.4 | 91.4 |

NOTE: Percent bias and MSE were computed by averaging $(\hat{\beta}-\beta)/\beta$ and $(\beta-\hat{\beta})^2$ over 1000 simulations. Empirical standard error (SE) is the square root of the empirical variance over all replicates. Coverage rate represents the coverage of a 95% confidence interval. We used the sandwich variances for the $\widehat{\text{SE}}$'s and confidence intervals of the naive and cut-off calibration methods. For the exact calibration method, we applied both the sandwich and hessian matrix approaches for estimating the variances in the coverage rate and estimated standard error evaluation, denoting by $\hat{\beta}^{(E)}$ and $\hat{\beta}^{(E*)}$ in the "Coverage Rate" columns respectively. In the SE×100($\widehat{\text{SE}} \times 100$) section, the numbers in the brackets denote the average estimated standard error ($\widehat{\text{SE}}$) of $\hat{\beta}$ over all replicates; for the exact calibration method, the first and second number denote the $\widehat{\text{SE}}$ by standard variance and pseudo-likelihood hessian matrix methods respectively.

3.2. *When X Does not Follow a Normal Distribution.* In the simulation study in Section 3.1, we assumed $\epsilon_{x_{jk}}$ follows a normal distribution. In this section, we compare the three methods when $X_{jk}$ (and thus $\boldsymbol{H}_{jk}$) does not follow a normal distribution. Specifically, we considered two specific distributions for $\epsilon_{x_{jk}}$: (1) Uniform distribution; (2) Skew Normal distribution [Fernández and Steel (1998)]. We first generated $X_{jk}$ assuming $X_{jk} = \alpha_{0j} + \tau W_{jk} + \epsilon_{x_{jk}}$, where $\epsilon_{x_{jk}}$ followed either an uniform or a skew normal distribution, with all the parameters adjusted to satisfying mean 0 and variance 3. For the skew normal distribution, we set the skew parameter $\gamma = 10$, which makes the moment coefficient of skewness approximates 1; and $\alpha_{0j}$, $\tau$ and $W_{jk}$ are the same as those in Section 3.1. We generate $H_{jk,d}$ based on $H_{jk,d} \sim N(X_{jk}, \sigma_d^2)$, $\quad j = 0, 1, \ldots, 5$. All the other parameters are also same as those in Section 3.1.

Table S11 in the supplementary material showed the simulation results. The results were similar to the scenarios when $\epsilon_{x_{jk}}$ followed a normal distribution; the ECM outperformed

the other two methods in the aspects of percent bias and coverage rate, and the CCM achieved relatively small percent bias under rare disease prevalences and small effect sizes. In Section 5, we will provide a theoretical adjustment of the proposed methods for the scenarios when $X$ does not follow a normal distribution.

**4. Applied Example.** As an illustrative example, we applied our methods to investigate the association between circulating vitamin D (25(OH)D) and risk of colorectal cancer. Specifically, our example is based on the Nurses' Health Study (NHS) [Colditz, Manson and Hankinson (1997)] and Health Professionals Follow-up Study (HPFS) [Choi et al. (2005)], two large cohort studies in the United States. The NHS enrolled 121,701 female nurses, aged 30 to 55 in 1976. The HPFS was established in 1986 with the enrollment of 51,529 male health professionals, aged 40 to 75 years in 1986. From 1989 to 1995, both studies selected a subset of participants, obtained their blood samples and then completed assays for a host of biomarkers, which included 25(OH)D. Individuals who did not have colorectal cancer outcome or 25(OH)D measurement available were excluded from the pooling analysis. A total of 1876 subjects constituted our pooling analysis (Table 4), which was extracted from the studies above with a nested case-control design (1 to 2 matching). For illustrative purpose, we used the unconditional logistic regression, adjusting for the matching factors by including them in the regression model, to evaluate the biomarker-disease association. A COCS design was implemented by randomly select 29 controls in each study and re-assaying their blood samples at Heartland Assays, LLC (Ames, IA). We will refer these laboratory measurements as reference laboratory measurements.

TABLE 4
*Descriptive characteristics of NHS and HPFS cohorts*

| Study | Cases/Controls (Disease prevalence) | Median 25(OH)D nmol/L (10-90%) | Median age at blood draw, years (10-90%) |
|---|---|---|---|
| NHS | 348/694 (34.0%) | 59 (36,85) | 60 (49,67) |
| HPFS | 267/519 (33.4%) | 70 (45,100) | 68 (53,75) |
| Total | 615/1213 (33.6%) | 64 (40,93) | 62 (50,73) |

NOTE: Quantiles of 25(OH)D levels were obtained from the distribution of local laboratory measurements.

