HHS Public Access

Author manuscript

Stat Methods Med Res. Author manuscript; available in PMC 2021 September 21.

Published in final edited form as:

Stat Methods Med Res. 2021 August ; 30(8): 1944–1959. doi:10.1177/09622802211025992.

Statistical methods for analysis of combined biomarker data from multiple nested case–control studies

Chao Cheng[1], **Abigail Sloan**[2], **Molin Wang**[2,3,4]

[1]Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA

[2]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

[3]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

[4]Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Abstract

By combining data across multiple studies, researchers increase sample size, statistical power, and precision for pooled analyses of biomarker–disease associations. However, researchers must adjust for between-study variability in biomarker measurements. Previous research often treats the biomarker measurements from a reference laboratory as a gold standard, even though those measurements are certainly not equal to their true values. This paper addresses measurement error and bias arising from both the reference and study-specific laboratories. We develop two calibration methods, the exact calibration method and approximate calibration method, for pooling biomarker data drawn from nested or matched case–control studies, where the calibration subset is obtained by randomly selecting controls from each contributing study. Simulation studies are conducted to evaluate the empirical performance of the proposed methods. We apply the proposed methods to a pooling project of nested case–control studies to evaluate the association between circulating 25-hydroxyvitamin D (25(OH)D) and colorectal cancer risk.

Keywords

Between-study variability; calibration; measurement error; pooling biomarker data

Corresponding author: Molin Wang, Harvard T.H. Chan School of Public Health, Room 828, Kresge Building, Boston, MA 02115, USA. stmow@channing.harvard.edu.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Supplementary material

Supplementary appendices and tables: "Pooling Data supp.pdf" presents additional appendices (Supplementary material Appendices A to E), Supplementary tables (Tables S1 to S4), and Supplementary figure (Figure S1) unshown in the manuscript.

R-software for the proposed method: The R software and an illustrative example for implementation of the proposed methods can be found at "Pooling Data code.zip" in the supplementary files.

1 Introduction

Pooling biomarker data across different studies to analyze biomarker–disease associations is a common strategy in epidemiological research since individual studies are often not large enough for precise estimation. By including more biomarker measurements from various studies, investigators achieve more statistical power to improve the estimates of the effect of biomarker exposure. Examples of pooling projects examining biomarker–disease relationships include the Endogenous Hormones, Nutritional Biomarkers, and Prostate Cancer Collaborative Group,[1,2] Cohort Consortium Vitamin D Pooling Project of Rarer Cancers,[3] COPD Biomarkers Qualification Consortium Database,[4] Vitamin D Pooling Project of Breast and Colorectal Cancer,[5] and the Pooling Project of Prospective Studies of Diet and Cancer.[6]

Between-laboratory variation in biomarker data may exist if not all samples are assayed at the same laboratory at the same time, and this variability will impair estimation of the biomarker–disease association. For example, the between-laboratory coefficient of variation, a measure of laboratory measurement error, was generally large (>25%) for measurements of estrone and estradiol.[7] Under such circumstances, between-study/laboratory variation in biomarker measurements should be addressed in statistical analyses for evaluating biomarker–disease relationship. Generally, calibration is implemented to harmonize measurements from different laboratories and assays by re-assaying a subset of non-case biospecimens randomly from each contributing study at a designated reference laboratory.[8–10] This calibration procedure can be utilized to adjust for the between-study measurement variability. In practice, investigators typically use only non-cases in the calibration study subsets due to potential concerns about the availability of case biospecimens.[8]

The methodology in this paper is motivated by the Vitamin D Pooling Project of Breast and Colorectal Cancer (VDPP),[5] which combined information from 17 studies to investigate the association between the circulating vitamin D (25(OH)D) and colorectal cancer. Previous researchers observed that the measurements of 25(OH)D can vary up to 40% among laboratories and assays.[11,12] This variability motivates the development of statistical models that can address the between-study variation in 25(OH)D measurements. From 2011 to 2013, the VDPP selected a random subset of controls in each study to serve as a calibration subset and remeasured their 25(OH)D values at Heartland Assays, LLC (i.e., the reference laboratory). Previous statistical methods treat the measurements from the reference laboratory as the “gold standard” measurements.[8,9] However, the reference laboratory measurements in many pooling projects only provide a benchmark value for all study-specific laboratories and are not necessarily closer to the underlying truth. Hence, treating the observed or calibrated measurements from the reference laboratory as the underlying truth may result in biased estimates of the biomarker–disease association.

In this paper, we will adjust for measurement error in both the reference and study-specific laboratories, allowing flexible calibration and measurement error models. 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 uses the methods to evaluate the association between circulating 25(OH)D levels and colorectal cancer in data pooled from VDPP, including the Nurses' Health Study (NHS) and Health Professionals Follow-Up Study (HPFS). Section 5 offers our conclusions.

2 Methods

2.1 Notation 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 jth study consists of K_j strata. The kth stratum consists of a total of M_{jk} individuals where the first $M_{jk}^{(1)}$ individuals are cases. Let X_{jkm} denote the unobserved true value of the continuous biomarker for the mth individual in the kth stratum from the jth study, Y_{jkm} denote the binary disease outcome, and $\mathbf{Z}_{jkm}$ denote other potential confounders for the X–Y relationship. Without further specification, all vectors are column vectors throughout the paper. We consider the following logistic regression model for the biomarker–disease association

$$\text{logit}\left(P\left(Y_{jkm} \mid X_{jkm}, \mathbf{Z}_{jkm}\right)\right) = \beta_{0jk} + \beta_x X_{jkm} + \boldsymbol{\beta}_z^T \mathbf{Z}_{jkm} \quad (1)$$

where β_{0jk} is a stratum-specific intercept and $\boldsymbol{\beta}_z$ is a vector of covariate effects. The parameter of interest is β_x, which denotes the log odds ratio (OR) representing the biomarker–disease association or log relative risk if the jth nested case–control study has the incidence density sampling design.

Suppose a total of N individuals contribute to the analysis across all studies. Each study j has N_j total individuals ($\sum_{j=1}^{J} N_j = N$), n_j, of whom were included in the calibration subset. Biospecimens from individuals in the calibration subset are re-assayed at a reference laboratory. Since case biospecimens may be unavailable 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 $d = 0$ indicates the reference laboratory and $d = j > 0$ indicates the local laboratory of study j. Individuals who are not selected into the calibration subset have only the local laboratory measurement $H_{jkm,j}$ available. For brevity, we use $\mathbf{H}_{jkm}$ to denote all measurements of X_{jkm}; i.e., for individuals in the calibration subset, $\mathbf{H}_{jkm} = [H_{jkm,0}, H_{jkm,j}]^T$, and for individuals out of the calibration subset, $\mathbf{H}_{jkm} = H_{jkm,j}$.

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

$$H_{jkm,d} = \xi_d + (1 + \gamma_d) X_{jkm} + \epsilon_{jkm,d} \quad (2)$$

where ξ_d and γ_d are the zero-mean random effects representing laboratory-specific intercept and slope bias, respectively, and $\epsilon_{jk,d}$ is the measurement error. The $\epsilon_{jkm,d}$ terms are independent of X_{jkm} and follow mean-zero normal distribution with laboratory-specific

variances; i.e., $\epsilon_{jkm,d} \sim N(0, \sigma_d^2)$, $d = 0, 1, \ldots, J$. Here, we also assume $\xi_d \overset{iid}{\sim} N(0, \sigma_\xi^2)$ and $\gamma_d \overset{iid}{\sim} N(0, \sigma_\gamma^2)$ for $d = 0, 1, \ldots, J$, and ξ_d and γ_d are mutually independent.

2.2 Approximate conditional likelihood

Let $\boldsymbol{Y}_{jk}$, $\boldsymbol{X}_{jk}$, $\boldsymbol{H}_{jk}$, and $\boldsymbol{Z}_{jk}$ denote their respective measurements from all individuals from stratum k of study j (i.e., $\boldsymbol{Y}_{jk} = [Y_{jk1}, \ldots, Y_{jkM_{jk}}]^T$, $\boldsymbol{X}_{jk} = [X_{jk1}, \ldots, X_{jkM_{jk}}]^T$, $\boldsymbol{H}_{jk} = [\boldsymbol{H}_{jk1}^T, \ldots, \boldsymbol{H}_{jkM_{jk}}^T]^T$, $\boldsymbol{Z}_{jk} = [\boldsymbol{Z}_{jk1}^T, \ldots, \boldsymbol{Z}_{jkM_{jk}}^T]^T$). The observed data likelihood can be written as $L = \prod_{j=1}^{J} \prod_{k=1}^{K_j} L_{jk}$, where

$$\begin{aligned} L_{jk} &= P\left(\boldsymbol{Y}_{jk} \mid \boldsymbol{H}_{jk}, \boldsymbol{Z}_{jk}, \sum_{m=1}^{M_{jk}} Y_{jkm} = M_{jk}^{(1)}\right) \\ &= \int P\left(\boldsymbol{Y}_{jk} \mid \boldsymbol{X}_{jk}, \boldsymbol{H}_{jk}, \boldsymbol{Z}_{jk}, \sum_{m=1}^{M_{jk}} Y_{jkm} = M_{jk}^{(1)}\right) f \\ &\left(\boldsymbol{X}_{jk} \mid \boldsymbol{H}_{jk}, \boldsymbol{Z}_{jk}, \sum_{m=1}^{M_{jk}} Y_{jkm} = M_{jk}^{(1)}\right) \mathrm{d}\boldsymbol{X}_{jk} \end{aligned} \tag{3}$$

