

©Copyright 2012

Shanshan Zhao

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

Shanshan Zhao

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

University of Washington

2012

Reading Committee:

Ross L. Prentice, Chair

Ying Qing Chen

Susanne May

Program Authorized to Offer Degree:
Biostatistics

University of Washington

Abstract

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

Shanshan Zhao

Chair of the Supervisory Committee:
Professor Ross L. Prentice
Department of Biostatistics

Mediation analysis is important in understanding the mechanisms of one variable causing changes in another. Measurement error could be obscuring the ability of the potential mediator to explain this mechanism. Existing correction methods in mediation analysis literature are not directly applicable to failure time data.

This dissertation focuses on developing correction methods for measurement error in the potential mediator with time-to-event outcome. We consider two specifications of measurement errors: one with technical measurement error only, and one with both technical measurement error and temporal variation. The underlying model with the true mediator values is assumed to be the Cox proportional hazards model. The hazard function induced by the observed mediator value no longer corresponds to a simple partial likelihood independent of the baseline hazard function, due to the conditioning event $\{\tilde{T} \geq t\}$. We propose a mean-variance regression calibration and a follow-up time calibration approach to approximate the induced partial likelihood. Both methods demonstrate successes in recovering treatment effect estimates with both types of measurement error in simulation studies. Variance estimators are derived for both approaches. These two methods can be generalized to multiple biomarkers and case-cohort design. We apply these correction methods to the Women's Health Initiative hormone therapy trials to understand the mediation effect of biomarker IGFBP4 in the relationship between hormone therapy and stroke.

TABLE OF CONTENTS

	Page
List of Figures	iv
List of Tables	v
Chapter 1: Introduction	1
1.1 Motivating Example	1
1.2 Failure Time Data and the Cox Proportional Hazards Model	3
1.3 Existing Measurement Error Correction Methods with Failure Time Data	4
1.3.1 Naive Approach	5
1.3.2 Induced Hazard Approach	6
1.3.3 Regression Calibration	7
1.3.4 Risk Set Calibration	8
1.3.5 Other Approaches	9
1.4 Mediation Analysis	9
1.5 Measurement Error Model	12
1.6 Overview of Dissertation	15
Chapter 2: Induced Hazard Approach—Uncorrelated Measurement Error	18
2.1 Overview	18
2.2 Naive Approach	20
2.3 Mean-Variance Regression Calibration	23
2.4 Follow-up Time Calibration	26
2.5 Biomarker Process Modeling	29
2.6 Simulation Studies	33
2.6.1 Simulation Studies with Distribution Parameters (ρ_0, ρ_1) Known	34
2.6.2 Simulation studies with Distribution Parameters (ρ_0, ρ_1) Estimated	40
2.6.3 Simulation Studies of Sensitivity on Censoring Mechanism	41
2.6.4 Simulation Studies of Violation to Normality Assumption	43
2.6.5 Summary of Simulation Studies	45

Chapter 3:	Induced Hazard Approach— Correlated Measurement Errors	47
3.1	Overview	47
3.2	Naive Approach	48
3.3	Mean-Variance Regression Calibration and Follow-up Time Calibration	51
3.4	Simulation Studies	56
3.4.1	Simulation Studies with Distribution Parameters (ρ, r_0, r_1) Known	56
3.4.2	Simulation Studies with Distribution Parameters (ρ, r_0, r_1) Estimated	61
3.4.3	Simulation Studies of Sensitivity on Censoring Mechanism	62
3.4.4	Simulation Studies of Violation to Normality Assumption	63
3.4.5	Summary of Simulation Studies	65
Chapter 4:	Distribution Theory	67
4.1	Introduction	67
4.2	Notation and Assumptions	67
4.2.1	Formulation of the Model	67
4.2.2	Regularity Conditions	72
4.2.3	Some Basic Tools	73
4.3	Consistency to the Target β^*	76
4.4	Asymptotic Normality	81
4.5	Details with Uncorrelated and Correlated Measurement Errors	90
4.5.1	Uncorrelated Measurement Errors	90
4.5.2	Correlated Measurement Errors	95
4.6	Simulation Studies	100
Chapter 5:	Multiple Biomarkers and Cohort Sampling Generalizations	102
5.1	Multiple Biomarkers	102
5.2	Case-Cohort Study Design	105
Chapter 6:	Applications	111
6.1	Data Description	111
6.2	Results	113
6.3	Summary	116
Chapter 7:	Discussion and Future Work	117
7.1	Dissertation Contributions	117
7.2	Future Work	120

7.2.1	Interaction Models	120
7.2.2	Relaxation of the Proportional Hazard Assumption	120
7.2.3	Generalization to Other Study Designs	121
7.2.4	Measurement Error in the Independent Variable	122
7.2.5	Other Approaches	123
Appendix A:	Biases Associated with Naive Approach	128

LIST OF FIGURES

Figure Number	Page
1.1 Three-variable mediation analysis diagram.	10
2.1 Mediation analysis diagram, with pre- and post-randomization biomarkers. . .	19
2.2 Biases of naive approach with uncorrelated measurement error.	24
3.1 Biases of naive approach with correlated measurement error.	50

LIST OF TABLES

Table Number	Page
2.1 Model assumptions on each component of the biomarker process model.	29
2.2 Biases with uncorrelated ME after correction: $\sigma_1^2, \sigma_e^2, \beta_Z$ vary	35
2.3 SEs with uncorrelated ME after correction: $\sigma_1^2, \sigma_e^2, \beta_Z$ vary	36
2.4 Biases with uncorrelated ME after correction: d, β_Z vary	38
2.5 Biases with uncorrelated ME after correction: $\sigma_{e0}^2, \sigma_{e1}^2$ relationship varies . . .	39
2.6 Biases and SEs with uncorrelated ME after correction: (ρ_0, ρ_1) estimated . . .	41
2.7 Biases with uncorrelated ME after correction: censoring mechanism varies . . .	44
2.8 Biases with uncorrelated ME after correction: non-normal ME.	45
3.1 Biases with correlated ME after correction: $\sigma_1^2, \sigma_e^2, \beta_Z$ vary	58
3.2 SEs with correlated ME after correction: $\sigma_1^2, \sigma_e^2, \beta_Z$ vary	59
3.3 Biases with correlated ME after correction: d, β_Z vary	60
3.4 Biases and SEs with correlated ME after correction: (ρ, r_0, r_1) estimated . . .	62
3.5 Biases with correlated ME after correction: censoring mechanism varies . . .	64
3.6 Biases with correlated ME after correction: non-normal ME.	65
4.1 Mean standard error estimates with uncorrelated ME.	101
4.2 Mean standard error estimates with correlated ME.	101
5.1 Biases with multiple biomarkers and uncorrelated ME.	106
5.2 Biases with multiple biomarkers and correlated ME.	106
5.3 Biases and SEs with uncorrelated ME after correction: case-cohort sampling .	109
5.4 Biases and SEs with correlated ME after correction: case-cohort sampling . .	110
6.1 Summary of hazard ratios, adjusting for the uncorrelated ME.	115
6.2 Summary of hazard ratios, adjusting for the correlated ME.	116

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my dissertation advisor, Ross Prentice, for his inspiration and guidance. I could not have asked for a more brilliant and supportive mentor. I would also like to thank the members of my supervisory committee, Ying Qing Chen, Susanne May, Charles Kooperberg, and David Yanez for their invaluable inputs into this dissertation. I want to thank Susan Heckbert and Johanna Lampe for serving as my biology advisor and Graduate School Representative. I also want to extend my thanks to the Women's Health Initiative Study Group for providing the data.

Also, I would like to acknowledge my family. Thanks to my parents for being supportive and having confidence in me. I want to thank my husband Wei. Without his constant love and support, I could not have completed this journey.

DEDICATION

To my parents.

Chapter 1

INTRODUCTION

1.1 Motivating Example

The motivation for developing covariate measurement error correction methods in mediation analysis with failure time data can be illustrated by the Women's Health Initiative hormone therapy trials.

The Women's Health Initiative (WHI) addressed the most common causes of death, disability and impaired quality of life in postmenopausal women. The WHI randomized, placebo controlled hormone therapy trials aimed to provide important information about the risks and benefits of long-term postmenopausal hormone therapy and to offer women some guidance about hormone use. A total of 10,739 post-hysterectomy women were randomized to 0.625 mg/d conjugated equine estrogens (E-alone) or placebo, and 16,608 women with an intact uterus were randomized to 0.625 mg/d conjugated equine estrogens plus 2.5 mg/d medroxyprogesterone acetate (E+P) or placebo. Both hormone therapy trials were stopped early due to disease risk elevations, including an elevated stroke incidence (Anderson et al., 2004; Rossouw et al., 2002). The estimated hazard ratio for stroke was 1.31 with 95% CI (1.02, 1.68) for women assigned to active E+P compared to placebo, and 1.37 with 95% CI (1.09, 1.73) for active E-alone compared to placebo. Early elevations in coronary heart disease (CHD) incidence and a later major elevation in breast cancer incidence with E+P, and a more sustained elevation in venous thromboembolism (VTE) with both preparations, were also among the clinical effects that led to the early termination of both trials.

To understand the early adverse effect of hormone therapy seen in the WHI hormone therapy trials, several follow-up studies were conducted, including a Cardiovascular Disease Biomarker (CVD) study (Kooperberg et al., 2007; Rossouw et al., 2008). Each of stroke, CHD, and VTE incidences was investigated in relation to blood biomarkers that had been reported to associate with cardiovascular disease risk. Many baseline biomarkers were found

to be related to stroke and CHD risks in this nested case-control study, including several inflammation markers, blood lipids and lipoprotein markers, and thrombosis markers. Another aim of this study was to explore whether post-treatment changes in these biomarkers could provide an explanation to the observed hormone therapy effect on these diseases. Most of the inflammation markers, blood lipids markers and some thrombosis markers changed significantly with active hormone therapy compared to that of the placebo group. However, when these biomarker changes were added into the statistical models used for data analysis, none of them appeared to mediate the E+P or E-alone effect on stroke and CHD.

A recent plasma proteomic discovery study identified several novel stroke risk factors (Prentice et al., 2010). Among these risk factors, insulin-like growth factor-binding protein 4 (IGFBP4) has been subject to a validation study. Baseline IGFBP4 concentrations were found to be 16.6% ($p = 0.005$) higher in cases than in controls. The use of either E+P or E-alone was associated with about 20% increase in IGFBP4 concentration at 1 year post randomization, which corresponded to about 40% increase in stroke risk (95% CI (6%, 85%)). However, when we included both baseline and 1-year IGFBP4 concentrations along with hormone therapy indicator in the data analysis, there appeared to be little evidence that E+P or E-alone effects on stroke were mediated by IGFBP4 concentration change during the first year of hormone therapy use.

The studies mentioned above all suggest that some biomarkers, such as IGFBP4, are related to stroke risk, and biomarker concentrations change significantly with active hormone therapy. However, with the standard statistical method that compares treatment effects estimated from models with and without these biomarkers changes, we were unable to identify their mediation effects in the relationship between hormone therapy and the stroke outcome. Measurement error in biomarker assessments may help to explain this issue. Biomarkers are measured based on blood drawn during clinic visits. Their assessments may differ from the long-term pre- or post- randomization biomarker levels because of diurnal or seasonal variation, or other factors. And long-term biomarker levels may be more relevant to disease risk than measurements at specific clinic visits. In addition there may be technical measurement errors related to the biomarker assay or specimen handling. Measurement errors could be obscuring the ability of available biomarkers to explain the observed effects

of hormone therapy on stroke. Novel statistical methods are needed to reliably examine this research topic.

1.2 Failure Time Data and the Cox Proportional Hazards Model

In this dissertation, we focus on failure time data. Failure time data analysis has important applications in medicine, epidemiology, biology, and social sciences. In failure time data analysis, our primary interest is the time until an event (e.g. stroke in the above illustration) occurs. We often refer to this event as a “failure”. However, in many studies, we are unable to follow subjects long enough until failures occur for various reasons, such as patients drop out or study ends. In these situations, we call these subjects as “right censored”. This “censoring” feature distinguishes failure time data analysis from other types of data analysis.