The cut-off points ($< 30$, $[30, 50)$, $[50, 75)$, $\geq 75$ nmol/L) used for 25(OH)D were the IOM Standard [Gail et al. (2016)]. The covariates in model (2.3) for 25(OH)D included week of the year at blood draw (1-52), age of blood draw (ranged 43-82), physical activity (continuous), smoking (ever/never) and BMI (greater or less than 25 kg/m$^2$). Because 25(OH)D is influenced by seasonal variation, we also included a periodic function $\tau_1 \sin(2\pi t/52) + \tau_2 \cos(2\pi t/52) + \tau_3 \sin(4\pi t/52) + \tau_4 \cos(4\pi t/52)$ in model (2.3), where $t$ denotes week of the year at blood draw, to fit the seasonal trend [Gail et al. (2016)]. Since the association of the covariates above with 25(OH)D may change over gender, we fitted model (2.3) separately for NHS and HPFS. For comparison, we considered two other models for 25(OH)D. The first model included study-specific intercepts, seasonal trend and physical activity, and the second model included the study-specific intercepts only. We name them

Model I, II, III for the original and two additional models respectively.

TABLE 5
*Parameter estimates for model (2.3) in the calibration step based on the NHS and HPFS*

| Models | Study | (intercept) | sin $\frac{2\pi}{52}$ | cos $\frac{2\pi}{52}$ | sin $\frac{4\pi}{52}$ | cos $\frac{4\pi}{52}$ | Physical | Age(Blood) | Smoke | BMI |
|---|---|---|---|---|---|---|---|---|---|---|
| Model I | NHS | 52.083* | -5.367* | -1.765 | -0.407 | -0.517 | 0.126* | 0.163 | -0.078* | -4.279* |
| | | (5.757) | (0.964) | (1.038) | (0.904) | (1.060) | (0.032) | (0.101) | (0.033) | (1.399) |
| | HPFS | 72.296* | -7.498* | -6.774* | 1.338 | 0.466 | -0.049* | 0.088 | -0.023 | -4.327* |
| | | (6.571) | (1.076) | (1.133) | (1.089) | (1.072) | (0.022) | (0.094) | (0.043) | (1.506) |
| Model II | NHS | 58.278* | -5.237* | -1.596 | -0.569 | -0.513 | 0.150* | | | |
| | | (0.923) | (0.957) | (1.042) | (0.915) | (1.068) | (0.031) | | | |
| | HPFS | 65.788* | -7.395* | -6.816* | 1.338 | 0.365 | 0.096* | | | |
| | | (1.141) | (1.072) | (1.150) | (1.095) | (1.063) | (0.022) | | | |
| Model III | NHS | 60.860* | | | | | | | | |
| | | (0.702) | | | | | | | | |
| | HPFS | 71.871* | | | | | | | | |
| | | (0.786) | | | | | | | | |

NOTE: 1 The variance estimates of Model I are $\hat{\sigma}_x^2 = 447.387$, $\hat{\sigma}_0^2 = 436.220$, $\hat{\sigma}_1^2 = 35.342$, $\hat{\sigma}_2^2 < 0.001$.
2 The variance estimates of Model II are $\hat{\sigma}_x^2 = 453.336$, $\hat{\sigma}_0^2 = 452.976$, $\hat{\sigma}_1^2 = 37.204$, $\hat{\sigma}_2^2 < 0.001$.
3 The variance estimates of Model III are $\hat{\sigma}_x^2 = 507.491$, $\hat{\sigma}_0^2 = 401.664$, $\hat{\sigma}_1^2 = 8.186$, $\hat{\sigma}_2^2 < 0.001$.

The point estimates with standard errors for the regression coefficients in Model I, II, III are shown in Table 5. We applied constraints $\sigma_x^2 \geq 0$, $\sigma_0^2 \geq 0$, $\sigma_1^2 \geq 0$, and $\sigma_2^2 \geq 0$ to ensure positive variances. In an analysis (not showing here), without these constraints, the estimates $\hat{\sigma}_2^2$ was negative. Finally, we applied the naive method, CCM and ECM to the pooling dataset, adjusting for age when the case control status was determined, physical activity total (continuous), family history of colorectal cancer (yes/no), smoking (ever/never), BMI (greater or less than 25 kg/m$^2$). Because using sandwich variance in the ECM could widen the confidence interval when negative variance exists (see Supplementary Material Appendix E and Tables S7–S9.), we used the pseudo-likelihood hessian matrix method to obtain the estimated variance of $\hat{\boldsymbol{\beta}}_x$. The results are displayed in Table 6. In Table 6, we also provide the OR estimates based on the NHS and HPFS separately, where the local laboratory measurements were treated as gold standard and the potential confounders that had been adjusted were identical with those in pooled analyses. All analytic approaches, except the HPFS-specific analyses, demonstrated that increasing 25(OH)D levels tended to have a protective effect against colorectal cancer, which was not statistically significant. The OR estimates from the pooled analyses provided narrower confidence interval due to larger sample size. The estimated ORs by the CCM and ECM were typically slightly smaller than the naive estimating due to attenuation in the naive estimates.