We make a surrogacy assumption $P(\boldsymbol{Y}_{jk} \mid \boldsymbol{X}_{jk}, \boldsymbol{H}_{jk}, \boldsymbol{Z}_{jk}, \sum_{m=1}^{M_{jk}} Y_{jkm} = M_{jk}^{(1)}) = P(\boldsymbol{Y}_{jk} \mid \boldsymbol{X}_{jk}, \boldsymbol{Z}_{jk}, \sum_{m=1}^{M_{jk}} Y_{jkm} = M_{jk}^{(1)})$; i.e., the biomarker measurement $\boldsymbol{H}_{jk}$ will not provide any information to predict the disease status if we already know the true biomarker value $\boldsymbol{X}_{jk}$, conditional on the covariates and the matching scheme. Let $\boldsymbol{W}_{jkm}$ contain all the variables in $\boldsymbol{Z}_{jkm}$ that could be associated with X_{jkm}; i.e., we assume $P(\boldsymbol{X}_{jk} \mid \boldsymbol{H}_{jk}, \boldsymbol{Z}_{jk}, \sum_{m=1}^{M_{jk}} Y_{jkm} = M_{jk}^{(1)}) = P(\boldsymbol{X}_{jk} \mid \boldsymbol{H}_{jk}, \boldsymbol{W}_{jk}, \sum_{m=1}^{M_{jk}} Y_{jkm} = M_{jk}^{(1)})$, 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

$$\begin{aligned} L_{jk} &= \int \frac{\exp\left\{\sum_{m=1}^{M_{jk}^{(1)}} \left(\beta_x \boldsymbol{X}_{jkm} + \beta_z^T \boldsymbol{Z}_{jkm}\right)\right\}}{\sum_{\mathscr{M} \in \mathscr{C}_{jk}} \exp\left\{\sum_{m \in \mathscr{M}} \left(\beta_x \boldsymbol{X}_{jkm} + \beta_z^T \boldsymbol{Z}_{jkm}\right)\right\}} f\left(\boldsymbol{X}_{jk} \mid \boldsymbol{H}_{jk}, \boldsymbol{W}_{jk}, \sum_{m=1}^{M_{jk}} Y_{jkm} = M_{jk}^{(1)}\right) \mathrm{d}\boldsymbol{X}_{jk}, \\ &= E_{\boldsymbol{X}_{jk} \mid \boldsymbol{H}_{jk}, \boldsymbol{W}_{jk}, \sum_{m=1}^{M_{jk}} Y_{jkm} = M_{jk}^{(1)}} \left[\frac{\exp\left\{\sum_{m=1}^{M_{jk}^{(1)}} \left(\beta_x X_{jkm} + \beta_z^T \boldsymbol{Z}_{jkm}\right)\right\}}{\sum_{\mathscr{M} \in \mathscr{C}_{jk}} \exp\left\{\sum_{m \in \mathscr{M}} \left(\beta_x X_{jkm} + \beta_z^T \boldsymbol{Z}_{jkm}\right)\right\}} \right] \end{aligned} \tag{4}$$

where C_{jk} contains all the subsets of size $M_{jk}^{(1)}$ in set $\{1,\ldots, M_{jk}\}$. Generally, the probability density function (p.d.f.), $f(\boldsymbol{X}_{jk} \mid \boldsymbol{H}_{jk}, \boldsymbol{W}_{jk}, \sum_{m=1}^{M_{jk}} Y_{jkm} = M_{jk}^{(1)})$, has a complex form which contains complicated integrals. We use the following approximation to simplify the calculation of above p.d.f.

$$f\left(X_{jkm} \mid H_{jkm,d}, \boldsymbol{W}_{jkm}, Y_{jkm}\right) \approx f\left(X_{jkm} \mid H_{jkm,d}, \boldsymbol{W}_{jkm}\right) \tag{5}$$

which assumes that when $H_{jkm,d}$ and $\boldsymbol{W}_{jkm}$ are known, the disease outcome does not add much additional information for predicting X_{jkm}. The approximation performs best when (i) the association between X and Y is not strong and/or (ii) the disease is rare. Further details about the proof of these conditions are deferred to Supplementary Material Appendix A. Under assumption (5), the conditional likelihood contribution (4) becomes

$$L_{jk} \approx E_{\boldsymbol{X}_{jk} \mid \boldsymbol{H}_{jk}, \boldsymbol{W}_{jk}}\left[\frac{\exp\left\{\sum_{m=1}^{M_{jk}^{(1)}}\left(\beta_x X_{jkm} + \beta_z^T \boldsymbol{Z}_{jkm}\right)\right\}}{\sum_{\mathcal{M} \in \mathcal{C}_k} \exp\left\{\sum_{m \in \mathcal{M}}\left(\beta_x X_{jkm} + \beta_z^T \boldsymbol{Z}_{jkm}\right)\right\}}\right] \tag{6}$$

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

Practically, there could be variables that are not included in $\boldsymbol{Z}_{jk}$ and are also associated with the biomarker value $\boldsymbol{X}_{jk}$. We can take advantage of the availability of these variables when constructing the model for $\boldsymbol{X}_{jk}$ to achieve a more accurate β_x estimate. Hereafter, we use $\boldsymbol{W}_{jk}^*$ to denote all the available variables that are informative about $\boldsymbol{X}_{jk}$, possibly including variables not in $\boldsymbol{Z}_{jk}$. Then the likelihood contribution can still be written as in equation (6), with $\boldsymbol{W}_{jk}$ replaced by $\boldsymbol{W}_{jk}^*$. We discuss the benefit of considering the additional variables in Section 3 simulation studies.

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

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

$$X_{jkm} = \alpha_{0j} + \tau^T \boldsymbol{W}_{jkm} + \epsilon_{x_{jkm}} \tag{7}$$

where α_{0j} terms denote the study-specific intercepts, τ represents unknown parameters commonly for all the studies, and ϵ_{xjkm} is the error term with distribution $N\left(0, \sigma_x^2\right)$. If $\boldsymbol{W}_{jkm}$ is null, the regression (7) degenerates to $X_{jkm} = \alpha_{0j} + \epsilon_{xjkm}$. Under the specification of equations (2) and (7), $\left(X_{jkm}\middle|\boldsymbol{W}_{jkm}, \tilde{\boldsymbol{H}}_{jkm,0}\middle|\boldsymbol{W}_{jkm}, \tilde{\boldsymbol{H}}_{jkm,j} \mid \boldsymbol{W}_{jkm}\right)^T$ follows a multivariate normal distribution such that

$$\begin{pmatrix} X_{jkm} \mid \boldsymbol{W}_{jkm} \\ \tilde{H}_{jkm,0} \mid \boldsymbol{W}_{jkm} \\ \tilde{H}_{jkm,j} \mid \boldsymbol{W}_{jkm} \end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix} \mu_{X_{jkm} \mid \boldsymbol{W}_{jkm}} \\ \mu_{X_{jkm} \mid \boldsymbol{W}_{jkm}} \\ \mu_{X_{jkm} \mid \boldsymbol{W}_{jkm}} \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_x^2 & \sigma_x^2 \\ . & \sigma_x^2 + \dfrac{\sigma_0^2}{(1+\gamma_0)^2} & \sigma_x^2 \\ . & . & \sigma_x^2 + \dfrac{\sigma_f^2}{(1+\gamma_j)^2} \end{pmatrix} \right) \tag{8}$$

where $\tilde{H}_{jkm,d} = \frac{H_{jkm,d} - \xi_d}{1+\gamma_d}$ is the centralized value of $H_{jkm,d}$, for $d = 0, 1, \ldots j$, and $\mu_{X_{jkm}|\boldsymbol{W}_{jkm}}$ is the abbreviation of $\alpha_{0j} + \boldsymbol{\tau}^T \boldsymbol{W}_{jkm}$. Derivation of distribution (8) are deferred to Supplementary Material Appendix B. It follows that, for individuals out of the calibration subset

$$X_{jkm} \mid \boldsymbol{H}_{jkm}, \boldsymbol{W}_{jkm} \sim N(\rho_j \tilde{H}_{jk,j} + (1-\rho_j)\mu_{X_{jkm} \mid \boldsymbol{W}_{jkm}}, \rho_j \delta_j^2) \tag{9}$$

where $\rho_j = \frac{\sigma_x^2}{\sigma_x^2 + \delta_j^2}$, $\delta_j^2 = \frac{\sigma_j^2}{(1+\gamma_j)^2}$, and for individuals in the calibration subset

$$X_{jkm} \mid \boldsymbol{H}_{jkm}, \boldsymbol{W}_{jkm} \sim N(\rho_j^*(w_j \tilde{H}_{jk,j} + (1-w_j)\tilde{H}_{jk,0}) + (1-\rho_j^*)\mu_{X_{jkm} \mid \boldsymbol{W}_{jkm}}, \rho_j^* w_j \delta_j^2) \tag{10}$$

where $\rho_j^* = \sigma_x^2/(\sigma_x^2 + \delta_j^2 w_j)$ and $w_j = \delta_0^2/(\delta_j^2 + \delta_0^2)$. Next, we describe the procedures for estimating the parameters involved in the conditional mean of $X_{jkm}|\boldsymbol{H}_{jkm}, \boldsymbol{W}_{jkm}$.

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 $\sigma^2 = \left[\sigma_\xi^2, \sigma_\gamma^2, \sigma_x^2, \sigma_0^2, \sigma_1^2, \ldots, \sigma_J^2\right]^T$ denote the unknown parameters in the means and variances of equations (9) and (10). We rewrite $\mu_{X_{jkm}|\boldsymbol{W}_{jkm}}$ as $\tilde{\boldsymbol{W}}_{jkm}^T \boldsymbol{\theta}$, where $\tilde{\boldsymbol{W}}_{jkm} = [\boldsymbol{E}_{j,J}^T, \boldsymbol{W}_{jkm}^T]^T$ and $\boldsymbol{E}_{j,J}$ is a $J \times 1$ vector with one on jth element and zeros elsewhere. Combining equations (2) and (7) yields the following mixed-effects model

$$H_{jkm,d} = \underbrace{\tilde{\boldsymbol{W}}_{jkm}^T \boldsymbol{\theta}}_{\text{fixed effect}} + \underbrace{\xi_d + \epsilon_{x_{jkm}} + \epsilon_{jkm,d}}_{\text{random effects}} + \underbrace{\gamma_d\left(\tilde{\boldsymbol{W}}_{jkm}^T \boldsymbol{\theta}\right) + \epsilon_{x_{jkm}} \gamma_d}_{\text{interaction terms}} \tag{11}$$

where $\tilde{\boldsymbol{W}}_{jkm}^T \boldsymbol{\theta}$ is the fixed-effect term, ξ_d, $\epsilon_{x_{jkm}}$ and $\epsilon_{jkm,d}$ are the random-effect terms, $\gamma_d\left(\tilde{\boldsymbol{W}}_{jkm}^T \boldsymbol{\theta}\right)$ is an interaction term between the fixed-effect term $\tilde{\boldsymbol{W}}_{jkm}^T \boldsymbol{\theta}$ and the random-effect term γ_d, and $\epsilon_{x_{jkm}} \gamma_d$ is another interaction term between two random-effect terms, γ_d and $\epsilon_{x_{jkm}}$. It follows that $\boldsymbol{H}_{jkm}$, which contains all the measurements of $\boldsymbol{X}_{jkm}$, can be written in the matrix form

$$\boldsymbol{H}_{jkm} = \boldsymbol{U}_{jkm}^T\boldsymbol{\theta} + \boldsymbol{D}_{jkm}^T\boldsymbol{r} + \boldsymbol{P}_{jkm}^T\boldsymbol{\epsilon}_{jkm} \tag{12}$$

where if X_{jkm} is in the calibration subset

$$\boldsymbol{U}_{jkm} = \left[\tilde{\boldsymbol{W}}_{jkm}, \tilde{\boldsymbol{W}}_{jkm}\right], \boldsymbol{D}_{jkm} = \begin{bmatrix} \boldsymbol{E}_{1,J+1} & \boldsymbol{E}_{j+1,J+1} \\ \boldsymbol{Q}_{jkm,0} & \boldsymbol{Q}_{jkm,j} \end{bmatrix}, \boldsymbol{P}_{jkm} = \begin{bmatrix} 1+\gamma_0 & 1+\gamma_j \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \boldsymbol{\epsilon}_{jkm} = \begin{bmatrix} \epsilon_{x_{jkm}} \\ \epsilon_{jkm,0} \\ \epsilon_{jkm,j} \end{bmatrix}$$

