Biometrics 63, 1–25 DOI: xx.xxxx/j.xxxx-xxxx.xxxx.xxxxx.x
December 2007

Statistical Methods for Analysis of Combined Biomarker Data from Multiple Nested Case-Control Studies*

Chao Cheng

Center for Methods in Implementation and Prevention Science, Yale School of Public Health, New Haven

email: c.cheng@yale.edu

and

Molin Wang

Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston

email: stmow@channing.harvard.edu

Summary: Combining biomarker data across multiple studies allows for more precise estimation of biomarker-disease association as more statistcial power has been gained by increased sample size, whereas between-study variability exists in biomarker measurements and generally calibration to a reference assay is required prior to the pooling analysis. Previous researches treated the biomarker measurements from the reference laboratory as the gold standard, even through these biomarker measurements are not certainly equal to the true value. This paper accepted a linear measurement error model, which considered the measurement errors and biases from both the reference and study-specific laboratories. Specifically, we developed two calibration methods, the exact calibration method (ECM) and approximate calibration method (ACM), for the pooling biomarker data drawn from nested or matched case-control studies, where calibration subset was obtained by randomly selecting a number of controls from each contributing study. Our results illustrated that, under rare disease prevalence and/or small exposure effect size, the proposed methods can provide less biased estimates and accurate confidence intervals, significantly outperforming the naive method, which did not calibrate the biomarker measurements. For illustrative purposes, we applied the proposed methods in an application to evaluate the association between circulating 25-hydroxyvitamin D (25(OH)D) and colorectal cancer risk in a pooling project of nested case-control studies.

Key words: Between-study variability; Calibration; Case-control study; Measurement error.

*Note: This is a manuscript to be submitted, and the authorship line may change

1. Introduction

Pooling biomarker data across different studies to analyse biomarker-disease associations has become an increasingly popular strategy in epidemiological researches, since one single study is usually not large enough to provide precise estimates of associations. By including more biomarker measurements from various studies, this pooling analysis utilizes more statistical power to improve the estimates of biomarker exposure effect. There are a bunch of pooling projects examining biomarker-disease relationships, including the Endogenous Hormones, Nutritional Biomarkers, and Prostate Cancer Collaborative Group [Crowe et al (2014); Key et al (2015)], Cohort Consortium Vitamin D Pooling Project of Rarer Cancers [Gallicchio et al (2010)], COPD Biomarkers Qualification Consortium Database [Tabberer et al (2017)], and Vitamin D Pooling Project of Breast and Colorectal Cancer [McCullough et al (2018)].

However, between-laboratory variation in biomarker data may exist if not all measurements are assayed at the same laboratory at the same time, which will impair estimates of biomarker-disease associations. For example, measurements of estrone and testosterone, two specific hormones, have highly variable measurements between laboratories and assays [Tworoger and Hankinson (2006); Key et al (2010)]. Furthermore, measurements of circulating vitamin D (25(OH)D) can also vary up to 40% among laboratories and assays [Lai et al (2012); Snellman et al (2010)]. Under such circumstances, investigators must address between-study variation in biomarker measurements when develop statictical methods to evaluate the biomarker-disease relationships. Usually, calibration is implemented to harmonize different laboratories and assays by re-assaying a subset of non-case biospecimens randomly from each contributing study at a designated reference laboratory [Sloan et al (2018); Gail et al (2016)]. This calibration procedure can be utilized to correct the between-study measurement variability. In practice, investigators typically use only none-cases in the

calibration study subsets due to potential concerns about the availability of case biospecimens [Sloan et al (2018)].

Two views exist for analyzing biomarker measurements from the reference laboratory. The first view treats the biomarker measurements from the reference laboratory as the "gold standard" measurements [Sloan et al (2018); Gail et al (2016)]. However, biomarker measurements in the reference laboratory is not necessialy closer to the underlying truth in reality. The second view relaxes this "gold standard" rule in the reference laboratory, and considers the biomarker measurement errors in the reference laboratory. For example, the exact calibration method in Cheng, Roser and Wang (2019+) assumes $H = X + \epsilon$, where H is the biomarker measurements from the reference or study-specific laboratories, X is the true biomarker, and ϵ is the measurement error term. Its real data analysis also suggests that the measurement errors in the reference laboratory is even larger than those from some study-specific laboratories. Hence, treating the observed or calibrated measurements from the reference laboratory as the underlying truth may result in biased odd ratio (OR) estimates, which may not the best choice when there are sufficient data across multiple laboratories available in one pooling project.

In this paper, we accept the second view and assumes the both the study-specific and reference laboratories have measurement errors. We introduce laboratory-specific intercept and slope measurement biases into our calibration models, i.e., $H = \xi + (1+\gamma)X + \epsilon$, where ξ and γ represent the laboratory-specific intercept and slope measurement biases. Specifically, this paper develops two calibration methods, the approximate calibration method and exact calibration method, for pooled biomarker data from nested or matched case-control studies. The framework of this paper is as follows: Section 2 presents the models and statistical methods. In Section 3, we compare the methods via Monte Carlo simulation. Section 4 illustrates the methods in a real data example involving the circulating vitamin D pooled

from Nurses' Health Study (NHS) and Health Professionals Follow-up Study (HPFS) for colorectal cancer outcome, and we offer a conclusion in Section 5.

2. Methods

2.1 *Notations and Models*

Suppose that there are J nested case-control studies, each associated with a study-specific local laboratory j, where $j = 1, 2, \ldots, J$. Suppose the j^{th} study has been subdivided into K_j matched sets; the k^{th} matched set consists of M_{jk} individuals and the first $M_{jk}^{(1)}$ individuals are cases, where $M_{jk}^{(1)}$ $(\geqslant 1)$ and M_{jk} $(> M_{jk}^{(1)})$ may vary among different studies and matched sets. Let X_{jkm} be the unobserved true value of the continuous biomarker for the m^{th} individual in the k^{th} matched set from the j^{th} study, Y_{jkm} the binary disease outcome, $\boldsymbol{Z}_{jkm}$ other potential unmatched confunders for the X–Y relationship. Without further specification, all vectors are column vectors throughout the paper. We consider the following conditional logistic regression model for the biomarker-disease association

$$\text{logit}\Big(P(Y_{jkm}|X_{jkm}, \boldsymbol{Z}_{jkm})\Big) = \beta_{0jk} + \beta_x X_{jkm} + \boldsymbol{\beta}_z^T \boldsymbol{Z}_{jkm}, \tag{1}$$

where β_{0jk} is a stratum-specific intercept, $\boldsymbol{\beta}_z$ is a vector of covariate effects. The parameter of interest is β_x, which denotes the log odds ratio representing the biomarker-disease association.

Let $\boldsymbol{Y}_{jk}$, $\boldsymbol{X}_{jk}$ and $\boldsymbol{Z}_{jk}$ denotes their respective measurements from all individuals from matched set k of study j (i.e, $\boldsymbol{Y}_{jk} = [Y_{jk1}, \ldots, Y_{jkM_{jk}}]^T$, $\boldsymbol{X}_{jk} = [X_{jk1}, \ldots, X_{jkM_{jk}}]^T$, $\boldsymbol{Z}_{jk} = [\boldsymbol{Z}_{jk1}^T, \ldots, \boldsymbol{Z}_{jkM_{jk}}^T]^T$), the conditional likelihood contribution from matched set k of study j

with respect to the unknown parameter vector $\boldsymbol{\beta} = [\beta_x, \boldsymbol{\beta}_z^T]^T$ is

$$\begin{aligned} L_{jk} &= P\Big(\boldsymbol{Y}_{jk}\Big|\boldsymbol{X}_{jk}, \boldsymbol{Z}_{jk}, \sum_{m=1}^{M_{jk}} Y_{jkm} = M_{jk}^{(1)}\Big) \\ &= P\Big(Y_{jkm} = 1 \text{ for } m \leqslant M_{jk}^{(1)}, Y_{jkm} = 0 \text{ for } m > M_{jk}^{(1)}\Big|\boldsymbol{X}_{jk}, \boldsymbol{Z}_{jk}, \sum_{m=1}^{M_{jk}} Y_{jkm} = M_{jk}^{(1)}\Big) \quad (2) \\ &= \frac{\exp\big\{\sum_{m=1}^{M_{jk}^{(1)}}(\beta_x X_{jkm} + \boldsymbol{\beta}_z^T \boldsymbol{Z}_{jkm})\big\}}{\sum_{\mathcal{M}\in\mathcal{C}_{jk}} \exp\big\{\sum_{m\in\mathcal{M}}(\beta_x X_{jkm} + \boldsymbol{\beta}_z^T \boldsymbol{Z}_{jkm})\big\}}, \end{aligned}$$

where $\mathcal{C}_{jk}$ contains all the subsets of size $M_{jk}^{(1)}$ in set $\{1, \ldots, M_{jk}\}$. As the true biomarker measurement $\boldsymbol{X}_{jk}$ is unavailable, we can not use the conditional logistic regression above to estimate β_x.

Suppose there are a total of N individuals, with N_j individuals in the j^{th} study, where n_j individuals in study j are included in the calibration subset. Biospecimens from individuals in the calibration subset are re-assayed at a reference laboratory. Due to concerns about the availability of case biospecimens for re-assay at the reference laboratory, the calibration subset consists of a random selection of control biospecimens. Let $H_{jkm,d}$ be the measurement of X_{jkm} from laboratory d, where $d = 0$ refers to the reference laboratory and $d = j > 0$ refers to the jth study's local laboratory. Individuals who are not selected into the calibration subset only have the local laboratory measurement $H_{jkm,j}$. For brevity, we use $\boldsymbol{H}_{jkm}$ to denote all the measurements of X_{jkm}; i.e, for the individuals in the calibration subset, $\boldsymbol{H}_{jkm} = [H_{jkm,0}, H_{jkm,j}]^T$, and for the individuals who only have local laboratory measurements, $\boldsymbol{H}_{jkm} = H_{jkm,j}$.

We assume that the measurement $H_{jkm,d}$ and the underlying truth X_{jkm} follow a linear measurement error model

$$H_{jkm,d} = \xi_d + (1 + \gamma_d)X_{jkm} + \epsilon_{jkm,d}, \quad (3)$$

where ξ_d's and γ_d's are the zero-mean random effects representing laboratory-specific intercept and slope biases, $\epsilon_{jk,d}$ is the the measurement error, which is independent from X_{jk}

and follows a mean-zero normal distribution with a laboratory-specific variance; i.e, $\epsilon_{jkm,d} \sim N(0, \sigma_d^2)$, $d = 0, j$ and $j = 1, \ldots, J$. Here, we suppose $\xi_d \overset{iid}{\sim} N(0, \sigma_\xi^2)$ and $\gamma_d \overset{iid}{\sim} N(0, \sigma_\gamma^2)$, for $d = 1, \ldots, J$.