**5. Adjustment of Methods When $X$ Does not Follow a Normal Distribution.** The normality is assumed in our analytical framework. Both the measurement errors of the local and reference laboratories and the error term in the biomarker-covariates regression are assumed to be normally distributed. The normality of measurement errors is typically a reasonable assumption. However, the biomarker data, $X_{jk}$, may be skewed and need transformation before performing the biomarker-covariates regression [Mitchell et al. (2014)].

TABLE 6

*OR-estimates and 95% confidence interval for the circulating 25(OH)D–colorectal cancer relationship, based on the NHS and HPFS, adjusting for age when the case control status was determined, physical activity total (continuous), family history of colorectal cancer (yes/no), smoking (ever/never), BMI (greater or less than 25 kg/m$^2$).*

| Models | ORs | Single Study (Naive analysis) | | Pooled Analysis | | |
|---|---|---|---|---|---|---|
| | | NHS | HPFS | Naive | CCM | ECM |
| Model I | $\exp(\beta_{x,2})$ | 0.881(0.478,1.626) | 1.114(0.268,4.632) | 0.881 (0.505,1.537) | 0.857 (0.475,1.547) | 0.771 (0.372,1.598) |
| | $\exp(\beta_{x,3})$ | 0.623(0.343,1.131) | 1.272(0.318,5.087) | 0.732 (0.427,1.255) | 0.723 (0.408,1.280) | 0.664 (0.344,1.284) |
| | $\exp(\beta_{x,4})$ | 0.615(0.327,1.157) | 0.945(0.234,3.809) | 0.629 (0.361,1.098) | 0.613 (0.340,1.106) | 0.537 (0.273,1.059) |
| Model II | $\exp(\beta_{x,2})$ | 0.881(0.478,1.626) | 1.114(0.268,4.632) | 0.881 (0.505,1.537) | 0.858 (0.476,1.549) | 0.771 (0.369,1.612) |
| | $\exp(\beta_{x,3})$ | 0.623(0.343,1.131) | 1.272(0.318,5.087) | 0.732 (0.427,1.255) | 0.722 (0.408,1.279) | 0.664 (0.341,1.291) |
| | $\exp(\beta_{x,4})$ | 0.615(0.327,1.157) | 0.945(0.234,3.809) | 0.629 (0.361,1.098) | 0.615 (0.341,1.109) | 0.537 (0.271,1.065) |
| Model III | $\exp(\beta_{x,2})$ | 0.881(0.478,1.626) | 1.114(0.268,4.632) | 0.881 (0.505,1.537) | 0.917 (0.524,1.607) | 0.821 (0.446,1.511) |
| | $\exp(\beta_{x,3})$ | 0.623(0.343,1.131) | 1.272(0.318,5.087) | 0.732 (0.427,1.255) | 0.763 (0.444,1.313) | 0.690 (0.389,1.225) |
| | $\exp(\beta_{x,4})$ | 0.615(0.327,1.157) | 0.945(0.234,3.809) | 0.629 (0.361,1.098) | 0.651 (0.371,1.140) | 0.579 (0.320,1.046) |

NOTE: Model I, II and III here correspond to Table 6. The "Single Study (Naive analysis)" columns denote the results for naive logistic model based on the NHS and HPFS separately.

Under transformation, the biomarker-covariates regression (2.3) becomes

$$T(X_{jk}) = \mu_{X_{jk}|\boldsymbol{W}_{jk}} + \epsilon_{x_{jk}}, \tag{5.1}$$

where $T(.)$ is a given transformation function satisfying regularity conditions and $\epsilon_{x_{jk}} \sim N(0, \sigma_x^2)$. If the measurement error, $\epsilon_{jk,d}$, is small, the transformed measurement, $T(H_{jk,d})$, can be approximated by

$$T(H_{jk,d}) \approx \mu_{X_{jk}|\boldsymbol{W}_{jk}} + \epsilon_{x_{jk}} + S(\mu_{X_{jk}|\boldsymbol{W}_{jk}})\epsilon_{jk,d}, \tag{5.2}$$

where, $S(.)$ denotes $T^{'}T^{-1}(.)$ and $T^{'}(.)$ is the first derivative with respect to the function $T(.)$. See Supplementary Material Appendix F for more details about this approximation. To ease notations, $T(H_{jk,d})$ and $T(X_{jk})$ are abbreviated as $\tilde{H}_{jk,d}$ and $\tilde{X}_{jk}$ henceforth. Based on models (5.1) and (5.2), $(\tilde{X}_{jk}|\boldsymbol{W}_{jk}, \tilde{H}_{jk,0}|\boldsymbol{W}_{jk}, \tilde{H}_{jk,j}|\boldsymbol{W}_{jk})^T$ approximately follows the multivariate normal distribution below