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

$$\boldsymbol{U}_{jkm} = \tilde{\boldsymbol{W}}_{jkm}, \boldsymbol{D}_{jkm} = \begin{bmatrix} \boldsymbol{E}_{j+1,J+1} \\ \boldsymbol{Q}_{jkm,j} \end{bmatrix}, \boldsymbol{P}_{jkm} = \begin{bmatrix} 1+\gamma_j \\ 1 \end{bmatrix}, \boldsymbol{\epsilon}_{\text{jkm}} = \begin{bmatrix} \epsilon_{x_{jkm}} \\ \epsilon_{jkm,j} \end{bmatrix}$$

Here, $\boldsymbol{Q}_{jkm,d}$ is a $(J+1) \times 1$ vector with $\tilde{\boldsymbol{W}}_{jkm}^T\boldsymbol{\theta}$ on the $(d+1)$th element and zeros elsewhere. Since $\boldsymbol{D}_{jkm}$ and $\boldsymbol{P}_{jkm}$ contain unknown parameters $\boldsymbol{\theta}$ and $\boldsymbol{r}$ respectively, they are not covariate matrices and we may name them pseudo-covariate matrices. The variance-covariance matrix of $\boldsymbol{r}$ is $\boldsymbol{R}$: $= \text{var}(\boldsymbol{r}) = \text{diag}(\sigma_\xi^2, \ldots, \sigma_\xi^2, \sigma_\gamma^2, \ldots, \sigma_\gamma^2)$.

Now, aggregating all the measurements from all the individuals in all the studies together, model (12) can be summarized as

$$\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\theta} + \boldsymbol{D}\boldsymbol{r} + \boldsymbol{P}\boldsymbol{\epsilon} \tag{13}$$

where the matrices are defined in Appendix A.1. Since $\boldsymbol{D}$ and $\boldsymbol{P}$ depend on unknown parameters $\boldsymbol{\theta}$ and $\boldsymbol{r}$, respectively, we use $\mathscr{D}(\boldsymbol{\theta})$ and $\mathscr{P}(\boldsymbol{r})$ in replacement of $\boldsymbol{D}$ and $\boldsymbol{P}$ hereafter. Also, note that $\mathscr{D}(\boldsymbol{\theta})\boldsymbol{r}$ and $\mathscr{P}(\boldsymbol{r})\boldsymbol{\epsilon}$ are not independent as both depend on the random effects $\boldsymbol{r}$. This poses computational difficulties because it is hard to explicitly express the covariance structure of model (13), i.e., $\text{var}(\mathscr{D}(\boldsymbol{\theta})\boldsymbol{r} + \mathscr{P}(\boldsymbol{r})\boldsymbol{\epsilon})$. Instead of the standard maximum likelihood estimation algorithm for a linear mixed-effects model, we propose an "iteratively reweighted" algorithm[13] to obtain the estimators of $\boldsymbol{\theta}$, $\boldsymbol{r}$, and $\boldsymbol{\sigma}^2$. This algorithm is described as follows.

Let $\hat{\boldsymbol{\theta}}^{(0)}$ and $\hat{\boldsymbol{r}}^{(0)}$ be preliminary estimators for $\boldsymbol{\theta}$ and $\boldsymbol{r}$. Here, $\hat{\boldsymbol{\theta}}^{(0)}$ can be the ordinary least squares (OLS) estimator of the regression $E(\boldsymbol{H}|\boldsymbol{U}) = \boldsymbol{U}\boldsymbol{\theta}$, i.e., $\hat{\boldsymbol{\theta}}^{(0)} = \left(\boldsymbol{U}^T\boldsymbol{U}\right)^{-1}\boldsymbol{U}^T\boldsymbol{H}$. The elements in $\hat{\boldsymbol{r}}^{(0)}$, namely, $\hat{\gamma}_d^{(0)}$ and $\hat{\xi}_d^{(0)}$, can be set as the OLS estimators of $E\left(H_{jkm,d}\right) = \xi_d + \left(1+\gamma_d\right)\hat{X}_{jkm}^{(0)}$ for all measurements drawn from laboratory d, where $\hat{X}_{jkm}^{(0)} = \tilde{\boldsymbol{W}}_{jkm}^T\hat{\boldsymbol{\theta}}^{(0)}$ is a preliminary estimated true biomarker value.

In the tth iteration, we replace the unknown parameters, $\boldsymbol{\theta}$ and $\boldsymbol{r}$, in $\mathscr{D}(\boldsymbol{\theta})$ and $\mathscr{P}(\boldsymbol{r})$, with their estimators in the $(t-1)$th iteration, $\hat{\boldsymbol{\theta}}^{(t-1)}$ and $\hat{\boldsymbol{r}}^{(t-1)}$. Since $\mathscr{D}(\hat{\boldsymbol{\theta}}^{(t-1)})$ and $\mathscr{P}(\hat{\boldsymbol{r}}^{(t-1)})$ are

fixed values and $\boldsymbol{r}$ is assumed to be independent of $\boldsymbol{\epsilon}$, $\mathscr{D}(\hat{\boldsymbol{\theta}}^{(t-1)})\boldsymbol{r}$ is independent of $\mathscr{P}(\hat{\boldsymbol{r}}^{(t-1)})\boldsymbol{\epsilon}$. As a result, model (13) can be approximated by the following extended marginal model

$$\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\theta} + \boldsymbol{\epsilon}^{*(t)} \tag{14}$$

where $\boldsymbol{\epsilon}^{*(t)} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{V}(\boldsymbol{\sigma}^2, \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)}))$ and $\boldsymbol{V}(\boldsymbol{\sigma}^2, \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)}) = \mathscr{D}(\hat{\boldsymbol{\theta}}^{(t-1)})\boldsymbol{R}\mathscr{D}(\hat{\boldsymbol{\theta}}^{(t-1)})^T + \mathscr{P}(\hat{\boldsymbol{r}}^{(t-1)})\boldsymbol{\Sigma}\mathscr{P}(\hat{\boldsymbol{r}}^{(t-1)})^T$. The matrix $\boldsymbol{\Sigma}$ is the variance-covariance matrix of $\boldsymbol{\epsilon}$, which is defined in Appendix A.1. Note that $\boldsymbol{V}(\boldsymbol{\sigma}^2, \hat{\boldsymbol{\theta}}^{(l-1)}, \hat{\boldsymbol{r}}^{(t-1)})$ only depends on $\boldsymbol{\sigma}^2$. The log-likelihood function for model (14) is

$$\begin{aligned} l(\boldsymbol{\theta}, \boldsymbol{\sigma}^2 \mid \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)}) = & -\frac{1}{2}\Big\{\log|\boldsymbol{V}(\boldsymbol{\sigma}^2, \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)})| \\ & + (\boldsymbol{H} - \boldsymbol{U}\boldsymbol{\theta})^T \boldsymbol{V}(\boldsymbol{\sigma}^2, \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)})^{-1} (\boldsymbol{H} - \boldsymbol{U}\boldsymbol{\theta})\Big\} + \text{ constant} \end{aligned} \tag{15}$$

One can then obtain $\hat{\boldsymbol{\theta}}^{(t)}$ and $\hat{\boldsymbol{\sigma}}^{2(t)}$ as the maximum likelihood estimators based on equation (15). See Appendix B.1 for technical details.

Finally, the empirical best linear unbiased predictor (EBLUP) of $\boldsymbol{r}$ in the tth iteration is

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

where $\hat{\boldsymbol{R}}^{(t)} = diag(\hat{\sigma}_{\xi}^{2(t)}, \ldots, \hat{\sigma}_{\xi}^{2(t)}, \hat{\sigma}_{\gamma}^{2(t)}, \ldots, \hat{\sigma}_{\gamma}^{2(t)})$. The iteration continues until convergence. The convergence criteria can depend on the relative difference $\frac{\| \hat{\pi}^{(t+1)} - \hat{\pi}^{(t)} \|}{\| \hat{\pi}^{(t)} \|}$ where $\hat{\pi}^{(t)} = [\hat{\boldsymbol{\sigma}}^{2(t)T}, \hat{\boldsymbol{\theta}}^{(t)T}, \hat{\boldsymbol{r}}^{(t)T}]^T$, 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. Using distributions (9) and (10), we evaluate the mean and variance of $X_{jkm}|\boldsymbol{H}_{jkm}, \boldsymbol{W}_{jkm}$, denoted as μ_{jkm} and s_{jkm}^2 at $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{r}}$, and $\hat{\boldsymbol{\sigma}}^2$ from Section 2.4, leading to $\hat{\mu}_{jkm}$ and $\hat{s}_{jkm}^2$. Denote the likelihood contribution in equation (6) with this substitution as $\tilde{L}_{jk}$. Estimates of $\boldsymbol{\beta}$ can be obtained by maximizing the pseudo-likelihood $\tilde{L} = \Pi_{j,k}\tilde{L}_{jk}$. Although this likelihood cannot be written as an explicit function, we can use a Monte Carlo approach or Gauss–Hermite Quadrature (GHQ) approach[14,15] to calculate it numerically. The GHQ approach was developed to integrate some functions with respect to the multivariate normal distribution. It approximates the integral as a weighted sum of the function values at selected knots, and can be less computationally intensive for lower-dimension integrals. Therefore, we propose an integration dimension reduction strategy before applying the GHQ approach to calculate $\tilde{L}$.