Suppose that we have n subjects in the study. For subject i , let \tilde{T}_i denote the ‘true’ failure time, C_i denote the censoring time, and $\mathbf{X}_i = (X_{0i}, X_{1i}, \dots, X_{pi})'$ denote a $p \times 1$ covariate vector. Since we are unable to observe subject i beyond censoring time C_i , the observation for this particular subject is $(T_i, \delta_i, \mathbf{X}_i)$, where $T_i = \min(\tilde{T}_i, C_i)$ is the observed time to event, and $\delta_i = I(\tilde{T}_i \leq C_i)$ is the censoring indicator. We assume that given \mathbf{X}_i , censoring time C_i is independent of \tilde{T}_i (independent censoring) throughout this dissertation. The distribution of the failure time \tilde{T}_i given \mathbf{X}_i is completely described by the hazard function

$$\lambda(t|\mathbf{X}_i) = \lim_{\Delta t \rightarrow 0} \frac{P(\tilde{T} \in [t, t + \Delta t] | \tilde{T} \geq t, \mathbf{X}_i)}{\Delta t}.$$

Cox (1972) proposed a semi-parametric regression model for censored failure time data. The Cox proportional hazards model expresses the hazard rate at time t as a product of an arbitrary baseline hazard function $\lambda_0(t)$ and a hazard ratio function $\exp(\boldsymbol{\beta}^T \mathbf{X})$,

$$\lambda(t; \mathbf{X}) = \lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{X}). \quad (1.1)$$

In the special case of a single covariate ($p = 1$), $\exp(\beta)$ is the hazard ratio corresponding to one unit change in X . Due to the flexibility of this model which includes the possibility of time-varying covariate $\mathbf{X}(t)$, it has become a standard approach in failure time data analysis. However, due to the nonparametric component $\lambda_0(t)$, the usual parametric

likelihood approach does not readily apply under this model. Instead, Cox (1975) proposed a partial likelihood approach. Suppose the sample comprises k uncensored failure times $t_1 < t_2 < \dots < t_k$, and the failure times are not tied (i.e. only one subject fails at each unique failure time). The remaining $n - k$ subjects are right censored. We denote the subject fails at t_i as i , and let $R(t) = \{i : T_i \geq t\}$ denote the set of subjects still at risk of failure at t . The partial likelihood for the parameter $\boldsymbol{\beta}$ is given by

$$PL(\boldsymbol{\beta}) = \prod_{i=1}^k \frac{\exp(\boldsymbol{\beta}^T \mathbf{X}_i)}{\sum_{j \in R(t_i)} \exp(\boldsymbol{\beta}^T \mathbf{X}_j)} \quad (1.2)$$

Maximized partial likelihood estimate, $\hat{\boldsymbol{\beta}}$, of (1.2) can be obtained as a solution to the score equation

$$U(\boldsymbol{\beta}) = \frac{\partial \log PL}{\partial \boldsymbol{\beta}} = \sum_{i=1}^k [\mathbf{X}_i - \epsilon(\boldsymbol{\beta}, t_i)], \quad (1.3)$$

where

$$\epsilon(\boldsymbol{\beta}, t_i) = \frac{\sum_{j \in R(t_i)} \mathbf{X}_j \exp(\boldsymbol{\beta}^T \mathbf{X}_j)}{\sum_{j \in R(t_i)} \exp(\boldsymbol{\beta}^T \mathbf{X}_j)}$$

is the expectation of \mathbf{X} with respect to the covariate distribution of the risk set $R(t_i)$ conditional on a failure occurs at t_i . The value $\hat{\boldsymbol{\beta}}$ can be obtained through Newton-Raphson approach. The covariance matrix $\mathcal{I}(\boldsymbol{\beta})$ can be estimated as minus the second derivative of log partial likelihood. Andersen and Gill (1982) showed asymptotic results based on (1.2) are similar to those with standard likelihood approaches under mild assumptions. In the absence of ties, the asymptotic distribution of $\hat{\boldsymbol{\beta}}$ estimated from (1.3) is

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \rightarrow N(0, \mathcal{I}(\boldsymbol{\beta})).$$

Partial likelihood can be justified as marginal likelihood and profile likelihood (e.g., Kalbfleisch and Prentice, 2002). As noted above, the Cox model can be extended to allow time-varying covariate $\mathbf{X}(t)$ in the model.

1.3 Existing Measurement Error Correction Methods with Failure Time Data

The measurement error issue has been investigated in failure time settings. From now on, we denote the true covariate as \mathbf{X} , and the observed covariate as \mathbf{W} . A classical measurement

error is assumed,

$$\mathbf{W} = \mathbf{X} + \mathbf{U}, \quad (1.4)$$

where \mathbf{U} is the random error not related to \mathbf{X} value.

1.3.1 Naive Approach

A naive approach to estimate β in (1.1) simply uses \mathbf{W} as a surrogate of \mathbf{X} , and plugs \mathbf{W} into the score equation (1.3) to get an estimate $\hat{\beta}_W$. In a simple linear regression with classical measurement error (1.4), it can be shown that

$$E(\hat{\beta}_W) = \frac{\beta}{1 + \delta^2},$$

where $\delta^2 = \sigma_\epsilon^2 / \sigma_X^2$ is the ratio of σ_ϵ^2 , the measurement error variance and σ_X^2 , the variance of the underlying value of \mathbf{X} in the population sampled. Hughes (1993) investigated this naive approach in the context of Cox proportional hazards model with one mis-measured covariate. He derived the exact relationship between β_W and the true β , under three different censoring mechanisms: no censoring, administrative censoring at a fixed time point, and a form of random censoring. This relationship is expressed in terms of a complicated integrate, thus requires a numerical rather than analytical solution. When there is no censoring, the relationship between β_W and β does not depend on the form of the underlying baseline hazard function $\lambda_0(t)$. Simulation study shows that as variance ratio and the true β value get larger, the bias of the naive estimate becomes more substantial. Using the adjustment factor $1 + \delta^2$ as derived from the simple linear regression gives an underestimate of β by a considerable amount, unless β is close to 0. When censorship is present, the baseline hazard function $\lambda_0(t)$ starts to affect the relationship between β_W and β . Adjustment factors are substantially different from $1 + \delta^2$, and they tend to increase with the true β value and the variance ratio δ^2 , and decrease with the censoring probability. This analysis demonstrates that, although naive approach is easy to apply in practice, it may result in a big bias with the Cox proportional hazards model. Unlike in a simple linear regression, a uniform adjustment factor is not adequate under the Cox model. Adjustment factor depends on the magnitude of the true β , the variance ratio δ^2 , the censoring mechanism, and the functional form of

baseline hazard function $\lambda_0(t)$. In a mediation analysis setting, there could be multiple covariates with measurement error, making the form of bias and correction procedure much more complex.

1.3.2 Induced Hazard Approach

Prentice (1982) introduced the induced hazard function for the Cox proportional hazards model with classical measurement error as

$$\begin{aligned}\lambda(t; \mathbf{W}) &= E\{\lambda(t; \mathbf{X})|\tilde{T} \geq t, \mathbf{W}\} \\ &= \lambda_0(t)E\{\exp(\boldsymbol{\beta}^T \mathbf{X})|\tilde{T} \geq t, \mathbf{W}\},\end{aligned}\tag{1.5}$$

With this model, the hazard ratio

$$E\{\exp(\boldsymbol{\beta}^T \mathbf{X})|\tilde{T} \geq t, \mathbf{W}\}$$

usually depends on the baseline hazard function $\lambda_0(t)$ through the presence of $\{\tilde{T} \geq t\}$ in the conditioning event. Partial likelihood (1.2) can be modified as

$$PL(\boldsymbol{\beta}) = \prod_{i=1}^k \frac{E\{\exp(\boldsymbol{\beta}^T \mathbf{X}_i)|\tilde{T} \geq t_i, \mathbf{W}_i\}}{\sum_{j \in R(t_i)} E\{\exp(\boldsymbol{\beta}^T \mathbf{X}_j)|\tilde{T} \geq t_i, \mathbf{W}_j\}},\tag{1.6}$$

which also involves the baseline hazard function $\lambda_0(t)$. Since the primary reason for using partial likelihood approach is to avoid the nuisance function $\lambda_0(t)$, conditions under which (1.6) is exactly or approximately independent of $\lambda_0(t)$ are of interest. It turns out that when $\boldsymbol{\beta} = 0$, and approximately if $f(\mathbf{x}|\mathbf{w})$ is concentrated, or if $P(\tilde{T} \geq t|\mathbf{w}) \approx 1$ over the follow-up, (1.6) is approximately independent of $\lambda_0(t)$. The last condition $P(\tilde{T} \geq t|\mathbf{w}) \approx 1$ can be interpreted as failure events (or disease events in many applications) are rare, which is typically a reasonable approximation in prevention trials. In this case, we can approximate the induced hazard by

$$\begin{aligned}\lambda(t; \mathbf{W}) &= E\{\lambda(t; \mathbf{X})|\tilde{T} \geq t, \mathbf{W}\} \\ &\approx E\{\lambda(t; \mathbf{X})|\mathbf{W}\}.\end{aligned}$$

A special case is when \mathbf{X} given \mathbf{W} is normally distributed with mean $\mathbf{E}(\mathbf{W}, \mathcal{A})$, and covariance matrix $\mathbf{V}(\mathbf{W}, \mathcal{A})$ (\mathcal{A} is a set of distribution parameters). The induced hazard (1.5)

is then

$$\begin{aligned}\lambda(t; \mathbf{W}) &= \lambda_0(t) E\{\exp(\boldsymbol{\beta}^T \mathbf{X}) | \tilde{T} \geq t, \mathbf{W}\} \\ &\approx \lambda_0(t) E\{\exp(\boldsymbol{\beta}^T \mathbf{X}) | \mathbf{W}\} \\ &= \lambda_0(t) \exp\{\boldsymbol{\beta}^T \mathbf{E}(\mathbf{W}, \mathcal{A}) + \frac{1}{2} \boldsymbol{\beta}^T \mathbf{V}(\mathbf{W}, \mathcal{A}) \boldsymbol{\beta}\},\end{aligned}$$

and $\boldsymbol{\beta}$ can be estimated by plugging in the observed \mathbf{W} along with an estimate of \mathcal{A} .

When the rare disease assumption does not hold, ignoring the conditioning event $\{\tilde{T} \geq t\}$ can lead to serious bias. Zhou and Pepe (1995) proposed a nonparametric approach to estimate the induced hazard (1.5) without the rare disease assumption, when covariates were discrete. Zhou and Wang (2000) extended this method to continuous covariates using kernel smoothing. However, this method depends heavily on the bandwidth of the kernel function, and it is not practical if the dimension of \mathbf{X} is high. Another limitation of the two methods is that they both require a validation subsample in which covariates of interest are measured without error, which may not be a possibility, even conceptually, in many applications.

1.3.3 Regression Calibration

Regression calibration is a simple and intuitive method to handle measurement error. This approach approximates the true covariate value \mathbf{X} by $\mathbf{E}(\mathbf{X} | \mathbf{W})$, which is the conditional expectation of \mathbf{X} given the observed covariate \mathbf{W} . Often, distribution parameters \mathcal{A} involved in this conditional expectation are estimated from a validation subsample. With the Cox proportional hazards model, this conditional expectation $\mathbf{E}(\mathbf{X} | \mathbf{W})$ can replace \mathbf{X} in the usual partial likelihood (1.2) and the corresponding score function (1.3) to estimate $\boldsymbol{\beta}$. Wang et al. (1997) studied the asymptotic distribution of this estimator, and proposed a suitable variance estimator. Their simulation study shows that the bias associated with this approach is modest in situations of practical interest. A major advantage of regression calibration is that, after getting an estimate of $\mathbf{E}(\mathbf{X} | \mathbf{W})$, one can use existing computer software to fit the Cox model without modification, though a non-standard variance estimator is needed. Also, since a validation subsample may often not be a possibility, the regression

calibration procedure typically uses, instead, a reliability subsample that involves repeated observations of \mathbf{W} from (1.4) on some study subjects for the estimation of distribution parameters \mathcal{A} .

1.3.4 Risk Set Calibration

When a reliability sample is available, Xie et al. (2001) extended the regression calibration by improving the approximation of $E\{\exp(\boldsymbol{\beta}^T \mathbf{X}) | \tilde{T} \geq t, \mathbf{W}\}$. Instead of providing a constant estimate $\mathbf{E}(\mathbf{X} | \mathbf{W})$, they suggested recalibrating within each risk set across the failure time axis. Let

$$\mathbf{W}_{ij} = \mathbf{X}_i + \epsilon_{ij}, i = 1, 2, \dots, n; j = 1, \dots, k_i$$

denote the j^{th} measurement of the i^{th} subject, and $\overline{\mathbf{W}}_i$ denote the mean of $\mathbf{W}_{i1}, \dots, \mathbf{W}_{ik_i}$. Under the classical measurement error model, the measurement error variance matrix $\boldsymbol{\Delta}$, mean $\boldsymbol{\mu}(t)$ and variance $\boldsymbol{\Sigma}(t)$ of the true \mathbf{X} at time t can be estimated from replication data of subjects still at risk at t . If $(\mathbf{X}_i, \overline{\mathbf{W}}_i)$ has a joint normal distribution in the risk set at time t , the expectation of \mathbf{X}_i given $\overline{\mathbf{W}}_i$ in this risk set is

$$\hat{\mathbf{X}}_i(t) = \hat{\boldsymbol{\Delta}} k_i^{-1} \{\hat{\boldsymbol{\Sigma}}(t) + \hat{\boldsymbol{\Delta}} k_i^{-1}\}^{-1} \hat{\boldsymbol{\mu}}(t) + \hat{\boldsymbol{\Sigma}}(t) \{\hat{\boldsymbol{\Sigma}}(t) + \hat{\boldsymbol{\Delta}} k_i^{-1}\}^{-1} \overline{\mathbf{W}}_i.$$

This estimate $\hat{\mathbf{X}}_i(t)$ is used at time t in the partial likelihood. Instead of assuming the covariate distribution is fixed, this approach allows the covariate distribution to change over time, which may provide a better approximation to the induced hazard $\lambda(t; \mathbf{W})$. However, this estimator typically has some asymptotic bias, but this bias is often negligible in situations of practical importance.

Xie extended the risk set calibration in her Department of Biostatistics, University of Washington dissertation (Xie, 1997) by a second-order approximation. The first-order approximation mentioned above is

$$E[\exp(\boldsymbol{\beta}^T \mathbf{X}) | \tilde{T} \geq t, \mathbf{W}] \approx \exp[\boldsymbol{\beta}^T E(\mathbf{X} | \tilde{T} \geq t, \mathbf{W})].$$

The second-order approach further approximates the induced hazard function by both the first and the second moments of the conditional distribution of \mathbf{X} given \mathbf{W} :

$$E[\exp(\boldsymbol{\beta}^T \mathbf{X}) | \tilde{T} \geq t, \mathbf{W}] \approx \exp[\boldsymbol{\beta}^T \mathbf{E}(\mathbf{X} | \tilde{T} \geq t, \mathbf{W})] + \frac{1}{2} \boldsymbol{\beta}^T \mathbf{V}(\mathbf{X} | \tilde{T} \geq t, \mathbf{W}) \boldsymbol{\beta}.$$

When the conditional distribution of $(\mathbf{X}|\tilde{T} \geq t, \mathbf{W})$ is not normal, this can be viewed as a second-order Taylor expansion. Both the mean and the variance of the \mathbf{X} distribution contribute to the approximation. This method works well under a broad range of β values and some complicated censoring mechanisms, but asymptotic bias is still present. Another limitation of both risk set calibration approaches is that they require a reliability sample, which may not be available in some biomarker studies.

1.3.5 Other Approaches

Huang and Wang (2000, 2006) proposed a nonparametric approach when error-assessment data is available. They avoided distributional assumptions on \mathbf{X} or \mathbf{W} , and proposed a corrected score function approach. This approach allows a wide selection of internal or external error-assessment data, which includes the situation of validation subsample and reliability subsample. This approach provides a consistent estimator. However, this nonparametric approach may not have satisfactory small-sample performance, due to inflated parameter estimate variances when measurement error is substantial. It is also numerically challenging.

