



Published in final edited form as:

Biometrics. 2014 December ; 70(4): 835–844. doi:10.1111/biom.12205.

Covariate Measurement Error Correction Methods in Mediation Analysis with Failure Time Data

Shanshan Zhao and Ross L. Prentice

Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, U.S.A

Ross L. Prentice: rprentic@whi.org

Summary

Mediation analysis is important for understanding the mechanisms whereby one variable causes changes in another. Measurement error could obscure the ability of the potential mediator to explain such changes. This paper focuses on developing correction methods for measurement error in the mediator with failure time outcomes. We consider a broad definition of measurement error, including technical error and error associated with temporal variation. The underlying model with the ‘true’ mediator is assumed to be of the Cox proportional hazards model form. The induced hazard ratio for the observed mediator no longer has a simple form independent of the baseline hazard function, due to the conditioning event. We propose a mean-variance regression calibration approach and a follow-up time regression calibration approach, to approximate the partial likelihood for the induced hazard function. Both methods demonstrate value in assessing mediation effects in simulation studies. These methods are generalized to multiple biomarkers and to both case-cohort and nested case-control sampling design. We apply these correction methods to the Women's Health Initiative hormone therapy trials to understand the mediation effect of several serum sex hormone measures on the relationship between postmenopausal hormone therapy and breast cancer risk.

Keywords

Cox model; Mean-variance estimating functions; Measurement error; Mediation analysis; Regression calibration

1 Introduction

Mediation analysis is important in biomedical and social sciences research to understand the mechanisms whereby one variable causes changes in another (MacKinnon, 2008). A classical mediation analysis compares coefficients of the independent variable Z in two linear models: one regresses the outcome Y on Z and other covariates C , while the other regresses Y on Z , C and the potential mediator X . There is evidence of X mediating the

Correspondence to: Ross L. Prentice, rprentic@whi.org.

Web Appendix A, B referenced in Sections 2 and 4 and the R code used to implement the simulations are available with this paper at the *Biometrics* website on Wiley Online Library.

relationship between Z and Y , if the coefficient of Z in the second model is substantially closer to the null compared to that in the first. With failure time data, Lin et al. (1997) considered the mediation by comparing two Cox proportional hazards models, and they discussed conditions under which the two Cox models are approximately compatible. Lange and Hansen (2011) proposed a decomposition of the total treatment effect into ‘natural’ direct and indirect effects under the Aalen additive hazards model, assuming that X can be modeled by a linear regression on Z and C . In this paper, we extend the methods in Lin et al. (1997) to settings in which mediator measurement error needs to be taken into account.

Covariate measurement error methods have been investigated in failure time data settings. Hughes (1993) examined the naive approach, which replaces X with an observed error prone W in the Cox model, and found that the bias depends on true coefficient value, measurement error magnitude, censoring mechanism and others factors. Prentice (1982) considered the induced hazard function as

$$\lambda(t;W, Z, C) = E\{\lambda(t;X, Z, C) | \tilde{T} \geq t; W, Z, C\},$$

where T denotes the failure time. It was noted that when $\lambda(t; X, Z, C)$ follows a Cox model, the corresponding induced hazard ratio involves the baseline hazard function due to the conditioning event $\{T \geq t\}$. However the induced hazard can typically be approximated in the rare disease setting by replacing X with $E(X|W, Z, C)$, the so-called regression calibration approach. Wang et al. (1997) provided a suitable variance estimator for resulting regression parameter estimates. Xie et al. (2001) extended this method to risk set regression calibration, which recalibrates at each failure time. Zhou and Pepe (1995), Zhou and Wang (2000) and Carroll et al. (1995) investigated nonparametric approaches to estimate the model the induced hazard. Other measurement error correction approaches include a nonparametric corrected score approach proposed by (Huang and Wang 2000, 2006), and a full likelihood approach proposed by Hu et al. (1998). None of these methods has been investigated in the mediation analysis setting.

Here we propose two correction methods based on the induced partial likelihood in Section 2. We describe procedures to estimate parameters needed for the correction methods in Section 3. The performances of the proposed methods are demonstrated through simulation studies in Section 4. Section 5 applies our methods to the Women’s Health Initiative (WHI) hormone therapy trials. We conclude with discussion in Section 6.

2 Calibration Approaches

2.1 Model Assumptions

We assume an underlying causal diagram is as in Figure 1. The outcome $Y = (T, \delta)$ relates directly with pre- and post-randomization biomarker values $\mathbf{X} = (X_0, X_1)$ and with treatment assignment $Z \in \{0, 1\}$, where $T = \min(\tilde{T}, C)$ is the censored failure time, and \tilde{T}, C are the underlying failure and censoring times, δ is a non-censoring indicator. \tilde{T} and C are assumed to be independent given (\mathbf{X}, Z) . In addition, the post-randomization biomarker value X_1 (or equivalently, the change due to treatment $X_1 - X_0$) may mediate the relationship between Z

and Y . Thus treatment Z can have both a direct effect and an indirect effect through the biomarker change $X_1 - X_0$ on Y . For now, we do not consider other covariates C , but all the methods described can be extended readily to include covariates. To assess the mediation, we compare treatment effects α_Z and β_Z from the following two Cox models:

$$\lambda(t; X_0, Z) = \lambda_0(t) \exp(\alpha_Z Z + \alpha_0 X_0) \quad (1)$$

$$\lambda(t; X_0, X_1, Z) = \lambda_1(t) \exp(\beta_Z Z + \beta_0 X_0 + \beta_1 X_1). \quad (2)$$

Although the two models may be technically incompatible, as discussed in Lin et al. (1997), (1) is a good approximation of the marginal hazard induced from (2) if the failure time outcome is rare, that is, $\Lambda_2(t) = \int_0^t \lambda_2(s) ds$ is small, or otherwise if β_1 is small. Hence, if β_Z is much closer to 0 compared to α_Z , we can reasonably conclude that X substantially mediates the relationship between Z and Y .

We assume that biomarker values (X_0, X_1) are measured with mean zero classical measurement error, so that $W_j = X_j + U_j$, where U_j is independent of X_j given Z_j , $j = 0, 1$. As a naive approach, we replace $\mathbf{X} = (X_0, X_1)$ with $\mathbf{W} = (W_0, W_1)$ in the above models:

$$\begin{aligned} \lambda(t; W_0, Z) &= \lambda_2(t) \exp(a_Z Z + a_0 W_0) \\ \lambda(t; W_0, W_1, Z) &= \lambda_3(t) \exp(b_Z Z + b_0 W_0 + b_1 W_1) \end{aligned}$$

Since (X_0, U_0) are pre-randomization variables and typically independent of Z , a_Z is expected to be close to α_Z . However, (X_1, U_1) are post-randomization variables whose distributions may depend on Z . In this case, using b_Z to approximate β_Z may involve a large bias, and lead to incorrect conclusions about mediation. We will focus on reducing bias in β_Z estimation.