More details about the Monte Carlo and GHQ approaches were deferred to Supplementary Material Appendix C.

We name this method *Exact Calibration Method* (ECM), as it aims to exactly calculate the likelihood contribution (6). And we denote the β-estimator from the Monte Carlo and GHQ approaches as $\hat{\beta}^{(E1)}$ and $\hat{\beta}^{(E2)}$, respectively. We abbreviated the Monte Carlo and GHQ Exact Calibration Methods as ECM1 and ECM2 hereafter. Each approach has merits and shortcomings. The GHQ method can provide accurate approximations for lower-dimensional integrations compared to the Monte Carlo approach, but it bears "curse of dimensionality" when M_{jk} is large since the number of knots grows exponentially with M_{jk}. Conversely, the accuracy of a Monte Carlo approach is typically lower than the GHQ approach, but it is robust with the dimension of integration. In fact, any numerical integration approach can be employed to calculate likelihood contribution (6) if it favors accuracy and computational efficiency. For example, one may try the quasi-Monte Carlo integration[16] to achieve a higher rate of convergence to the true integral value compared to the standard Monte Carlo integration approach when the integral is high-dimensional.

2.6 Approximate calibration method

Alternatively, we can use a second-order Taylor expansion with respect to $\boldsymbol{X}_{jk}$ about $E(\boldsymbol{X}_{jk}|\boldsymbol{H}_{jk}, \boldsymbol{W}_{jk})$ to approximate the likelihood contribution in equation (6). This yields the following approximate likelihood contribution

$$\tilde{L}_{jk}^{(A)} = \prod_{j=1}^{J}\prod_{k=1}^{K_j} \frac{\exp\left\{\sum_{m=1}^{M_{jk}^{(1)}} (\beta_x \hat{X}_{jkm} + \beta_z^T \boldsymbol{Z}_{jkm})\right\}}{\sum_{\mathcal{M} \in \mathscr{C}_{jk}} \exp\left\{\sum_{m \in \mathcal{M}} (\beta_x \hat{X}_{jkm} + \beta_z^T \boldsymbol{Z}_{jkm})\right\}} \tag{16}$$

where $\hat{X}_{jkm} = \hat{\mu}_{jkm}$ is the estimated value for X_{jkm}. We name it *Approximate Calibration Method* (ACM) since it uses a Taylor expansion to approximate the likelihood contribution (6) rather than numerically calculating it as in the ECM. We denote the estimates from the ACM as $\hat{\beta}^{(A)}$. 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 strong. Further details on the derivation of these conditions are available in Supplementary Material Appendix D.

2.7 Variance estimation of $\hat{\beta}$

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

- Generate new variance estimates for pseudo-data set i via $\tilde{\boldsymbol{\sigma}}^{2(i)} \sim N(\hat{\boldsymbol{\sigma}}^2, \widehat{\mathrm{Var}}(\hat{\boldsymbol{\sigma}}^2))$, where $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\sigma}}^2) = [-\frac{d^2 l_p(\vartheta(\boldsymbol{\sigma}^2 \mid \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{r}}), \boldsymbol{\sigma}^2)}{d(\boldsymbol{\sigma}^2)^2}|_{\boldsymbol{\sigma}^2 = \hat{\boldsymbol{\sigma}}^2}]^{-1}$.

- Generate new fixed and random effects for pseudo-data set i via $\tilde{\boldsymbol{\theta}}^{(i)} \sim N(\hat{\boldsymbol{\theta}}, \widehat{\mathrm{Var}}(\hat{\boldsymbol{\theta}}))$ and $\tilde{\boldsymbol{r}}^{(i)} \sim N(\hat{\boldsymbol{r}}, \widehat{\mathrm{Var}}(\hat{\boldsymbol{r}}))$. (Specific forms of $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\theta}})$ and $\widehat{\mathrm{Var}}(\hat{\boldsymbol{r}})$ are deferred to Supplementary Material Appendix E.)
- Compute new conditional distributions $X_{jkm}^{(i)} | \boldsymbol{H}_{jkm}, \boldsymbol{W}_{jkm}$ in equations (9) and (10) for each individual based on the new pseudo-calibration parameters $\tilde{\boldsymbol{\sigma}}^{2(i)}$, $\tilde{\boldsymbol{\theta}}^{(i)}$, and $\tilde{\boldsymbol{r}}^{(i)}$. Replace $\hat{\boldsymbol{\mu}}_{jk}$ and $\hat{s}_{jk}$ in $\tilde{L}_{jk}$ in Section 2.5 and equation (16) with $\tilde{\boldsymbol{\mu}}_{jk}^{(i)} = [\tilde{\mu}_{jk1}^{(i)}, \ldots, \tilde{\mu}_{jkM_{jk}}^{(i)}]^T$ and $\tilde{\boldsymbol{s}}_{jk}^{(i)} = diag(\tilde{s}_{jk1}^{(i)}, \ldots, \tilde{s}_{jkM_{jk}}^{(i)})$, leading to $\tilde{L}_{jk}^{(i)}$ and $\tilde{L}_{jk}^{(A),(i)}$, where $\tilde{\mu}_{jkm}^{(i)}$ and $\tilde{s}_{jkm}^{2(i)}$ are the mean and variance of $X_{jkm}^{(i)} | \boldsymbol{H}_{jkm}, \boldsymbol{W}_{jkm}$.
- Obtain point estimator $\hat{\boldsymbol{\beta}}^{(B),(i)}$ and the corresponding naive estimated variance $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}^{(B),(i)}) = [-\frac{d^2 ln\tilde{L}^{(B),(i)}}{d\boldsymbol{\beta}^2} | _{\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}^{(B),(i)}}]^{-1}$, where B could be $E1$, $E2$, and A representing ECM1, ECM2, and ACM, respectively.
- Repeat Steps 1 to 4 I times to obtain I estimates of $\hat{\boldsymbol{\beta}}^{(B),(i)}$ and $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}^{(B),(i)})$. In the following simulation studies and real data example, we choose $I = 20$.
- $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}})$ can be estimated as

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

where $\bar{\boldsymbol{\beta}} = \sum_{i=1}^{I} \frac{\hat{\boldsymbol{\beta}}^{(B),(i)}}{I}$.

Comparing with the variance estimation of $\hat{\boldsymbol{\beta}}$ using the Hessian matrix of log-likelihood $\tilde{L}$ in Section 2.5 or equation (16), this resampling estimator considers additional variations of the calibration parameter estimators, i.e., $\hat{\boldsymbol{\sigma}}^2$, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{r}}$. The 95% confidence interval for the parameter of interest, β_x, can be estimated based on normal approximation as $\left(\hat{\beta}_x - 1.96 \times \sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_x)}, \hat{\beta}_x + 1.96 \times \sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_x)}\right)$, where $\widehat{\mathrm{Var}}(\hat{\beta}_x)$ is the variance estimator of $\hat{\beta}_x$ obtained from Step 6. Our simulation studies in Section 3 show that the proposed resampling method provides satisfactory confidence interval coverage rates.

3 Simulation studies

3.1 Simulation setup and results

We first describe the data generating mechanism for the unobserved biomarker X_{jkm}, local and reference laboratory measurements $H_{jkm,0}$ and $H_{jkm,k}$, and the binary disease outcome Y_{jkm}.

To mimic a large pooling project with multiple contributing studies, we assume $J = 10$ matched case–control studies such that each includes 250 individuals (125 cases and 125

matched controls). For the jth case–control study in the pooling analysis, we first generated a large source population ($N_j = 5000$) as follows. First, we generated a standardized exposure; i.e., X has a one unit standard deviation. Specifically, we drew $W \sim N(0, 2)$ and then drew $X \sim N\left(\alpha_{0j} + \tau W, \sigma_x^2\right)$, where study-specific interprets, α_j, were generated from $N(3, 0.5^2)$, for $j = 1,\ldots, 10$. We set $\tau = 0.5$ and $\sigma_x^2 = 0.5$ such that X in each contributing study has variance of 1. Next, we generated reference and local laboratory measurements, H_0 and H_j, per measurement error model (2), where laboratory-specific intercept and slope biases, ξ_d and γ_d, were generated from $N(0, 0.5^2)$ and $N(0, 0.1^2)$, respectively, for $d = 0,\ldots, 10$. This indicates that the laboratory-specific intercept and slope, ξ_d and $1 + \gamma_d$, had the 2.5th to the 97.5th percentile (−0.98,0.98) and (0.80, 1.20), respectively. Considering that some biomarker measurements can vary up to 40% among different laboratories (such as vitamin D), we generated the variances of the measurement error term in each laboratory, σ_d^2, from Unif(0.15, 0.35), which makes the intra-laboratory correlation coefficient (ICC), computed as $\frac{\sigma_x^2}{\sigma_x^2 + \sigma_d^2}$ for laboratory d, ranges from 59% to 77%. Next, the binary outcome Y was generated from logit($P(Y = 1|X)$) = $\beta_{0j} + \beta_x X$, where we considered β_x = (log(1.25), log(1.5), log(1.75), log(2)) to represent weak to medium 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 the cases ($Y = 1$) and 125 quintuples from the controls ($Y = 0$). We first assumed a sample size of 25 for each calibration subset, where the calibration subsamples were randomly selected from the controls in the original case–control data. Since 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 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 did not implement any matching variables in our simulation study, so cases were randomly matched with controls to obtain 125 pairs in each study.

At each β_x and calibration design considered, we completed 1000 simulation replicates and investigated the performance of ACM, ECM1, and ECM2. We considered a naive method as a benchmark, which replaced X_{jkm} in model (3) with the average of $H_{jkm,0}$ and $H_{jkm,j}$ if $H_{jkm,0}$ was available and with $H_{jkm,j}$ otherwise, and fit a conditional logistic regression model to obtain a β_x estimate, denoted by $\hat{\beta}_x^{(N)}$ henceforth. For the purpose of comparison, we also included the full calibration method from Sloan et al.[17] in the simulation study. This method utilizes ordinary least square method to fit the model $H_{jkm} = \alpha_d + \beta_d H_{jkm,j} + \epsilon_{jkm,j}$ in each study-specific calibration subset. The estimated reference laboratory measurements, $\widehat{H}_{jkm,0}$, then replace X when fitting the logistic regression model in equation (1). We denote the estimated exposure effect by $\hat{\beta}_x^{(N)}$ and apply a standard sandwich method to obtain its standard error.