$$\begin{pmatrix} \tilde{X}_{jk}|\boldsymbol{W}_{jk} \\ \tilde{H}_{jk,0}|\boldsymbol{W}_{jk} \\ \tilde{H}_{jk,j}|\boldsymbol{W}_{jk} \end{pmatrix} \dot{\sim} \text{MVN} \left( \begin{pmatrix} \mu_{X_{jk}|\boldsymbol{W}_{jk}} \\ \mu_{X_{jk}|\boldsymbol{W}_{jk}} \\ \mu_{X_{jk}|\boldsymbol{W}_{jk}} \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_x^2 & \sigma_x^2 \\ . & \sigma_x^2 + \phi_{x_{jk}}\sigma_0^2 & \sigma_x^2 \\ . & . & \sigma_x^2 + \phi_{x_{jk}}\sigma_j^2 \end{pmatrix} \right),$$

where $\phi_{x_{jk}} = S^2(\mu_{X_{jk}|\boldsymbol{W}_{jk}})$. Now, for individuals who only have local laboratory measurements,

$$\tilde{X}_{jk}|\boldsymbol{H}_{jk}, \boldsymbol{W}_{jk} \dot{\sim} N\Big(\tilde{\rho}_j \tilde{H}_{jk,j} + (1-\tilde{\rho}_j)\mu_{X_{jk}|\boldsymbol{W}_{jk}}, \tilde{\rho}_j \phi_{x_{jk}} \sigma_j^2\Big), \tag{5.3}$$

where $\tilde{\rho}_j = \frac{\sigma_x^2}{\sigma_x^2 + \phi_{x_{jk}}\sigma_j^2}$, and for individuals in the calibration subset

$$\tilde{X}_{jk}|\boldsymbol{H}_{jk}, \boldsymbol{W}_{jk} \dot{\sim} N\Big(\tilde{\rho}_j^*(w_j \tilde{H}_{jk,j} + (1-w_j)\tilde{H}_{jk,0}) + (1-\tilde{\rho}_j^*)\mu_{X_{jk}|\boldsymbol{W}_{jk}}, \tilde{\rho}_j^* w_j \phi_{x_{jk}} \sigma_j^2\Big), \tag{5.4}$$

where $\tilde{\rho}_j^* = \sigma_x^2/(\sigma_x^2+\phi_{x_{jk}}\sigma_j^2 w_j)$. Hereafter, the mean and standard deviation of $\tilde{X}_{jk}|(\boldsymbol{H}_{jk}, \boldsymbol{W}_{jk})$ are denoted as $\tilde{\mu}_{jk}$ and $\tilde{s}_{jk}$. Given $\tilde{\mu}_{jk}$ and $\tilde{s}_{jk}$, $p_{jk,l}$ in the likelihood contribution $L_{jk}$ can be expressed as

$$\begin{aligned} p_{jk,l} &= P(g_{l-1} \le X_{jk} < g_l | \boldsymbol{H}_{jk}, \boldsymbol{W}_{jk}) \\ &= P(T(g_{l-1}) \le \tilde{X}_{jk} < T(g_l) | \boldsymbol{H}_{jk}, \boldsymbol{W}_{jk}) = \Phi(\tfrac{T(g_l)-\tilde{\mu}_{jk}}{\tilde{s}_{jk}}) - \Phi(\tfrac{T(g_{l-1})-\tilde{\mu}_{jk}}{\tilde{s}_{jk}}). \end{aligned} \tag{5.5}$$

If we have a estimator of $\tilde{\mu}_{jk}$ and $\tilde{s}_{jk}$, we can apply the ECM and CCM in section 2.4 and 2.5 respectively to obtain $\hat{\boldsymbol{\beta}}$ and $\text{var}(\hat{\boldsymbol{\beta}})$. Next, we discuss how to estimate $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}^2$, which, as defined in Section 2.4, are unknown parameters in $\tilde{\mu}_{jk}$ and $\tilde{s}_{jk}$.

Combining (5.1) and (5.2), we have

$$\tilde{H}_{jk,d} = \mu_{X_{jk}|\boldsymbol{W}_{jk}} + \tilde{\delta}_{x_{jk},d}, \tag{5.6}$$

where $\tilde{\delta}_{x_{jk},d} \sim N(0, \sigma_x^2 + \phi_{x_{jk}}\sigma_d^2)$. Its first- and second-order estimating equations now become