The induced hazard from model (2) is

$$\lambda(t; \mathbf{W}, Z) = \lambda_1(t) E\{\exp(\beta_Z Z + \beta_X^T \mathbf{X}) | \tilde{T} \geq t; \mathbf{W}, Z\}, \quad (3)$$

where $\beta_X = (\beta_0, \beta_1)^T$. We denote the k distinct failure times in a cohort study by $\{t_1, t_2, \dots, t_k\}$, and let i be the index of the individual failing at t_i . The corresponding partial likelihood can be written as

$$PL(\beta) = \prod_{i=1}^k \frac{E\{\exp(\beta_Z Z_i + \beta_X^T \mathbf{X}_i) | \tilde{T} \geq t_i, \mathbf{W}_i, Z_i\}}{\sum_{j \in R(t_i)} E\{\exp(\beta_Z Z_j + \beta_X^T \mathbf{X}_j) | \tilde{T} \geq t_i, \mathbf{W}_j, Z_j\}}, \quad (4)$$

which typically depends on $\lambda_1(t)$.

2.2 Mean-Variance Regression Calibration

Under the rare disease assumption $\Pr(T \geq t | \mathbf{X}, Z) \approx 1$ for all follow-up times t , the induced hazard $\lambda(t; \mathbf{W}, Z)$ in (3) can be approximated by

$$\lambda(t; \mathbf{W}, Z) \approx \lambda_1(t) E\{\exp(\beta_z Z + \beta_x^T \mathbf{X}) | \mathbf{W}, Z\}, \quad (5)$$

and we replace “ \approx ” in (5) by “=” subsequently. This approximation implies that the joint distribution of $(\mathbf{X}, \mathbf{U} | T \geq t, Z)$ is constant over time. If $(\mathbf{X}, \mathbf{U} | Z)$ is jointly normal

$$(\mathbf{X}^T, \mathbf{U}^T) | Z \sim N((\mathbf{M}_z^T, 0)^T, \text{diag}(\Sigma_z, \Delta_z)). \quad (6)$$

the conditional distribution of $(\mathbf{X} | \mathbf{W}, Z)$ is also normal with mean $\mathbf{E}(\mathbf{X} | \mathbf{W}, Z)$ and variance $\mathbf{V}(\mathbf{X} | \mathbf{W}, Z)$. Then the induced hazard can be written as

$$\lambda(t; \mathbf{W}, Z) = \lambda_1(t) \exp\{\beta_z Z + \beta_x^T \mathbf{E}(\mathbf{X} | \mathbf{W}, Z) + \frac{1}{2} \beta_x^T \mathbf{V}(\mathbf{X} | \mathbf{W}, Z) \beta_x\}, \quad (7)$$

which can be written as a Cox model with covariates \mathbf{W}, Z and their interactions:

$$\lambda(t; \mathbf{W}, Z) = \lambda_4(t) \exp(\gamma_0 W_0 + \gamma_1 W_1 + \gamma_2 Z + \gamma_3 W_0 Z + \gamma_4 W_1 Z). \quad (8)$$

Here $\boldsymbol{\gamma} = \{\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4\}$ is a function of $\boldsymbol{\beta} = (\beta_z, \beta_0, \beta_1)$ and distribution parameters $\mathcal{A} = \{\mathbf{M}_z, \Sigma_z, Z; Z = 0, 1\}$. When \mathcal{A} is known, maximizing the partial likelihood for (8) as a function of $\boldsymbol{\beta}$ using, for example, the Newton-Raphson method gives estimates of $\boldsymbol{\beta}$. Otherwise, a consistent estimate $\hat{\mathcal{A}}$ is needed for plugging into the partial likelihood. We discuss how to obtain $\hat{\mathcal{A}}$ in Section 3.

This approach is similar to that proposed in Wang et al. (2001), except that we assume normality to avoid higher order moments of the distribution of \mathbf{X} given (\mathbf{W}, Z) . Compared to conventional regression calibration, expression (7) makes use of both the conditional mean and variance. Hence, we refer to this method as a mean-variance regression calibration (MVC). Under the rare disease assumption, this approach is expected to provide hazard ratio estimates with reduced biases compared to either a naive approach or a conventional regression calibration approach without the conditional variance term.

2.3 Follow-up Time Regression Calibration

While we expect MVC to provide better estimates compared to other approaches just mentioned, its performance may deteriorate under departure from the rare disease assumption. In this section, we modify MVC in an attempt to reduce any such deterioration.

To compute the exact partial likelihood (4), the joint distributions $(\mathbf{X}, \mathbf{W} | T \geq t, Z)$ at all failure times would be needed. When the number of failures is large, it is computationally intensive to calibrate at every failure time. Also, at later failure times, calibration accuracy

may be low due to limited risk set sizes. In contrast, with MVC, we assume the conditional distribution of $(\mathbf{X}, \mathbf{W}|T \geq t, Z)$ is constant over time t , then only one calibration is needed. There can be remaining biases in parameter estimates due to differential changes in the covariate distribution over time between treatment arms.

We propose a follow-up time regression calibration (FUC) to avoid the two extremes described above. We divide the time axis into L intervals: $[I_1, I_2), [I_2, I_3), \dots, [I_L, I_{L+1})$, where $I_1 = 0$ and $I_{L+1} = \infty$; then calibrate at each $I_i, i = 1, 2, \dots, L$. This way, we assume covariate distribution is constant within each interval, but may differ between intervals. By adjusting L , we can balance between accuracy and computational burden. If $L = 1$, this is the MVC. If $L = k + 1$ and $I_{i+1} = t_i, i = 1, 2, \dots, k$, calibration is done at time 0 and at each failure time. This corresponds to a special case of risk set regression calibration.

Specifically, we approximate the partial likelihood by

$$PL(\beta) \approx \prod_{l=1:L} \prod_{t_i \in [I_l, I_{l+1})} \frac{E\{\exp(\beta_Z Z_i + \beta_X^T \mathbf{X}_i) | \tilde{T}_i \geq I_l, \mathbf{W}_i, Z_i\}}{\sum_{j \in R(t_i)} E\{\exp(\beta_Z Z_j + \beta_X^T \mathbf{X}_j) | \tilde{T}_j \geq I_l, \mathbf{W}_j, Z_j\}}$$

We further assume that $(\mathbf{X}, \mathbf{U}|T \geq I_l, Z_i)$ is jointly normal with distribution parameters $\mathcal{A}_l = \{\mathbf{M}_Z(I_l), \Sigma_Z(I_l), z; Z = 0, 1\}, l = 1, 2, \dots, L$, and

$$(\mathbf{X}^T, \mathbf{U}^T)^T | \tilde{T} \geq I_l, Z \sim N((\mathbf{M}_Z(I_l)^T, 0)^T, \text{diag}(\sum_z (I_l), \Delta_z)). \tag{9}$$