We compared the performance among the naive method, the full calibration method, and the proposed methods with regard to the following four operating characteristics: mean percent

bias, mean squared error (MSE), empirical standard error, and coverage rate of 95% confidence interval. The simulation results are shown in Table 1. The naive method performed poorly: all percent biases exceeded −19% and the coverage rates dropped to less than 20% as exposure effect increased. All the calibration approaches improved the coverage rates. Comparing to the full calibration method, the ACM and ECMs provided coverage rates closer to the nominal 95%, where the coverage rates under a strong biomarker–disease association (OR = 2.0) dropped to around 92% due to depression of point estimates. Even applying the sandwich variance method to correct the confidence interval, the coverage rates of the full calibration estimates were typically less than 92% under all exposure effects, and as exposure effect increased, its coverage rate dropped significantly. The full calibration method and ECM1 typically minimized the percent bias to less than 1% for all ORs considered, while the ACM and ECM2 estimates were biased downwards by roughly 0.5% to 2%. 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. Estimates from the full calibration method had larger MSEs in comparison with other calibration methods. For example, the MSE of the estimates from the full calibration method was approximately twice as large as the corresponding MSE from the ACM as OR = 2.

In consideration of the fact that the variable W_{jkm} may be unavailable in practice, we also conducted a simulation experiment where the model for X was misspecified as $X_{jkm} = \alpha_{0j} + \epsilon_{x_{jkm}}$ ($j = 1,\ldots, 10$), i.e., W_{jkm} was not used in the analysis. The results (Table 2) are similar to those in Table 1 except that the percent biases and MSEs of all the calibration methods were larger, implying that including all available variables associated with the biomarker data into the model for X can improve the estimation accuracy. Several additional simulation studies were conducted to check the performance of the proposed calibration methods, and the results are summarized in Supplementary Material Appendix F.

In summary, the proposed calibration methods demonstrated significant advantages over the naive method in terms of the percent bias, MSE, and confidence interval coverage rates. The ACM performed best with regard to MSE and percent bias under most parameter settings. All the proposed methods could provide satisfactory confidence interval for small effect sizes (OR ≤ 1.75). In contrast, the naive method was heavily biased in most simulation settings. Moreover, the performance of all the proposed methods improved when the variance of σ_γ^2 decreased and when the calibration subset sample size increased.

3.2 When X does not follow a normal distribution

The previous simulation experiments assumed that $\epsilon_{x_{jkm}}$ was normally distributed. However, the biomarker data could be skewed or fat-tailed in reality, violating the normality assumption. In this section, we investigated the proposed calibration methods when the error term in the model for X_{jkm} (and thus $\boldsymbol{H}_{jkm}$) does not follow a normal distribution. Two specific distributions for $\epsilon_{x_{jkm}}$ were investigated: the uniform distribution and the skew normal distribution.[18] Specifically, we first generated X_{jkm} based on $X_{jkm} = \alpha_{0j} + \tau W_{jkm} + \epsilon_{x_{jkm}}$, where $\epsilon_{x_{jkm}}$ followed either the uniform or the skew normal distribution, and all the parameters were adjusted to satisfy mean 0 and variance 0.5. For the skew normal distribution scenario, we set the skew parameter, s, at 1.5 and 10, leading to a moment

coefficient of skewness of approximately 0.5 and 1.0, respectively. The density plots of both skew normal distributions are visualized in Figure S1 in Supplementary Material. We generated $H_{jkm,d}$ based on $H_{jkm,d} \sim N\left(\xi_d + (1+\gamma_d)X_{jkm}, \sigma_d^2\right)$, $j = 0, 1, \ldots, 10$, where ξ_d terms, γ_d terms, and all the other design parameters were identical to those in the previous section.

Table 3 provides the simulation results, which are similar to the results in Section 3.1. The proposed methods exhibit robust performance when X is not normally distributed, especially for smaller effect sizes (OR ≤ 1.5). Across all exposure effects, the ACM and ECMs performed better in the aspects 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 in the effect estimate.

4 Application and case study

We conducted one real data example to illustrate the proposed methods. In this example, we investigated the impact of 25(OH)D levels on risk of colorectal cancer based on two studies in the VDPP. Specifically, we combined data from the NHS[19] and HPFS.[20] The NHS began enrollment in 1976 and included 121,701 female nurses aged 30 to 55 years at baseline. The HPFS began enrollment in 1986 and included 51,529 male health professionals aged 40 to 75 years at baseline. Between 1989 and 1995, both studies completed laboratory assays on blood samples for a host of biomarkers including 25(OH)D for a subset of participants. A total of 1876 participants, extracted from both studies, constituted the population for our pooling analysis. Each study randomly selected 29 controls and re-assayed their blood samples at Heartland Assays, LLC. These laboratory measurements were treated as reference laboratory measurements in this example. Our pooled analysis consisted of 615 case–control pairs matched on age and sex (1:1 or 1:2 matching), with 348 matched sets from the NHS and 267 matched sets from the HPFS.

For comparison, three models for 25(OH)D were considered. In the first model, we considered all factors potential related to 25(OH)D as covariates in model (7), including age of blood draw (continuous, ranged 43–82), week of the year at blood draw (integers, 1–52), physical activity (continuous, ≥ 0), smoking (ever/never), and body mass index (BMI) (greater or less than 25 kg/m^2). Considering the seasonal fluctuations of 25(OH)D, a periodic function $\tau_1\sin(2\pi t/52) + \tau_2\cos(2\pi t/52) + \tau_3\sin(4\pi t/52) + \tau_4\cos(4\pi t/52)$ was utilized in model (7) to fit the seasonal variation,[9] where t represents week of the year at blood draw. The second model included study-specific intercepts and the seasonal periodic function in model (7), and the third model included the study-specific intercepts only. These three models were named as Models I, II, and III, respectively.

Table 4 shows the pooled point estimates with standard errors for the regression coefficients corresponding to Models I, II, and III. The HPFS-specific intercept is slightly larger than the NHS-specific intercept in all of the three models. Moreover, the 25(OH)D measurements exhibits strong seasonal fluctuations based on Models I and II. Estimates of the parameters in the measurement error models from equation (2) for the study-specific and reference laboratories are presented in Table 5. The HPFS-specific laboratory provides the smallest

measurement error ($\sigma_2^2 < 0.01$ in Models I, II, and III), followed by the NHS-specific laboratory. However, the reference laboratory exhibits the largest measurement error, where the variance of its measurement error (i.e., σ_0^2) was greater than 30 in all three models.

Intercept and slope biases are also detected by the calibration analysis, but they are generally less noticeable comparing to the measurement errors. Figure 1 shows the estimated measurement error model of equation (2) with pointwise 95% confidence bands for the reference and study-specific laboratories, when the true 25(OH)D values are in the range of 50 to 80 nmol/L. At the same true 25(OH)D value, the reference and HPFS-specific laboratories tend to have similar 25(OH)D measurements. However, the confidence bands of the reference laboratory measurements are significantly wider than those for the HPFS-specific laboratory, indicating that the measurement error for the reference laboratory was larger. Moreover, the NHS-specific laboratory measurements tend to be slightly lower than the measurements from the other two participating laboratories. This is due to the negative intercept and slope biases in the estimated measurement error model for the NHS-specific laboratory.

Next, using a representative individual in the NHS calibration subset as an example, we illustrate how the proposed calibration model combines the information in the laboratory measurements and covariates to obtain the distribution of the underlying true 25(OH)D. As shown in Figure 2, conditional on the covariates age of blood draw, week of the year at blood draw, physical activity, smoking, and BMI, the distribution of 25(OH)D values (black area) has a large variance. Combining the information from the NHS-specific laboratory measurement, the distribution of 25(OH)D becomes more concentrated (blue area). The 95% confidence interval for the true 25(OH)D is (61, 79) nmol/L. Finally, incorporating the reference laboratory measurement, the distribution of 25(OH)D is further compressed into the red area in Figure 2. Now, the 95% confidence interval for the true 25(OH)D becomes (63, 76) nmol/L.

Finally, we applied the naive method, ACM, ECM1, and ECM2 to the pooling data set, 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 OR estimates for the association between 25(OH)D and colorectal cancer and their 95% confidence intervals are displayed in Table 6. All the analytic approaches indicate that increased 25(OH)D levels are associated with a statistically significant (based on 95% confidence level) protective effect against colorectal cancer. In either Model I, II, or III, the OR estimates among all approaches are quite similar, which is due to the fact that the measurement errors for all participating laboratories were significantly smaller than the variance of the true biomarker in our example. Specifically, the intra-laboratory correlation coefficients, defined as $\mathrm{ICC}_d = \frac{\hat{\sigma}_x^2}{\hat{\sigma}_d^2 + \hat{\sigma}_x^2}$ ($d = 0, 1, 2$, indexing the reference, NHS-specific and HPFS-specific laboratories, respectively), are generally large (e.g., in Model I: $\mathrm{ICC}_0 = 92.8\%$, $\mathrm{ICC}_1 = 93.9\%$, $\mathrm{ICC}_2 = 99.9\%$). The difference between the proposed calibration approaches and the naive method may have been larger if the ICCs are smaller.

We also investigated the relative goodness-of-fit among Models I, II, and III, three models that use different covariates for describing the measurement error process, by comparing their Akaike information criterion (AIC) values. Under the approach of Ten Eyck and Cavanaugh[21] for pseudo likelihood statistical methods, the AIC is given by $\mathrm{AIC} = -2\log\left(\widehat{\tilde{L}}(\hat{\theta},\hat{\pi})\right) + 2(p+q)$, where $\widehat{\tilde{L}}(\hat{\theta},\hat{\pi})$ denotes the pseudo-likelihood function of the ECM or ACM after substituting the point estimators, $\hat{\theta}$ and $\hat{\pi}$, and p and q denote the number of estimated parameters in θ and π, respectively. A lower AIC value indicates a better-fit model. The AIC values for the ACM for Models I, II, and III are −1261.9, −1269.9, and −1277.9, respectively, which indicates that Model III fits better than the other two candidates. In addition, the AIC values for the ECM1 and ECM2 also favor Model III (not shown here).