Hu et al. (1998) proposed using a full likelihood instead of a partial likelihood to handle measurement error. They assumed the conditional distribution function $f(\mathbf{W}|\mathbf{X})$ was of a known form, and used parametric, semiparametric and nonparametric approaches to estimate $f(\mathbf{X})$. Their simulation study shows full likelihood approach has better performance than regression calibration in some settings, and it is robust to distributional assumptions. However, the method is designed for one variable with measurement error, and it is difficult to be generalized to multiple mis-measured variables. Also the computational burden is prohibitive when considering the use of this method to large data sets.

1.4 Mediation Analysis

Mediation analysis is important in social and behavioral sciences to understand the mechanisms that changes in one variable causes changes in another. It is also important in biomedical research, for example in the design and evaluation of prevention and treatment intervention trials (MacKinnon, 2008). A simple mediation diagram is shown in Figure 1.1,

where Z is an independent variable, M is a mediation variable (or mediator), and Y is a dependent variable (or response variable). It is of interest to understand the process by which Z affects Y . Z may explain some variability in Y directly, and may also explain some variability in Y due to Z 's influence on the mediation variable, M .

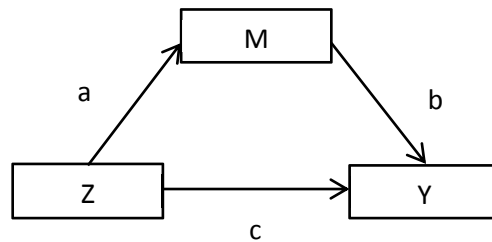


Figure 1.1: Three-variable mediation analysis diagram, with independent variable Z , dependent variable Y and mediator M .

When the relationships are linear, a classic mediation analysis involves fitting the following three regression models:

$$M = \beta_1 + aZ + \epsilon_1$$

$$Y = \beta_2 + cZ + \epsilon_2$$

$$Y = \beta_3 + c'Z + bM + \epsilon_3$$

As discussed in Baron and Kenny (1986), evidence for mediation is said to be likely if,

1. there is a linear relationship between the independent variable Z and the mediator M ; which means, the parameter a is significant;
2. there is a linear relationship between the independent variable Z and the dependent variable Y ; which means, the parameter c is significant;
3. in the third regression model, the mediator M is useful in predicting the dependent variable Y , and the direct effect of the independent variable Z on the dependent variable Y becomes much closer to 0 relative to that not adjusted for the mediator; which means, parameter b is significant, and c' is much closer to 0 comparing to c .

Preacher and Hayes (2004) discussed formal tests for the hypothesis $H_0 : c' - c = 0$.

Hoyle and Kenny (1999) identified biases in both the direct and indirect effect estimates in this three-variable mediation analysis with linear model (Figure 1.1) when the mediator was measured with error. The biases increase as the reliability of measurements decreases. Baron and Kenny (1986) commented that “the presence of measurement error in the mediator tends to produce an underestimate of the effect of the mediator and an overestimate of the effect of the independent variable on the dependent variable when all coefficients are positive. Obviously this is not a desirable outcome, because successful mediators may be overlooked”. They suggested using a latent-variable structural equation modeling (SEM) approach to correct for biases introduced by measurement error in mediator. This approach involves solving a system of equations which represent the relationship between the independent variable, the mediator and the dependent variable. When there is a latent variable (here the true value of the mediator), at least two construct variables are required to estimate the latent variable. The construct variables should be caused by the latent variable (i.e. true value of mediator) but are not related to any other variables in the pathway. Hoyle and Kenny (1999) demonstrated through simulation studies that the latent-variable SEM approach corrected the biases well in linear regression setting with a reasonable sample size. However, this method may not be applicable in our motivating example. First, these methods evidently acknowledge only the technical part of measurement error, but not the temporal variation of the underlying measurement from the long-term “average” biomarker value that is typically relevant to disease risk. Second, available mediation methods have not dealt with censored failure time outcomes. New methods are needed to allow for measurement error in mediation analysis with failure time data.

There are many recent developments in causal inference. The counterfactual (potential outcome) framework is one method that has attracted some research interest. Lange and Hansen (2011) proposed a measure of the mediation effect in the failure time data analysis setting based on counterfactual framework. Instead of the Cox proportional hazards model, they assume the hazard rate follows an Aalen additive hazard model

$$\lambda(t; Z, M, \mathbf{C}) = \lambda_0(t) + \lambda_1(t)Z + \lambda_2(t)\mathbf{C} + \lambda_3(t)M, \quad (1.7)$$

where Z is the treatment, M is the mediator, and \mathbf{C} are other covariates. They further assume that the mediator is a normal variable that can be modeled by a linear regression on Z and \mathbf{C}

$$M = \alpha_0 + \alpha_1 Z + \alpha_2 \mathbf{C} + e. \quad (1.8)$$

Under these models and some other conditions, the total treatment effect of Z can be decomposed into a natural direct effect and a natural indirect effect. The natural direct and indirect effects can both be expressed as functions of the parameters in model (1.7) and (1.8). This approach provides a way to quantify the mediation effect in the survival analysis context. However, it assumes the underlying model is an additive hazard model rather than the most widely used Cox proportional hazards model. Extending this method to the Cox model is not straightforward.

In this dissertation, we are interested in methods to assess the role of biomarker changes in explaining treatment effects, rather than evaluating causal effects. Hence, we will use the classical methods of comparing the treatment effects estimated from the models with and without the potential mediator, in the context of a Cox proportional hazards model for hazard rates.

1.5 Measurement Error Model

As mentioned before, we are interested in understanding how measurement error in the mediator affects the hazard ratio estimates (especially treatment effect estimate) under the Cox proportional hazards model. For subject i , $i = 1, 2, \dots, n$, let $Z_i = I(\text{treatment})$ be a 0/1 indicator of the treatment assignment; $\mathbf{X}_i = (X_{i0}, X_{i1})$ denote the baseline and post-randomization ‘true’ biomarker values, and $\mathbf{W}_i = (W_{i0}, W_{i1})$ be the corresponding observed biomarker values. Suppose classical measurement error model holds

$$\mathbf{W}_i = \mathbf{X}_i + \mathbf{U}_i,$$

where $\mathbf{U}_i = (U_{i0}, U_{i1})$ are random errors independent of \mathbf{X}_i . Before specifying the joint distribution of (\mathbf{X}, \mathbf{W}) , we need to carefully decide some relevant ways that measurement error can be conceptualized.

Although modeling observed \mathbf{W} is not our primary interest, we can usefully decompose \mathbf{W} to help us understand its variability. When covariates are measured longitudinally (at baseline $t=0$ and 1-year post-randomization $t=1$ in WHI setting), there are at least three sources of random variations that may contribute to the variability of the observed W_{ij} , ($i = 1, 2, \dots, n$, $j = 0, 1, \dots, m$, and $k = m$ in WHI example) (Diggle et al., 2002, Chapter 5): subject-specific random effects, temporal variation and technical measurement error. Hence W_{ij} can be written as

$$W_{ij} = \mu(Z_i, t_j) + b_i(Z_i, t_j) + S_i(Z_i, t_j) + \epsilon_{ij}, \quad (1.9)$$

Here, $\mu(Z_i, t_j)$ is a fixed population mean, which may differ by treatment assignment Z_i and time t_j . And $b_i(Z_i, t_j)$ is a subject-specific random effect. This represents the difference between the mean of the i^{th} subject's measures and the population mean. When units are sampled at random from a population, some of them might have consistently high biomarker values and some might have consistently low values. This subject-specific random effect differs between subjects, and by definition its mean value is zero. $S_i(Z_i, t_j)$ is the temporal variation for the biomarker process. It has mean zero, and the correlation is expected to become weaker as the time separation increases. ϵ_{ij} is the noise aspect of the measurement error, 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 measurement error in the i^{th} subject's biomarker assessment at time t_j , 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 they are uncorrelated with variabilities of other subjects. We will discuss the model structure and the estimates in detail later. With this decomposition, we can specify two formulations of measurement errors: uncorrelated measurement errors and correlated measurement errors.

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

$$\begin{aligned} X_{ij} &= \mu(Z_i, t_j) + b_i(Z_i, t_j) + S_i(Z_i, t_j) \\ U_{ij} &= \epsilon_{ij} \end{aligned}$$

We consider the technical measurement error as the only source of the measurement error,

and the targeted X_{ij} is the true biomarker value of subject i at time t_j . Measurement errors are independent between and within subjects. The joint distribution of $(\mathbf{X}, \mathbf{U}) = (X_0, X_1, U_0, U_1)$ given Z is

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{U} \end{pmatrix} | Z \sim N\left(\begin{pmatrix} \mathbf{M}_Z \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_Z & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}_Z \end{pmatrix}\right), \quad (1.10)$$

where

$$\begin{aligned} \mathbf{M}_0 &= \begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix}, & \boldsymbol{\Sigma}_0 &= \begin{pmatrix} \sigma_0^2 & \rho_0 \sigma_0^2 \\ \rho_0 \sigma_0^2 & \sigma_0^2 \end{pmatrix}, & \boldsymbol{\Delta}_0 &= \begin{pmatrix} \sigma_{e0}^2 & 0 \\ 0 & \sigma_{e0}^2 \end{pmatrix}, \\ \mathbf{M}_1 &= \begin{pmatrix} \mu_0 \\ \mu_1 + d \end{pmatrix}, & \boldsymbol{\Sigma}_1 &= \begin{pmatrix} \sigma_0^2 & \rho_1 \sigma_0 \sigma_1 \\ \rho_1 \sigma_0 \sigma_1 & \sigma_1^2 \end{pmatrix}, & \boldsymbol{\Delta}_1 &= \begin{pmatrix} \sigma_{e0}^2 & 0 \\ 0 & \sigma_{e1}^2 \end{pmatrix}. \end{aligned} \quad (1.11)$$

By assuming this distribution, we allow biomarker values to change with time (μ_0, μ_1 unequal), and treatment to change the mean biomarker values by d , but variances of biomarker values only depend on whether the subject has already been treated. Measurement errors always have mean zero and are uncorrelated. We usually assume their distribution does not depend on \mathbf{X} or Z , which means $\sigma_{e0}^2 = \sigma_{e1}^2$. However, we use the above notation to allow a more general situation that the measurement error variance can change with treatment status. In this scenario, distribution parameters are $\mathcal{A} = (\mu_0, \mu_1, d, \sigma_0^2, \sigma_1^2, \rho_0, \rho_1, \sigma_{e0}^2, \sigma_{e1}^2)$. We assume that the underlying hazard function $\lambda(t|\mathbf{X}, Z)$ is

$$\lambda(t|\mathbf{X}, Z) = \lambda_0(t) \exp(\beta_0 X_0 + \beta_1 X_1 + \beta_Z Z). \quad (1.12)$$

This model allows both baseline and 1-year post-randomization biomarker levels and treatment in relation to disease risk.

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

$$\begin{aligned} X_{ij} &= \mu(Z_i, t_j) + b_i(Z_i, t_j) \\ U_{ij} &= S_i(Z_i, t_j) + \epsilon_{ij}. \end{aligned}$$

With this definition, measurement error includes both the technical measurement error and the temporal variation. The measurement errors within a subject are correlated because of

the $S_i(Z_i, t_j)$ term. The corresponding X_{ij} is a subject-specific mean biomarker value. It may change with time and treatment assignment. In the special case when both $\mu(Z_i, t_j)$ and $b_i(Z_i, t_j)$ do not change with t_j , X_{ij} becomes the subject's long-term average of the biomarker value. The joint distribution of $(\mathbf{X}, \mathbf{U}) = (X_0, X_1, U_0, U_1)$ is then

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{U} \end{pmatrix} | Z \sim N\left(\begin{pmatrix} \mathbf{M}_Z \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_Z & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}_Z \end{pmatrix}\right), \quad (1.13)$$

where

$$\begin{aligned} \mathbf{M}_0 &= \begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix}, & \boldsymbol{\Sigma}_0 &= \begin{pmatrix} \sigma_0^2 & \sigma_0^2 \\ \sigma_0^2 & \sigma_0^2 \end{pmatrix}, & \boldsymbol{\Delta}_0 &= \begin{pmatrix} \sigma_{e0}^2 & r_0 \sigma_{e0}^2 \\ r_0 \sigma_{e0}^2 & \sigma_{e0}^2 \end{pmatrix}, \\ \mathbf{M}_1 &= \begin{pmatrix} \mu_0 \\ \mu_1 + d \end{pmatrix}, & \boldsymbol{\Sigma}_1 &= \begin{pmatrix} \sigma_0^2 & \rho \sigma_0 \sigma_1 \\ \rho \sigma_0 \sigma_1 & \sigma_1^2 \end{pmatrix}, & \boldsymbol{\Delta}_1 &= \begin{pmatrix} \sigma_{e0}^2 & r_1 \sigma_{e0} \sigma_{e1} \\ r_1 \sigma_{e0} \sigma_{e1} & \sigma_{e1}^2 \end{pmatrix}. \end{aligned} \quad (1.14)$$

By this distribution, we assume similar mean and variance structure for (\mathbf{X}, \mathbf{U}) as in the uncorrelated measurement error scenario. The differences are that the correlation between X_0 and X_1 now becomes exactly 1 in the control group ($Z = 0$), and measurement errors are now correlated with each other ($r_0, r_1 \neq 0$). In this scenario, the distribution parameters are $\mathcal{A} = (\mu_0, \mu_1, d, \sigma_0^2, \sigma_1^2, \rho, \sigma_{e0}^2, \sigma_{e1}^2, r_0, r_1)$. The hazard function $\lambda(t|\mathbf{X}, Z)$ is still specified as (1.12) that

$$\lambda(t|\mathbf{X}, Z) = \lambda_0(t) \exp(\beta_0 X_0 + \beta_1 X_1 + \beta_Z Z).$$

The choice between the 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 the correlated measurement error should be used. If one simply wishes to conduct mediation analysis that is adjusted for technical measurement error, then the uncorrelated measurement error is more appropriate. We will discuss the biases associated with both specifications of measurement errors and methods to correct for them in the subsequent chapters.