Now the partial likelihood reduces approximately to

$$PL(\beta) = \prod_{l=1:L} \prod_{t_i \in [I_l, I_{l+1})} \frac{\exp\{\beta_Z Z_i + \beta_X^T \mathbf{E}(\mathbf{X}_i | \mathcal{A}_l, \mathbf{W}_i, Z_i) + \frac{1}{2} \beta_X^T \mathbf{V}(\mathbf{X}_i | \mathcal{A}_l, \mathbf{W}_i, Z_i) \beta_X\}}{\sum_{j \in R(t_i)} \exp\{\beta_Z Z_j + \beta_X^T \mathbf{E}(\mathbf{X}_j | \mathcal{A}_l, \mathbf{W}_j, Z_j) + \frac{1}{2} \beta_X^T \mathbf{V}(\mathbf{X}_j | \mathcal{A}_l, \mathbf{W}_j, Z_j) \beta_X\}}, \tag{10}$$

where $\mathbf{E}(\mathbf{X}_j | \mathcal{A}_l, \mathbf{W}_j, Z_j)$ and $\mathbf{V}(\mathbf{X}_j | \mathcal{A}_l, \mathbf{W}_j, Z_j)$ are the corresponding conditional mean and variance. If the joint distribution $(\mathbf{X}, \mathbf{U}|T \geq I_l, Z)$ is not normal, equation (10) can be considered as a second-order Taylor approximation. With this approximated partial likelihood, we first derive the conditional mean and variance of \mathbf{X} at each $I_l, l = 1, 2, \dots, L$, and then plug them into the partial likelihood to get MLE $\hat{\beta}$. Theoretically, dividing time into shorter intervals may lead to a less biased $\hat{\beta}$. However, we do not recommend choosing a large L due to the increasing computation time and unstable performance at later intervals. From numerical evaluation, it is preferable to choose I_l as the l^{th} L-quantile of all failure times, to have similar information accumulation within each time interval. The procedures of estimating $\mathcal{A}_l, l = 1, 2, \dots, L$ are discussed in detail in Section 3.

The idea of FUC was mentioned in Liao et al. (2011) without a detailed development. This approach relaxes the constant covariate distribution assumption, thus is expected to be less sensitive to the rare disease assumption. Allowing control of the number of calibrations (L) opens the possibility of estimates that are both reliable and computationally efficient.

2.4 Asymptotic Properties

We use techniques similar to those discussed in Andersen and Gill (1982) to develop asymptotic properties for the two calibration approaches. MVC can be considered as a special case of FUC with $L = 1$. Under some mild regularity conditions, we have Theorem 1 for consistency and Theorem 2 for asymptotic normality:

Theorem 1: Under regularity conditions, $\hat{\beta} \xrightarrow{P} \beta^*$, where β^* is the true value of β in the approximate induced hazard model (10).

Theorem 2: Under regularity conditions,

$$n^{1/2}(\hat{\beta} - \beta^*) \xrightarrow{D} N(0, \Omega(\beta^*)^{-1}\{B(\beta^*)+D(\beta^*)\}\Omega(\beta^*)^{-1}), \text{ as } n \rightarrow \infty.$$

Theorem 1 shows that $\hat{\beta}$ is consistent for a value β^* , that typically differs somewhat from β . However, the bias $|\beta^* - \beta|$ tends to be small in many contexts, as will be shown in Section 4 simulation studies. Theorem 2 states that $\hat{\beta}$ has a sandwich-form variance. The middle part of the variance arises from two sources: one from the regular estimating equation, another from the variability in estimating distribution parameters $\alpha_l, l = 1, 2, \dots, L$. Detailed regularity conditions, proof of both theorems and explicit formula for $\Omega(\cdot)$, $B(\cdot)$ and $D(\cdot)$ are given in Web Appendix A.

2.5 Extension to Multiple Mediators and to Other Sampling Designs

When there are $K(K > 1)$ biomarkers measured for each subject, one can ask whether these biomarkers jointly mediate the relationship between Z and Y . It is straightforward to extend MVC and FUC to multiple mediators: $\mathbf{X}, \mathbf{U}, \mathbf{W}$ become $2K \times 1$ vectors. The approximate induced hazards (7) and (10) are the same, with joint conditional means and variances of the K markers plugged in. All the other steps follow as for a univariate biomarker.

Prentice (1986) proposed the case-cohort design as a way to reduce data collection burden for large cohort studies with infrequent failures. A subcohort of sample size n_{sc} is randomly selected from the entire cohort of sample size n at the beginning of the study. Covariate histories are only assembled for the subcohort members and the cases. Barlow (1994) viewed this design as a weighted cohort study with pseudo-partial likelihood

$$PL^*(\beta|\mathbf{X}, Z) = \prod_{i=1}^k \frac{w_i(t_i)\exp(\beta_X^T \mathbf{X}_i + \beta_Z Z_i)}{\sum_{j \in R(t_i)} w_j(t_i)\exp(\beta_X^T \mathbf{X}_j + \beta_Z Z_j)},$$

where at time t_i , case i has weight 1, at risk members in the subcohort have weight equal to the inverse sampling rate n/n_{sc} , and other subjects have weight 0. MVC and FUC can be easily adopted. The induced pseudo-partial likelihood is approximated by

$$\prod_{l=1}^L \prod_{i:t_i \in [I_l, I_{l+1})} \prod_{j \in R(t_i)} \frac{w_i(t_i) \exp[\beta_z Z_i + \beta_x^T E(\mathbf{X}_i | \mathbf{W}_i, Z_i, \mathcal{A}_{I_l}) + \frac{1}{2} \beta_x^T V(\mathbf{X}_i | \mathbf{W}_i, Z_i, \mathcal{A}_{I_l}) \beta_x]}{\sum_{j \in R(t_i)} w_j(t_i) \exp[\beta_z Z_j + \beta_x^T E(\mathbf{X}_j | \mathbf{W}_j, Z_j, \mathcal{A}_{I_l}) + \frac{1}{2} \beta_x^T V(\mathbf{X}_j | \mathbf{W}_j, Z_j, \mathcal{A}_{I_l}) \beta_x]}$$

with FUC of L intervals: $[I_1, I_2), [I_2, I_3), \dots, [I_L, I_{L+1})$, and MVC as a special case with $L = 1$.

A nested case-control study can be viewed as a cohort study with outcome-dependent weighting, and analyzed through the inverse probability weight estimator framework (Cai and Zheng, 2011). Both MVC and FUC can be applied similarly to the weighted partial likelihood as in the case-cohort design.

3 Measurement Error Model and Biomarker Process Modeling

So far, we restricted U to be normally distributed mean zero classical measurement error. In this section, we consider a class of measurement error models, and discuss the estimation of corresponding distribution parameters under data structures arising in mediation analysis.