5 Discussion

In this paper, we proposed and evaluated statistical methods for pooling biomarker data from multiple nested case–control studies. We focused on evaluating the OR describing the association between a continuous biomarker and a binary disease outcome. In line with the common practice, we randomly selected biospecimens from the controls in each contributing study for the calibration subsets. We considered the measurement errors and biases of the observed biomarker data in both the reference and study-specific laboratories. The R software and an illustrative example for implementation of the proposed methods can be found at “Pooling_Data_code.zip” in the Supplementary Files.

The major messages of this work are as follows: First, all the proposed calibration methods, including ACM, ECM1 and ECM2, were able to obtain a less biased point and interval estimates than the naive approach, which did not adjust for the measurement error or bias. Second, across the proposed calibration methods, the ACM is preferred approach in consideration of its minimal biased point estimates and MSEs under all simulation scenarios. Third, we observed that the OR point estimates were slightly biased for strong biomarker–disease association in the simulation studies, but all the proposed calibration methods yielded satisfactory estimates under smaller exposure effects (OR ≤ 1.75).

We used the following two conditions to simplify the likelihood contribution L_{jk} in equation (4): the disease is rare and/or the association between the exposure and outcome is weak. These settings are frequently encountered in epidemiologic studies. In fact, the rare outcome condition is commonly undertaken in epidemiologic methodologies when correcting for the bias due to a mis-measured exposure with respect to a binary outcome.[22–24] In addition, weak exposure–outcome associations are frequently observed in cancer epidemiology. Many studies demonstrate that associations between cancers and nutritional factors are generally of a small magnitude.[25,26] The simulation studies also establish that the proposed methods work well for moderate effect sizes. In summary, both conditions may be plausible in many epidemiologic settings.

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 the mixed-effects model $H_d = \xi_d + (1 + \gamma_d)X + \epsilon_d$, where ξ_d and γ_d are zero-mean

random effects representing the measurement bias in laboratory d, and ϵ_d is the corresponding measurement error term. All calibration methods discussed in this work can be applied to prospective or retrospective cohort studies with a binary disease outcome.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors thank the Circulating Biomarkers and Breast and Colorectal Cancer Consortium team (R01 CA152071, 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.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: MW was supported in part by NIH/NCI grant R03 CA212799. This work was supported in part by the Intramural Program of the National Cancer Institute, Division of Cancer Epidemiology and Genetics.

Appendix 1

A.1. Matrices in model (13)

$\boldsymbol{H}$ includes all $\boldsymbol{H}_{jkm}$, sorting first by study number j, followed by matched set number k and individual number m. Specifically, $\boldsymbol{H}^T$ is defined as

$$\begin{bmatrix} \boldsymbol{H}_{111}^T, \boldsymbol{H}_{112}^T, \ldots, \boldsymbol{H}_{11M_{11}}^T, \boldsymbol{H}_{121}^T, \boldsymbol{H}_{122}^T, \ldots, \boldsymbol{H}_{12M_{12}}^T, \ldots, \boldsymbol{H}_{1K_1 1}^T, \boldsymbol{H}_{1K_1 2}^T, \ldots, \boldsymbol{H}_{1K_1 M_{1K_1}}^T, \\ \boldsymbol{H}_{211}^T, \boldsymbol{H}_{212}^T, \ldots, \boldsymbol{H}_{21M_{21}}^T, \boldsymbol{H}_{221}^T, \boldsymbol{H}_{222}^T, \ldots, \boldsymbol{H}_{22M_{22}}^T, \ldots, \boldsymbol{H}_{2K_2 1}^T, \boldsymbol{H}_{2K_2 2}^T, \ldots, \boldsymbol{H}_{2K_2 M_{2K_2}}^T, \\ \vdots \\ \boldsymbol{H}_{J11}^T, \boldsymbol{H}_{J12}^T, \ldots, \boldsymbol{H}_{J1M_{J1}}^T, \boldsymbol{H}_{J21}^T, \boldsymbol{H}_{J22}^T, \ldots, \boldsymbol{H}_{J2M_{J2}}^T, \ldots, \boldsymbol{H}_{JK_J 1}^T, \boldsymbol{H}_{JK_J 2}^T, \ldots, \boldsymbol{H}_{JK_J M_{JK_J}}^T \end{bmatrix}$$

Following the same indexing scheme in $\boldsymbol{H}$, the matrices $\boldsymbol{U}$, $\boldsymbol{D}$, and $\boldsymbol{\epsilon}$ are defined as $\boldsymbol{U}^T = [\boldsymbol{U}_{111}, \boldsymbol{U}_{112}, \ldots, \boldsymbol{U}_{JK_J M_{JK_J}}]$, $\boldsymbol{D}^T = [\boldsymbol{D}_{111}, \boldsymbol{D}_{112}, \ldots, \boldsymbol{D}_{JK_J M_{JK_J}}]$, $\boldsymbol{P} = Diag(\boldsymbol{P}_{111}^T, \boldsymbol{P}_{112}^T, \ldots, \boldsymbol{P}_{JK_J M_{JK_J}}^T)$, and $\boldsymbol{\epsilon}^T = [\boldsymbol{\epsilon}_{111}^T, \boldsymbol{\epsilon}_{112}^T, \ldots, \boldsymbol{\epsilon}_{JK_J M_{JK_J}}^T]$. Here, Diag($\boldsymbol{A}_1$, $\boldsymbol{A}_2$, …, $\boldsymbol{A}_p$) denotes the block diagonal matrix generated by matrices $\boldsymbol{A}_1$, $\boldsymbol{A}_2$,…, $\boldsymbol{A}_p$.

Similarly, $\boldsymbol{\Sigma}$, the variance-covariance matrix of $\boldsymbol{\epsilon}$, is defined as

$$\boldsymbol{\Sigma} = Diag(\boldsymbol{\Sigma}_{111}, \boldsymbol{\Sigma}_{112}, \ldots, \boldsymbol{\Sigma}_{JK_J M_{JK_J}})$$

where $\boldsymbol{\Sigma}_{jkm}$, is the variance-covariance matrix of $\boldsymbol{\epsilon}_{jkm}$ with $\boldsymbol{\Sigma}_{jkm} = \boldsymbol{D}iag(\sigma_x^2, \sigma_0^2, \sigma_j^2)$ if X_{jkm} is in the calibration subset and $\boldsymbol{\Sigma}_{jkm} = \text{diag}(\sigma_x^2, \sigma_j^2)$ if X_{jkm} is outside of the calibration subset.

B. 1. Expressions of $\hat{\theta}^{(t)}$ and $\hat{\sigma}^{2(t)}$

Maximizing equation (15) for fixed $\boldsymbol{\sigma}^2$ with respect to $\boldsymbol{\theta}$ leads to

$$\hat{\boldsymbol{\theta}}^{(t)} = \vartheta(\boldsymbol{\sigma}^2 \mid \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)}) := (\boldsymbol{U}^T \boldsymbol{V}(\boldsymbol{\sigma}^2, \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)})^{-1} \boldsymbol{U})^{-1} \boldsymbol{U}^T \boldsymbol{V}(\boldsymbol{\sigma}^2, \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)})^{-1} \boldsymbol{H}$$

It follows that the profile log-likelihood is

$$l_p(\vartheta(\boldsymbol{\sigma}^2 \mid \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)}), \boldsymbol{\sigma}^2) = = -\frac{1}{2}\begin{Bmatrix} \log|\boldsymbol{V}(\boldsymbol{\sigma}^2, \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)})| + (\boldsymbol{H} - \boldsymbol{U}\vartheta(\boldsymbol{\sigma}^2 \mid \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)}))^T \\ \boldsymbol{V}(\boldsymbol{\sigma}^2, \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)})^{-1} (\boldsymbol{H} - \boldsymbol{U}\vartheta(\boldsymbol{\sigma}^2 \mid \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)})) \end{Bmatrix}$$

+ constant

and the MLE estimators of $\boldsymbol{\sigma}^2$ and $\boldsymbol{\theta}$ are $\hat{\boldsymbol{\sigma}}^{2(t)} = \underset{\boldsymbol{\sigma}^2}{\text{argmax}}\, l_p(\vartheta(\boldsymbol{\sigma}^2 \mid \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)}), \boldsymbol{\sigma}^2)$ and $\hat{\boldsymbol{\theta}}^{(t)} = \vartheta(\hat{\sigma}^{2^{(t)}} \mid \hat{\boldsymbol{\theta}}^{(t-1)}, \hat{\boldsymbol{r}}^{(t-1)})$, respectively.

References

1. Crowe FL, Appleby PN, Travis RC, et al.Circulating fatty acids and prostate cancer risk: individual participant meta-analysis of prospective studies. J Natl Cancer Inst2014; 106.
2. Key TJ, Appleby PN, Travis RC, et al.Carotenoids, retinol, tocopherols, and prostate cancer risk: pooled analysis of 15 studies. Am J Clin Nutr2015; 102: 1142–1157. [PubMed: 26447150]
3. Gallicchio F, Helzlsouer KJ, Chow WH, et al.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 Epidemiol2010; 172: 10–20. [PubMed: 20562188]
4. Tabberer M, Benson VS, Gelhorn H, et al.The COPD biomarkers qualification consortium database: baseline characteristics of the St George's respiratory questionnaire dataset. Chron Obstruct Pulmon Dis: J COPD Found2017; 4: 112.
5. McCullough MF, Zoltick ES, Weinstein SJ, et al.Circulating vitamin D and colorectal cancer risk: an international pooling project of 17 cohorts. J Natl Cancer Inst2019; 111: 158–169. [PubMed: 29912394]
6. Smith-Warner SA, Spiegelman D, Ritz J, et al.Methods for pooling results of epidemiologic studies: the pooling project of prospective studies of diet and cancer. Am J Epidemiol2006; 163: 1053–1064. [PubMed: 16624970]
7. Hankinson SE, Manson J, Fondon SJ, et al.Faboratory reproducibility of endogenous hormone levels in postmenopausal women. Cancer Epidemiol Prev Biomark1994; 3: 51–56.
8. Sloan A, Song Y, Gail MH, et al.Design and analysis considerations for combining data from multiple biomarker studies. Star Med2019; 38: 1303–1320.
9. Gail MH, Wu J, Wang M, et al.Calibration and seasonal adjustment for matched case-control studies of vitamin D and cancer. Stat Med2016; 35: 2133–2148. [PubMed: 27133461]
10. Rosner B, Spiegelman D and Willett W. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. Am J Epidemiol 1990; 132: 734–745. [PubMed: 2403114]
11. Fai JK, Fucas RM, Banks Eet al.Variability in vitamin D assays impairs clinical assessment of vitamin d status. Intern Med J2012; 42: 43–50. [PubMed: 21395958]
12. Snellman G, Melhus H, Gedeborg R, et al.Determining vitamin d status: a comparison between commercially available assays. PLoS One2010; 5: el 1555.