$$\begin{aligned} \Psi_{\boldsymbol{\theta}} &= \sum_{j=1}^{M}\sum_{k=1}^{n_j} \frac{d}{d\boldsymbol{\theta}} \begin{bmatrix} \mu(\boldsymbol{W}_{jk};\alpha_{0j},\boldsymbol{\tau}) \\ \mu(\boldsymbol{W}_{jk};\alpha_{0j},\boldsymbol{\tau}) \end{bmatrix}^T \begin{bmatrix} \sigma_x^2+\phi_{x_{jk}}\sigma_0^2 & \sigma_x^2 \\ \sigma_x^2 & \sigma_x^2+\phi_{x_{jk}}\sigma_j^2 \end{bmatrix}^{-1} \begin{bmatrix} e_{jk,0} \\ e_{jk,j} \end{bmatrix} \\ &\quad + \sum_{j=1}^{M}\sum_{k=n_j+1}^{N_j} \frac{d\mu(\boldsymbol{W}_{jk};\alpha_{0j},\boldsymbol{\tau})}{d\boldsymbol{\theta}} (\sigma_x^2+\phi_{x_{jk}}\sigma_j^2)^{-1} e_{jk,j} = \mathbf{0}, \\ \Psi_{\boldsymbol{\sigma}^2} &= \sum_{j=1}^{M}\sum_{k=1}^{n_j} \frac{d}{d\boldsymbol{\sigma}^2} \begin{bmatrix} \sigma_x^2+\phi_{x_{jk}}\sigma_0^2 \\ \sigma_x^2 \\ \sigma_x^2+\phi_{x_{jk}}\sigma_j^2 \end{bmatrix}^T \begin{bmatrix} e_{jk,0}^2-(\sigma_x^2+\phi_{x_{jk}}\sigma_0^2) \\ e_{jk,0}e_{jk,j}-\sigma_x^2 \\ e_{jk,j}^2-(\sigma_x^2+\phi_{x_{jk}}\sigma_j^2) \end{bmatrix} \\ &\quad + \sum_{j=1}^{M}\sum_{k=n_j+1}^{N_j} \frac{d(\sigma_x^2+\phi_{x_{jk}}\sigma_j^2)}{d\boldsymbol{\sigma}^2} \left(e_{jk,j}^2-(\sigma_x^2+\phi_{x_{jk}}\sigma_j^2)\right) = \mathbf{0}. \end{aligned} \tag{5.7}$$

We propose a three-step iteration method to estimate $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}^2$. The three steps in the $m^{th}$ iteration are

- First step: calculate $\hat{\phi}_{x_{jk}}^{(m)}$ by replacing the unknown $\boldsymbol{\theta}$ in $S^2(\mu_{X_{jk}|\boldsymbol{W}_{jk}})$ by $\hat{\boldsymbol{\theta}}^{(m)}$;
- Second step: fix $\boldsymbol{\sigma}^2 = \hat{\boldsymbol{\sigma}}^{2(m)}$ and $\phi_{x_{jk}} = \hat{\phi}_{x_{jk}}^{(m)}$, and obtain $\hat{\boldsymbol{\theta}}^{(m+1)}$ by solving $\Psi_{\boldsymbol{\theta}} = \mathbf{0}$;
- Third step: fix $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(m+1)}$ and $\phi_{x_{jk}} = \hat{\phi}_{x_{jk}}^{(m)}$, and obtain $\hat{\boldsymbol{\sigma}}^{2(m+1)}$ by solving $\Psi_{\boldsymbol{\sigma}^2} = \mathbf{0}$.

The initial value $\hat{\boldsymbol{\sigma}}^{2(0)}$ and convergence criteria can be the same as those in section 2.4, whereas the initial value $\hat{\boldsymbol{\theta}}^{(0)}$ can be set as the Ordinary Least Square estimate of the regression (5.6). The iteration continues until convergence to obtain the final $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\sigma}}^2$. The estimates $\hat{\tilde{\mu}}_{jk}$ and $\hat{\tilde{s}}_{jk}$ can be obtained by plugging $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\sigma}}^2$ into distribution (5.3) or (5.4). With $\hat{\tilde{\mu}}_{jk}$ and $\hat{\tilde{s}}_{jk}$, we can obtain $\hat{p}_{jk,l}$ according to (5.5), and utilize the ECM in section 2.5 and CCM in section 2.6 (with $\hat{X}_{jk} = \hat{\tilde{\mu}}_{jk}$ in likelihood (2.13)) to obtain $\hat{\boldsymbol{\beta}}$ and its standard error.

We evaluated the performance of this method in the following simulation study. We assumed a natural logarithm transformation function, i.e., $T(.) = log(.)$. We first generated $log(X_{jk})$ according to $log(X_{jk}) = \alpha_{0j} + \epsilon_{x_{jk}}$, with $\alpha_{0j} \sim N(4.5, 0.1^2)$, $\sigma_x^2 = 0.01$, and $j = 1, 2 \ldots, 5$. Then, $H_{jk,d}$ was generated from $N(X_{jk}, \sigma_d^2)$ with $\sigma_d^2 \sim Unif(25, 75)$, $j = 0, \ldots, 5$, which resulted in ICC, defined as $\frac{\text{var}(X_{jk})}{\text{var}(X_{jk})+\sigma_d^2}$ for laboratory $d$, ranging from 77% to 91%. All the other simulation parameters and procedures were identical with those in Section 3.1. Showed in Table S12 are the simulation results. Similar to the simulation results in section 3; the ECM outperformed the the naive method and CCM in terms of percent bias and coverage rate, and the CCM produced satisfactory point and interval estimates under rare disease prevalences and small effect sizes.