2.2 *Approximate Conditional Likelihood*

The likelihood function corresponding to the logistic regression model in Model (1) is

$$L = \prod_{j=1}^{J} \prod_{k=1}^{K_j} \frac{\exp\left\{ \sum_{m=1}^{M_{jk}^{(1)}} (\beta_x X_{jkm} + \beta_z^T \boldsymbol{Z}_{jkm}) \right\}}{\sum_{\mathcal{M} \in \mathcal{C}_{jk}} \exp\left\{ \sum_{m \in \mathcal{M}} (\beta_x X_{jkm} + \beta_z^T \boldsymbol{Z}_{jkm}) \right\}}.$$

This likelihood, however, cannot be computed because X_{jk} is not available. In fact, the contribution of the observed likelihood from a matched set can be shown as

$$\begin{aligned} L_{jk} &= P\Big(\boldsymbol{Y}_{jk} \Big| \boldsymbol{H}_{jk}, \boldsymbol{Z}_{jk}, \sum_{m=1}^{M_{jk}} Y_{jkm} = M_{jk}^{(1)}\Big) \\ &= \int P\Big(\boldsymbol{Y}_{jk} \Big| \boldsymbol{X}_{jk}, \boldsymbol{H}_{jk}, \boldsymbol{Z}_{jk}, \sum_{m=1}^{M_{jk}} Y_{jkm} = M_{jk}^{(1)}\Big) f\Big(\boldsymbol{X}_{jk} \Big| \boldsymbol{H}_{jk}, \boldsymbol{Z}_{jk}, \sum_{m=1}^{M_{jk}} Y_{jkm} = M_{jk}^{(1)}\Big) \mathrm{d}\boldsymbol{X}_{jk}, \end{aligned} \tag{4}$$

where $\boldsymbol{H}_{jk} = [\boldsymbol{H}_{jk1}^T, \ldots, \boldsymbol{H}_{jkM_{jk}}^T]^T$. We make the surrogacy assumption that the distribution of Y_{jkm} given $X_{jkm}, \boldsymbol{H}_{jkm}, \boldsymbol{Z}_{jkm}$ depends only on X_{jkm} and $\boldsymbol{Z}_{jk}$; i.e, $P(Y_{jkm}|X_{jkm}, \boldsymbol{H}_{jkm}, \boldsymbol{Z}_{jkm}) = P(Y_{jkm}|X_{jkm}, \boldsymbol{Z}_{jkm})$, which further implies $P(\boldsymbol{Y}_{jk}|\boldsymbol{X}_{jk}, \boldsymbol{H}_{jk}, \boldsymbol{Z}_{jk}, \sum_{m=1}^{M_{jk}} Y_{jkm} = M_{jk}^{(1)}) = P(\boldsymbol{Y}_{jk}|\boldsymbol{X}_{jk}, \boldsymbol{Z}_{jk}, \sum_{m=1}^{M_{jk}} Y_{jkm} = M_{jk}^{(1)})$. Let $\boldsymbol{W}_{jkm}$ contain the variables in $\boldsymbol{Z}_{jkm}$ that could be associated with X_{jkm}; that is, we assume $P(\boldsymbol{X}_{jk}|\boldsymbol{H}_{jk}, \boldsymbol{Z}_{jk}, \sum_{m=1}^{M_{jk}} Y_{jkm} = M_{jk}^{(1)}) = P(\boldsymbol{X}_{jk}|\boldsymbol{H}_{jk}, \boldsymbol{W}_{jk}, \sum_{m=1}^{M_{jk}} Y_{jkm} = M_{jk}^{(1)})$, where $\boldsymbol{W}_{jk} = [\boldsymbol{W}_{jk1}^T, \ldots, \boldsymbol{W}_{jkM_{jk}}^T]^T$.

It follows that the likelihood contribution L_{jk} can be written as

$$\begin{aligned} L_{jk} &= \int P\Big(\boldsymbol{Y}_{jk}\Big|\boldsymbol{X}_{jk},\boldsymbol{Z}_{jk},\sum_{m=1}^{M_{jk}}Y_{jkm}=M_{jk}^{(1)}\Big)f\Big(\boldsymbol{X}_{jk}\Big|\boldsymbol{H}_{jk},\boldsymbol{W}_{jk},\sum_{m=1}^{M_{jk}}Y_{jkm}=M_{jk}^{(1)}\Big)\mathrm{d}\boldsymbol{X}_{jk}, \\ &= \int \frac{\exp\big\{\sum_{m=1}^{M_{jk}^{(1)}}(\beta_x X_{jkm}+\beta_z^T\boldsymbol{Z}_{jkm})\big\}}{\sum_{\mathcal{M}\in\mathcal{C}_{jk}}\exp\big\{\sum_{m\in\mathcal{M}}(\beta_x X_{jkm}+\beta_z^T\boldsymbol{Z}_{jkm})\big\}}f\Big(\boldsymbol{X}_{jk}\Big|\boldsymbol{H}_{jk},\boldsymbol{W}_{jk},\sum_{m=1}^{M_{jk}}Y_{jkm}=M_{jk}^{(1)}\Big)\mathrm{d}\boldsymbol{X}_{jk}, \\ &= E_{\boldsymbol{X}_{jk}|\boldsymbol{H}_{jk},\boldsymbol{W}_{jk},\sum_{m=1}^{M_{jk}}Y_{jkm}=M_{jk}^{(1)}}\Big[\frac{\exp\big\{\sum_{m=1}^{M_{jk}^{(1)}}(\beta_x X_{jkm}+\beta_z^T\boldsymbol{Z}_{jkm})\big\}}{\sum_{\mathcal{M}\in\mathcal{C}_{jk}}\exp\big\{\sum_{m\in\mathcal{M}}(\beta_x X_{jkm}+\beta_z^T\boldsymbol{Z}_{jkm})\big\}}\Big] \end{aligned} \tag{5}$$

Generally, the probability density function (p.d.f), $f\Big(\boldsymbol{X}_{jk}\Big|\boldsymbol{H}_{jk},\boldsymbol{W}_{jk},\sum_{m=1}^{M_{jk}}Y_{jkm}=M_{jk}^{(1)}\Big)$, has a complex form which contains complicated intergals. We make the following assumption to simplify the calculation of above p.d.f, that is

$$f(X_{jkm}|H_{jkm,d},\boldsymbol{W}_{jkm},Y_{jkm})=f(X_{jkm}|H_{jkm,d},\boldsymbol{W}_{jkm}), \tag{6}$$

which considers that $H_{jkm,d}$ and $\boldsymbol{W}_{jkm}$ have provided enough information to predict X_{jkm}, regardless of the disease outcome. The assumption performs best when (i) σ_d's defined in Model (3) are small for $d=0,\ldots,J$, i.e there is not much noise in the relationship between X_{jkm} and $\boldsymbol{H}_{jkm}$, and/or (ii) small exposure effect, i.e., the association between X and Y is not too strong, and/or (iii) rare disease prevalence. Further details about the proof of these conditions are defered in the Appendix A. Under assumption (7), we can show

$$\begin{aligned} f\Big(\boldsymbol{X}_{jk}\Big|\boldsymbol{H}_{jk},\boldsymbol{W}_{jk},\sum_{m=1}^{M_{jk}}Y_{jkm}=M_{jk}^{(1)}\Big) &\approx f\Big(\boldsymbol{X}_{jk}\Big|\boldsymbol{H}_{jk},\boldsymbol{W}_{jk}\Big) \\ &= \prod_{m=1}^{M_{jk}} f\Big(X_{jkm}\Big|\boldsymbol{H}_{jkm},\boldsymbol{W}_{jkm}\Big), \end{aligned} \tag{7}$$

and the conditional likelihood contribution (5) becomes

$$\begin{aligned} L_{jk} &\approx \int \frac{\exp\big\{\sum_{m=1}^{M_{jk}^{(1)}}\beta_x X_{jkm}+\beta_z^T\boldsymbol{Z}_{jkm}\big\}}{\sum_{\mathcal{M}\in\mathcal{C}_{jk}}\exp\big\{\sum_{m\in\mathcal{M}}\beta_x X_{jkm}+\beta_z^T\boldsymbol{Z}_{jkm}\big\}}\prod_{m=1}^{M_{jk}} f\Big(X_{jkm}\Big|\boldsymbol{H}_{jkm},\boldsymbol{W}_{jkm}\Big)\mathrm{d}\boldsymbol{X}_{jk} \\ &= E_{\boldsymbol{X}_{jk}|\boldsymbol{H}_{jk},\boldsymbol{W}_{jk}}\Big[\frac{\exp\big\{\sum_{m=1}^{M_{jk}^{(1)}}\beta_x X_{jkm}+\beta_z^T\boldsymbol{Z}_{jkm}\big\}}{\sum_{\mathcal{M}\in\mathcal{C}_{jk}}\exp\big\{\sum_{m\in\mathcal{M}}\beta_x X_{jkm}+\beta_z^T\boldsymbol{Z}_{jkm}\big\}}\Big], \end{aligned} \tag{8}$$

where the p.d.f of $\boldsymbol{X}_{jk}|\boldsymbol{H}_{jk}, \boldsymbol{W}_{jk}$ is $\prod_{m=1}^{M_{jk}} f\Big(X_{jkm}\Big|\boldsymbol{H}_{jkm}, \boldsymbol{W}_{jkm}\Big)$.

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 $L_{jk} = P\Big(\boldsymbol{Y}_{jk}\Big|\boldsymbol{H}_{jk}, \boldsymbol{W}^*_{jk}, \sum_{m=1}^{M_{jk}} Y_{jkm} = M^{(1)}_{jk}\Big)$, which can still be written as in (5). A discussion about the potential benefit of including the additional variables $\boldsymbol{W}^*_{jk}$ is in Section 3.1 of the simulation study.

Next, we derive the analytic forms of $f\Big(X_{jkm}\Big|\boldsymbol{H}_{jkm}, \boldsymbol{W}_{jkm}\Big)$.