3.1 Measurement Error Model

Although modeling \mathbf{W} is not our primary interest, we can usefully decompose it to understand its variability. There are at least three sources of random variations that could be associated with the observed W_{ij} , which is the j^{th} measure on the i^{th} subject (Diggle et al., 2002, Chapter 5): subject-specific random effects, temporal variation and technical error:

$$W_{ij} = \mu(Z_i, t_j) + b_i(Z_i, t_j) + S_{ij}(Z_i, t_j) + \varepsilon_{ij}. \quad (11)$$

Here, $\mu(Z_i, t_j)$ is a fixed population mean, which may differ by treatment Z_i and time t_j . Also $b_i(Z_i, t_j)$ is a subject-specific random effect. It represents the difference between the mean of the i^{th} subject's measures and the population mean and has mean zero. $S_{ij}(Z_i, t_j)$ is the temporal variation, which also has mean zero. The within-subject correlation is typically weaker as the time separation increases. Finally, ε_{ij} is the noise, which is assumed to have mean zero and to be uncorrelated with ε_{ik} if $j \neq k$. We refer to ε_{ij} as the technical error, even though ε_{ij} may incorporate local temporal variation beyond that attributable to the measurement technology. These four parts are assumed to be independent of each other given (Z_i, t_j) and independent between subjects. With this decomposition, we specify two formulations of measurement errors: uncorrelated and correlated measurement errors.

By uncorrelated measurement errors, we are considering the following specification:

$$X_{ij} = \mu(Z_i, t_j) + b_i(Z_i, t_j) + S_{ij}(Z_i, t_j), \quad U_{ij} = \varepsilon_{ij}$$

We consider the technical error as the only source of measurement error, and the targeted X_{ij} is the true biomarker value of subject i at time t_j . Under this definition, measurement errors are independent between and within subjects: with $z = 0, 1$,

$$M_z = (\mu_0, \mu_{z+1})^T, \Sigma_z = \begin{pmatrix} \sigma_0^2 & \rho_z \sigma_0 \sigma_{z+1} \\ \rho_z \sigma_0 \sigma_{z+1} & \sigma_{z+1}^2 \end{pmatrix}, \Delta_z = \text{diag}(\sigma_{e0}^2, \sigma_{e(z+1)}^2), \quad (12)$$

This distribution may be further simplified. For example, it may be appropriate to assume that the variance of X is constant over time (i.e., $\sigma_1^2 = \sigma_2^2$). Also, one could consider assumptions that the measurement error distribution does not depend on X or Z (i.e.,

$$\sigma_{e0}^2 = \sigma_{e1}^2 = \sigma_{e2}^2), \text{ or that variance ratios are constant (i.e., } k_0 = k_1 = k_2, \text{ where } k_i = \frac{\sigma_{e_i}^2}{\sigma_i^2}.$$

By correlated measurement errors, we are considering the following model:

$$X_{ij} = \mu(Z_i, t_j) + b_i(Z_i, t_j), \quad U_{ij} = S_{ij}(Z_i, t_j) + \varepsilon_{ij}.$$

With this specification, measurement error includes both the technical error and the temporal variation, thus measurement errors within a subject are correlated. The targeted X_{ij} is a subject-specific mean biomarker level, which may change with time and treatment. In the important special case where both $\mu(Z_i, t_j)$ and $b_i(Z_i, t_j)$ do not change with t_j , X_{ij} is considered as the subject's long-term average biomarker value. Under this specification,

$$M_z = (\mu_0, \mu_{z+1})^T, \Sigma_z = \begin{pmatrix} \sigma_0^2 & \rho_z \sigma_0 \sigma_{z+1} \\ \rho_z \sigma_0 \sigma_{z+1} & \sigma_{z+1}^2 \end{pmatrix}, \Delta_z = \begin{pmatrix} \sigma_{e0}^2 & r_z \sigma_{e0} \sigma_{e(z+1)} \\ r_z \sigma_{e0} \sigma_{e(z+1)} & \sigma_{e(z+1)}^2 \end{pmatrix}, \quad (13)$$

with $z = 0, 1$ and $\rho_0 = 1$. Note the correlation between X_0 and X_1 becomes exactly 1 in the control group, and measurement errors are correlated with each other (i.e., $r_0, r_1 \neq 0$). Again, further constraints may be suitable in applications.

The choice between uncorrelated and correlated measurement errors depends substantially on the research question of interest. If the long-term average biomarker level is more relevant to disease risk mediation, then correlated measurement error model is of interest. If one simply wishes to conduct mediation analysis that is adjusted for technical error, then uncorrelated measurement error is more appropriate.

3.2 Biomarker Process Modeling

To estimate $E(X|W, Z, A)$ and $V(X|W, Z, A)$ in MVC, one needs to estimate A . The likelihood of the observed W can be written based on the joint normal distribution in (6) and detailed parameter specifications in (12) and (13). Notice that with uncorrelated measurement error, there are 8 variance-covariance parameters (i.e.,

$\sigma_0^2, \sigma_1^2, \sigma_2^2, \sigma_{e0}^2, \sigma_{e1}^2, \sigma_{e2}^2, \rho_0, \rho_1$), but only 5 unique components in the covariance matrix of W (i.e., $\sigma_0^2 + \sigma_{e0}^2, \sigma_1^2 + \sigma_{e1}^2, \sigma_2^2 + \sigma_{e2}^2, \rho_0 \sigma_0 \sigma_1, \rho_1 \sigma_0 \sigma_2$). Similarly, with correlated measurement error, there are more parameters than unique components in the covariance matrix. To solve this identifiability problem, additional information is needed. First, we can consider some constraints as discussed above. Second, we can sometimes obtain estimates of some

parameters from external data. For example, one may be able to assume that (ρ_0, ρ_1) in the uncorrelated measurement error scenario, and (ρ_1, r_0, r_1) in the correlated measurement error scenario, or variance ratios k_i are similar across studies. If there is a study with the same biomarker measured longitudinally, we can estimate these parameters to be plugged into the likelihood of \mathbf{W} . If unfortunately there is no external information, sensitivity studies on the above mentioned distribution parameters will be needed to cover a range of possible values.

With FUC, additional steps are needed to estimate $\hat{\alpha}_l, l = 2, \dots, L$. Notice that we let both $\mathbf{M}_Z(\cdot)$ and $\Sigma_Z(\cdot)$ vary with time, but \mathbf{Z} is not time-varying. This is because study subjects with longer survival time may have different characteristics in \mathbf{X} , but measurement error \mathbf{U} does not affect survival time. In addition, we expect differential changes of \mathbf{X} distribution in the two treatment arms. Thus we allow common parameters in the distribution specification (12) and (13), such as μ_0 and σ_0^2 , to be updated separately within treatment groups. At each cutoff time $t_l, l = 2, \dots, L$, we maximize the likelihood within each treatment arm on subjects still at risk, with $\hat{\alpha}_z$ estimated at $t = 0$ plugged in. This way, we get estimate $\hat{\alpha}_l$, and can subsequently estimate $\hat{\mathbf{E}}(\mathbf{X}|\mathbf{W}, \mathbf{Z}, \hat{\alpha}_l), \mathbf{V}(\hat{\mathbf{X}}|\mathbf{W}, \mathbf{Z}, \hat{\alpha}_l)$.