1.6 Overview of Dissertation

In this dissertation, we focus on developing correction methods for measurement error in the potential mediator with failure time outcome. We are interested in the special case that a

biomarker is measured at both baseline and post-randomization, and the change from baseline to post-randomization potentially mediates the relationship between the failure time and the treatment assignment. Both measures are subject to measurement errors. In this dissertation, we restrict study subjects to those who survive to the post-randomization clinic visit. For well-maintained prevention trials, this usually does not lead to problems, since study subjects are unlikely to develop disease or quit study before the post-randomization visit.

In Chapter 2 and Chapter 3, we assume the biomarker measures are associated with the uncorrelated measurement error (i.e., technical measurement error) and the correlated measurement error (i.e., technical measurement error and temporal variation), respectively. In each chapter, we first investigate how the specific type of measurement error affects the parameter estimates if the observed biomarker values are plugged into the model without any adjustment (i.e., naive approach). We then base our correction methods on the hazard function induced by the observed biomarker values. This induced hazard function no longer corresponds to a partial likelihood independent of the baseline hazard function, due to the conditioning event $\{\tilde{T} \geq t\}$. We propose two methods to simplify the induced partial likelihood, including a mean-variance regression calibration (MVC) which ignores the conditioning event $\{\tilde{T} \geq t\}$ under the rare disease assumption, and a follow-up time calibration (FUC) which assumes biomarker distributions are constant within each pre-specified time interval. We present simulation results to evaluate and compare the performances of these two proposed methods and their sensitivities on parameter specification, censoring mechanism and normality assumptions.

In Chapter 4, we develop the asymptotic distribution theory through the multivariate counting process and the martingale theory. Variance estimators are provided, with simulation studies to compare them with simulation standard errors.

In Chapter 5, we generalize our discussion to allow multiple biomarkers and the case-cohort study design. When the relationship between the treatment and the outcome is potentially mediated by multiple biomarkers, MVC and FUC still apply, with some modification in calibration. We propose to compute the conditional means and variances separately for each biomarker, ignoring the correlations between the biomarkers. We perform a simu-

lation study to demonstrate that separate calibration is efficient in recovering the treatment effect. For large prevention trials, case-cohort design is cost-efficient. MVC and FUC can be generalized to such study designs, and the performance is stable when the subcohort size is relatively large.

In Chapter 6, we apply our correction methods to the WHI hormone therapy trials data. A re-analysis of the mediation effect of IGFBP4 in the relationship between active hormone therapy and the time to stroke is provided.

In Chapter 7, we summarize the dissertation with a discussion of the direction of future research.

Chapter 2

**INDUCED HAZARD APPROACH— UNCORRELATED
MEASUREMENT ERROR**

2.1 Overview

When covariates are measured with error, the simplest method is to use the observed \mathbf{W} as a surrogate for the true covariate values \mathbf{X} . However, this method may lead to biased hazard ratio estimates. As discussed in the motivating example, our main interest is in recovering the true treatment effect in the presence of measurement error in the mediator. Hence, we focus on developing correction methods that provide good estimate for the treatment effect, while hoping that biases of other parameters are reasonably small. In this chapter, we first consider the hazard function that is induced as a function of \mathbf{W} and the distribution parameters \mathcal{A} under the rare disease assumption, in a setting as show in Figure 2.1. The outcome Y relates directly with both pre- and post- randomization biomarker values $\mathbf{X} = (X_0, X_1)$, and the treatment assignment Z . In addition, the post-randomization biomarker value X_1 (or equivalently, $X_1 - X_0$) may contribute to the mediation of the relationship between Z and Y . This means that the treatment Z can have both a direct effect on the outcome Y and an indirect effect through the change of biomarker value $X_1 - X_0$. With classical measurement error $\mathbf{W} = \mathbf{X} + \mathbf{U}$ and under the rare disease assumption $P(\tilde{T} \geq t | \mathbf{X}, Z) \approx 1$, the induced hazard $\lambda(t; \mathbf{W}, Z)$ can be simplified and the true parameters can be estimated based on the parameters in the naive model. In a large prevention trial, the rare disease assumption is usually a good approximation. For example, in the WHI hormone therapy trial mentioned in the motivating example, by the end of study, less than 1% subjects had stroke (205 out of 27347 subjects). In this chapter, we first discuss the naive approach, and then discuss the estimates based on the induced hazard approach under the rare disease assumption (mean-variance regression calibration). We further extend our method to allow covariate distribution changing over time and between treatment arms (follow-up

time calibration), which is more robust to the departure from the rare disease assumption. Performances of the two proposed estimates are compared with estimates from the naive approach through simulation studies. In this chapter, we focus on the uncorrelated measurement error (technical measurement error) case, and discuss the correlated measurement error case in Chapter 3. For now, we only deal with cohort studies. More complicated study designs, such as case-cohort design, are discussed in Chapter 5.

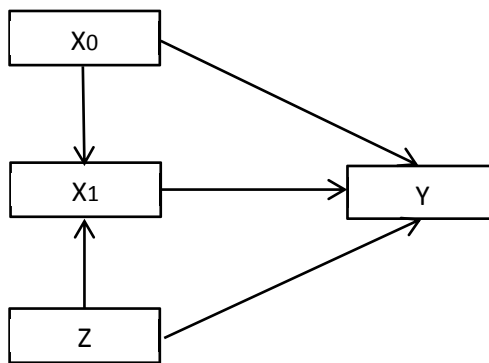


Figure 2.1: Mediation analysis diagram, with pre- and post-randomization biomarker values X_0 , X_1 , treatment assignment Z and outcome Y .

Assume that there are n subjects in total, among which n_0 subjects are not treated and n_1 are treated. As before, for subject i , $i = 1, 2, \dots, n$, let \tilde{T}_i denote the failure time, C_i the censoring time, $T_i = \min(\tilde{T}_i, C_i)$ the observed event time, and $\delta_i = I(\tilde{T}_i \geq C_i)$ the censoring indicator for subject i . Let $Z_i = I(\text{treatment})$ be a 0/1 indicator of the treatment assignment. For now, we deal with a single biomarker. Let $\mathbf{X}_i = (X_{i0}, X_{i1})$ denote the baseline and post-randomization values of this biomarker, and $\mathbf{W}_i = (W_{i0}, W_{i1})$ be the observed biomarker values. For subject i , we observe $(T_i, \delta_i, \mathbf{W}_i, Z_i)$ instead of $(T_i, \delta_i, \mathbf{X}_i, Z_i)$ due to measurement errors. Note that only subjects with complete (W_0, W_1) measures are included in the analysis. That means, subjects have to survive to the post-randomization clinic visit to be eligible for study participation. As discussed before, this usually does not cause any problem in prevention trials. The joint distribution of $(\mathbf{X}, \mathbf{U}|Z)$ is as described

in (1.10) and (1.11):

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{U} \end{pmatrix} | Z \sim N\left(\begin{pmatrix} \mathbf{M}_Z \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_Z & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}_Z \end{pmatrix}\right), \quad (2.1)$$

where

$$\begin{aligned} \mathbf{M}_0 &= \begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix}, & \boldsymbol{\Sigma}_0 &= \begin{pmatrix} \sigma_0^2 & \rho_0 \sigma_0^2 \\ \rho_0 \sigma_0^2 & \sigma_0^2 \end{pmatrix}, & \boldsymbol{\Delta}_0 &= \begin{pmatrix} \sigma_{e0}^2 & 0 \\ 0 & \sigma_{e0}^2 \end{pmatrix}, \\ \mathbf{M}_1 &= \begin{pmatrix} \mu_0 \\ \mu_1 + d \end{pmatrix}, & \boldsymbol{\Sigma}_1 &= \begin{pmatrix} \sigma_0^2 & \rho_1 \sigma_0 \sigma_1 \\ \rho_1 \sigma_0 \sigma_1 & \sigma_1^2 \end{pmatrix}, & \boldsymbol{\Delta}_1 &= \begin{pmatrix} \sigma_{e0}^2 & 0 \\ 0 & \sigma_{e1}^2 \end{pmatrix}. \end{aligned} \quad (2.2)$$

Technical measurement error is usually assumed to be independent of treatment assignment Z . For this chapter of uncorrelated measurement errors, we mainly focus on this situation (i.e., $\sigma_{e0}^2 = \sigma_{e1}^2 = \sigma_e^2$), but keep the general notation to allow more flexibility. Distribution parameters are $\mathcal{A} = (\mu_0, \mu_1, d, \sigma_0^2, \sigma_1^2, \rho_0, \rho_1, \sigma_{e0}^2, \sigma_{e1}^2)$.

In the motivating example, all subjects were randomized to either control or treatment ($Z = 0, 1$) and their baseline IGFBP4 values (W_0) were measured at the beginning of the study. One year after randomization, their IGFBP4 values were measured again (W_1). Subjects were followed until stroke or a defined end to the study follow-up period (T, δ). In the preliminary analyses, time to stroke was seen to be closely associated with treatment and IGFBP4 values at both baseline and 1 year after randomization, and post-randomization IGFBP4 value was affected by both its baseline value and the treatment.

2.2 Naive Approach

We assume that the Cox proportional hazards model holds for $(T, \delta, \mathbf{X}, Z)$. For mediation analysis, we compare the following two models:

$$\lambda(t; X_0, Z) = \lambda_0(t) \exp(\alpha_0 X_0 + \alpha_Z Z) \quad (2.3)$$

$$\lambda(t; X_0, X_1, Z) = \lambda_1(t) \exp(\beta_0 X_0 + \beta_1 X_1 + \beta_Z Z) \quad (2.4)$$

If β_Z moves substantially towards null compared to α_Z , we conclude that there is evidence that X_1 (or equivalently, $X_1 - X_0$) contributes to the mediation of the relationship between T and Z . In practice, treatment Z is usually measured accurately, while $\mathbf{X} = (X_0, X_1)$ may be

subject to technical measurement error. When X_0 is measured with technical measurement error, naive estimate of α_Z in model (2.3) usually is not associated with large bias, because both baseline biomarker value X_0 and its measurement error U_0 distributions are similar in both treatment groups (i.e., (X_0, U_0) and Z are statistically independent). However, in model (2.4), β_Z can be biased. This is because post-randomization biomarker value X_1 and its corresponding measurement error U_1 distributions may highly depend on the treatment assignment Z . Hence, the naive approach

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

results in estimate $\hat{\alpha}_Z$ close to α_Z , but $\hat{\beta}_Z$ biased away from β_Z . This potentially leads to wrong conclusions about the mediation effect of X_1 .

Suppose that the sample comprises k uncensored failure times $t_1 < t_2 < \dots < t_k$, where i denotes the subject failing at t_i . The other $n - k$ subjects are right censored. To estimate $(\beta_0, \beta_1, \beta_Z)$, the score equations, if the underlying X_0 and X_1 were available, are

$$\begin{cases} \sum_{i=1}^k (\mathbf{X}_i - \frac{\sum_{j \in R(t_i)} \mathbf{X}_j \exp(\boldsymbol{\beta}_X^T \mathbf{X}_j + \beta_Z Z_j)}{\sum_{j \in R(t_i)} \exp(\boldsymbol{\beta}_X^T \mathbf{X}_j + \beta_Z Z_j)}) = 0 \\ \sum_{i=1}^k (Z_i - \frac{\sum_{j \in R(t_i)} Z_j \exp(\boldsymbol{\beta}_X^T \mathbf{X}_j + \beta_Z Z_j)}{\sum_{j \in R(t_i)} \exp(\boldsymbol{\beta}_X^T \mathbf{X}_j + \beta_Z Z_j)}) = 0 \end{cases}$$

where $\boldsymbol{\beta}_X = (\beta_0, \beta_1)$. Following similar derivation in Hughes (1993), when replacing \mathbf{X} by \mathbf{W} , the score equations become

$$\begin{cases} \sum_{i=1}^k (\mathbf{W}_i - \frac{\sum_{j \in R(t_i)} \mathbf{W}_j \exp(\mathbf{b}_W^T \mathbf{W}_j + b_Z Z_j)}{\sum_{j \in R(t_i)} \exp(\mathbf{b}_W^T \mathbf{W}_j + b_Z Z_j)}) = 0 \\ \sum_{i=1}^k (Z_i - \frac{\sum_{j \in R(t_i)} Z_j \exp(\mathbf{b}_W^T \mathbf{W}_j + b_Z Z_j)}{\sum_{j \in R(t_i)} \exp(\mathbf{b}_W^T \mathbf{W}_j + b_Z Z_j)}) = 0 \end{cases}$$

where $\mathbf{b}_W = (b_0, b_1)$. Let sample size $n \rightarrow \infty$, and we assume a simple censoring mechanism that all subjects are censored at the study end C_{end} , then the score equations imply

$$\begin{cases} E_{\mathbf{W}, Z | \tilde{T} \leq C_{end}} \mathbf{W} - E_{t | t \leq C_{end}} \frac{E_{\mathbf{W}, Z | \tilde{T} \geq t} \mathbf{W} \exp(\mathbf{b}_W^T \mathbf{W} + b_Z Z)}{E_{\mathbf{W}, Z | \tilde{T} \geq t} \exp(\mathbf{b}_W^T \mathbf{W} + b_Z Z)} = 0 \\ E_{\mathbf{W}, Z | \tilde{T} \leq C_{end}} Z - E_{t | t \leq C_{end}} \frac{E_{\mathbf{W}, Z | \tilde{T} \geq t} Z \exp(\mathbf{b}_W^T \mathbf{W} + b_Z Z)}{E_{\mathbf{W}, Z | \tilde{T} \geq t} \exp(\mathbf{b}_W^T \mathbf{W} + b_Z Z)} = 0 \end{cases}$$

Details for evaluating these score equations are provided in Appendix A. The resulting score equations involve double integrals and usually do not have explicit solutions. Numeric solutions can be provided through iterative procedures.