2.3 *Conditional Distribution of the Unknown True Biomarker Value*

Even though X_{jkm} is unobservable, we can derive the conditional distribution of X_{jkm} given $\boldsymbol{H}_{jkm}$ and $\boldsymbol{W}_{jkm}$. First, we assume the following X–W relationship

$$X_{jkm} = \alpha_{0j} + \boldsymbol{\tau}^T \boldsymbol{W}_{jkm} + \epsilon_{x_{jkm}}, \tag{9}$$

where α_{0j}'s are the study-specific intercepts, $\boldsymbol{\tau}$ represents unknown parameters common for all the studies, and $\epsilon_{x_{jkm}}$ is the error term following $N(0, \sigma_x^2)$. If $\boldsymbol{W}_{jkm}$ is null, the regression (9) degenerates to $X_{jkm} = \alpha_{0j} + \epsilon_{x_{jkm}}$. According to (3) and (9), $(X_{jkm}|\boldsymbol{W}_{jkm}$, $\tilde{H}_{jkm,0}|\boldsymbol{W}_{jkm}, \tilde{H}_{jkm,j}|\boldsymbol{W}_{jkm})^T$ follows the multivariate normal distribution below

$$\begin{pmatrix} X_{jkm}|\boldsymbol{W}_{jkm} \\ \tilde{H}_{jkm,0}|\boldsymbol{W}_{jkm} \\ \tilde{H}_{jkm,j}|\boldsymbol{W}_{jkm} \end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix} \mu_{X_{jkm}|\boldsymbol{W}_{jkm}} \\ \mu_{X_{jkm}|\boldsymbol{W}_{jkm}} \\ \mu_{X_{jkm}|\boldsymbol{W}_{jkm}} \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_x^2 & \sigma_x^2 \\ . & \sigma_x^2 + \frac{\sigma_0^2}{(1+\gamma_0)^2} & \sigma_x^2 \\ . & . & \sigma_x^2 + \frac{\sigma_j^2}{(1+\gamma_j)^2} \end{pmatrix} \right), \tag{10}$$

where $\tilde{H}_{jkm,d} = \frac{H_{jkm,d}-\xi_d}{1+\gamma_d}$ for $d = 0, j$, $\mu_{X_{jkm}|\boldsymbol{W}_{jkm}}$ is the abbreviation of $\alpha_{0j} + \boldsymbol{\tau}^T\boldsymbol{W}_{jkm}$. It follows that, for individuals who only have local laboratory measurements,

$$X_{jkm}|\boldsymbol{H}_{jkm}, \boldsymbol{W}_{jkm} \sim N\Big(\rho_j\tilde{H}_{jk,j} + (1-\rho_j)\mu_{X_{jkm}|\boldsymbol{W}_{jkm}}, \rho_j\tilde{\sigma}_j^2\Big), \tag{11}$$

where $\rho_j = \frac{\sigma_x^2}{\sigma_x^2+\tilde{\sigma}_j^2}$ for $j \in \{1, \ldots, J\}$, $\tilde{\sigma}_d^2 = \frac{\sigma_d^2}{(1+\gamma_d)^2}$ for $d \in \{0, \ldots, J\}$, and for individuals in the calibration subset

$$X_{jkm}|\boldsymbol{H}_{jkm}, \boldsymbol{W}_{jkm} \sim N\Big(\rho_j^*\big(w_j\tilde{H}_{jk,j} + (1-w_j)\tilde{H}_{jk,0}\big) + (1-\rho_j^*)\mu_{X_{jkm}|\boldsymbol{W}_{jkm}}, \rho_j^*w_j\tilde{\sigma}_j^2\Big), \tag{12}$$

where $\rho_j^* = \sigma_x^2/(\sigma_x^2+\tilde{\sigma}_j^2w_j)$ and $w_j = \tilde{\sigma}_0^2/(\tilde{\sigma}_j^2+\tilde{\sigma}_0^2)$. Hearafter, we use μ_{jkm} and s_{jkm} to denote the mean and standard deviation of $X_{jkm}|\boldsymbol{H}_{jkm}, \boldsymbol{W}_{jkm}$. Next, we describe the procedures for estimating the parameters involved in the conditional mean, μ_{jkm}.

2.4 *Estimation of Parameters in the Conditional Mean*

Let $\boldsymbol{\theta} = [\alpha_{01}, \alpha_{02}, \ldots, \alpha_{0J}, \boldsymbol{\tau}^T]^T$, $\boldsymbol{r} = [\xi_0, \ldots, \xi_J, \gamma_0, \ldots, \gamma_J]$ 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 (11) and (12), and $\mu_{X_{jkm}|\boldsymbol{W}_{jkm}}$ can be rewritten as $\widetilde{\boldsymbol{W}}_{jkm}^T\boldsymbol{\theta}$, where $\widetilde{\boldsymbol{W}}_{jkm} = [\boldsymbol{D}_{jkm}^T, \boldsymbol{W}_{jkm}^T]^T$, $\boldsymbol{D}_{jkm}$ is a $J \times 1$ vector with one on j^{th} element and zeros elsewhere. Combining (3) and (9), we have the following mixed-effect model:

$$H_{jkm,d} = \underbrace{\widetilde{\boldsymbol{W}}_{jkm}^T\boldsymbol{\theta}}_{\text{fixed effect}} + \underbrace{\xi_d + \epsilon_{x_{jkm}} + \epsilon_{jkm,d}}_{\text{random effects}} + \underbrace{\gamma_d\big(\widetilde{\boldsymbol{W}}_{jkm}^T\boldsymbol{\theta}\big) + \epsilon_{x_{jkm}}\gamma_d}_{\text{interaction terms}}, \tag{13}$$

where $\widetilde{\boldsymbol{W}}_{jkm}^T\boldsymbol{\theta}$ is the fixed-effect term, ξ_d, $\epsilon_{x_{jkm}}$ and $\epsilon_{jkm,d}$ are the random-effect terms, $\gamma_d\big(\widetilde{\boldsymbol{W}}_{jkm}^T\boldsymbol{\theta}\big)$ is another interaction term between the fixed-effect term $\widetilde{\boldsymbol{W}}_{jkm}^T\boldsymbol{\theta}$ and a random-effect term γ_d, and $\epsilon_{x_{jkm}}\gamma_d$ is an interaction term between two random-effect terms, γ_d and $\epsilon_{x_{jkm}}$. Succinctly, model (13) can be rewitten as

$$H_{jkm,d} = \widetilde{\boldsymbol{W}}_{jkm}^T\boldsymbol{\theta} + \widetilde{\boldsymbol{D}}_{jkm,d}^T\boldsymbol{r} + (1+\gamma_d)\epsilon_{x_{jkm}} + \epsilon_{jkm,d},$$

where $\widetilde{\boldsymbol{D}}_{jkm,d} = [\boldsymbol{C}_d^T, \boldsymbol{E}_{jkm,d}^T]^T$, $\boldsymbol{C}_d$ is a $(J+1)\times 1$ vector with one on the $(d+1)^{th}$ element and zeros elsewhere, $\boldsymbol{E}_{jkm,d}$ is a $(J+1)\times 1$ vector with $\widetilde{\boldsymbol{W}}_{jkm}^T\boldsymbol{\theta}$ on the $(d+1)^{th}$ element and zeros elsewhere. Noting that the matrix $\widetilde{\boldsymbol{D}}_{jkm,d}$ is not a covariate matrix as it contains unknown parameters, $\boldsymbol{\theta}$, we may name it pseudo-covariate matrix. The variance-covariance matrix of $\boldsymbol{r}$ is $\boldsymbol{R} := \text{var}(\boldsymbol{r}) = diag(\sigma_\xi^2, \ldots, \sigma_\xi^2, \sigma_\gamma^2, \ldots, \sigma_\gamma^2)$. Then, for all the measurements of X_{jkm}, i.e., $\boldsymbol{H}_{jkm}$, we have

$$\boldsymbol{H}_{jkm} = \boldsymbol{U}_{jkm}^T\boldsymbol{\theta} + \widetilde{\boldsymbol{D}}_{jkm}^T\boldsymbol{r} + \boldsymbol{P}_{jkm}^T\tilde{\boldsymbol{\epsilon}}_{jkm}, \tag{14}$$

where if X_{jkm} is in the calibration subset

$$\boldsymbol{U}_{jkm} = [\widetilde{\boldsymbol{W}}_{jkm}, \widetilde{\boldsymbol{W}}_{jkm}]^T, \widetilde{\boldsymbol{D}}_{jkm} = [\widetilde{\boldsymbol{D}}_{jkm,0}, \widetilde{\boldsymbol{D}}_{jkm,j}]^T,$$

$$\boldsymbol{P}_{jkm} = [\boldsymbol{P}_{jkm,0}, \boldsymbol{P}_{jkm,j}]^T \text{ with } \boldsymbol{P}_{jkm,0} = [1+\gamma_0, 1, 0]^T \text{ and } \boldsymbol{P}_{jkm,0} = [1+\gamma_j, 0, 1]^T,$$

$$\tilde{\boldsymbol{\epsilon}}_{jkm} = [\epsilon_{x_{jkm}}, \epsilon_{jkm,0}, \epsilon_{jkm,j}]^T \text{ with } \boldsymbol{\Sigma}_{jkm} := \text{var}\big(\tilde{\boldsymbol{\epsilon}}_{jkm}\big) = diag(\sigma_x^2, \sigma_0^2, \sigma_j^2);$$

if X_{jkm} is outside of the calibration subset, we have

$$\boldsymbol{U}_{jkm} = \widetilde{\boldsymbol{W}}_{jkm}, \widetilde{\boldsymbol{D}}_{jkm} = \widetilde{\boldsymbol{D}}_{jkm,j}, \boldsymbol{P}_{jkm} = [1+\gamma_j, 1]^T,$$

$$\tilde{\boldsymbol{\epsilon}}_{jkm} = [\epsilon_{x_{jkm}}, \epsilon_{jkm,j}]^T \text{ with } \boldsymbol{\Sigma}_{jkm} := \text{var}\big(\tilde{\boldsymbol{\epsilon}}_{jkm}\big) = diag(\sigma_x^2, \sigma_j^2).$$

Now, aggregating all the measurements from all the individuals in all studies together, model (14) can be summarized as

$$\boldsymbol{H} = \boldsymbol{U}^T\boldsymbol{\theta} + \widetilde{\boldsymbol{D}}^T\boldsymbol{r} + \boldsymbol{P}^T\tilde{\boldsymbol{\epsilon}}, \tag{15}$$

where

$$\boldsymbol{H} = [\boldsymbol{H}_1, \ldots, \boldsymbol{H}_J]^T, \boldsymbol{H}_j = [\boldsymbol{H}_{j1}, \ldots, \boldsymbol{H}_{jK_j}]^T, \boldsymbol{H}_{jk} = [\boldsymbol{H}_{jk1}, \ldots, \boldsymbol{H}_{jkM_{jk}}]^T,$$

$$\boldsymbol{U} = [\boldsymbol{U}_1, \ldots, \boldsymbol{U}_J]^T, \boldsymbol{U}_j = [\boldsymbol{U}_{j1}, \ldots, \boldsymbol{U}_{jK_j}]^T, \boldsymbol{U}_{jk} = [\boldsymbol{U}_{jk1}, \ldots, \boldsymbol{U}_{jkM_{jk}}]^T,$$

$$\widetilde{\boldsymbol{D}} = [\widetilde{\boldsymbol{D}}_1, \ldots, \widetilde{\boldsymbol{D}}_J]^T, \widetilde{\boldsymbol{D}}_j = [\widetilde{\boldsymbol{D}}_{jK_j}, \ldots, \widetilde{\boldsymbol{D}}_{jK_j}]^T, \widetilde{\boldsymbol{D}}_{jk} = [\widetilde{\boldsymbol{D}}_{j1}, \ldots, \widetilde{\boldsymbol{D}}_{jkM_{jk}}]^T$$

$$\boldsymbol{P} = Diag(\boldsymbol{P}_1, \ldots, \boldsymbol{P}_J), \boldsymbol{P}_j = Diag(\boldsymbol{P}_{j1}, \ldots, \boldsymbol{P}_{jK_j}), \boldsymbol{P}_{jk} = Diag(\boldsymbol{P}_{jk1}, \ldots, \boldsymbol{P}_{jkM_{jk}})$$