When joint mediation effect of multiple biomarkers is of interest, ideally one would model their joint distribution. Thus in addition to the covariance matrix of each marker, one needs to specify the between-biomarker correlation structures. With relatively limited external information, fitting such a model may lead to unstable performance, which can adversely influence the calibration performance. Hence, it may be preferable to calibrate biomarkers individually. This individual calibration approach, however, could result in some efficiency loss. More comprehensive biomarker process models can be applied when the external dataset is large and longitudinal.

With a case-cohort design, similar methods can be applied to estimate distribution parameters, but only subcohort members at risk at the time of calibration are used. This approach is expected to provide approximately unbiased distribution parameter estimates, as the subcohort is a random sample of the population. However the performance can be unstable if the subcohort is small. With a nested case-control design, we may estimate distribution parameters similarly as in a cohort design with inverse probability weights.

4 Simulation Studies

In this section, we conduct several simulation studies to investigate the performances of the two proposed mediation analysis measurement error correction methods.

We specify $\mu_0 = \mu_1 = 0, \mu_2 = 0.5, \sigma_0^2 = \sigma_1^2 = 1, \sigma_2^2 = 1.2$, and $\beta_0 = \beta_1 = 1, \beta_Z = \log(1.5) \approx 0.41, \lambda_1(t) = 1$. In the uncorrelated measurement error setting, we assume $\sigma_{e0}^2 = \sigma_{e1}^2 = \sigma_{e2}^2 = 0.5$, and $(\rho_0, \rho_1) = (0.95, 0.9)$. In the correlated measurement error setting, we assume $\sigma_{e0}^2 = \sigma_{e1}^2 = 0.5, \sigma_{e2}^2 = 1, (\rho_1, r_0, r_1) = (0.9, 0.7, 0.5)$. Half of the subjects are assigned to active treatment. We compare the performance of the naive approach which replaces \mathbf{X} with \mathbf{W} (Naive), MVC and FUC with 4 and 8 intervals (FUC4, FUC8). Censoring probabilities are chosen as 80% to 95% under two censoring mechanisms. Under the first mechanism, all subjects are censored at a fixed time point C_{end} (Censor I). We define intervals as $[0,$

$Q_{T,1/L}, [Q_{T,1/L}, Q_{T,2/L}), \dots, [Q_{T,(L-1)/L}, \infty)$, where $Q_{T,k/L}$ is the k^{th} L-quantile of all failure times. Under the second mechanism, censoring time follows an exponential distribution within each arm, and the censoring rates differ between arms (Censor II). We let the censoring probability in the treatment group to be 5% higher than that in the control group. For example, censoring probabilities in treatment and control group are 97.5% and 92.5% respectively, to achieve 95% overall censoring probability. In this setting, we define intervals as $[0, Q_{T,1/L}), [Q_{T,1/L}, Q_{T,2/L}), \dots, [Q_{T,(L-1)/L}, \infty)$, where $Q_{T,k/L, z=1}$ is the k^{th} L-quantile of failure times in the treatment group, to ensure there are enough treated subjects at each calibration. Cohort size varies with censoring probability to achieve 500 expected failures. Simulation results are based on 1000 replications.

First assume that distribution parameters (ρ_0, ρ_1) and (ρ_1, r_0, r_1) in the two settings are known. Simulation results of β_Z are summarized in Table 1. Naive biases in β_Z are generally non-ignorable. MVC does not reduce bias with 80% censoring probability, but its performance improves with higher censoring probability. This is expected from the underlying rare disease assumption of MVC. FUC4 and FUC8 provide considerably smaller biases in all scenarios. Theoretically, more calibrations will result in smaller biases. However, FUC8 does not improve much upon FUC4. This suggests that with 500 events, dividing time into 4 intervals is accurate enough for β_Z estimation. With uncorrelated measurement error and censoring probability 0.95, we actually observe that the bias of β_Z tends to increase from negative to 0 as number of calibrations increases, and more calibrations has the potential to make it further increase over 0, resulting in an over-correction. MVC and FUC are associated with slightly larger biases with the second censoring mechanism. This is because the time span is longer with this mechanism, and censoring time distributions are differential between the two arms. The proposed estimated standard errors agree well with simulation standard errors, especially when bias is small. Coverage probabilities are generally close to 95%.

To investigate the robustness of results to distribution parameters specification, a simulation study with (ρ_0, ρ_1) and (ρ_1, r_0, r_1) in the two measurement error settings estimated from an external data was conducted. Model specification is similar as before, and we assumed a common censoring time with 95% of subjects censored. External datasets are simulated as described in (11), (12) and (13). Simulation results of β_Z are summarized in Table 2. Compared to when (ρ_0, ρ_1) are known, both MVC and FUC still reduce β_Z bias, and the bias decreases as the number of intervals increases. As distribution parameter estimates become less precise, bias tend to increase, especially with correlated measurement error, and more calibrations are likely to cause over-correction. Standard errors tend to increase as well, which agrees with our proposed variance estimates. This increase is generally quite small with uncorrelated measurement error, but can be large with correlated measurement error. The differences between the simulated standard errors and estimated mean standard errors in correlated measurement error scenarios are related to the remaining biases in β_0 and β_1 (see Web Appendix B). These remaining biases are mostly due to the variability in $\hat{\rho}_1$. Hence, for correlated measurement error, we recommend a careful evaluation of distribution parameters, as results can be quite sensitive to their specification.

Next we investigate the robustness of MVC and FUC to non-normality. We assume that \mathbf{X} is normally distributed, while \mathbf{U} is non-normal. This corresponds to the situation where, under a suitable transformation, \mathbf{X} becomes approximately normal while \mathbf{U} is generated from a multivariate non-normal distributions with mean 0 and unit variance using methods described in Vale and Maurelli (1983). Two skewness and kurtosis combinations: (0,6/5) and (1,3) are investigated. Here (0,6/5) is chosen to resemble a logistic distribution, which is close to a normal distribution but has heavier tails. The combination of (1, 3) is chosen to allow the distributions to be both skewed and with heavy tails. Other model assumptions are the same with 95% censoring at a common time. Results of β_z are summarized in Table 3. As measurement error distribution becomes further away from normal, the naive bias in β_z tends to increase. Both MVC and FUC show the ability to reduce bias, but over-correction can become serious with severe violation of normality. This is because FUC assumes normality at each calibration, which evidently relies on the normal assumption more heavily than MVC. In this example with relatively high censoring probability, MVC is expected to perform well. The proposed variance estimator is not accurate when violation of normality is severe, as it is derived under normality assumption.