**6. Discussion.** In this paper, we proposed and evaluated two data analysis methods, the CCM and ECM, for pooling biomarker data under the COCS and RSCS designs. We focus on the population-averaged association of the underlying true biomarker categories with disease outcome, based on the logistic regression with fixed effects. In order to estimate this association for true biomarker levels, calibration studies are required for each lab/study and the between-lab/study variation is taken into account by allowing lab/study-specific measurement error distributions. R functions for the proposed methods are available at https://www.hsph.harvard.edu/molin-wang/software.

Different from previous researches, we do not treat the reference laboratory or any local laboratory as the gold standard. Several practical recommendations and conclusions follow from our work. First, both of the CCM and ECM yield less biased point estimates and more accurate coverage rates of confidence intervals than unadjusted analysis (i.e naive method). Second, under both the COCS and RSCS designs, the ECM provides significantly less biased estimates and more accurate confidence intervals than the CCM since it maximally utilizes available information, but it tends to present larger variance than the CCM. Third, the CCM can yield satisfactory estimates under small measurement errors and/or small exposure effects, but simulation study shows the CCM performs increasingly poorly as exposure effects increase. Fourth, generally, the parameter estimates are asymptotically biased under a COCS design, but the bias is minimal if the exposure effects and/or the disease prevalence are small.

The proposed methods are mainly for epidemiological researches that study disease etiology. In these researches, measurement error-caused bias could lead to misleading scientific conclusions, and thus it is crucial to obtain valid point and interval estimates that take into account measurement errors. In clinical applications, it could be of interest to learn the relationship between the biomarker measurement from the reference or local laboratory and the disease outcome (i.e, $H$–$Y$ association). In this case, under classical additive measurement error model (2.2), the odds ratio with respect to one unit increase in the biomarker measurement from study $j$ and laboratory $d$, denoted as $H_{j,d}$, can be estimated as $\widehat{\text{OR}}_{H_{j,d}}(h) = \frac{[\sum_{l=2}^{G} \hat{\Delta}_{j,k}(h+1,l)]/[1-\sum_{l=2}^{G} \hat{\Delta}_{j,k}(h+1,l)]}{[\sum_{l=2}^{G} \hat{\Delta}_{j,k}(h,l)]/[1-\sum_{l=2}^{G} \hat{\Delta}_{j,k}(h,l)]}$, where $\hat{\Delta}_{j,k}(h,l) =$

$\frac{\exp\{\hat{\beta}_{0j}+\hat{\beta}_{x,l}+\hat{\boldsymbol{\beta}}_z^T\boldsymbol{Z}_{jk}\}}{1+\exp(\hat{\beta}_{0j}+\hat{\beta}_{x,l}+\hat{\boldsymbol{\beta}}_z^T\boldsymbol{Z}_{jk})}\hat{P}(g_{l-1} \le X < g_l|H_{j,d}=h,\boldsymbol{W}_{jk})$, $[\hat{\beta}_{0j},\hat{\boldsymbol{\beta}}_x^T,\hat{\boldsymbol{\beta}}_z^T]^T$ are the estimated parameters in logistic regression (2.1) by ECM or CCM ($\hat{\beta}_{x,1}$ is fixed as 0), and $\hat{P}(g_{l-1} \le X < g_l|H_{j,d}=h,\boldsymbol{W}_{jk})$ is derived in (2.11). If the disease prevalence is low, $\widehat{\mathrm{OR}}_{H_{j,d}}(h)$ can be approximated by $\frac{\sum_{l=2}^{G}\exp(\hat{\beta}_{x,l})\hat{P}(g_{l-1}\le X<g_l|H_{j,d}=h+1,\boldsymbol{W}_{jk})}{\sum_{l=2}^{G}\exp(\hat{\beta}_{x,l})\hat{P}(g_{l-1}\le X<g_l|H_{j,d}=h,\boldsymbol{W}_{jk})}$. See Supplementary Material Appendix G for details, including the variance estimate of $\widehat{\mathrm{OR}}_{H_{j,d}}(h)$.

While our simulation study showed satisfactory performance of the ECM and CCM under small calibration studies (e.g., calibration study sizes 10 and 20 in Supplementary Material Table S10), the size of the calibration subset leading to satisfactory estimates of the biomarker-disease association depends on a series of factors, including variance of the true biomarker data, variance of the measurement error, disease prevalence and magnitude of the exposure effect. With the exposure effect decreasing or disease prevalence increasing, smaller calibration study may be sufficient to provide good performance. Similarly, if the variance of the measurement error is smaller, less subjects in the calibration study subset may be sufficient.