$$\tilde{\boldsymbol{\epsilon}} = [\tilde{\boldsymbol{\epsilon}}_1, \ldots, \tilde{\boldsymbol{\epsilon}}_J]^T, \tilde{\boldsymbol{\epsilon}}_j = [\tilde{\boldsymbol{\epsilon}}_{j1}, \ldots, \tilde{\boldsymbol{\epsilon}}_{jK_j}]^T, \tilde{\boldsymbol{\epsilon}}_{jk} = [\tilde{\boldsymbol{\epsilon}}_{jk1}, \ldots, \tilde{\boldsymbol{\epsilon}}_{jkM_{jk}}]^T,$$

$$\boldsymbol{\Sigma} := \mathrm{var}\big(\tilde{\boldsymbol{\epsilon}}\big) = Diag(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_J), \boldsymbol{\Sigma}_j := \mathrm{var}\big(\tilde{\boldsymbol{\epsilon}}_j\big) = Diag(\boldsymbol{\Sigma}_{j1}, \ldots, \boldsymbol{\Sigma}_{jK_j}),$$

$$\boldsymbol{\Sigma}_{jk} := \mathrm{var}\big(\tilde{\boldsymbol{\epsilon}}_{jk}\big) = Diag(\boldsymbol{\Sigma}_{jk1}, \ldots, \boldsymbol{\Sigma}_{jkM_{jk}}),$$

and $Diag(\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_p)$ denotes the block diagonal matrix generated by square matrix $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_p$. As $\widetilde{\boldsymbol{D}}$ and $\boldsymbol{P}$ depend on unknown parameters $\boldsymbol{\theta}$ and $\boldsymbol{r}$ respectively, we use $\widetilde{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})$ and $\boldsymbol{\mathcal{P}}(\boldsymbol{r})$ replace $\widetilde{\boldsymbol{D}}$ and $\boldsymbol{P}$ hereafter. Also, note that $\widetilde{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})^T\boldsymbol{r}$ and $\boldsymbol{\mathcal{P}}(\boldsymbol{r})^T\tilde{\boldsymbol{\epsilon}}$ are not independent as both depend on the random effects $\boldsymbol{r}$. This could pose computational difficults as it is hard to express the covariance structure of model (15), i.e., $\mathrm{var}\big(\widetilde{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})^T\boldsymbol{r} + \boldsymbol{\mathcal{P}}(\boldsymbol{r})^T\tilde{\boldsymbol{\epsilon}}\big)$, explicitly. Instead of the standard estimation algorithm for a linear mixed-effect model, based on Pinheiro and Bates (2006) Section 5.2, we propose an "iteratively reweighted" algorithm to obtain the estimators of $\boldsymbol{\theta}$, $\boldsymbol{r}$, and $\boldsymbol{\sigma}^2$. This algorithm is discribed as follows.

Let $\hat{\boldsymbol{\theta}}^{(0)}$ and $\hat{\boldsymbol{r}}^{(0)}$ be preliminary estimators for $\boldsymbol{\theta}$ and $\boldsymbol{r}$, where $\hat{\boldsymbol{\theta}}^{(0)}$ could be the ordinary least squares (OLS) estimator of the regression $E(\boldsymbol{H}|\boldsymbol{U}) = \boldsymbol{U}^T\boldsymbol{\theta}$, i.e., $\hat{\boldsymbol{\theta}}^{(0)} = (\boldsymbol{U}\boldsymbol{U}^T)^{-1}\boldsymbol{U}\boldsymbol{H}$; and $\hat{\gamma}_d^{(0)}$'s and $\hat{\xi}_d^{(0)}$'s, the elements in $\hat{\boldsymbol{r}}^{(0)}$, could be the OLS estimator of $E(H_{jkm,d}) = \xi_d + (1+\gamma_d)\boldsymbol{U}^T\boldsymbol{\theta}^{(0)}$, for all measurements in labortory d. In the t^{th} iteration, we replace the unknown parameters, $\boldsymbol{\theta}$ and $\boldsymbol{r}$, in $\widetilde{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})$ and $\boldsymbol{\mathcal{P}}(\boldsymbol{r})$, with their estimators in the $(t-1)^{th}$ iteration, $\hat{\boldsymbol{\theta}}^{(t-1)}$ and $\hat{\boldsymbol{r}}^{(t-1)}$. It follows that $\widetilde{\boldsymbol{\mathcal{D}}}(\hat{\boldsymbol{\theta}}^{(t-1)})^T\boldsymbol{r}$ can be assumed to be independent of $\boldsymbol{\mathcal{P}}(\hat{\boldsymbol{r}}^{(t-1)})^T\tilde{\boldsymbol{\epsilon}}$, and Model (15) can be approximated by the following extended marginal model:

$$\boldsymbol{H} = \boldsymbol{U}^T\boldsymbol{\theta} + \boldsymbol{\epsilon}^{*(t)}, \tag{16}$$

where $\boldsymbol{\epsilon}^{*(t)} \sim N\big(\boldsymbol{\mu}, \mathcal{V}(\boldsymbol{\sigma}^2, \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)})\big)$, and $\mathcal{V}(\boldsymbol{\sigma}^2, \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)}) = \widetilde{\boldsymbol{\mathcal{D}}}(\hat{\boldsymbol{\theta}}^{(t-1)})^T \boldsymbol{R} \widetilde{\boldsymbol{\mathcal{D}}}(\hat{\boldsymbol{\theta}}^{(t-1)}) + \boldsymbol{\mathcal{P}}(\hat{\boldsymbol{r}}^{(t-1)})^T \boldsymbol{\Sigma} \boldsymbol{\mathcal{P}}(\hat{\boldsymbol{r}}^{(t-1)})$ depends on $\boldsymbol{\sigma}^2$. The log-likelihood function for model (16) is

$$\begin{aligned} l(\boldsymbol{\theta}, \boldsymbol{\sigma}^2 | \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)}) = \; & -\tfrac{1}{2}\Big\{ \log \big|\mathcal{V}(\boldsymbol{\sigma}^2, \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)})\big| + \\ & \big(\boldsymbol{H} - \boldsymbol{U}^T\boldsymbol{\theta}\big)^T \mathcal{V}(\boldsymbol{\sigma}^2, \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)})^{-1} \big(\boldsymbol{H} - \boldsymbol{U}^T\boldsymbol{\theta}\big) \Big\} + \text{constant}. \end{aligned} \tag{17}$$

Maximizing (17) for fixed $\boldsymbol{\sigma}^2$ with respect to $\boldsymbol{\theta}$ leads to

$$\boldsymbol{\theta}(\boldsymbol{\sigma}^2 | \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)}) = \big(\boldsymbol{U}\mathcal{V}(\boldsymbol{\sigma}^2, \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)})^{-1}\boldsymbol{U}^T\big)^{-1} \boldsymbol{U}\mathcal{V}(\boldsymbol{\sigma}^2, \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)})^{-1}\boldsymbol{H}.$$

It follows that the profile loglikelihood is

$$\begin{aligned} l_p\big(\boldsymbol{\theta}(\boldsymbol{\sigma}^2 | \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)}), \boldsymbol{\sigma}^2\big) = \; & -\tfrac{1}{2}\Big\{ \log \big|\mathcal{V}(\boldsymbol{\sigma}^2, \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)})\big| + \Big(\boldsymbol{H} - \boldsymbol{U}^T\boldsymbol{\theta}(\boldsymbol{\sigma}^2 | \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)})\Big)^T \\ & \mathcal{V}(\boldsymbol{\sigma}^2, \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)})^{-1} \Big(\boldsymbol{H} - \boldsymbol{U}^T\boldsymbol{\theta}(\boldsymbol{\sigma}^2 | \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)})\Big) \Big\} + \text{constant}, \end{aligned}$$

and the MLE estimator of $\boldsymbol{\sigma}^2$ and $\boldsymbol{\theta}$ can be $\widehat{\boldsymbol{\sigma}}^{2(t)} = \underset{\boldsymbol{\sigma}^2}{\operatorname{argmax}}\, l_p\big(\boldsymbol{\theta}(\boldsymbol{\sigma}^2 | \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)}), \boldsymbol{\sigma}^2\big)$ and $\hat{\boldsymbol{\theta}}^{(t)} = \boldsymbol{\theta}(\widehat{\boldsymbol{\sigma}}^{2(t)} | \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)})$ respectively. Finally, the empirical best linear unbiased predictor (EBLUP) of $\boldsymbol{r}$ is

$$\hat{\boldsymbol{r}}^{(t)} = \widehat{\boldsymbol{R}}^{(t)} \widetilde{\boldsymbol{\mathcal{D}}}(\hat{\boldsymbol{\theta}}^{(t)}) \mathcal{V}(\hat{\boldsymbol{\sigma}}^{2(t)}, \hat{\boldsymbol{\theta}}^{(t)}, \hat{\boldsymbol{r}}^{(t)})^{-1} \big(\boldsymbol{H} - \boldsymbol{U}^T\hat{\boldsymbol{\theta}}^{(t)}\big),$$

where $\widehat{\boldsymbol{R}}^{(t)} = diag(\hat{\sigma}_\xi^{2(t)}, \ldots, \hat{\sigma}_\xi^{2(t)}, \hat{\sigma}_\gamma^{2(t)}, \ldots, \hat{\sigma}_\gamma^{2(t)})$. In summary, there are three stages in the t^{th} iteration:

- Stage 1: $\widehat{\boldsymbol{\sigma}}^{2(t)} = \underset{\boldsymbol{\sigma}^2}{\operatorname{argmax}}\, l_p\big(\boldsymbol{\theta}(\boldsymbol{\sigma}^2 | \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)}), \boldsymbol{\sigma}^2\big)$.
- Stage 2: $\hat{\boldsymbol{\theta}}^{(t)} = \boldsymbol{\theta}(\widehat{\boldsymbol{\sigma}}^{2(t)} | \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)})$.
- Stage 3: $\hat{\boldsymbol{r}}^{(t)} = \widehat{\boldsymbol{R}}^{(t)} \widetilde{\boldsymbol{\mathcal{D}}}(\hat{\boldsymbol{\theta}}^{(t)}) \mathcal{V}(\hat{\boldsymbol{\sigma}}^2, \hat{\boldsymbol{\theta}}^{(t)}, \hat{\boldsymbol{r}}^{(t)})^{-1} \big(\boldsymbol{H} - \boldsymbol{U}^T\hat{\boldsymbol{\theta}}^{(t)}\big)$

The iteration continues until convergence. The convergence criteria can depend on the relative difference $\dfrac{||\hat{\boldsymbol{\pi}}^{(t+1)} - \hat{\boldsymbol{\pi}}^{(t)}||}{||\hat{\boldsymbol{\pi}}^{(t)}||}$, where $\hat{\boldsymbol{\pi}}^{(t)} = [\widehat{\boldsymbol{\sigma}}^{2(t)^T}, \hat{\boldsymbol{\theta}}^{(t)^T}, \hat{\boldsymbol{r}}^{(t)^T}]^T$, and $||\bullet||$ denotes the Euclidean norm.