To illustrate this relationship, we consider a cohort study of sample size 10,000 with half of the study participants randomized to active treatment, and all subjects are censored at a fixed time C_{end} (administrative censoring). The underlying joint distribution of covariates (X_0, X_1) is normal in each treatment arm, having mean 0 without treatment and 0.5 with treatment and correlation 0.95 and 0.9, and variance of pre-treatment biomarker is standardized to 1 (i.e., $\mu_0 = \mu_1 = 0, d = 0.5, \sigma_0^2 = 1, \rho_0 = 0.95, \rho_1 = 0.9$). Suppose that the baseline hazard $\lambda_0(t)$ is constantly 1 over time, and $\beta_0 = \beta_1 = 1$. We investigate how biases of each parameter vary with $\sigma_1^2, \sigma_{e0}^2, \sigma_{e1}^2, \beta_Z$ and the censoring probability. Results are summarized in Figure 2.2.

In Figure 2.2(a)–(c), we constrain measurement error distribution to be the same in the two treatment arms (i.e., $\sigma_{e0}^2 = \sigma_{e1}^2 = \sigma_e^2$), and investigate the effects of σ_1^2, β_Z and censoring probability on the relationship between σ_e^2 and the biases in β . In all the scenarios, it is clear that as the measurement error variances σ_e^2 gets larger, biases for all the three parameters get further away from 0. When β_Z is equal to $\log(1.5)$ (i.e., hazard ratio of treatment is 1.5), with 95% censoring probability, biases of all the three parameters change with σ_1^2 (Figure 2.2(a)). As X_1 distribution depends stronger on Z (i.e., σ_1^2 gets away from σ_0^2), bias of β_Z gets further away from 0. In Figure 2.2(b), we fix $\sigma_1^2 = 1.5$ and censoring probability as 95%. Biases of β_0 and β_1 do not seem to depend on the true value of β_Z , while that of β_Z does differ with its true value. With this combination of parameters, larger β_Z value corresponds to smaller biases in β_Z . Figure 2.2(c) shows that biases of all the three parameters change with censoring probability, but in different directions, when we fix $\sigma_1^2 = 1.5$ and $\beta_Z = \log(1.5)$. Larger censoring probability results in smaller number of events, which corresponds to larger biases for β_0 and β_1 , and smaller bias for β_Z .

In Figure 2.2(d), we allow the measurement error variance to differ with treatment status (i.e., $\sigma_{e0}^2 \neq \sigma_{e1}^2$). When fixing σ_1^2 to be 1.5 and censoring probability as 95%, biases of all the three parameters change with the values of $(\sigma_{e0}^2, \sigma_{e1}^2)$. Biases are larger when σ_{e1}^2 are relatively more different from σ_{e0}^2 . Together with Figure 2.2(a), this confirms that the dependence degree of (X_1, U_1) distribution on treatment Z influences the bias of the β_Z estimate.

To sum up, the biases of $(\beta_0, \beta_1, \beta_Z)$ rely on both the true hazard function and the dis-

tribution parameters \mathcal{A} in a complicated fashion with this simple administrative censoring. Ignoring measurement error may lead to big biases in hazard ratio estimates. There is no simple expression of the adjustment factors. With other censoring mechanisms, the relationship can be even more complicated. New methods are needed to provide valid hazard ratio estimates. Note also that bias of β_0 and β_1 are larger than that of β_Z . This is simply due to the fact that the true values of β_0 and β_1 are larger than that of β_Z . Smaller absolute bias does not necessarily mean β_Z tends to be less influenced by measurement error. In fact, in Figure 2.2(b), we find β_Z to be the only parameter influenced by the true value of β_Z .

2.3 Mean-Variance Regression Calibration

Under the rare disease assumption

$$\Pr(\tilde{T} \geq t | \mathbf{X}, Z) \approx 1$$

for all follow-up time, the induced hazard $\lambda(t; \mathbf{W}, Z)$ corresponding to model (2.4) can be simplified as

$$\begin{aligned} \lambda(t; \mathbf{W}, Z) &= E[\lambda(t; \mathbf{X}, Z) | \tilde{T} \geq t; \mathbf{W}, Z] \\ &= \lambda_1(t) E[\exp(\boldsymbol{\beta}_X^T \mathbf{X} + \beta_Z Z) | \tilde{T} \geq t; \mathbf{W}, Z] \\ &\approx \lambda_1(t) E[\exp(\boldsymbol{\beta}_X^T \mathbf{X} + \beta_Z Z) | \mathbf{W}, Z], \end{aligned} \quad (2.5)$$

and we replace “ \approx ” in (2.5) by “=” subsequently. With this assumption, the approximation above implies that the joint distribution of $(\mathbf{X}, \mathbf{U} | \tilde{T} \geq t, Z)$ is constant over time. When $(\mathbf{X}, \mathbf{U} | Z)$ has a joint normal distribution with uncorrelated measurement error as described in (2.1) and (2.2), the conditional distribution of $(\mathbf{X} | \mathbf{W}, Z)$ is also jointly normal

$$\mathbf{X} | \mathbf{W}, Z \sim N(\mathbf{E}(\mathbf{X} | \mathbf{W}, Z), \mathbf{V}(\mathbf{X} | \mathbf{W}, Z)).$$

The induced hazard can be further simplified as

$$\lambda(t; \mathbf{W}, Z) = \lambda_1(t) \exp[\beta_Z Z + \boldsymbol{\beta}_X^T \mathbf{E}(\mathbf{X} | \mathbf{W}, Z) + \frac{1}{2} \boldsymbol{\beta}_X^T \mathbf{V}(\mathbf{X} | \mathbf{W}, Z) \boldsymbol{\beta}_X] \quad (2.6)$$

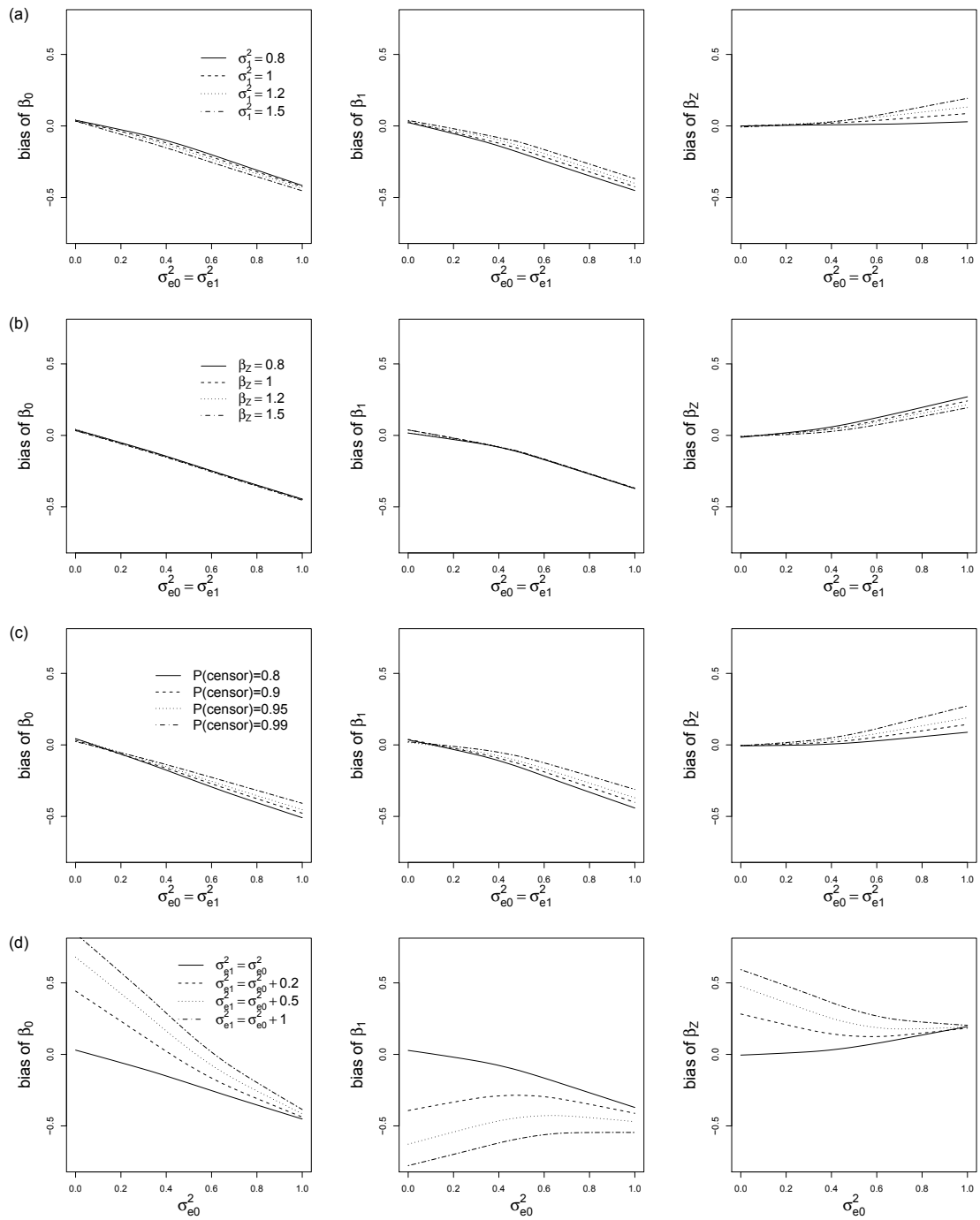


Figure 2.2: Summary of estimated biases of naive approach with uncorrelated measurement error, for different configurations of parameters: (a) different σ_1^2 , (b) different β_Z , (c) different censoring probabilities, (d) different $\sigma_{e0}^2, \sigma_{e1}^2$ relationship.

to a good approximation. The conditional expectations and variances are

$$\begin{aligned}
\mathbf{E}(\mathbf{X}|\mathbf{W}, Z = 0) &= \frac{1}{k_0} \begin{pmatrix} a_1 W_0 + a_2 W_1 + a_3 \\ a_2 W_0 + a_1 W_1 + a_4 \end{pmatrix} \\
\mathbf{E}(\mathbf{X}|\mathbf{W}, Z = 1) &= \frac{1}{k_1} \begin{pmatrix} a_5 W_0 + a_6 W_1 + a_7 \\ a_8 W_0 + a_9 W_1 + a_{10} \end{pmatrix} \\
\mathbf{V}(\mathbf{X}|\mathbf{W}, Z = 0) &= \frac{1}{k_0} \begin{pmatrix} v_1 & v_2 \\ v_2 & v_1 \end{pmatrix} \\
\mathbf{V}(\mathbf{X}|\mathbf{W}, Z = 1) &= \frac{1}{k_1} \begin{pmatrix} v_3 & v_4 \\ v_4 & v_5 \end{pmatrix}, \tag{2.7}
\end{aligned}$$

where

$$\begin{aligned}
k_0 &= (\sigma_0^2 + \sigma_{e0}^2)^2 - \rho_0^2 \sigma_0^4, \quad k_1 = (\sigma_0^2 + \sigma_{e0}^2)(\sigma_1^2 + \sigma_{e1}^2) - \rho_1^2 \sigma_0^2 \sigma_1^2, \\
a_1 &= \sigma_0^2[(1 - \rho_0^2)\sigma_0^2 + \sigma_{e0}^2], \quad a_2 = \rho_0 \sigma_0^2 \sigma_{e0}^2, \quad a_3 = \sigma_{e0}^2(\sigma_0^2 + \sigma_{e0}^2)\mu_0 - \rho_0 \sigma_0^2 \sigma_{e0}^2 \mu_1, \\
a_4 &= -\rho_0 \sigma_0^2 \sigma_{e0}^2 \mu_0 + \sigma_{e0}^2(\sigma_0^2 + \sigma_{e0}^2)\mu_1, \quad a_5 = \sigma_0^2[(1 - \rho_1^2)\sigma_1^2 + \sigma_{e1}^2], \quad a_6 = \rho_1 \sigma_0 \sigma_1 \sigma_{e0}^2, \\
a_7 &= \sigma_{e0}^2(\sigma_1^2 + \sigma_{e1}^2)\mu_0 - \rho_1 \sigma_0 \sigma_1 \sigma_{e0}^2(\mu_1 + d), \quad a_8 = \rho_1 \sigma_0 \sigma_1 \sigma_{e1}^2, \\
a_9 &= \sigma_1^2[(1 - \rho_1^2)\sigma_0^2 + \sigma_{e0}^2], \quad a_{10} = -\rho_1 \sigma_0 \sigma_1 \sigma_{e1}^2 \mu_0 + \sigma_{e1}^2(\sigma_0^2 + \sigma_{e0}^2)(\mu_1 + d), \\
v_1 &= \sigma_0^2 \sigma_{e0}^2[(1 - \rho_0^2)\sigma_0^2 + \sigma_{e0}^2], \quad v_2 = \rho_0 \sigma_0^2 \sigma_{e0}^4, \\
v_3 &= \sigma_0^2 \sigma_{e0}^2[(1 - \rho_1^2)\sigma_1^2 + \sigma_{e1}^2], \quad v_4 = \rho_1 \sigma_0 \sigma_1 \sigma_{e0}^2 \sigma_{e1}^2, \quad v_5 = \sigma_1^2 \sigma_{e1}^2[(1 - \rho_1^2)\sigma_0^2 + \sigma_{e0}^2].
\end{aligned}$$

Since conditional means and variances are different for the two treatment groups, when plugging these conditional means and variances into the induced hazard (2.6), $\lambda(t; \mathbf{W}, Z)$ involves interactions between \mathbf{W} and Z :

$$\lambda(t; \mathbf{W}, Z) = \lambda_2(t) \exp(b_0 W_0 + b_1 W_1 + b_2 Z + b_3 W_0 Z + b_4 W_1 Z), \tag{2.8}$$

and the corresponding partial likelihood with $\mathbf{b} = (b_0, b_1, b_2, b_3, b_4)$ becomes

$$PL(\mathbf{b}) = \prod_{i=1}^k \frac{\exp[b_0 W_{i0} + b_1 W_{i1} + b_2 Z_i + b_3 W_{i0} Z_i + b_4 W_{i1} Z_i]}{\sum_{j \in R(t_i)} \exp[b_0 W_{j0} + b_1 W_{j1} + b_2 Z_j + b_3 W_{j0} Z_j + b_4 W_{j1} Z_j]}. \tag{2.9}$$