We need to notice that the CCM and naive method can perform quite close when measurement errors are significantly smaller than the errors in the biomarker-covariates regression (i.e, $\sigma_d^2 \ll \sigma_x^2$ for all laboratories). The CCM would fully present its superiority over the naive method when the measurement errors are relatively larger than or on the same level with the errors in the biomarker-covariates regression.

This paper focuses on the classical additive measurement error model where the local and reference laboratory measurements are the summation of underlying true value and random errors [e.g., Wang, Carroll and Liang (1996); Ganguli, Staudenmayer and Wand (2005); Liang et al. (2008); Hu and Ridder (2012)]. Extension of our methods to more flexible measurement error models, for example, those defined by $H = \alpha_d + \beta_d X + \epsilon_d$, where $\alpha_d$ and $\beta_d$ are study-specific intercept and coefficient, is a topic for future research. To summarize, under both the RSCS and COCS designs, the ECM provides smallest percent bias and most accurate confidence interval but relatively larger MSEs. Specifically, the ECM presents significant advantages over others for larger measurement errors and stronger biomarker-disease relationships. Besides, the CCM may be used under rare disease prevalences and small effect sizes. These observations could help researchers in selecting appropriate analyses to fulfill their goals.

**Acknowledgements.** We are grateful to Mitchell H. Gail for his helpful discussions. We also thank the Circulating Biomarkers and Breast and Colorectal Cancer Consortium team (R01CA152071, PI: Stephanie Smith-Warner; Intramural Research Program, Division of Cancer Epidemiology and Genetics, National Cancer Institute: Regina Ziegler) for conducting the calibration study in the vitamin D example. Molin Wang was supported in part by NIH/NCI grant R03CA212799. This work was supported in part by the Intramural Program of the National Cancer Institute, Division of Cancer Epidemiology and Genetics.

## SUPPLEMENTARY MATERIAL

**Supplement A: Appendix and Supplementary Tables**
(http://www.e-publications.org/ims/support/dowload/imsart-ims.zip). The appendixes (Appendixes A–G) and supplementary tables (Tables S1–S12) in this paper.

**Supplement B: R Code**
(https://www.hsph.harvard.edu/molin-wang/software). R code for implementing the methods used in this paper.

## References.


Choi, H. K., Atkinson, K., Karlson, E. W. and Curhan, G. (2005). Obesity, weight change, hypertension, diuretic use, and risk of gout in men: the health professionals follow-up study. *Archives of internal medicine* **165** 742–748.

Colditz, G. A., Manson, J. E. and Hankinson, S. E. (1997). The Nurses' Health Study: 20-year contribution to the understanding of health among women. *J. Womens Health* **6** 49–62.

Crowe, F. L., Appleby, P. N., Travis, R. C., Barnett, M., Brasky, T. M., Bueno-de Mesquita, H. B., Chajes, V., Chavarro, J. E., Chirlaque, M.-D., English, D. R. et al. (2014). Circulating fatty acids and prostate cancer risk: individual participant meta-analysis of prospective studies. *JNCI: Journal of the National Cancer Institute* **106**.

El Leithy, H. A., Wahed, Z. A. A. and Abdallah, M. S. (2016). On non-negative estimation of variance components in mixed linear models. *Journal of advanced research* **7** 59–68.

Fernández, C. and Steel, M. F. (1998). On Bayesian modeling of fat tails and skewness. *J. Am. Stat. Assoc.* **93** 359–371.

Fletcher, D. and Underwood, A. (2002). How to cope with negative estimates of components of variance in ecological field studies. *Journal of Experimental Marine Biology and Ecology* **273** 89–95.

Gail, M. H., Wu, J., Wang, M., Yaun, S.-S., Cook, N. R., Eliassen, A. H., McCullough, M. L., Yu, K., Zeleniuch-Jacquotte, A., Smith-Warner, S. A. et al. (2016). Calibration and seasonal adjustment for matched case–control studies of vitamin D and cancer. *Stat. Med.* **35** 2133–2148.

Gallicchio, L., Helzlsouer, K. J., Chow, W.-H., Freedman, D. M., Hankinson, S. E., Hartge, P., Hartmuller, V., Harvey, C., Hayes, R. B., Horst, R. L. et al. (2010). Circulating 25-hydroxyvitamin D and the risk of rarer cancers: design and methods of the Cohort Consortium Vitamin D Pooling Project of Rarer Cancers. *Am. J. Epidemiol.* **172** 10–20.

Ganguli, B., Staudenmayer, J. and Wand, M. (2005). Additive models with predictors subject to measurement error. *Australian & New Zealand Journal of Statistics* **47** 193–202.