2.5 *Exact Calibration Method*

In this section, we propose a likelihood-based method for the estimation of exposure effects based on the data in the pooling project. According to the distribution of $X_{jkm}|\boldsymbol{H}_{jkm}, \boldsymbol{W}_{jkm}$ in (11) and (12), we can show that $\boldsymbol{X}_{jk}|\boldsymbol{H}_{jk}, \boldsymbol{W}_{jk}$ follows a multivariate normal distribution with mean $\boldsymbol{\mu}_{jk} = [\mu_{jk1}, \ldots, \mu_{jkM_{jk}}]^T$ and variance $\boldsymbol{s}^2_{jk} = diag(s^2_{jk1}, \ldots, s^2_{jkM_{jk}})$. We plug the estimators $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{\sigma}}^2$ in section 2.4 into $\boldsymbol{\mu}_{jk}$ and $\boldsymbol{s}^2_{jk}$, leading to $\hat{\boldsymbol{\mu}}_{jk}$ and $\hat{\boldsymbol{s}}^2_{jk}$. The likelihood contribution in (8) with this substitution becomes

$$\begin{aligned}\tilde{L}_{jk} &= \int \frac{\exp\left\{\sum_{m=1}^{M_{jk}^{(1)}} \beta_x X_{jkm} + \beta_z^T \boldsymbol{Z}_{jkm}\right\}}{\sum_{\mathcal{M}\in\mathcal{C}_{jk}} \exp\left\{\sum_{m\in\mathcal{M}} \beta_x X_{jkm} + \beta_z^T \boldsymbol{Z}_{jkm}\right\}} f\big(\boldsymbol{X}_{jk}|\hat{\boldsymbol{\mu}}_{jk}; \hat{\boldsymbol{s}}^2_{jk}\big)\mathrm{d}\boldsymbol{X}_{jk} \\ &= E_{\boldsymbol{X}_{jk}|\hat{\boldsymbol{\mu}}_{jk};\hat{\boldsymbol{s}}^2_{jk}}\Big[\frac{\exp\left\{\sum_{m=1}^{M_{jk}^{(1)}} \beta_x X_{jkm} + \boldsymbol{\beta}_z^T \boldsymbol{Z}_{jkm}\right\}}{\sum_{\mathcal{M}\in\mathcal{C}_{jk}} \exp\left\{\sum_{m\in\mathcal{M}} \beta_x X_{jkm} + \boldsymbol{\beta}_z^T \boldsymbol{Z}_{jkm}\right\}}\Big],\end{aligned} \tag{18}$$

where $f\big(\boldsymbol{X}_{jk}|\hat{\boldsymbol{\mu}}_{jk}; \hat{\boldsymbol{s}}^2_{jk}\big)$ denotes the p.d.f of a multivariate normal distribution with mean $\hat{\boldsymbol{\mu}}_{jk}$ and variance $\hat{\boldsymbol{s}}^2_{jk}$. Note that the likelihood contribution in (18) is the expectation of the observed likelihood (2) based on the distribution $\boldsymbol{X}_{jk}|\hat{\boldsymbol{\mu}}_{jk}; \hat{\boldsymbol{s}}^2_{jk}$. Estimates of $\boldsymbol{\beta}$ can be obtained by maximizing the pseudo-likelihood $\tilde{L} = \Pi_{j,k}\tilde{L}_{jk}$. However, the likelihood contribution in (18) can not be written as an explicit function, but we can use Monte Carlo approach or Gauss-Hermite Quadrature (GHQ) approach to calculate it numerically. The GHQ approach is developed to integrate some function with respect to the multivariate normal distribution, which approximate this integral as a weighted summation of knots and can be less computationally expensive for lower-dimension integrals. Therefore, we implement an intergration dimension reduction strategy first before applying the GHQ approach to calculate (18). More details about the Monte Carlo and GHQ approaches were defered to Supplementary Material Appendix B.

We name this method as *Exact Calibration Method* (ECM) and denote the $\boldsymbol{\beta}$-estimator from Monte Carlo and GHQ approaches as $\hat{\boldsymbol{\beta}}^{(E1)}$ and $\hat{\boldsymbol{\beta}}^{(E2)}$ respectively. For simplicity, we abbreiated the Monte Carlo and GHQ Exact Calibration Methods as ECM1 and ECM2

respectively. Each approach has merits and shortcomings: the GHQ method can present very accurate approximations with lower dimensions comparing to the Monte Carlo approach, but it bears "curse of dimensionality" when M_{jk} is large as the number of knots grows exponentially with M_{jk}; conversely, the accuracy of Monte Carlo approach is lower than the GHQ approach but it is pretty robust with the dimension of intergration.

2.6 *Approximate Calibration Method*

Alternatively, we can use a second order Taylor series to approximate the likelihood contribution (8) with respect to $\boldsymbol{X}_{jk}$ about $E(\boldsymbol{X}_{jk}|\boldsymbol{H}_{jk}, \boldsymbol{W}_{jk})$ yielding the following approximate likelihood contribution

$$\tilde{L}_{jk}^{(A)} = \prod_{j=1}^{J}\prod_{k=1}^{K_j} \frac{\exp\left\{\sum_{m=1}^{M_{jk}^{(1)}}(\beta_x \hat{X}_{jkm} + \beta_z^T \boldsymbol{Z}_{jkm})\right\}}{\sum_{\mathcal{M}\in\mathcal{C}_{jk}} \exp\left\{\sum_{m\in\mathcal{M}}(\beta_x \hat{X}_{jkm} + \beta_z^T \boldsymbol{Z}_{jkm})\right\}}. \tag{19}$$

where $\hat{X}_{jkm}$ is the estimated value for X_{jkm} with $\hat{X}_{jkm} = \hat{\mu}_{jkm}$. We name it *Approximate Calibration Method* (ACM) as it substitutes X_{jkm} in (2) with its estimated biomarker values directly. We denote the estimates from the ACM as $\hat{\beta}_x^{(A)}$. The ACM performs best when σ_d^2, $d = 0, 1, \ldots, M$, are small and/or the association between Y_{jk} and X_{jk} is not too strong. Further details on the derivation of these conditions are available in Appendix C in the Supplementary Material.

2.7 *Variance Estimation of* $\hat{\boldsymbol{\beta}}$

We utilize a resampling approach to obtain $\widehat{\text{Var}}(\hat{\boldsymbol{\beta}})$ in the ECM and ACM via the following steps:

- 1, Generate new variance estimates for pseudo-dataset i via $\tilde{\boldsymbol{\sigma}}^{2^{(i)}} \sim N(\hat{\boldsymbol{\sigma}}^2, \widehat{\text{Var}}(\hat{\boldsymbol{\sigma}}^2))$, where $\widehat{\text{Var}}(\hat{\boldsymbol{\sigma}}^2) = -\frac{d^2 l_p\left(\boldsymbol{\theta}(\boldsymbol{\sigma}^2|\hat{\boldsymbol{\theta}},\hat{\boldsymbol{r}}),\boldsymbol{\sigma}^2\right)}{d(\boldsymbol{\sigma}^2)^2}\Big|\boldsymbol{\sigma}^2 = \hat{\boldsymbol{\sigma}}^2$.
- 2, Generate new fixed and random effects for pseudo-dataset i via $\tilde{\boldsymbol{\theta}}^{(i)} \sim N(\hat{\boldsymbol{\theta}}, \widehat{\text{Var}}(\hat{\boldsymbol{\theta}}))$, $\tilde{\boldsymbol{r}}^{(i)} \sim N(\hat{\boldsymbol{r}}, \widehat{\text{Var}}(\hat{\boldsymbol{r}}))$, where $\widehat{\text{Var}}(\hat{\boldsymbol{\theta}}) = \left(\boldsymbol{U}\widehat{\boldsymbol{V}}^{-1}\boldsymbol{U}^T\right)^{-1}$ and $\widehat{\text{Var}}(\hat{\boldsymbol{r}}) = \widehat{\boldsymbol{R}}\widehat{\boldsymbol{D}}\big(\widehat{\boldsymbol{V}}^{-1} - \widehat{\boldsymbol{V}}^{-1}\boldsymbol{U}^T(\boldsymbol{U}\widehat{\boldsymbol{V}}^{-1}\boldsymbol{U}^T)\boldsymbol{U}\widehat{\boldsymbol{V}}^{-1}\big)\widehat{\boldsymbol{D}}^T\widehat{\boldsymbol{R}}$ with $\widehat{\boldsymbol{V}} = \mathcal{V}(\tilde{\boldsymbol{\sigma}}^{2^{(i)}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{r}})$, $\widehat{\boldsymbol{D}} = \widetilde{\boldsymbol{\mathcal{D}}}(\hat{\boldsymbol{\theta}})$, and $\widehat{\boldsymbol{R}} = diag(\hat{\sigma}_\xi^2, \ldots, \hat{\sigma}_\xi^2, \hat{\sigma}_\gamma^2, \ldots, \hat{\sigma}_\gamma^2)$.

- 3, Compute new conditional distributions $\tilde{X}_{jkm}^{(i)}|\boldsymbol{H}_{jkm}, \boldsymbol{W}_{jkm}$ in (11) and (12) for each individual based on the new pseudo calibration paramters $\tilde{\boldsymbol{\sigma}}^{2^{(i)}}$, $\tilde{\boldsymbol{\theta}}^{(i)}$ and $\tilde{\boldsymbol{r}}^{(i)}$. Denote its mean and variance as $\tilde{\mu}_{jkm}^{(i)}$ and $\tilde{s}_{jkm}^{2(i)}$ respectively.
- 4, Compute log-likelihood function (18) where $\hat{\mu}_{jk}$ and $\hat{\boldsymbol{s}}_{jk}$ are replaced by $\tilde{\boldsymbol{\mu}}_{jk}^{(i)} = [\tilde{\mu}_{jk1}^{(i)}, \ldots, \tilde{\mu}_{jkM_{jk}}^{(i)}]^T$ and $\tilde{\boldsymbol{s}}_{jk}^{(i)} = [\tilde{s}_{jk1}^{(i)}, \ldots, \tilde{s}_{jkM_{jk}}^{(i)}]^T$; then, calculate and optimize the corresponding $\tilde{L}^{(E1),(i)}$, $\tilde{L}^{(E2),(i)}$ and $\tilde{L}^{(A),(i)}$ based on the Monte Carlo ECM, GHQ ECM and ACM respectively to obtain $\hat{\boldsymbol{\beta}}^{(E1),(i)}$, $\hat{\boldsymbol{\beta}}^{(E2),(i)}$, $\hat{\boldsymbol{\beta}}^{(A),(i)}$ and their naive estimated variances