Comparing expression (2.8) with (2.6) after substituting (2.7) into it, we get the following correspondence between \mathbf{b} and $\boldsymbol{\beta}$:

$$\begin{cases} b_0 = \frac{1}{k_0}(a_1\beta_0 + a_2\beta_1) \\ b_1 = \frac{1}{k_0}(a_2\beta_0 + a_1\beta_1) \\ b_0 + b_3 = \frac{1}{k_1}(a_5\beta_0 + a_8\beta_1) \\ b_1 + b_4 = \frac{1}{k_1}(a_6\beta_0 + a_9\beta_1) \\ b_2 = \beta_Z + \frac{1}{k_1}(a_7\beta_0 + a_{10}\beta_1) + \frac{1}{2k_1}(v_3\beta_0^2 + 2v_4\beta_0\beta_1 + v_5\beta_1^2) \\ \quad - \frac{1}{k_0}(a_3\beta_0 + a_4\beta_1) - \frac{1}{2k_0}(v_1\beta_0^2 + 2v_2\beta_0\beta_1 + v_1\beta_1^2) \end{cases}$$

Partial likelihood (2.9) can be approximated by plugging $\mathbf{b} = \mathbf{b}(\boldsymbol{\beta})$ into it,

$$PL(\boldsymbol{\beta}) = \prod_{i=1}^k \frac{\exp[b_0(\boldsymbol{\beta})W_{0i} + b_1(\boldsymbol{\beta})W_{i1} + b_2(\boldsymbol{\beta})Z_i + b_3(\boldsymbol{\beta})W_{0i}Z_i + b_4(\boldsymbol{\beta})W_{i1}Z_i]}{\sum_{j \in R(t_i)} \exp[b_0(\boldsymbol{\beta})W_{0j} + b_1(\boldsymbol{\beta})W_{1j} + b_2(\boldsymbol{\beta})Z_j + b_3(\boldsymbol{\beta})W_{0j}Z_j + b_4(\boldsymbol{\beta})W_{1j}Z_j]}.$$

Maximizing $PL(\boldsymbol{\beta})$ through Newton-Raphson method gives a set of estimates of $\boldsymbol{\beta}$.

Note that the distribution parameters involved in the conditional means and variances, i.e., $\mathcal{A} = (\mu_0, \mu_1, d, \sigma_0^2, \sigma_1^2, \rho_0, \rho_1, \sigma_{e0}^2, \sigma_{e1}^2)$, are needed for this estimation procedure. We discuss how to estimate these parameters in Section 2.5.

Compared to the conventional regression calibration, which replaces \mathbf{X} by $\mathbf{E}(\mathbf{X}|\mathbf{W}, Z)$ to get hazard ratio parameter estimates, our approach 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 comparing to both the naive approach and the regular regression calibration approach.

2.4 Follow-up Time Calibration

While we expect the mean-variance regression calibration to provide better estimates compared to other approaches just mentioned, its performance may depend strongly on the rare disease assumption. In this section, we modify the mean-variance regression calibration in an attempt to reduce such sensitivity.

Without the rare disease assumption, the induced hazard is

$$\lambda(t; \mathbf{W}, Z) = \lambda_1(t) E[\exp(\boldsymbol{\beta}_X^T \mathbf{X} + \beta_Z Z) | \tilde{T} \geq t, \mathbf{W}, Z], \quad (2.10)$$

and the corresponding induced partial likelihood becomes

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

To compute the exact partial likelihood, we need to know the joint distribution of $(\mathbf{X}, \mathbf{W} | \tilde{T} \geq t, Z)$ at each failure time $t_i, i = 1, 2, \dots, k$. But when the number of failures is large, it is computationally intensive to calibrate at every failure time. With the mean-variance regression calibration, we assume that the conditional distribution of $(\mathbf{X}, \mathbf{W} | \tilde{T} \geq t, Z)$ is constant over time t . Hence only one calibration is needed at the beginning of the study. But results may be quite sensitive to this assumption. There can be remaining biases in the treatment effect estimate due to differential changes in the covariate distribution over time between treatment arms. One approach to avoid the two extreme solutions is to do the following: first, divide time 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 within each time interval. That is, we assume covariate distribution is constant within each time interval, but could differ between intervals. At the beginning of each interval, we use all subjects at risk at that time point to calibrate. By changing the number of intervals, we can control how accurate and fast our estimate is. When $L = 1$, only one calibration is done at the beginning of the study. This is the mean-variance regression calibration special case. If $L = k + 1$ and let $I_1 = 0, I_{i+1} = t_i, i = 1, 2, \dots, k$, calibration is done at the beginning of the study and at each failure time. This corresponds to the special case of risk set calibration.

Mathematically, we approximate the partial likelihood as

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

If we further assume that $(\mathbf{X}, \mathbf{U} | \tilde{T} \geq I_l, Z)$ is jointly normal with similar mean and variance structure as at the beginning of the study, but allow the distribution parameters \mathcal{A} to change

with time ($\mathcal{A}(t)$), then

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{U} \end{pmatrix} | \tilde{T} \geq I_l, Z \sim N\left(\begin{pmatrix} \mathbf{M}_Z(I_l) \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_Z(I_l) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}_Z(I_l) \end{pmatrix}\right).$$

Now the conditional distribution becomes

$$\mathbf{X} | \tilde{T} \geq I_l; \mathbf{W}, Z \sim N(\mathbf{E}(\mathbf{X} | I_l, \mathbf{W}, Z), \mathbf{V}(\mathbf{X} | I_l, \mathbf{W}, Z)),$$

and the induced hazard further simplifies to

$$PL(\boldsymbol{\beta}) = \prod_{l=1}^L \prod_{i: t_i \in [I_l, I_{l+1})} \frac{\exp\{\beta_Z Z_i + \boldsymbol{\beta}_X^T \mathbf{E}(\mathbf{X}_i | I_l, \mathbf{W}_i, Z_i) + \frac{1}{2} \boldsymbol{\beta}_X^T \mathbf{V}(\mathbf{X}_i | I_l, \mathbf{W}_i, Z_i) \boldsymbol{\beta}_X\}}{\sum_{j \in R(t_i)} \exp\{\beta_Z Z_j + \boldsymbol{\beta}_X^T \mathbf{E}(\mathbf{X}_j | I_l, \mathbf{W}_j, Z_j) + \frac{1}{2} \boldsymbol{\beta}_X^T \mathbf{V}(\mathbf{X}_j | I_l, \mathbf{W}_j, Z_j) \boldsymbol{\beta}_X\}}. \quad (2.11)$$

If the joint distribution ($\mathbf{X}, \mathbf{U} | \tilde{T} \geq I_l, Z$) is not normal, equation (2.11) becomes a second-order Taylor approximation. With this partial likelihood, we can first work out the conditional means and variances of \mathbf{X} at each time interval cutoff point, and then plug them into the partial likelihood. Maximizing this partial likelihood gives the parameter estimates. Dividing time into shorter intervals may lead to a less biased maximum partial likelihood estimate. However, this also increases computational time and may increase hazard ratio parameter variances. We use methods similar to those used in the mean-variance regression calibration to estimate the conditional means $\mathbf{E}(\mathbf{X} | I_l, \mathbf{W}, Z)$ and variances $\mathbf{V}(\mathbf{X} | I_l, \mathbf{W}, Z)$, but only consider subjects who are at risk at the beginning of each time interval. We also allow the common parameters in the two treatment arms to have different values to account for the potential differential changes in the covariate distributions. The procedure of estimating $\mathbf{E}(\mathbf{X} | I_l, \mathbf{W}, Z)$ and $\mathbf{V}(\mathbf{X} | I_l, \mathbf{W}, Z)$ is discussed in detail in Section 2.5. When L is large, calibrations at later time intervals can be unstable, due to the fact that the number of subjects still at risk at the beginning of these intervals can be small. We do not recommend choosing a large L due to the increasing computation time and unstable performance at later intervals, even though theoretically it might result in smaller biases.

To sum up, this method relaxes the assumption that the covariate distribution does not change over time in the mean-variance regression calibration. We refer to this method as a follow-up time calibration (FUC). Performance of this method is expected to be less

Component	Assumptions
$\mu(Z_i, t_j)$	$\mu(Z_i, t_j) = c_0 + c_1 Z_i + c_2 t_j + c_3 Z_i t_j$
$b_i(Z_i, t_j)$	$b_i(Z_i, t_j) \sim N(0, \sigma_{b0}^2)$ if $Z_i t_j = 0$ $b_i(Z_i, t_j) \sim N(0, \sigma_{b1}^2)$ if $Z_i t_j \neq 0$ $cor(b_i(Z_i, t_j), b_i(Z_i, t_k)) = \rho_b$ if $Z_i t_j = 0, Z_i t_k \neq 0$ $cor(b_i(Z_i, t_j), b_i(Z_i, t_k)) = 1$ if $Z_i t_j \neq 0, Z_i t_k \neq 0$
$S_i(Z_i, t_j)$	$S_i(Z_i, t_j) \sim N(0, \sigma_{s0}^2)$ if $Z_i t_j = 0$ $S_i(Z_i, t_j) \sim N(0, \sigma_{s1}^2)$ if $Z_i t_j \neq 0$ $cor(S_i(Z_i, t_j), S_i(Z_i, t_k)) = \rho_{s0}^{ t_j - t_k }$ if $Z_i = 0$ $cor(S_i(Z_i, t_j), S_i(Z_i, t_k)) = \rho_{s1}^{ t_j - t_k }$ if $Z_i \neq 0$
ϵ_{ij}	$\epsilon \sim N(0, \sigma_\epsilon^2)$

Table 2.1: Model assumptions on each component of the biomarker process model.

sensitive to the rare disease assumption. Also this method is flexible in that we can control the number of calibrations (L) to get both reliable and fast estimates.

2.5 Biomarker Process Modeling

For both the mean-variance regression calibration and the follow-up time calibration, distribution parameters $\mathcal{A} = (\mu_0, \mu_1, d, \sigma_0^2, \sigma_1^2, \rho_0, \rho_1, \sigma_{e0}^2, \sigma_{e1}^2)$ are needed in the calibration procedure. We discuss how to estimate these parameters from available data in this section.

For the mean-variance regression calibration, the mean parameters (μ_0, μ_1, d) can be estimated easily as sample means:

$$\hat{\mu}_0 = \frac{1}{n} \sum_{i=1}^n W_{i0}, \quad \hat{\mu}_1 = \frac{1}{n_0} \sum_{i:Z_i=0} W_{i1}, \quad \hat{d} = \frac{1}{n_1} \sum_{i:Z_i=1} W_{i1} - \hat{\mu}_1.$$

For the variance part, if only two data points are available for each subject, there is not enough data to identify the 6 variance parameters $(\sigma_0^2, \sigma_1^2, \rho_0, \rho_1, \sigma_{e0}^2, \sigma_{e1}^2)$. Additional data is needed to estimate these parameters.

If there is a subset of subjects with more than 2 longitudinal measurements of the biomarker, they can be used to estimate the variance components as specified in the

biomarker process (1.9)

$$W_{ij} = \mu(Z_i, t_j) + b_i(Z_i, t_j) + S_i(Z_i, t_j) + \epsilon_{ij},$$

and thus we can estimate all the 6 variance parameters needed for calibration based on this model. We summarize the distribution assumptions of each component in Table 2.1. $b_i(Z_i, t_j)$, $S_i(Z_i, t_j)$ and ϵ_{ij} are independent of each other given (Z_i, t_j) , and they are also independent between subjects. With this model, we allow the population mean to change with both treatment, time and their interactions, and allow the distribution of $b_i(Z_i, t_j)$ and $S_i(Z_i, t_j)$ to depend on whether the subject is ever treated ($Z_i t_j = 0$ or $Z_i t_j \neq 0$). One can reduce the number of parameters by making more restricted assumptions, such as assuming treatment effect on population mean does not change with time ($c_2 = 0$), or $b_i(Z_i, t_j)$, $S_i(Z_i, t_j)$ distributions do not change with treatment ($\sigma_{b0}^2 = \sigma_{b1}^2, \sigma_{s0}^2 = \sigma_{s1}^2, \rho_b = 1, \rho_{s0} = \rho_{s1}$). The correspondences between $(\sigma_0^2, \sigma_1^2, \rho_0, \rho_1, \sigma_{e0}^2, \sigma_{e1}^2)$ and the parameters in Table 2.1 are

$$\begin{aligned} \sigma_0^2 &= \sigma_{b0}^2 + \sigma_{s0}^2 & \sigma_1^2 &= \sigma_{b1}^2 + \sigma_{s1}^2 & \sigma_{e0}^2 &= \sigma_{e1}^2 = \sigma_e^2 \\ \rho_0 &= \frac{\sigma_{b0}^2 + \rho_{s0}\sigma_{s0}^2}{\sigma_{b0}^2 + \sigma_{s0}^2} & \rho_1 &= \frac{\rho_b\sigma_{b0}\sigma_{b1} + \rho_{s1}\sigma_{s0}\sigma_{s1}}{\sqrt{(\sigma_{b0}^2 + \sigma_{s0}^2)(\sigma_{b1}^2 + \sigma_{s1}^2)}} \end{aligned}$$

If we further allow ϵ_{ij} distribution to depend on Z ($\epsilon|Z \sim N(0, \sigma_{eZ}^2)$), then $\sigma_{e0}^2 = \sigma_{e1}^2$ and $\sigma_{e1}^2 = \sigma_{e1}^2$. In this section, we restrict our discussion to $\sigma_{e0}^2 = \sigma_{e1}^2 = \sigma_e^2$.