Simulation results for β_0 and β_1 are provided in Web Appendix B. Naive biases in these two parameters are generally larger than those of β_z in our simulation settings, partially due to larger β_0 and β_1 values. MVC and FUC both reduce biases greatly. However, the remaining bias is typically non-negligible, suggesting that more calibrations may be needed if these parameters are of substantial interest.

5 Postmenopausal Hormone Therapy Application

We now re-examine the mediation of postmenopausal hormone therapy effects on breast cancer by serum sex hormones, in the WHI randomized controlled trials. There were two major trials: 16,608 women with uterus were randomized to 0.625 mg/day conjugated equine estrogen plus 2.5 mg/day medroxyprogesterone acetate (E+P) or placebo; and 10,739 post-hysterectomy women were randomized to this same estrogens preparation without the progestin (E-alone) or placebo. Both hormone therapy trials stopped early due to adverse health events. An elevation in invasive breast cancer risk was pivotal in the early stopping of the E+P trial (Rossouw et al., 2002; Chlebowski et al., 2009), while the E-alone trial showed a surprising reduction in breast cancer incidence with treatment (Anderson et al., 2004, 2012).

A nested case-control study was conducted within each hormone therapy trial toward understanding the divergent trial results, with the major changes in plasma hormones induced by these regimens as natural candidates for breast cancer effects mediation. The study included 348 and 235 cases in the E+P trial and E-alone trial, and corresponding 1-1 matched controls. Concentrations of sex hormones were measured at baseline and 1-year following randomization. Major serum estrogens, including estradiol, estrone, and estrone sulfate, were approximately doubled by these hormone therapies, as was the sex hormone binding globulin (SHBG) (Edlefsen et al., 2010; Farhat et al., 2013). However, the sex hormone changes were nearly identical with E-alone and with E+P, presumably reducing the likelihood that these changes can substantially explain the divergent hormone therapy effects

on breast cancer. The mediation analyses given here exclude cases occurring during the first year following randomization. Cox model analyses of the randomization indicator variable and log-transformed baseline sex hormone variables were carried out without, and with, the addition of 1-year log-transformed sex hormone concentrations to the model. Matching variables age and race, were included in the regression for confounding control.

As an external dataset, blind duplicate sex hormone assessments were available from 120 women who were screened for WHI participation, but did not enroll. Differences between the log-transformed assessments from the duplicate samples are assumed to be due to technical error. Variance ratios, k_0 , are estimated to be 0.13, 0.07, 0.19, and 0.02 for estradiol, estrone, estrone sulfate, and SHBG. We assume that variance ratios in the uncorrelated measurement error model only change with hormone therapy treatment assignment, and provide sensitivity analyses with $k_0 = k_1$, while k_2 varies. Only MVC is applied to these data, since only about 2% of women developed breast cancer during the intervention phases of these clinical trials. Table 4 presents some results from mediation analyses, on the left with estradiol as potential mediator, and on the right with the four sex hormones considered as possible joint mediators.

Serum estradiol appears to partially mediate a substantial effect of E+P on breast cancer risk, even without measurement error correction, with estimated hazard ratio (HR) dropping from 1.64 to 1.35 when 1-year estradiol was added to the analysis. Allowance for uncorrelated measurement error gave HR estimates a little closer to the null. Allowing measurement errors to be correlated, as may be necessary to address mediation by sex hormone levels over the entire trial intervention period, shows the possibility of rather complete mediation by estradiol with HR estimates in the vicinity of one. These are purely sensitivity analyses, however. In order to maintain a non-negative correlation between measurement errors, the smallest possible k_0 for estradiol is 0.95 for the E+P trial and 1.1 for the E-alone trial. We denote these numbers by k_s and vary k_i , $i = 0, 1, 2$ from k_s to $2k_s$. Note that confidence intervals are quite wide, in line with simulation study findings of sensitivity to error distribution parameter specifications. Mediation of the E+P effect on breast cancer was not enhanced by bringing in the other sex hormones. In contrast, the reduced HR with E-alone is evidently minimally explained by the change in serum estradiol, regardless of measurement error correction. However, when the four sex hormones are considered simultaneously as potential mediators, there is evidence of partial mediation without measurement error correction, and rather complete mediations when allowing for technical measurement error. The interpretation of mediation analyses can be complex. Here, specifically, mediation of the risk elevation with E+P seems to be related to removal of a baseline estradiol effect following treatment, whereas the E-alone risk reduction may reflect SHBG increase that offsets the serum estrogen increase (Zhao et al., 2013).

6 Discussion

This article discusses covariate measurement error correction methods in the context of mediation analysis with failure time data. The proposed mean-variance regression calibration is suitable under a rare disease assumption, and the follow-up time regression calibration further extends this applicability. Simulation studies demonstrate that both

measurement error correction methods have desirable performances under various biomarker process scenarios.

A requirement of these methods is that some additional information about the biomarker process be available. In application, it might be challenging to obtain such information for novel biomarkers, such as in our WHI hormone therapy trial example. The need for a reliability dataset has always been important in measurement error area. In our more complicated mediation analysis setting, investigators need to plan the reliability study to have sufficient sample size with suitable longitudinal measures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank the Women's Health Initiative investigator group for access to the hormone therapy trial data to illustrate the methods proposed here. This work was supported by NIH grants HL109527, CA53996 and CA155340.

References

- Andersen PK, Gill RD. Cox's regression model for counting processes: A large sample study. *Annals of Statistics*. 1982; 10:1100–1120.
- Anderson G, Chlebowski R, Aragaki A, Kuller L, Manson J, Gass M, Bluhm E, Connelly S, Hubbell F, Lane D, Martin L, Ockene J, Rohan T, Schenken R, Wactawski-Wende J. Conjugated equine oestrogens and breast cancer incidence and mortality in postmenopausal women with hysterectomy: extended follow-up of the Women's Health Initiative randomized placebo-controlled trial. *Lancet Oncology*. 2012; 13:475–486.
- Anderson GL, Limacher M, Assaf AR, Bassford T, Beresford SA, Black H, et al. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. *Journal of the American Medical Association*. 2004; 291:1701–1712. [PubMed: 15082697]
- Barlow WE. Robust variance estimation for the case-cohort design. *Biometrics*. 1994; 50:1064–1072. [PubMed: 7786988]
- Cai T, Zheng Y. Evaluating prognostic accuracy of biomarkers in nested case-control studies. *Biostatistics*. 2011; 106:569–580.
- Carroll R, Knickerbocker R, Wang C. Dimension reduction in semiparametric measurement error models. *Annals of Statistics*. 1995; 23:161–181.
- Chlebowski R, Kuller L, Prentice R, Stefanick M, Manson J, Gass M, Aragaki A, Ockene J, Lane D, Sarto G, Rajkovic A, Schenken R, Hendrix S, Ravdin P, Rohan T, Yasmeen S, Anderson G. WHI investigators. Breast cancer after use of estrogen plus progestin in postmenopausal women. *New England Journal of Medicine*. 2009; 360:573–587.
- Diggle, PJ.; Heagerty, P.; Liang, KY.; Zeger, SL. *Analysis of Longitudinal Data*. Oxford University Press; 2002.
- Edlfsen K, Jackson R, Prentice R, Janssen I, Rajkovic A, O'Sullivan M, Anderson G. The effects of postmenopausal hormone therapy on serum estrogen, progesterone and sex-hormone binding globulin levels in healthy postmenopausal women. *Menopause*. 2010; 17:622–629. [PubMed: 20215977]
- Farhat G, Parimi N, Chlebowski R, Manson J, Anderson G, H AJ, V E, Lee J, Lacroix A, Cauley J, Jackson R, Grady D, Lane D, Phillips L, Simon M, Cummings S. Sex hormone levels and risk of breast cancer with estrogen plus progestin. *Journal of National Cancer Institute*. 2013; 105:1496–1503.