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}^{(E1),(i)}) = \frac{d^2 ln\tilde{L}^{(E1),(i)}}{d\boldsymbol{\beta}^2}\Big|\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}^{(E1),(i)},$$
$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}^{(E2),(i)}) = \frac{d^2 ln\tilde{L}^{(E2),(i)}}{d\boldsymbol{\beta}^2}\Big|\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}^{(E2),(i)},$$
$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}^{(A),(i)}) = \frac{d^2 ln\tilde{L}^{(A),(i)}}{d\boldsymbol{\beta}^2}\Big|\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}^{(A),(i)}.$$

- 5, Repeat Step 1 to 4 I times to obtain I estimates of $\hat{\boldsymbol{\beta}}^{(E1),(i)}$, $\hat{\boldsymbol{\beta}}^{(E2),(i)}$, $\hat{\boldsymbol{\beta}}^{(A),(i)}$, and $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}^{(E1),(i)})$, $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}^{(E2),(i)})$, $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}^{(A),(i)})$.
- 6, Obtain a final estimate of $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}})$, where $\hat{\boldsymbol{\beta}}$ could be $\hat{\boldsymbol{\beta}}^{(E1)}$, $\hat{\boldsymbol{\beta}}^{(E2)}$ or $\hat{\boldsymbol{\beta}}^{(A)}$, by taking an average of the naive variance and empirical variance from the pseudo-datasets such that

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{I} \frac{\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}^{(i)})}{I} + \sum_{i=1}^{I} \frac{(\hat{\boldsymbol{\beta}}^{(i)} - \bar{\boldsymbol{\beta}})^T(\hat{\boldsymbol{\beta}}^{(i)} - \bar{\boldsymbol{\beta}})}{I-1},$$

in which $\bar{\boldsymbol{\beta}} = \sum_{i=1}^{I} \frac{\hat{\boldsymbol{\beta}}^{(i)}}{I}$.

3. Simulation Studies

3.1 *Simulation Setup and Results*

In this section, we first describe how the unoberved biomarker X_{jkm}, local and reference laboratory mearsurements $H_{jkm,0}$ and $H_{jkm,j}$, and the binary disease outcome Y_{jkm} were generated.

We assumed there were 10 matched case-control studies with 1:1 case-control ratios, i.e, $J = 10$, each with 250 individuals (125 cases and 125 controls). Firstly, we generated $\xi_d \sim N(0, 0.5^2)$ and $\gamma_d \sim N(0, 0.1^2)$, for $d = 0, 1. \ldots, 10$, representing the intercept and slope

biases for each reference and local laboratory. Moreover, the variance of the measurement error term in every laboratory, σ_d^2, were generated by Unif$(0.15, 0.35)$. For the j^{th} case-control study utilized in the pooling analysis, we at first generated a large source population ($N = 5000$) as follows. First, draw a data W following one-dimentional normal distribition with mean 0 and variance 2, and then draw a triple $(X, \tilde{H}_0, \tilde{H}_j)^T$ according to the multivariate normal distribution in (10), where $\mu_{X|\boldsymbol{W}} = \alpha_{0j} + \tau W$, α_{0j}'s were generated from $N(3, 0.5^2)$ and τ was set as 0.5. In the variance-covariance matrix, we set $\sigma_x^2 = 0.5$. 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 59% to 77%. The measurement in the reference lab, H_0, then could be calulated by $H_0 = \xi_0 + (1 + \gamma_0)\tilde{H}_0$; similarily, the measurement in the local laboratory, H_j, could be calulated by $H_j = \xi_j + (1 + \gamma_j)\tilde{H}_j$. The binary outcome Y was generated based on $\text{logit}(P(Y = 1|X)) = \beta_{0j} + \beta_x X$, where five values of β_x were considered, including log(1.25), log(1.5), log(1.75), log(2), representing weak, medium, strong and very strong biomarker-disease relationship. At this stage, we had N quintuples of $(Y, X, H_0, H_j, W)^T$.

To obtain the case-control data, we randomly selected 125 quintuples from among the triples with $Y = 1$ (cases) and 125 quintuples from among those with $Y = 0$ (controls). We first assumed a sample size of 25 for each calbration subset, where the calibration subsamples were randomly selected from the controls in the orginal case-control data. Due to X is unavailable and H_0 (the measurement in the reference laboratory) is only available for the individuals in the calibration subset, we observed $(Y, H_0, H_j, W)^T$ for the the individuals in the calibration subset, and observed $(Y, H_j, W)^T$ for all other individuals. These quadruples and triplets constituted the case-control data available for analysis. We assumed there are no matching variables in our simulation study, thus we randomly matched cases with controls to obtain 125 pairs in each study. Now, we have 10 case-control data for analysis.

At each β_x and calibration design considered, we completed 1000 simulation replicates and

compared the ACM, ECM1 and ECM2 with regard to the following operating characteristics: mean percent bias, mean squared error (MSE), empirical standard error, and coverage rate of 95% confidence interval. We considered a naive method as a benchmark, which replaced X_{jkm} in model (2) with the average of $H_{jkm,0}$ and $H_{jkm,j}$ if $H_{jk,0}$ was avaliable and with $H_{jkm,j}$ otherwise, and fitted the conditional logistic regression model to obtain the estimated β_x, denoted by $\hat{\beta}_x^{(N)}$ henceforth. For purpose of comparison, we also applied the Full Calibration method in Sloan and Wang (2019) in the simulation. This method utilized the ordinary least square method to fit the model $H_{jkm,0} = \alpha_d + \beta_d H_{jkm,j} + \epsilon_{jkm,j}$ for the calibration subset in each study. Then, the estimated reference laboratory measurement $\hat{H}_{jkm,0}$ for all individuals was obtained and treated as unobserved X_{jkm} in the model for the disease outcome, i.e., fitting the logistic regression model (1) by substituting X_{jkm} with $\hat{H}_{jkm,0}$. The estimated exposure effect was denoted by $\hat{\beta}_x^{(F)}$ in this paper and a standard sandwich method was applied to obtain its standard error.

The simulation results were shown in Table 1. The naive method performed poorly over all effects in consideration of that all percent biases exceeded -19% and the coverage rates dropped to less than 20% as exposure effect increased, indicating the necessity of calibration. All the calibration approaches reduced the percent bias within $\pm 2\%$. The full calibration method and ECM1 typically minimized the percent bias with values less than 1% for all ORs considered, while the ACM and ECM2 estimate were biased downwards by roughly 0.5 to 2 percents. The ACM typically minimized the MSEs of $\hat{\beta}_x$ estimates, while MSEs of the ECM1 and ECM2 estimates were slightly larger than those of the ACM estimates. The full calibration method performed larger MSEs in comparison with other calibration methods, where a approximately twice higher MSE for the full calibration method was observed compared with the ACM as the OR increased. The coverage rates for the ACM and ECMs were closer to the nominal 95%, where the coverage rate under strong biomarker-

disease association (OR=2.0) dropped to around 92% due to depression of point estimates. Even applying the sandwich method to correct the confidence interval, the coverage rates of the full calibration estimates were typicially less than 92% under all exposure effects, and as exposure effect increased, its coverage rate dropped significantly.

[Table 1 about here.]

Several additional simulation experiments were conducted to check the performance of the proposed calibration methods; the results were summarized as below: (I) In consideration of that the variable W_{jkm} may be unavailable in practice, we conducted a simulation experiment which assumed $X_{jkm} = \alpha_{0j} + \epsilon_{x_{jkm}}$ $(j = 1, \ldots, 10)$ in the model for X; i.e $\boldsymbol{W}_{jkm}$ was not in the analysis. The results shown in Table 2 were similar with what discussed above, whereas the percent biases and MSEs of all the calibration methods were larger than those in Table 1, which implies that including variables associated with the biomarker data into the model for X could improve the estimation accuracy. (II) We considered that W_{jkm} was also associated with the disease directly; i.e., assuming $\boldsymbol{Z}_{jkm} = W_{jkm}$ with $\beta_z = log(1.25)$ in model (1). The results were provided in Supplementary Table S1. The percent bias from the naive method was generally larger than those in Table 1, whereas the percent bias from the proposed calibration methods was slightly affected. However, the coverages rate from all methods was affected by this change, but ACM and ECMs still presented considerable interval estimates for smaller effect sizes (OR $\leqslant 1.5$). (III) When the variance of the slope randome effect, σ^2_γ, decreased, all calibration methods improved with regard to all charactistics (Supplementary Material Table S2), while the naive method did not change much in terms of the percent bias. (IV) As the percentage of the subjects selected in the calibration subset increased (Supplementary Material Table S3), the MSE and percent bias of the naive method were worse, whereas the estimates from the calibration methods slightly improved in terms of the MSE and coverage rate.

In summary, the proposed calibration methods expressed significant advantages over the naive method in terms of the percent bias, MSE and confidence interval. Moreover, the ACM performed best with regard to MSE and percent bias under most parameter settings. Besides, all the proposed methods could provide relatively accurate confidence interval for samll effect sizes (OR $\leqslant$ 1.5). In contrast, the naive method was heavily biaed in most simulation experiments.

[Table 2 about here.]

3.2 *When X Does not Follow a Normal Distribution*

The previous simulation experiments considered a normal distribution of $\epsilon_{x_{jkm}}$. However, in reality, the biomarker data may be skewed or fat-tailed, indicating the discrepancy of the normality assumption. In this section, we investiaged the proposed calibration methods when X_{jkm} (and thus $\boldsymbol{H}_{jkm}$) does not follow a normal distribution. Two specific distributions for $\epsilon_{x_{jk}}$ were intestigated: uniform distribution and skew normal distribution [Fernandez and Steel (1998)]. Specifically, we first generated X_{jkm} based on $X_{jkm} = \alpha_{0j} + \tau W_{jkm} + \epsilon_{x_{jkm}}$, where $\epsilon_{x_{jkm}}$ followed either an uniform or a skew normal distribution, and all the paramters were adjusted to satisfing mean 0 and variance 0.5. For the skew normal distribution scienario [Fernandez and Steel (1998)], we set the skew parameter equal to 1.5, leading to the moment coefficient of skewness approximately 0.5. We generate $H_{jk,d}$ based on $H_{jkm,d} \sim N(\xi_d + (1 + \gamma_d)X_{jkm}, \sigma_d^2)$, $\quad j = 0, 1, \ldots, 10$, where ξ_d's, γ_d's and all the other design parameters were identical with those in the previous section.

Table 3 provided the simulation results, which were similar with the results in Section 3.1. Across all exposure effects, the ACM and ECMs presented better performance in terms of percent bias and coverage rate, where the ACM typically minimized the MSE and SE. The naive method was still undesirable due to large percent biases.

[Table 3 about here.]