13. Pinheiro J and Bates D. Mixed-effects models in S and S-PLUS. Berlin: Springer Science & Business Media, 2006.
14. Jackel PA note on multivariate Gauss-Hermite quadrature. Tech Rep2005.
15. Fiu Q and Pierce DA. A note on Gauss-Hermite quadrature. Biometrika 1994; 81: 624–629.
16. Dick J, Kuo FY and Sloan IH. High-dimensional integration: the quasi-Monte Carlo way. Acta Numerica 2013; 22: 133.
17. Sloan A, Smith-Warner SA, Ziegler RG, et al.Statistical methods for biomarker data pooled from multiple nested case–control studies. Biostatistics2019.
18. Fernandez C and Steel MF. On Bayesian modeling of fat tails and skewness. J Am Stat Assoc 1998; 93: 359–371.
19. Eliassen AH, Warner ET, Rosner B, et al.Plasma 25-hydroxyvitamin D and risk of breast cancer in women followed over 20 years. Cancer Res2016; 76: 5423–5430. [PubMed: 27530324]
20. Choi HK, Atkinson K, Karlson EW, et al.Obesity, weight change, hypertension, diuretic use, and risk of gout in men: the health professionals follow-up study. Arch Intern Med2005; 165: 742–748. [PubMed: 15824292]
21. Ten Eyck P and Cavanaugh JE. An alternate approach to pseudo-likelihood model selection in the generalized linear mixed modeling framework. Sankhya B: Indian J Stat 2018; 80: 98–122.
22. Spiegelman D, McDermott A and Rosner B. Regression calibration method for correcting measurement-error bias in nutritional epidemiology. Am J Clin Nutr 1997; 65: 1179S–1186S. [PubMed: 9094918]
23. Wang C, Hsu F, Feng Z, et al.Regression calibration in failure time regression. Biometrics1997; 53: 131–145. [PubMed: 9147589]
24. Rosner B, Willett W and Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. Stat Med 1989; 8: 1051–1069. [PubMed: 2799131]
25. Boffetta PCausation in the presence of weak associations. Crit Rev Food Sci Nutr2010; 50: 13–16.
26. Dal Maso F, Bosetti C, Fa Vecchia C, et al.Risk factors for thyroid cancer: an epidemiological review focused on nutritional factors. Cancer Causes Control2009; 20: 75–86. [PubMed: 18766448]

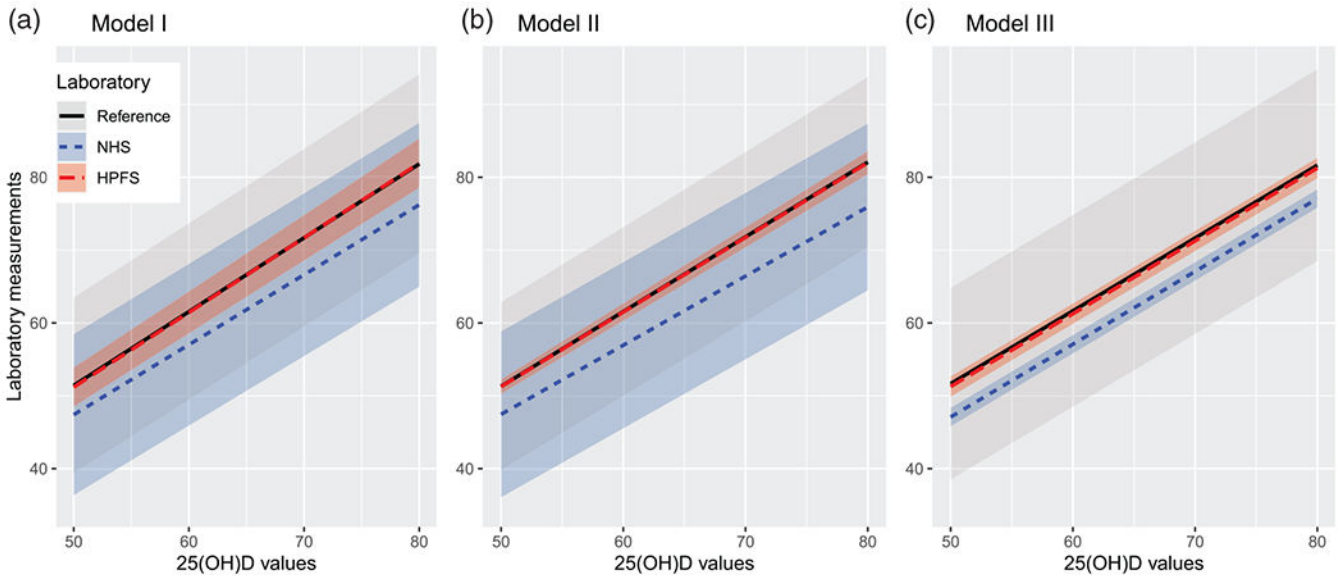

Figure 1.

Underlying true 25(OH)D values versus estimated laboratory measurements in Models I, II, and III. The black, blue, and red lines are for the reference, NHS-specific, and HPFS-specific laboratories, respectively. Black, blue, and red ribbons were added to represent the 95% pointwise confidence bands of the 25(OH)D measurements from the three participating laboratories. The distribution of the estimated measurements were based on $\widehat{H}_{jkm,d} \mid X \sim N(\hat{\mu}_{H_{jkm,d}}, \hat{\sigma}^2_{H_{jkm,d}})$, where

$\hat{\mu}_{H_{jkm,d}} = \hat{E}(\widehat{H}_{jkm,d} \mid X) = \hat{E}\left(\hat{\xi}_d + (1+\hat{\gamma}_d)X + \epsilon_{jkm,d} \mid X\right) = \hat{\xi}_d + \left(1+\hat{\gamma}_d\right)X$, and

$\hat{\sigma}^2_{H_{jkm,d}} = \widehat{\mathrm{Var}}(\widehat{H}_{jkm,d} \mid X) = \widehat{\mathrm{Var}}(\hat{\xi}_d + \left(1+\hat{\gamma}_d\right)X + \epsilon_{jkm,d} \mid X) = \widehat{\mathrm{Var}}(\hat{\xi}_d) + X^2\widehat{\mathrm{Var}}\left(\hat{\gamma}_d\right) + 2X\widehat{\mathrm{Cov}}$, $(\hat{\xi}_d, \hat{\gamma}_d) + \hat{\sigma}^2_d$

and $d = 0, 1, 2$ index the reference, NHS-specific, and HPFS-specific laboratories.

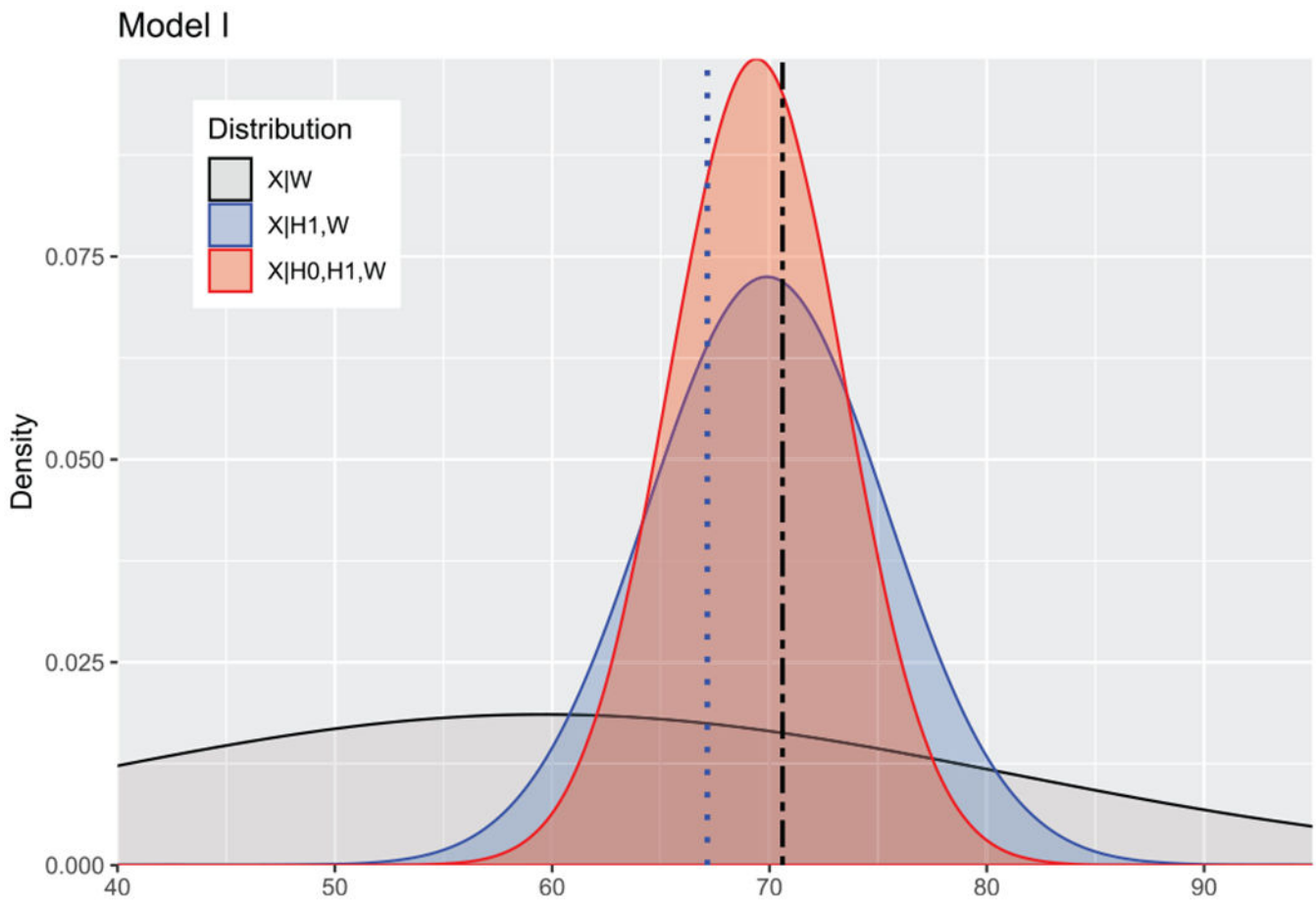

Figure 2.