- Hu P, Tsiatis AA, Davidian M. Estimating the parameters in the Cox model when covariate variables are measured with error. *Biometrics*. 1998; 54:1407–1419. [PubMed: 9883541]
- Huang Y, Wang CY. Cox regression with accurate covariates unascertainable: a nonparametric correction approach. *Journal of the American Statistical Association*. 2000; 95:1209–1219.
- Huang Y, Wang CY. Error-in-covariates effect on estimating functions: Additivity in limit and nonparametric correction. *Statistica Sinica*. 2006; 16:861–881.
- Hughes MD. Regression dilution in the proportional hazards model. *Biometrics*. 1993; 49:1056–1066. [PubMed: 8117900]
- Lange T, Hansen JV. Direct and indirect effects in a survival context. *Epidemiology*. 2011; 22:575–581. [PubMed: 21552129]
- Liao X, Zucker DM, Li Y, Speigelman D. Survival analysis with error-prone time-varying covariates: a risk set calibration approach. *Biometrics*. 2011; 67:50–58. [PubMed: 20486928]
- Lin DY, Fleming TR, De Gruttola V. Estimating the proportion of treatment effect explained by a surrogate marker. *Statistics in Medicine*. 1997; 16:1515–1527. [PubMed: 9249922]
- MacKinnon, DP. *Introduction to Statistical Mediation Analysis*. Taylor & Francis; 2008.
- Prentice RL. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*. 1982; 69:331–342.
- Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*. 1986; 73:1–11.
- Rossouw J, Anderson G, Prentice R, LaCroix A, Kooperberg C, Stefanick M, Jackson R, Beresford S, Howard B, Johnson K, Kotchen J, Ockene J. Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *Journal of the American Medical Association*. 2002; 288:321–333. [PubMed: 12117397]
- Vale C, Maurelli V. Simulating multivariate nonnormal distributions. *Psychometrika*. 1983; 48:465–471.
- Wang CY, Hsu L, Feng ZD, Prentice RL. Regression calibration in failure time regression. *Biometrics*. 1997; 53:131–145. [PubMed: 9147589]
- Wang CY, Xie CX, Prentice RL. Recalibration based on an approximate relative risk estimator in cox regression with missing covariates. *Statistica Sinica*. 2001; 11:1081–1104.
- Xie SX, Wang CY, Prentice RL. A risk set calibration method for failure time regression by using a covariate reliability sample. *Journal of the Royal Statistical Society: Series B*. 2001; 63:855–870.
- Zhao S, Chlebowski R, Anderson G, Kuller L, Manson J, Gass M, Patterson R, Rohan T, Lane D, Beresford S, Lavasani S, Rossouw J, Prentice R. Substantial mediation of postmenopausal hormone therapy effects on breast cancer by circulating sex hormones. *Breast Cancer Research*. 2013; 16:R30. [PubMed: 24670297]
- Zhou H, Pepe MS. Auxiliary covariate data in failure time regression. *Biometrika*. 1995; 82:139–149.
- Zhou H, Wang CY. Failure time regression with continuous covariates measured with error. *Journal of the Royal Statistical Society: Series B*. 2000; 62:657–665.

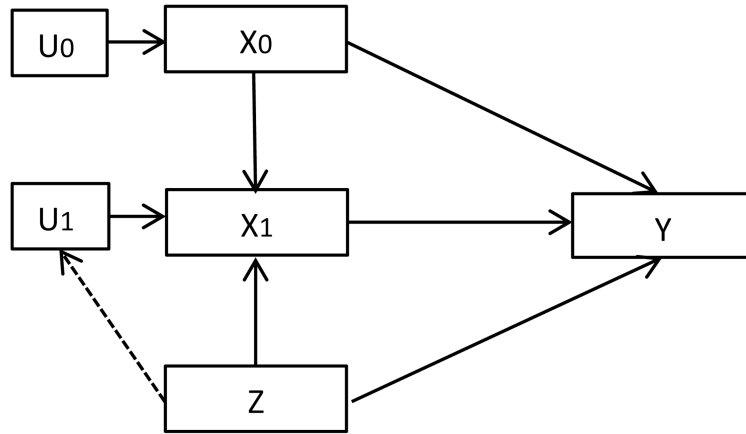


Figure 1.
Causal diagram of the underlying model.

Table 1

Summary statistics for β_z . Distribution parameters (ρ_0, ρ_S) in the uncorrelated measurement error setting and (ρ_1, r_0, r_1) in the correlated measurement error setting are assumed to be known.