4. Applied Example

We completed one real data example to illustrate the proposed methods. In this example, we investigate the impact of circulating 25-hydroxyvitamin D (25(OH)D) levels on risk of colorectal cancer, based on the data combined from two prospective cohort studies in the United States, containing the Nurses's Health Study (NHS) [Eliassen et al (2016)] and Health Professionals Follow-up Study (HPFS) [Choi et al (2005)]. The NHS began enrollment in 1976 and included 121,701 female nurses aged 30 to 55 years at baseline. The HPFS enrolled 51,529 male health professionals aged 40 to 75 years at baseline. Between 1989 to 1995, both studies completed laboratory assays on blood samples for a host of biomarkers, which inluded 25(OH)D, from a subset of participants. Subjects with a previous cancer diagnosis were excluded from this random selection. A total of 1876 participants, extracted from both studies, constituted our pooling analysis, where individuals were excluded from the pooling analysis if he or she did not have 25(OH)D measurements available or colorectal cancer outcome data. In the pooling analysis, we match each case to one or two controls (1:1 or 1:2 matching), based on sex and age at blood draw. From 2011 to 2013, each study randomly selected 29 controls and re-assayed their blood samples at Heartland Assays, LLC (Ames, IA). These laboratory measurements were treated as reference laboratory measurements in this example.

In the pooling analysis, the covariates $\boldsymbol{W}$ in model (9) for 25(OH)D contained age of blood draw (ranged 43-82), week of the year at blood draw (1-52), physical activity (continuous), smoking (ever/never) and BMI (greater or less than 25 kg/m^2). Due to the seasonal variation of 25(OH)D, we utilized 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 (9) to fit the seasonal trend [Gail et al (2016)], where t represents week of the year at blood draw. In consideration of comparison, we applied two other models for 25(OH)D, where the first model included study-specfic intercepts and seasonal

trend, the second model included the study-specific intercepts only. The original and two additional models were renamed as Model I, II and III respectively. Table 4 showed the point estimates with standard errors for the regression coefficients corresponding to Model I, II, III. Information about the study-specific and reference laboratories, including the intercept and slope biases, was presented in Table 5.

Our pooled analysis consists of 615 case-control pairs matched coarsely on age and sex, with 348 pairs from NHS and 267 from HPFS. Then, we applied the naive method, ACM, ECM1 and ECM2 to the pooling dataset, adjusting for physical activity total (continuous), family history of colorectal cancer (yes/no), smoking (ever/never) and BMI (greater or less than 25 kg/m^2). The results for the OR and its 95% confidence interval were displayed in Table 6. All the analytic approaches presented that increased 25(OH)D levels correlated with a statistically significant (based on 95% confidence level) protective effect against colorectal cancer. The point estimates of OR among all approaches were quite similar, possibly due to small measurement errors and intercept and slope biases for all laboratories.

[Table 4 about here.]

[Table 5 about here.]

[Table 6 about here.]

5. Discussion

In this paper, we proposed and evaluated statistical methods for pooling biomarker data drawn from multiple nested case-control studies. This work focused on evaluating the odds ratio representing the association between a continuous biomarker and a binary disease outcome. Being consistent with the common practice, we require a calibration subset of controls randomly selected from each contribution study and re-assayed in a reference lab. Different with the previous researches, we considered the measuremnt errors and biases of

the observed biomarker data in both the reference and study-specific laboratories. The R software to implement the proposed methods can be found at `https://www.hsph.harvard.edu/molin-wang/software`.

Several parctial observations and recommendations follow from this paper. First, all the proposed calibration methods, inlcuding ACM, ECM1 and ECM2, were able to obtain a less biased point estimation and more precise confidence interval of the OR than those from the naive approach, which did not adjust the measurement error and bias. Second, across the calibration methods, the ACM is prefered approach in consideration of its minimal biased point estimates and MSEs under all simulation scenarios. Moreover, the ACM can be easy implemented by any standard software that can analyze conditional logsitic regression, whereas we need to write additional codes to implement the ECM1 or ECM2. Third, we observed the OR point estimates were slightly biased for strong biomarker-disease association in the simulation studies, but all the proposed calibration methods yield satisfactory estimates under small exposure effects (OR $\leqslant$ 1.75). Those observations could help investigators in conducting real data analysis.

This paper focused on a linear measurement error model, where the reference and study-specific laboratory measurements, H_d, and the underlying true biomarker, X, were assumed to satisfy a mixed-effect model, say $H_d = \xi_d + (1 + \gamma_d)X + \epsilon_d$, where ξ_d and γ_d are zero-mean random effects representing the measurement bias in laboratory d, and ϵ_d is the conresponding measurement error term. All calibration methods discussed in this work rely on the conditional regression model corresponding to nested case-control studies, which can be applied to prospective or retrospective corhort studies with a binary disease outcome.

In practical applications, it could be interest to investigate whether the random effects, ξ_d's or γ_d's, should be introduced into the measurement error model or not. In another word, we want to investigate $\mathrm{H}_0 : \sigma_\xi^2 = 0$ or $\mathrm{H}_0 : \sigma_\gamma^2 = 0$. A likelihood ratio test (LRT) can be applied to

test this hypothesis. Specifically, let l_1 denotes the log-likelihood for the mixed effect model (15), where σ_ξ^2 or σ_γ^2 were constrained as 0. And let l_2 denote the corresponding unrestricted log-likelihood, then the LRT statistic is $\chi^2 = 2(l_2 - l_1)$. Under the null hypothesis that the restricted model is adequate, χ^2 should follow a Chi-squared distribution with 1 degree of freedom, and the P-value can be calculated as $1 - F_{\chi_1^2}(\chi^2)$, where $F_{\chi_1^2}(.)$ is the cumulative distribution function of the Chi-squared distribution with 1 degree of freedom. However, the P-value calculated by this way could be slightly greater than it should be, because the constrain $\sigma_\xi^2 = 0$ in the null hypothesis lie s on the boundary of the parameter region $\sigma_\xi^2 \geqslant 0$, which could influence the asymptotic distribution χ^2 [Stram and Lee (1994)]. A more robust method is applying a parametric bootstrap procedure to simulate the empirical distrubution of χ^2, see Pinheiro and Bates (2006) Section 2.4.1 for more details.

Acknowledgements

We 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.

References

Cheng, C., Rosner, B. and Wang, M., (2019+). Statistical methods for analysis of combined categorical biomarker data from multiple studies, under review by *Annals of Applied Statistics*.

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(7),** pp.742-748.

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. and Gibson, R.A. (2014). Circulating fatty acids and prostate cancer risk: individual participant meta-analysis of prospective studies. *Journal of the National Cancer Institute* **106(9).**

Eliassen, A. H., Warner, E. T., Rosner, B., Collins, L. C., Beck, A. H., Quintana, L. M., ... & Hankinson, S. E (2016). Plasma 25-hydroxyvitamin D and risk of breast cancer in women followed over 20 years. *Cancer research* **76(18),** 5423-5430.

Fernandez C., Steel M.F.J. (1998). On Bayesian Modelling of Fat Tails and Skewness. *Publications of the American Statistical Association* **93(441),** pp.359-371.

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. and Ziegler, R.G. (2016). Calibration and seasonal adjustment for matched case–control studies of vitamin D and cancer. *Statistics in medicine* **35(13),** pp.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. and Koenig, K.L. (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. *American Journal of Epidemiology* **172(1),** pp.10-20.

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

Jackel, P. (2005). A note on multivariate Gauss-Hermite quadrature. *London: ABN-Amro. Re.*

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 **19(4),** pp 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. and Chen, C. (2015). Carotenoids, retinol, tocopherols, and prostate cancer risk: pooled analysis of 15 studies–3. *The American Journal of Clinical Nutrition* **102(5),** pp.1142-1157.

Lai, J.K., Lucas, R.M., Banks, E., Ponsonby, A.L. and Ausimmune Investigator Group (2012). Variability in vitamin D assays impairs clinical assessment of vitamin D status. *Internal Medicine Journal* **42(1),** pp.43-50.

Lanzkron, P.J., Rose, D.J. and Szyld, D.B. (1990). Convergence of nested classical iterative methods for linear systems. *Numerische Mathematik* **58(1),** pp.685-702.

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. and Barnett, M.J. (2018). Circulating Vitamin D and Colorectal Cancer Risk: An International Pooling Project of 17 Cohorts. *Journal of the National Cancer Institute*.

Pinheiro, J. and Bates, D. (2006). Mixed-effects models in S and S-PLUS. *Springer Science & Business Media*.

Rao, C.R. (1972). Estimation of variance and covariance components in linear models. *Journal of the American Statistical Association* **67(337),** pp.112-115.

Sloan, A., Song, Y., Gail, M.H., Ziegler, R.G., Smith-Warner, S.A., Wang, M. (2018). Design and analysis considerations for combining data from multiple biomarker studies. *Statistics in Medicine* **38,** pp.1303–1320.

Sloan, A., & Wang, M. (2019). Statistical methods for biomarker data pooled from multiple nested case-control studies. *arXiv preprint* arXiv:1906.00818.

Snellman, G., Melhus, H., Gedeborg, R., Byberg, L., Berglund, L., Wernroth, L. and Michaelsson, K. (2010). Determining vitamin D status: a comparison between commer-

cially available assays. *PloS One* **5(7),** e11555.

Stram, D. O. and Lee, J. W. (1994). Variance components testing in the longitudinal mixed-effects models, *Biometrics* **50,** 1171–1177.

Tabberer, M., Benson, V. S., Gelhorn, H., Wilson, H., Karlsson, N., Müllerova, H., ... & Jones, P. W. (2017). The COPD biomarkers qualification consortium database: baseline characteristics of the St George's respiratory questionnaire dataset. *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* **4(2),** 112.

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. and Berndt, S.I., 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). *International Journal of Cancer* **133(2),** pp.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(7),** pp.889-899.

Supporting Information

Web Appendix and supplementary tables are available with this paper at the Biometrics website on Wiley Online Library.

Received XX 20XX. *Revised XX* 20XX. *Accepted XX* 20XX.

Table 1

Comparison of operating characteristics for the naive method ($\hat{\beta}^{(N)}$), full calibration method ($\hat{\beta}^{(F)}$), approximate calibration method ($\hat{\beta}^{(A)}$), and Monte Carlo and GHQ exact calibration methods ($\hat{\beta}^{(E1)}$ and $\hat{\beta}^{(E2)}$).