The distributions of underlying true 25(OH)D of a representative individual in the NHS calibration subset, according to Model I. The black curve illustrates the distribution of 25(OH)D conditional on the covariates age of blood draw, week of the year at blood draw, physical activity, smoking, and BMI. The blue curve illustrates the distribution of 25(OH)D conditional on the covariates and the NHS-specific laboratory measurement. The red curve illustrates the distribution of 25(OH)D conditional on the covariates, the reference and NHS-specific laboratory measurements. Blue and black vertical lines were added to represent the NHS-specific and reference laboratory measurements, respectively.

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 $\hat{\beta}^{(E2)}$).

	Percent bias (coverage rate × 100)					MSE(× 100)					SE(× 100)				
β_x	$\hat{\beta}_x^{(N)}$	$\hat{\beta}_x^{(F)}$	$\hat{\beta}_x^{(A)}$	$\hat{\beta}_x^{(E1)}$	$\hat{\beta}_x^{(E2)}$	$\hat{\beta}_x^{(N)}$	$\hat{\beta}_x^{(F)}$	$\hat{\beta}_x^{(A)}$	$\hat{\beta}_x^{(E1)}$	$\hat{\beta}_x^{(E2)}$	$\hat{\beta}_x^{(N)}$	$\hat{\beta}_x^{(F)}$	$\hat{\beta}_x^{(A)}$	$\hat{\beta}_x^{(E1)}$	$\hat{\beta}_x^{(E2)}$
log(1.25)	−19.3 (77.2)	−0.5 (91.9)	1.0 (96.4)	0.7 (96.5)	1.4 (96.1)	0.43	0.34	0.28	0.28	0.28	4.93	5.83	5.26	5.29	5.32
log(1.5)	−19.8 (52.0)	−0.4 (86.2)	0.4 (95.3)	0.7 (94.9)	1.6 (95.0)	0.92	0.54	0.33	0.35	0.36	5.20	7.34	5.75	5.93	5.97
log(1.75)	−20.8 (32.3)	−0.5 (80.7)	−0.7 (93.6)	0.4 (93.0)	1.4 (93.2)	1.66	0.79	0.41	0.45	0.47	5.55	8.87	6.42	6.71	6.85
log(2)	−21.6 (18.3)	−0.9 (78.9)	−1.7 (91.8)	0.4 (92.1)	1.4 (92.1)	2.58	1.05	0.48	0.56	0.58	5.83	10.24	6.83	7.46	7.55

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

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 misspecified as $X_{jk} = \alpha_{0j} + \epsilon_{X_{jk}}$, $j = 1, \ldots, 10$.

	Percent bias (coverage rate × 100)					MSE(× 100)					SE(× 100)				
β_x	$\hat{\beta}_x^{(N)}$	$\hat{\beta}_x^{(F)}$	$\hat{\beta}_x^{(A)}$	$\hat{\beta}_x^{(E1)}$	$\hat{\beta}_x^{(E2)}$	$\hat{\beta}_x^{(N)}$	$\hat{\beta}_x^{(F)}$	$\hat{\beta}_x^{(A)}$	$\hat{\beta}_x^{(E1)}$	$\hat{\beta}_x^{(E2)}$	$\hat{\beta}_x^{(N)}$	$\hat{\beta}_x^{(F)}$	$\hat{\beta}_x^{(A)}$	$\hat{\beta}_x^{(E1)}$	$\hat{\beta}_x^{(E2)}$
log(1.25)	−19.8 (76.8)	−0.4 (91.3)	3.7 (97.2)	2.9 (97.4)	4.2 (97.3)	0.43	0.37	0.34	0.34	0.36	4.85	5.78	5.80	5.82	5.89
log(1.5)	−20.3 (50.3)	−0.2 (86.6)	3.2 (95.3)	3.5 (95.2)	4.7 (94.6)	0.95	0.56	0.51	0.54	0.59	5.23	7.47	6.99	7.21	7.41
log(1.75)	−21.3 (34.3)	−0.7 (81.9)	3.0 (95.6)	4.4 (95.7)	5.9 (95.2)	1.88	0.76	0.67	0.79	0.92	6.78	8.72	8.03	8.57	8.98
log(2)	−21.6 (17.2)	−0.8 (78.4)	2.1 (92.0)	4.7 (91.0)	6.4 (89.2)	2.58	1.11	0.85	1.14	1.34	5.80	10.51	9.11	10.19	10.72

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

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 one uniform and two skew normal distributions.

		Percent bias (coverage rate × 100)					MSE(× 100)					SE(× 100)				
$\epsilon_{X_{jk}}$	β_x	$\hat{\beta}_x^{(N)}$	$\hat{\beta}_x^{(F)}$	$\hat{\beta}_x^{(A)}$	$\hat{\beta}_x^{(E1)}$	$\hat{\beta}_x^{(E2)}$	$\hat{\beta}_x^{(N)}$	$\hat{\beta}_x^{(F)}$	$\hat{\beta}_x^{(A)}$	$\hat{\beta}_x^{(E1)}$	$\hat{\beta}_x^{(E2)}$	$\hat{\beta}_x^{(N)}$	$\hat{\beta}_x^{(F)}$	$\hat{\beta}_x^{(A)}$	$\hat{\beta}_x^{(E1)}$	$\hat{\beta}_x^{(E2)}$
Uniform	log(1.25)	−19.0 (76.0)	−0.2 (91.5)	1.0 (96.0)	0.7 (96.1)	1.4 (95.8)	0.42	0.34	0.27	0.28	0.28	4.94	5.86	5.22	5.25	5.28
	log(1.5)	−20.9 (48.7)	−1.3 (85.5)	−0.5 (95.0)	−0.1 (94.5)	0.7 (94.6)	0.99	0.54	0.34	0.35	0.36	5.21	7.33	5.79	5.96	6.00
	log(1.75)	−20.9 (30.8)	−0.9 (79.9)	−0.8 (93.4)	0.5 (92.9)	1.3 (92.5)	1.67	0.83	0.42	0.47	0.49	5.55	9.12	6.49	6.85	6.97
	log(2)	−21.1 (20.8)	−0.1 (76.4)	−0.9 (90.9)	1.2 (90.7)	2.3 (91.3)	2.50	1.10	0.52	0.62	0.66	5.93	10.51	7.20	7.81	7.97
Skew	log(1.25)	−20.6 (75.8)	−1.5 (94.4)	0.5 (97.1)	0.1 (96.7)	0.9 (97.0)	0.43	0.30	0.26	0.26	0.26	4.72	5.51	5.07	5.11	5.13
Normal	log(1.5)	−20.8 (48.8)	−0.8 (87.0)	0.3 (94.6)	0.7 (94.3)	1.5 (93.9)	0.99	0.54	0.36	0.38	0.39	5.27	7.36	5.97	6.15	6.20
(s= 1.5)	log(1.75)	−20.7 (32.4)	0.0 (79.7)	0.4 (92.9)	1.6 (92.6)	2.6 (92.5)	1.64	0.83	0.43	0.49	0.52	5.49	9.14	6.57	6.94	7.05
	log(2)	−21.6 (19.0)	0.6 (78.6)	−0.1 (92.2)	2.1 (91.8)	3.3 (91.3)	2.58	1.12	0.51	0.62	0.67	5.88	10.57	7.12	7.71	7.89
Skew	log(1.25)	−20.3 (74.1)	−0.5 (91.9)	0.1 (96.6)	−0.3 (97.0)	0.5 (96.4)	0.45	0.37	0.27	0.27	0.28	4.92	6.05	5.22	5.21	5.29
Normal	log(1.5)	−20.5 (50.8)	0.2 (86.8)	1.2 (95.8)	1.8 (95.6)	2.6 (95.4)	0.93	0.53	0.30	0.32	0.33	4.89	7.25	5.47	5.62	5.69
(s= 10)	log(1.75)	−20.5 (31.8)	1.5 (79.4)	1.5 (92.7)	2.9 (91.5)	4.0 (91.1)	1.65	0.91	0.46	0.53	0.57	5.72	9.49	6.73	7.10	7.23
	log(2)	−21.4 (19.3)	1.4 (78.5)	1.0 (92.9)	3.5 (90.9)	4.8 (90.1)	2.54	1.17	0.51	0.67	0.74	5.78	10.77	7.11	7.78	7.92

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

Table 4.

Parameter estimates for model (7) in the calibration step based on the NHS and HPFS.

Models	(Intercept)									
	NHS	HPFS	$\sin\frac{2\pi}{52}$	$\cos\frac{2\pi}{52}$	$\sin\frac{4\pi}{52}$	$\cos\frac{4\pi}{52}$	Physical	Age (blood)	Smoke	BMI
Model I	61.589 * (4.605)	64.586 * (6.697)	−6.040 * (0.748)	−3.801 * (0.745)	0.466 (0.699)	0.029 (0.752)	0.100 * (0.019)	0.056 (0.069)	−0.058 * (0.027)	−4.414 * (1.044)
Model II	63.897 * (2.079)	68.178 * (4.158)	−6.089 * (0.761)	−3.757 * (0.756)	0.394 (0.709)	0.027 (0.764)				
Model III	63.781 * (1.904)	70.603 * (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. HPFS: Health Professionals Follow-Up Study; NHS: Nurses' Health Study; BMI: body mass index; SE: standard error.

Note: The label "*"denote that the corresponding coefficient is statistically significant under a 5% significance level (i.e., p-value < 0.05).

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 (2).

		Model I			Model II			Model III		
Laboratory Name	**Laboratory Index (*d*)**	$\hat{\xi}_d$(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	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	1	−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	2	−0.061(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. HPFS: Health Professionals Follow-Up Study; NHS: Nurses' Health Study; SE: standard error.

Table 6.

OR-estimates and 95% confidence interval for the circulating 25(OH)Ddyimated intercept bias (he coverage of a 95% confidence intervausting 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).

	Model I			Model II			Model III		
Methods	$\hat{\beta}_x$	**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 correspond to a 20nmol/L increase in 25(OH)D. Models I, II, and III refers to Table 4. OR: odds ratio; CI: confidence interval; ACM: approximate calibration method; ECM: exact calibration method.