β_z	P(censor)	method	Censor I					Censor II				
			$\hat{\beta}_z$	bias	Sim SE ^a	Est SE ^b	Cp ^c	$\hat{\beta}_z$	bias	Sim SE	Est SE	CP
Uncorrelated Measurement Error												
0.8	Naive	0.435	0.029	0.100	-	-	0.438	0.033	0.104	-	-	
	MVC	0.356	-0.049	0.202	0.190	0.864	0.333	-0.072	0.210	0.254	0.954	
	FUC4	0.397	-0.008	0.204	0.192	0.944	0.375	-0.030	0.217	0.260	0.980	
	FUC8	0.404	-0.001	0.208	0.196	0.951	0.384	-0.021	0.219	0.265	0.983	
0.9	Naive	0.465	0.059	0.097	-	-	0.466	0.061	0.109	-	-	
	MVC	0.374	-0.032	0.191	0.176	0.916	0.333	-0.072	0.227	0.223	0.945	
	FUC4	0.399	-0.007	0.193	0.188	0.956	0.363	-0.042	0.231	0.247	0.950	
	FUC8	0.403	-0.002	0.195	0.191	0.961	0.369	-0.037	0.231	0.250	0.952	
0.95	Naive	0.485	0.079	0.099	-	-	0.516	0.111	0.132	-	-	
	MVC	0.393	-0.013	0.190	0.178	0.935	0.352	-0.053	0.273	0.251	0.937	
	FUC4	0.411	0.005	0.192	0.187	0.953	0.368	-0.038	0.275	0.269	0.950	
	FUC8	0.414	0.008	0.193	0.190	0.958	0.371	-0.034	0.276	0.269	0.950	
Correlated Measurement Error												
0.8	Naive	0.527	0.122	0.105	-	-	0.505	0.099	0.109	-	-	
	MVC	0.311	-0.094	0.283	0.364	1.000	0.288	-0.117	0.295	0.366	0.998	
	FUC4	0.384	-0.021	0.257	0.275	0.988	0.369	-0.036	0.273	0.316	0.976	
	FUC8	0.397	-0.008	0.260	0.272	0.980	0.391	-0.015	0.271	0.309	0.975	
0.9	Naive	0.576	0.171	0.102	-	-	0.534	0.128	0.118	-	-	
	MVC	0.330	-0.076	0.235	0.277	0.990	0.306	-0.100	0.264	0.283	0.973	
	FUC4	0.376	-0.029	0.228	0.241	0.985	0.360	-0.045	0.270	0.294	0.984	
	FUC8	0.385	-0.021	0.229	0.239	0.980	0.375	-0.031	0.269	0.292	0.982	
0.95	Naive	0.611	0.205	0.104	-	-	0.566	0.161	0.142	-	-	
	MVC	0.357	-0.049	0.216	0.227	0.972	0.316	-0.089	0.331	0.309	0.960	
	FUC4	0.389	-0.017	0.217	0.214	0.959	0.350	-0.055	0.326	0.329	0.967	

β_2	P(censor)	method	Censor I				Censor II					
			$\hat{\beta}_2$	bias	Sim SE ^a	Est SE ^b	Cp ^c	$\hat{\beta}_2$	bias	Sim SE	Est SE	CP
	FUC8		0.394	-0.011	0.218	0.213	0.957	0.359	-0.046	0.325	0.328	0.970

^aSimulation Standard Error

^bMean Estimated Standard Error

^cCoverage Probability

Table 2

Summary statistics for β_z . Distribution parameters (ρ_0, ρ_1) in uncorrelated measurement error setting and (ρ_1, r_0, r_1) in correlated measurement error setting are estimated from external datasets of sample size 1000 and 500.

β_z	$n_E = 1000$					$n_E = 500$					
	method	$\hat{\beta}_z$	bias	Sim SE	Est SE	CP	$\hat{\beta}_z$	bias	Sim SE	Est SE	CP
	Uncorrelated Measurement Error										
Naive	0.485	0.079	0.099	–	–	–	0.485	0.079	0.099	–	–
MVC	0.392	-0.014	0.191	0.189	0.944	0.394	-0.012	0.190	0.199	0.947	
FUC4	0.412	0.007	0.193	0.195	0.954	0.415	0.009	0.192	0.201	0.959	
FUC8	0.415	0.009	0.194	0.197	0.960	0.418	0.012	0.192	0.203	0.959	
	Correlated Measurement Error										
Naive	0.611	0.205	0.104	–	–	–	0.611	0.205	0.104	–	–
MVC	0.384	-0.021	0.757	0.741	0.958	0.512	0.106	1.376	1.407	0.958	
FUC4	0.409	0.004	0.591	0.535	0.963	0.460	0.054	0.576	0.669	0.974	
FUC8	0.408	0.003	0.594	0.509	0.968	0.463	0.057	0.584	0.651	0.968	

Table 3

Summary statistics for β_z . Measurement errors are generated from multivariate distribution with violation of normality assumption.

β_z	$(sk, ku) = (0, 6/5)$					$(sk, ku) = (1, 3)$					
	method	$\hat{\beta}_z$	bias	Sim SE	Est SE	CP	$\hat{\beta}_z$	bias	Sim SE	Est SE	CP
Uncorrelated Measurement Error											
Naive	0.491	0.086	0.102	–	–	0.517	0.111	0.106	–	–	–
MVC	0.403	-0.003	0.197	0.164	0.915	0.444	0.039	0.192	0.125	0.855	0.855
FUC4	0.424	0.018	0.201	0.173	0.933	0.464	0.058	0.198	0.134	0.897	0.897
FUC8	0.427	0.022	0.202	0.176	0.937	0.467	0.062	0.200	0.136	0.901	0.901
Correlated Measurement Error											
Naive	0.614	0.209	0.108	–	–	0.629	0.223	0.112	–	–	–
MVC	0.356	-0.050	0.220	0.215	0.970	0.399	-0.006	0.220	0.186	0.965	0.965
FUC4	0.389	-0.017	0.220	0.200	0.958	0.430	0.024	0.219	0.166	0.929	0.929
FUC8	0.395	-0.011	0.220	0.199	0.959	0.435	0.029	0.219	0.164	0.928	0.928

Table 4

Hazard ratios of treatment in E+P and E-alone trials, with and without measurement error correction. Matching variables, age and race, are adjusted in all models.

Potential Mediator Trial	Estradiol		Estradiol, Estrone, Estrone Sulfate, SHBG	
	E+P	E-alone	E+P	E-alone
	HR ^a (95% CI) ^b	HR (95% CI)	HR (95% CI)	HR (95% CI)
Baseline Biomarkers Only	1.64 (1.28, 2.10)	0.59 (0.45, 0.78)	1.71 (1.32, 2.20)	0.68 (0.51, 0.91)
Baseline+Year 1 Biomarkers				
No ME ^c Correction	1.35 (0.98, 1.84)	0.59 (0.41, 0.83)	1.47 (0.98, 2.20)	0.84 (0.51, 1.38)
Uncorrelated ME Correction				
$k_1 = k_2 = k_0$	1.31 (0.94, 1.82)	0.59 (0.41, 0.86)	1.43 (0.92, 2.22)	0.87 (0.50, 1.50)
$k_1 = k_0, k_2 = 1.5k_0$	1.29 (0.92, 1.81)	0.59 (0.41, 0.86)	1.39 (0.89, 2.16)	0.94 (0.55, 1.62)
$k_1 = k_0, k_2 = 2k_0$	1.27 (0.90, 1.80)	0.59 (0.40, 0.87)	1.36 (0.86, 2.15)	0.95 (0.54, 1.65)
Correlated ME Correction				
$k_0 = k_1 = k_2 = k_s$	1.09 (0.70, 1.68)	0.66 (0.35, 1.26)	–	–
$k_0 = k_1 = k_2 = 1.5k_s$	0.99 (0.54, 1.82)	0.65 (0.31, 1.34)	–	–
$k_0 = k_1 = k_2 = 2k_s$	0.89 (0.45, 1.77)	0.68 (0.28, 1.65)	–	–

^aHazard ratio

^bConfidence interval

^cMeasurement error