To estimate parameters in Table 2.1, one may use the existing functions in standard software, such as the R `lme()` function, or use a full likelihood approach to get MLEs for more flexibility. For either approach, at least 4 measurements for each subject are needed for identifiability. $(\hat{\sigma}_0^2, \hat{\sigma}_1^2, \hat{\rho}_0, \hat{\rho}_1, \hat{\sigma}_e^2)$ estimated from the subset of subjects with 4 or more measurements can be used to substitute $(\sigma_0^2, \sigma_1^2, \rho_0, \rho_1, \sigma_e^2)$ in the calibration procedure and is expected to have good performance. In an expensive biomarker study, samples with more than 4 measurements are usually rare. Efficiency loss of excluding subjects used in biomarker process modeling from the failure time data analysis is usually small.

Up to now, we discuss using a subset of the study population to estimate variance parameters. If such a subset can be considered as a random sample of the study population,

we can assume its distribution parameters are exactly the same as those of the main study population, and we refer to this subset as an internal subset. However, such a subset may not always be available. A more practical case is that we need to borrow information from an external data set. If there is a similar biomarker study with subjects measured repeatedly, we can follow exactly the same steps as discussed for internal subset to estimate $(\sigma_0^2, \sigma_1^2, \rho_0, \rho_1, \sigma_e^2)$ from the external data set. However, since the external population may have somewhat different characteristics, we cannot use all the information. Usually, it is reasonable to assume (ρ_0, ρ_1) are similar across closely related populations. With $(\hat{\rho}_0, \hat{\rho}_1)$ estimated from the external data and $(\hat{\mu}_0, \hat{\mu}_1, \hat{d})$ estimated from sample means of the study population, we estimate $(\sigma_0^2, \sigma_1^2, \sigma_e^2)$ as

$$\begin{aligned}\hat{\sigma}_0^2 &= \frac{1}{\hat{\rho}_0(n_0 - 1)} \sum_{i:Z_i=0} (W_{i0} - \hat{\mu}_0)(W_{i1} - \hat{\mu}_1) \\ \hat{\sigma}_1^2 &= \left[\frac{1}{\hat{\rho}_1(n_1 - 1)} \sum_{i:Z_i=1} (W_{i0} - \hat{\mu}_0)(W_{i1} - \hat{\mu}_1 - \hat{d}) \right]^2 / \hat{\sigma}_0^2 \\ \hat{\sigma}_e^2 &= \frac{1}{2} \left\{ \frac{1}{n-1} \sum_{i=1}^n (W_{i0} - \hat{\mu}_0)^2 - \hat{\sigma}_0^2 \right\} + \frac{n_0}{2n} \left\{ \frac{1}{n_0 - 1} \sum_{i:Z_i=0} (W_{i1} - \hat{\mu}_1)^2 - \hat{\sigma}_0^2 \right\} \\ &\quad + \frac{n_1}{2n} \left\{ \frac{1}{n_1 - 1} \sum_{i:Z_i=1} [W_{i1} - \hat{\mu}_1 - \hat{d}]^2 - \hat{\sigma}_1^2 \right\}\end{aligned}$$

If $\sigma_{e0}^2 \neq \sigma_{e1}^2$, we can estimate σ_{e0}^2 and σ_{e1}^2 as following:

$$\begin{aligned}\hat{\sigma}_{e0}^2 &= \frac{n}{n + n_0} \left\{ \frac{1}{n-1} \sum_{i=1}^n (W_{i0} - \hat{\mu}_0)^2 - \sigma_0^2 \right\} + \frac{n_0}{n + n_0} \left\{ \frac{1}{n_0 - 1} \sum_{i:Z_i=0} (W_{i1} - \hat{\mu}_1)^2 - \sigma_0^2 \right\}, \\ \hat{\sigma}_{e1}^2 &= \frac{1}{n_1 - 1} \sum_{i:Z_i=1} (W_{i1} - \hat{\mu}_1)^2 - \sigma_1^2.\end{aligned}$$

If the internal data set cannot be viewed as a random sample of the study population (for example, subjects survive longer are more likely to have longitudinal measures), we can handle it similarly as an external data set. If neither internal data nor external data are available, a sensitivity study covering a range of (ρ_0, ρ_1) values is needed.

For the follow-up time calibration, the first step is to get an initial estimate of $\mathcal{A}(0)$ with exactly the same method as described for the mean-variance regression calibration. For the following time intervals, we allow all distribution parameters to change except for $\hat{\sigma}_e^2$ (or $\hat{\sigma}_{e0}^2$ and $\hat{\sigma}_{e1}^2$ if they differ). This is because subjects surviving longer may have

some specific characteristics, such as smaller \mathbf{X} if β s are positive. But all subjects should have the same measurement error distribution. Hence measurement error variance should be kept the same over time, regardless of their failure time. In addition, we allow other distribution parameters for the two treatment arms to differ. For example, for subjects at the beginning of study, no matter which treatment group they are in, their mean baseline biomarker values are equal (i.e., $E(X_0|Z = 0) = E(X_0|Z = 1) = \mu_0$). However, as time goes, due to differential treatment effect, it is possible that subjects at risk in treatment group tend to have different baseline biomarker values than those not treated. Hence, we should allow $\hat{\mu}_0$ at time t ($t \neq 0$) to differ between the two treatment arms. Thus, we now denote the distribution parameters at time t as $\mathcal{A}(t)$, which is defined as

$$\mathcal{A}(t) = (\sigma_e^2, \mu_0(0, t), \mu_1(0, t), \sigma_0^2(0, t), \rho_0(0, t), \\ \mu_0(1, t), (\mu_1 + d)(1, t), \sigma_0^2(1, t), \sigma_1^2(1, t), \rho_1(1, t))$$

This definition is also valid at $t = 0$ with restriction that $\mu_0(0, t) = \mu_0(1, t), \sigma_0^2(0, t) = \sigma_0^2(1, t)$. For any $t \in [I_l, I_{l+1}), l \geq 1$, suppose n_l subjects are at risk at the beginning of the interval I_l , among which n_{0l} are in the control group and n_{1l} are in the treatment group, we have the following estimates of $\mathcal{A}(I_l)$:

$$\begin{aligned} \hat{\mu}_0(0, I_l) &= \frac{1}{n_{0l}} \sum_{i \in R(I_l), Z_i=0} W_{i0}, & \hat{\mu}_1(0, I_l) &= \frac{1}{n_{0l}} \sum_{i \in R(I_l), Z_i=0} W_{i1}, \\ \hat{\sigma}_0^2(0, I_l) &= \frac{1}{2(n_{0l} - 1)} \sum_{i \in R(I_l), Z_i=0} \{[W_{i0} - \hat{\mu}_0(0, I_l)]^2 + [W_{i1} - \hat{\mu}_1(0, I_l)]^2\} - \hat{\sigma}_e^2 \\ \hat{\rho}_0(0, I_l) &= \frac{1}{(n_{0l} - 1)\hat{\sigma}_0^2(0, I_l)} \sum_{i \in R(I_l), Z_i=0} [W_{i0} - \hat{\mu}_0(0, I_l)][W_{i1} - \hat{\mu}_1(0, I_l)]; \\ \hat{\mu}_0(1, I_l) &= \frac{1}{n_{1l}} \sum_{i \in R(I_l), Z_i=1} W_{i0}, & \widehat{\mu_1 + d}(1, I_l) &= \frac{1}{n_{1l}} \sum_{i \in R(I_l), Z_i=1} W_{i1}, \\ \hat{\sigma}_0^2(1, I_l) &= \frac{1}{n_{1l} - 1} \sum_{i \in R(I_l), Z_i=1} [W_{i0} - \hat{\mu}_0(1, I_l)]^2 - \hat{\sigma}_e^2 \\ \hat{\sigma}_1^2(1, I_l) &= \frac{1}{n_{1l} - 1} \sum_{i \in R(I_l), Z_i=1} [W_{i1} - \widehat{\mu_1 + d}(1, I_l)]^2 - \hat{\sigma}_e^2 \\ \hat{\rho}_0(1, I_l) &= \frac{1}{(n_{1l} - 1)\sqrt{\hat{\sigma}_0^2(1, I_l)\hat{\sigma}_1^2(1, I_l)}} \sum_{i \in R(I_l), Z_i=1} [W_{i0} - \hat{\mu}_0(1, I_l)][W_{i1} - \widehat{\mu_1 + d}(1, I_l)] \end{aligned}$$

and

$$\begin{aligned}\hat{\mathbf{M}}_0(I_l) &= \begin{pmatrix} \hat{\mu}_0(0, I_l) \\ \hat{\mu}_1(0, I_l) \end{pmatrix}, & \hat{\mathbf{\Sigma}}_0(I_l) &= \begin{pmatrix} \hat{\sigma}_0^2(0, I_l) & \hat{\rho}_0(0, I_l)\hat{\sigma}_0^2(0, I_l) \\ \hat{\rho}_0(0, I_l)\hat{\sigma}_0^2(0, I_l) & \hat{\sigma}_0^2(0, I_l) \end{pmatrix}, \\ \hat{\mathbf{M}}_1(I_l) &= \begin{pmatrix} \hat{\mu}_0(1, I_l) \\ \widehat{\mu_1 + d}(1, I_l) \end{pmatrix} & \hat{\mathbf{\Sigma}}_1(I_l) &= \begin{pmatrix} \hat{\sigma}_0^2(1, I_l) & \hat{\rho}_1(1, I_l)\hat{\sigma}_0(1, I_l)\hat{\sigma}_1(1, I_l) \\ \hat{\rho}_1(1, I_l)\hat{\sigma}_0(1, I_l)\hat{\sigma}_1(1, I_l) & \hat{\sigma}_1^2(1, I_l) \end{pmatrix}, \\ \hat{\mathbf{\Delta}}_0 &= \hat{\mathbf{\Delta}}_1 = \hat{\sigma}_e^2 I_{2 \times 2}.\end{aligned}$$

Conditional means and variances at each time interval can be easily computed from these parameters. For the following simulation studies and asymptotic distributions development, we always assume an external data set is available to provide (ρ_0, ρ_1) estimates.

2.6 Simulation Studies

The mean-variance regression calibration (MVC) and the follow-up time calibration (FUC) based on the induced hazard are simple correction methods for classical measurement error in mediation analysis. In this section, we study the performance of both approaches and compare them with the naive approach. We aim to investigate two situations:

1. when distribution parameters (ρ_0, ρ_1) are correctly specified;
2. when distribution parameters (ρ_0, ρ_1) are estimated from an external data set.

Within each situation, we investigate the biases and standard errors with MVC and FUC. There are four main purposes of this simulation study:

1. To compare the performance of our proposed MVC and FUC methods with the naive approach.
2. To investigate the sensitivity of MVC and FUC to (ρ_0, ρ_1) specification.
3. To investigate the sensitivity of MVC and FUC to censoring mechanisms and the rare disease assumption.
4. To investigate the sensitivity of MVC and FUC to the normality assumption.

2.6.1 Simulation Studies with Distribution Parameters (ρ_0, ρ_1) Known

For this part, we assume that distribution parameters (ρ_0, ρ_1) are known. We fix $\mu_0 = \mu_1 = 0$, $d = 0.5$, $\sigma_0^2 = 1$, $\rho_0 = 0.95$, $\rho_1 = 0.9$, and let the underlying $\lambda_0(t) = 1$, $\beta_0 = \beta_1 = 1$. Half of the cohort members are randomized to active treatment, and the rest are in the control group. For now, we assume a simple censoring mechanism that all subjects are censored at the end of study (C_{end}) and censoring probability is 95%. When measurement error distribution does not differ in the two treatment arms ($\sigma_{e0}^2 = \sigma_{e1}^2 = \sigma_e^2$), the biases are summarized in Table 2.2 and standard errors in Table 2.3. Sample size is 10,000, and results are based on 1,000 simulation replicates. We compare estimates from five scenarios: model with true \mathbf{X} known (True), the naive approach with \mathbf{W} replacing \mathbf{X} (Naive), the mean-variance regression calibration (MVC), the follow-up time calibration with $L = 4$ (FUC4) and $L = 8$ (FUC8). Intervals are defined 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.

Generally speaking, both MVC and FUC have the ability to reduce the biases of β_0 , β_1 and β_Z . If we consider MVC as a special case of FUC with $L = 1$, we find that the biases of β_0 and β_1 decrease as L increases. The improvements from MVC to FUC4 are larger than those from FUC4 to FUC8. For the naive biases of β_Z , as observed in Figure 2.2, they are much smaller than those in β_0 and β_1 . In most cases, MVC reduces the bias of β_Z to a reasonable range. FUC is expected to further reduce the biases of β_Z . It is true in most scenarios. However, in some combinations of parameters, such as $\beta_Z = 0$, $\sigma_1^2 = 1$, $\sigma_e^2 = 0.5$, larger L is associated with larger biases for β_Z , although still much smaller than the bias from the naive approach. When taking a closer look, we find that with positive β_Z , MVC usually provides negative biases for β_Z , and FUC can make the biases closer to 0. Once it crosses 0 and becomes positive, more calibrations result in bigger biases. If MVC already provides positive bias for β_Z , FUC usually results in larger biases. With negative β_Z , same trend is observed but with opposite direction. We consider this as an over-correction. Over-correction is very likely to be caused by the unstable calibrations with the later intervals. In some scenarios, biases of β_Z are quite small with naive approach, such as in $\beta_Z = \log(1.5)$ and $\sigma_1^2 = 0.8$. This might be due to the fact that naive bias of β_Z changes with both β_Z