Godambe, V. P. (1991). *Estimating functions*. Oxford University Press.

Hu, Y. and Ridder, G. (2012). Estimation of nonlinear models with mismeasured regressors using marginal information. *Journal of Applied Econometrics* **27** 347–385.

Key, T. J., Appleby, P. N., Allen, N. E. and Reeves, G. K. (2010). Pooling biomarker data from different studies of disease risk, with a focus on endogenous hormones. *Cancer Epidemiology and Prevention Biomarkers* **19** 960–965.

Key, T. J., Appleby, P. N., Travis, R. C., Albanes, D., Alberg, A. J., Barricarte, A., Black, A., Boeing, H., Bueno-de Mesquita, H. B., Chan, J. M. et al. (2015). Carotenoids, retinol, tocopherols, and prostate cancer risk: pooled analysis of 15 studies. *The American Journal of Clinical Nutrition* **102** 1142–1157.

Lai, J. K., Lucas, R. M., Banks, E., Ponsonby, A.-L. and Group, A. I. (2012). Variability in vitamin D assays impairs clinical assessment of vitamin D status. *Intern. Med. J.* **42** 43–50.

Lanzkron, P. J., Rose, D. J. and Szyld, D. B. (1990). Convergence of nested classical iterative methods for linear systems. *Numer. Math.* **58** 685–702.

Liang, H., Thurston, S. W., Ruppert, D., Apanasovich, T. and Hauser, R. (2008). Additive partial linear models with measurement errors. *Biometrika* **95** 667–678.

McCullough, M. L., Zoltick, E. S., Weinstein, S. J., Fedirko, V., Wang, M., Cook, N. R., Eliassen, A. H., Zeleniuch-Jacquotte, A., Agnoli, C., Albanes, D. et al. (2018). Circulating vitamin D and colorectal cancer risk: an international pooling project of 17 cohorts. *JNCI: Journal of the National Cancer Institute*.

Mitchell, E. M., Lyles, R. H., Manatunga, A. K., Danaher, M., Perkins, N. J. and Schisterman, E. F. (2014). Regression for skewed biomarker outcomes subject to pooling. *Biometrics* **70** 202–211.

Rao, C. R. (1972). Estimation of variance and covariance components in linear models. *J. Am. Stat. Assoc.* **67** 112–115.

Robinson, L. D. and Jewell, N. P. (1991). Some surprising results about covariate adjustment in logistic regression models. *Int. Stat. Rev.* **58** 227–240.

Sloan, A., Song, Y., Gail, M. H., Betensky, R., Rosner, B., Ziegler, R. G., Smith-Warner, S. A. and Wang, M. (2019). Design and analysis considerations for combining data from multiple biomarker studies. *Stat. Med.* **38** 1303–1320.

Snellman, G., Melhus, H., Gedeborg, R., Byberg, L., Berglund, L., Wernroth, L. and Michaelsson, K. (2010). Determining vitamin D status: a comparison between commercially available assays. *PloS one* **5** e11555.

Thompson, W. A. et al. (1962). The problem of negative estimates of variance components. *The Annals of Mathematical Statistics* **33** 273–289.

Tsilidis, K. K., Travis, R. C., Appleby, P. N., Allen, N. E., Lindström, S., Albanes, D., Ziegler, R. G., McCullough, M. L., Siddiq, A., Barricarte, A. et al. (2013). Insulin-like growth factor pathway genes and blood concentrations, dietary protein and risk of prostate cancer in the NCI Breast and Prostate Cancer Cohort Consortium (BPC3). *Int. J. Cancer* **133** 495–504.

Tworoger, S. S. and Hankinson, S. E. (2006). Use of biomarkers in epidemiologic studies: minimizing the influence of measurement error in the study design and analysis. *Cancer Causes Control* **17** 889–899.

Wang, N., Carroll, R. and Liang, K.-Y. (1996). Quasilikelihood estimation in measurement error models with correlated replicates. *Biometrics* 401–411.

C. Cheng
Center for Methods in Implementation and Prevention Science
Yale School of Public Health
135 College Street
New Haven, Connecticut 06511
and
Department of Mathematical Sciences
Tsinghua University
Beijing, China
E-mail: c.cheng@yale.edu

B. Rosner
Channing Laboratory
180 Longwood Avenue
Boston, Massachusetts 02115
and
Department of Biostatistics
Harvard T.H. Chan School of Public Health
Kresge Building
Boston, Massachusetts 02115
E-mail: stbar@channing.harvard.edu

M. Wang
Channing Division of Network Medicine
Brigham and Women's Hospital
Harvard Medical School
180 Longwood Avenue
Boston, Massachusetts 02115
and
Department of Epidemiology and Biostatistics
Harvard T.H. Chan School of Public Health
Kresge Building
Boston, Massachusetts 02115
E-mail: stmow@channing.harvard.edu