β_x	Percent Bias					MSE(×100)					SE(×100)					Coverage Rate(×100)				
	$\hat{\beta}_x^{(N)}$	$\hat{\beta}_x^{(F)}$	$\hat{\beta}_x^{(A)}$	$\hat{\beta}_x^{(E1)}$	$\hat{\beta}_x^{(E2)}$	$\hat{\beta}_x^{(N)}$	$\hat{\beta}_x^{(F)}$	$\hat{\beta}_x^{(A)}$	$\hat{\beta}_x^{(E1)}$	$\hat{\beta}_x^{(E2)}$	$\hat{\beta}_x^{(N)}$	$\hat{\beta}_x^{(F)}$	$\hat{\beta}_x^{(A)}$	$\hat{\beta}_x^{(E1)}$	$\hat{\beta}_x^{(E2)}$	$\hat{\beta}_x^{(N)}$	$\hat{\beta}_x^{(F)}$	$\hat{\beta}_x^{(A)}$	$\hat{\beta}_x^{(E1)}$	$\hat{\beta}_x^{(E2)}$
log(1.25)	-19.3	-0.5	1.0	0.7	1.4	0.43	0.34	0.28	0.28	0.28	4.93	5.83	5.26	5.29	5.32	77.2	91.9	96.4	96.5	96.1
log(1.5)	-19.8	-0.4	0.4	0.7	1.6	0.92	0.54	0.33	0.35	0.36	5.20	7.34	5.75	5.93	5.97	52.0	86.2	95.3	94.9	95.0
log(1.75)	-20.8	-0.5	-0.7	0.4	1.4	1.66	0.79	0.41	0.45	0.47	5.55	8.87	6.42	6.71	6.85	32.3	80.7	93.6	93.9	93.2
log(2)	-21.6	-0.9	-1.7	0.4	1.4	2.58	1.05	0.48	0.56	0.58	5.83	10.24	6.83	7.46	7.55	18.3	78.9	91.8	92.1	92.1

NOTE: Percent bias and MSE were computed by averaging $(\hat{\beta} - \beta)/\beta$ and $(\beta - \hat{\beta})^2$ over 1000 simulations. Standard error (SE) is the square root of the empirical variance over all replicates. Coverage rate represents the coverage of a 95% confidence interval.

Table 2

Comparison of operating characteristics for the naive method ($\hat{\beta}^{(N)}$), full calibration method ($\hat{\beta}^{(F)}$), approximate calibration method ($\hat{\beta}^{(A)}$), and Monte Carlo and GHQ exact calibration methods ($\hat{\beta}^{(E1)}$ and $\hat{\beta}^{(E2)}$), where the model for X_{jk} was assumed as $X_{jk} = \alpha_{0j} + \epsilon_{x_{jk}}$, $j = 1, \ldots, 10$.

β_x	Percent Bias					MSE(×100)					SE(×100)					Coverage Rate(×100)				
	$\hat{\beta}_x^{(N)}$	$\hat{\beta}_x^{(F)}$	$\hat{\beta}_x^{(A)}$	$\hat{\beta}_x^{(E1)}$	$\hat{\beta}_x^{(E2)}$	$\hat{\beta}_x^{(N)}$	$\hat{\beta}_x^{(F)}$	$\hat{\beta}_x^{(A)}$	$\hat{\beta}_x^{(E1)}$	$\hat{\beta}_x^{(E2)}$	$\hat{\beta}_x^{(N)}$	$\hat{\beta}_x^{(F)}$	$\hat{\beta}_x^{(A)}$	$\hat{\beta}_x^{(E1)}$	$\hat{\beta}_x^{(E2)}$	$\hat{\beta}_x^{(N)}$	$\hat{\beta}_x^{(F)}$	$\hat{\beta}_x^{(A)}$	$\hat{\beta}_x^{(E1)}$	$\hat{\beta}_x^{(E2)}$
log(1.25)	-19.8	-0.4	3.7	2.9	4.2	0.43	0.37	0.34	0.34	0.36	4.85	5.78	5.80	5.82	5.89	76.8	91.3	97.2	97.1	97.1
log(1.5)	-20.3	-0.2	3.2	3.5	4.7	0.95	0.56	0.51	0.54	0.59	5.23	7.47	6.99	7.21	7.41	50.3	86.6	95.3	95.2	94.6
log(1.75)	-21.3	-0.7	3.0	4.4	5.9	1.88	0.76	0.67	0.79	0.92	6.78	8.72	8.03	8.57	8.98	34.3	81.9	95.6	95.7	95.2
log(2)	-21.6	-0.8	2.1	4.7	6.4	2.58	1.11	0.85	1.14	1.34	5.80	10.51	9.11	10.19	10.72	17.2	78.4	92.0	91.0	89.2

NOTE: Percent bias and MSE were computed by averaging $(\hat{\beta} - \beta)/\beta$ and $(\beta - \hat{\beta})^2$ over 1000 simulations. Standard error (SE) is the square root of the empirical variance over all replicates. Coverage rate represents the coverage of a 95% confidence interval.

Table 3

Comparison of operating characteristics for the naive ($\hat{\beta}^{(N)}$), approximate calibration ($\hat{\beta}^{(A)}$) and exact calibration ($\hat{\beta}^{(E1)}$ and $\hat{\beta}^{(E2)}$) methods, with $\epsilon_{x_{jk}}$ following a uniform and skew normal distribution.

$\epsilon_{x_{jk}}$	β_x	Percent Bias					MSE(×100)					SE(×100)					Coverage Rate(×100)				
		$\hat{\beta}_x^{(N)}$	$\hat{\beta}_x^{(F)}$	$\hat{\beta}_x^{(A)}$	$\hat{\beta}_x^{(E1)}$	$\hat{\beta}_x^{(E2)}$	$\hat{\beta}_x^{(N)}$	$\hat{\beta}_x^{(F)}$	$\hat{\beta}_x^{(A)}$	$\hat{\beta}_x^{(E1)}$	$\hat{\beta}_x^{(E2)}$	$\hat{\beta}_x^{(N)}$	$\hat{\beta}_x^{(F)}$	$\hat{\beta}_x^{(A)}$	$\hat{\beta}_x^{(E1)}$	$\hat{\beta}_x^{(E2)}$	$\hat{\beta}_x^{(N)}$	$\hat{\beta}_x^{(F)}$	$\hat{\beta}_x^{(A)}$	$\hat{\beta}_x^{(E1)}$	$\hat{\beta}_x^{(E2)}$
Uniform	log(1.25)	-19.0	-0.2	1.0	0.7	1.4	0.42	0.34	0.27	0.28	0.28	4.94	5.86	5.22	5.25	5.28	76.0	91.5	96.0	96.1	95.8
	log(1.5)	-20.9	-1.3	-0.5	-0.1	0.7	0.99	0.54	0.34	0.35	0.36	5.21	7.33	5.79	5.96	6.00	48.7	85.5	95.0	94.5	94.6
	log(1.75)	-20.9	-0.9	-0.8	0.5	1.3	1.67	0.83	0.42	0.47	0.49	5.55	9.12	6.49	6.85	6.97	30.8	79.9	93.4	92.9	92.5
	log(2)	-21.1	-0.1	-0.9	1.2	2.3	2.50	1.10	0.52	0.62	0.66	5.93	10.51	7.20	7.81	7.97	20.8	76.4	90.9	90.7	91.3
Skew	log(1.25)	-20.6	-1.5	0.5	0.1	0.9	0.43	0.30	0.26	0.26	0.26	4.72	5.51	5.07	5.11	5.13	75.8	94.4	97.1	96.7	97.0
	log(1.5)	-20.8	-0.8	0.3	0.7	1.5	0.99	0.54	0.36	0.38	0.39	5.27	7.36	5.97	6.15	6.20	48.8	87.0	94.6	94.3	93.9
Normal	log(1.75)	-20.7	0.0	0.4	1.6	2.6	1.64	0.83	0.43	0.49	0.52	5.49	9.14	6.57	6.94	7.05	32.4	79.7	92.9	92.6	92.5
	log(2)	-21.6	0.6	-0.1	2.1	3.3	2.58	1.12	0.51	0.62	0.67	5.88	10.57	7.12	7.71	7.89	19.0	78.6	92.2	91.8	91.3

NOTE: Percent bias and MSE were computed by averaging $(\hat{\beta} - \beta)/\beta$ and $(\beta - \hat{\beta})^2$ over 1000 simulations. Standard error (SE) is the square root of the empirical variance over all replicates. Coverage rate represents the coverage of a 95% confidence interval.

Table 4

Parameter estimates for model (9) in the calibration step based on the NHS and HPFS

Models	(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
	NHS	HPFS								
Model I	61.589*	64.586*	-6.040*	-3.801*	0.466	0.029	0.100*	0.056	-0.058*	-4.414*
	(4.605)	(6.697)	(0.748)	(0.745)	(0.699)	(0.752)	(0.019)	(0.069)	(0.027)	(1.044)
Model II	63.897 *	68.178*	-6.089*	-3.757*	0.394	0.027				
	(2.079)	(4.158)	(0.761)	(0.756)	(0.709)	(0.764)				
Model III	63.781*	70.603*								
	(1.904)	(3.821)								

NOTE: The variance estimate, $\hat{\sigma}_x^2$, of Model I, II and III is 461.194, 477.446 and 503.000 respectively.

Table 5

Calibration parameter estimates for each laboratory, including the estimated intercept bias ($\hat{\xi}_d$), slope bias ($\hat{\gamma}_d$) and measurement error variance (σ_d^2), based on the pooled analysis and calibration model (3).

Laboratory	Model I			Model II			Model III		
	$\hat{\xi}_d$(SE)	$\hat{\gamma}_d$(SE)	σ_d^2	$\hat{\xi}_d$(SE)	$\hat{\gamma}_d$(SE)	σ_d^2	$\hat{\xi}_d$(SE)	$\hat{\gamma}_d$(SE)	σ_d^2
Reference	0.712(1.358)	0.014(0.023)	35.669	0.000(<0.001)	0.026(0.015)	34.133	1.657(0.951)	0.000(<0.001)	44.086
NHS	-0.651(1.316)	-0.039 (0.021)	29.776	-0.000(<0.001)	-0.051(0.010)	33.446	-2.923(0.630)	-0.000(<0.001)	0.013
HPFS	-1.061(1.347)	0.025 (0.021)	<0.001	-0.000(<0.001)	0.025(0.010)	0.002	1.266(0.703)	0.000(<0.001)	0.006

NOTE: Models I, II and III refers to Table 4. The $\hat{\sigma}_\xi^2$ corresponds to Models I, II, and III is 2.388, 1.907×10^{-7} and 7.299 respectively. The $\hat{\sigma}_\gamma^2$ corresponds to Models I, II, and III is 0.002, 0.002 and 1.239×10^{-8} respectively.

Table 6

OR-estimates and 95% confidence interval for the circulating 25(OH)D–colorectal cancer relationship based on the pooled dataset, adjusting for physical activity total (continuous), family history of colorectal cancer (yes/no), smoking (ever/never) and BMI (greater or less than 25 kg/m^2.

Methods	Model I			Model II			Model III		
	$\hat{\beta}_x$	OR	OR 95% CI	$\hat{\beta}_x$	OR	OR 95% CI	$\hat{\beta}_x$	OR	OR 95% CI
Naive	-0.125	0.882	(0.800,0.972)	-0.125	0.882	(0.800,0.972)	-0.125	0.882	(0.800,0.972)
ACM	-0.125	0.882	(0.797,0.976)	-0.125	0.882	(0.796,0.978)	-0.122	0.885	(0.804,0.975)
ECM1	-0.121	0.886	(0.801,0.980)	-0.125	0.883	(0.796,0.978)	-0.122	0.885	(0.804,0.975)
ECM2	-0.127	0.881	(0.796,0.975)	-0.126	0.881	(0.795,0.977)	-0.123	0.884	(0.803,0.973)

NOTE: Estimates correpond to a 20nmol/L increase in 25(OH)D. Models I, II and III refers to Table 4.