bias($\times 10^3$)			$\beta_Z = \log(0.8)$			$\beta_Z = 0$			$\beta_Z = \log(1.2)$			$\beta_Z = \log(1.5)$		
σ_1^2	σ_e^2	meth.	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$
0.8	0.2	True	2.6	0.3	1.0	4.2	-1.2	1.7	3.8	-1.1	1.6	2.4	-0.2	0.8
		Naive	-118.9	-157.5	32.1	-117.4	-160.0	23.3	-117.0	-161.5	15.5	-117.4	-162.9	5.7
		MVC	-36.7	-49.1	15.6	-37.0	-48.6	6.1	-39.1	-46.5	-2.6	-43.8	-42.4	-13.8
		FUC4	-32.1	-45.1	14.2	-31.9	-45.0	8.0	-33.5	-43.4	1.9	-37.6	-39.7	-6.1
		FUC8	-31.4	-44.3	14.1	-30.9	-44.5	8.4	-32.6	-42.8	2.8	-36.4	-39.4	-4.7
	0.5	True	2.6	0.3	1.0	4.2	-1.2	1.7	3.8	-1.1	1.6	2.4	-0.2	0.8
		Naive	-260.4	-298.6	62.0	-259.4	-301.2	44.9	-259.1	-303.1	30.5	-259.7	-305.2	12.6
		MVC	-58.6	-103.5	33.4	-64.8	-96.3	13.8	-71.9	-88.8	-3.2	-82.8	-78.0	-24.2
		FUC4	-51.7	-91.8	29.0	-55.8	-86.7	16.2	-61.0	-81.0	4.9	-70.0	-72.1	-9.9
		FUC8	-50.8	-89.4	28.5	-54.1	-85.1	17.1	-59.3	-79.5	6.6	-67.3	-71.5	-6.8
1	0.2	True	3.8	-1.0	1.9	2.4	0.6	1.3	1.1	1.5	0.1	0.9	0.1	0.1
		Naive	-133.4	-136.1	49.8	-133.8	-136.0	39.9	-134.9	-136.2	30.9	-136.1	-136.8	20.7
		MVC	-39.5	-49.4	13.6	-40.7	-48.5	3.3	-42.4	-48.0	-5.9	-44.0	-48.2	-16.0
		FUC4	-34.9	-45.6	14.0	-35.9	-44.9	7.0	-37.2	-44.5	0.4	-38.5	-44.9	-6.8
		FUC8	-33.7	-45.3	14.4	-34.8	-44.5	7.9	-36.1	-44.1	1.6	-37.4	-44.4	-5.1
	0.5	True	3.8	-1.0	1.9	2.4	0.6	1.3	1.1	1.5	0.1	0.9	0.1	0.1
		Naive	-272.3	-274.6	101.7	-272.8	-275.1	83.3	-273.9	-275.6	67.4	-275.2	-276.5	49.0
		MVC	-70.8	-96.6	26.9	-71.8	-96.3	8.5	-73.9	-95.7	-7.4	-76.5	-95.2	-26.0
		FUC4	-62.6	-86.9	26.2	-62.5	-87.5	14.6	-63.9	-87.5	3.9	-65.6	-87.5	-8.6
		FUC8	-60.5	-85.9	26.9	-60.8	-86.0	16.2	-61.8	-86.2	6.3	-63.9	-85.9	-5.4
1.2	0.2	True	2.3	0.2	1.5	1.4	0.9	0.1	0.8	1.2	-1.3	-1.4	2.7	-1.8
		Naive	-148.3	-117.1	60.0	-149.5	-116.6	48.2	-151.6	-115.2	38.1	-153.3	-114.7	28.4
		MVC	-40.1	-52.2	13.3	-38.7	-54.6	3.6	-41.5	-52.8	-6.7	-42.1	-53.7	-15.7
		FUC4	-35.4	-48.7	15.6	-34.1	-50.8	8.7	-36.6	-49.2	0.9	-37.0	-50.1	-5.0
		FUC8	-34.4	-48.2	16.2	-32.9	-50.5	9.9	-35.6	-48.6	2.3	-36.1	-49.4	-3.2
	0.5	True	2.3	0.2	1.5	1.4	0.9	0.1	0.8	1.2	-1.3	-1.4	2.7	-1.8
		Naive	-285.2	-253.2	131.4	-286.5	-252.7	110.9	-288.5	-251.7	93.8	-290.2	-251.5	75.8
		MVC	-65.7	-107.9	29.9	-60.3	-114.6	14.2	-62.2	-114.4	-2.1	-60.8	-118.1	-17.8
		FUC4	-56.5	-99.6	33.2	-52.1	-105.2	23.0	-53.3	-105.3	11.8	-51.8	-108.6	2.0
		FUC8	-55.1	-97.9	34.0	-50.1	-104.0	25.2	-52.0	-103.4	14.4	-50.6	-106.5	5.4

Table 2.2: Summary of estimated biases ($\times 10^3$) in parameter estimates ($\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_Z$) with 95% censoring at the end of study, with the uncorrelated measurement errors; with σ_1^2, σ_e^2 and true β_Z value changing.

SE			$\beta_Z = \log(0.8)$			$\beta_Z = 0$			$\beta_Z = \log(1.2)$			$\beta_Z = \log(1.5)$			
σ_1^2	σ_e^2	meth.	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	
0.8	0.2	True	0.117	0.125	0.097	0.115	0.124	0.098	0.114	0.123	0.098	0.112	0.122	0.098	
		Naive	0.065	0.068	0.096	0.064	0.067	0.096	0.064	0.068	0.096	0.063	0.067	0.096	
		MVC	0.216	0.229	0.120	0.211	0.226	0.118	0.208	0.224	0.117	0.204	0.220	0.116	
		FUC4	0.216	0.228	0.121	0.211	0.225	0.120	0.207	0.222	0.119	0.202	0.217	0.117	
		FUC8	0.216	0.228	0.121	0.211	0.225	0.120	0.208	0.222	0.119	0.202	0.217	0.118	
	0.5	True	0.117	0.125	0.097	0.115	0.124	0.098	0.114	0.123	0.098	0.112	0.122	0.098	
		Naive	0.048	0.049	0.096	0.047	0.049	0.096	0.046	0.049	0.096	0.046	0.049	0.097	
		MVC	0.306	0.323	0.144	0.300	0.320	0.142	0.295	0.316	0.140	0.290	0.312	0.138	
		FUC4	0.303	0.320	0.147	0.297	0.316	0.145	0.291	0.311	0.143	0.285	0.305	0.141	
		FUC8	0.304	0.320	0.148	0.298	0.317	0.147	0.293	0.312	0.144	0.285	0.305	0.142	
	1	0.2	True	0.116	0.118	0.105	0.114	0.115	0.105	0.112	0.113	0.106	0.109	0.111	0.106
			Naive	0.066	0.067	0.098	0.064	0.066	0.099	0.064	0.065	0.099	0.065	0.065	0.099
			MVC	0.213	0.214	0.138	0.206	0.208	0.136	0.203	0.204	0.135	0.202	0.203	0.136
			FUC4	0.212	0.212	0.138	0.205	0.206	0.137	0.201	0.201	0.136	0.200	0.200	0.137
			FUC8	0.212	0.213	0.139	0.205	0.206	0.138	0.201	0.202	0.137	0.200	0.200	0.137
0.5		True	0.116	0.118	0.105	0.114	0.115	0.105	0.112	0.113	0.106	0.109	0.111	0.106	
		Naive	0.048	0.049	0.096	0.047	0.048	0.098	0.047	0.048	0.098	0.047	0.048	0.099	
		MVC	0.309	0.310	0.175	0.300	0.302	0.173	0.296	0.298	0.171	0.296	0.297	0.171	
		FUC4	0.302	0.302	0.176	0.294	0.294	0.174	0.289	0.289	0.172	0.289	0.288	0.172	
		FUC8	0.304	0.304	0.178	0.295	0.295	0.175	0.290	0.290	0.173	0.290	0.289	0.173	
1.2		0.2	True	0.114	0.111	0.114	0.111	0.108	0.113	0.111	0.106	0.112	0.109	0.103	0.113
			Naive	0.066	0.064	0.101	0.065	0.064	0.101	0.066	0.064	0.100	0.067	0.064	0.103
			MVC	0.201	0.192	0.153	0.199	0.190	0.151	0.196	0.186	0.149	0.196	0.185	0.151
			FUC4	0.200	0.190	0.153	0.197	0.187	0.151	0.195	0.183	0.149	0.195	0.182	0.151
			FUC8	0.200	0.191	0.153	0.197	0.187	0.151	0.195	0.184	0.150	0.195	0.183	0.152
	0.5	True	0.114	0.111	0.114	0.111	0.108	0.113	0.111	0.106	0.112	0.109	0.103	0.113	
		Naive	0.048	0.047	0.099	0.048	0.047	0.099	0.048	0.047	0.098	0.049	0.048	0.101	
		MVC	0.289	0.276	0.195	0.287	0.273	0.194	0.282	0.268	0.190	0.285	0.269	0.193	
		FUC4	0.282	0.268	0.195	0.279	0.264	0.192	0.275	0.259	0.190	0.277	0.260	0.192	
		FUC8	0.284	0.269	0.196	0.280	0.265	0.193	0.277	0.260	0.191	0.279	0.261	0.194	

Table 2.3: Summary of standard errors in parameter estimates $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_Z)$ with 95% censoring at the end of study, with the uncorrelated measurement errors; with σ_1^2, σ_e^2 and true β_Z value changing.

and σ_1^2 , and the two trends may cancel out with each other. In these situations, biases of β_Z with MVC can be larger than those with naive approach. However, with FUC, biases of β_Z reduce to reasonable ranges.

Comparing to the naive approach, standard errors of all the three parameters are larger with MVC and FUC. As L increases, standard error of β_Z increases slightly.

To further evaluate the performances of MVC and FUC under other scenarios, we perform a simulation study with different d values. Biases are summarized in Table 2.4. Here we fix $\sigma_1^2 = 1.2$ and other parameters as before, but we allow d to change from 0 to 1. d represents the mean change in biomarker value due to treatment. Measurement error variance is 0.5 for both treatment arms ($\sigma_{e0}^2 = \sigma_{e1}^2 = \sigma_e^2 = 0.5$).

Here we find that β_0 and β_1 biases with naive approach does not change much with d value. And these biases reduce as the number of calibration intervals L increases. For β_Z , as d increases, the biases with naive approach increase significantly. With relatively large d ($d = 0.5, 1$), MVC provides much better β_Z estimates compared to the naive approach. Again we observe that in some scenarios, FUC may over-correct β_Z . However, FUC with 2 to 4 intervals still have small biases for β_Z , while β_0 and β_1 biases are smaller than those with MVC. To avoid over-correction, we do not recommend increasing the number of calibration intervals, L , beyond 4.

MVC and FUC can also be applied when measurement error distributions differ in the two treatment arms (i.e. $\sigma_{e0}^2 \neq \sigma_{e1}^2$). We now fix $\sigma_1^2 = 1.2$ and $d = 0.5$. Let $\sigma_{e0}^2 = 0.5$ and we vary the relationship between σ_{e0}^2 and σ_{e1}^2 . Results are summarized in Table 2.5. As σ_{e1}^2 gets larger, the naive biases of β_1 and β_Z increase while the bias of β_0 decreases. No matter whether measurement error distribution depends on the treatment or not, MVC and FUC reduce the biases in all the three parameters. This is due to the fact that these two calibration methods are both flexible enough to allow the measurement error distribution to differ between treatment arms. As the number of time intervals increases, biases of β_0 and β_1 decrease, while the bias of β_Z can be over-corrected by FUC with large L .

bias($\times 10^3$)		$\beta_Z = \log(0.8)$			$\beta_Z = 0$			$\beta_Z = \log(1.2)$			$\beta_Z = \log(1.5)$		
d	meth.	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$
0	True	2.6	0.3	1.2	1.5	0.4	1.8	2.0	0.6	2.0	2.2	-0.1	0.9
	Naive	-283.8	-254.5	48.1	-284.3	-254.0	29.2	-284.9	-253.3	13.5	-285.7	-253.1	-6.9
	MVC	-84.9	-88.2	15.0	-74.9	-98.4	-1.7	-69.1	-104.2	-16.0	-60.9	-113.4	-34.9
	FUC4	-74.1	-81.3	8.5	-65.4	-90.4	-2.0	-59.7	-96.2	-11.2	-52.2	-104.5	-23.8
	FUC8	-72.1	-80.3	7.8	-63.6	-89.3	-1.6	-58.5	-94.4	-10.0	-50.3	-103.3	-21.5
0.5	True	2.3	0.2	1.5	1.4	0.9	0.1	0.8	1.2	-1.3	-1.4	2.7	-1.8
	Naive	-285.2	-253.2	131.4	-286.5	-252.7	110.9	-288.5	-251.7	93.8	-290.2	-251.5	75.8
	MVC	-65.7	-107.9	29.9	-60.3	-114.6	14.2	-62.2	-114.4	-2.1	-60.8	-118.1	-17.8
	FUC4	-56.5	-99.6	33.2	-52.1	-105.2	23.0	-53.3	-105.3	11.8	-51.8	-108.6	2.0
	FUC8	-55.1	-97.9	34.0	-50.1	-104.0	25.2	-52.0	-103.4	14.4	-50.6	-106.5	5.4
1	True	0.3	1.6	-2.2	-1.5	2.3	-3.2	-1.7	2.4	-3.0	-0.2	1.9	-0.9
	Naive	-289.2	-251.5	211.8	-291.2	-251.6	193.4	-292.3	-251.9	179.4	-294.1	-251.7	163.1
	MVC	-61.2	-116.1	49.3	-61.4	-119.0	34.1	-59.4	-123.0	24.3	-59.9	-124.8	10.2
	FUC4	-52.5	-106.7	61.0	-52.4	-109.3	51.4	-50.2	-113.0	45.8	-50.5	-114.3	36.6
	FUC8	-51.5	-104.4	62.5	-51.4	-106.9	53.8	-49.3	-110.4	48.7	-49.7	-111.5	40.1

Table 2.4: Summary of estimated biases ($\times 10^3$) in parameter estimates ($\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_Z$) with 95% censoring at the end of study, with the uncorrelated measurement errors; with d and true β_Z value changing.

bias($\times 10^3$)		$\beta_Z = \log(0.8)$			$\beta_Z = 0$			$\beta_Z = \log(1.2)$			$\beta_Z = \log(1.5)$		
$\sigma_{e0}^2, \sigma_{e1}^2$	meth.	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$
$\sigma_{e1}^2 = \sigma_{e0}^2$	True	2.3	0.2	1.5	1.4	0.9	0.1	0.8	1.2	-1.3	-1.4	2.7	-1.8
	Naive	-285.2	-253.2	131.4	-286.5	-252.7	110.9	-288.5	-251.7	93.8	-290.2	-251.5	75.8
	MVC	-65.2	-104.1	36.3	-60.7	-109.6	20.0	-63.3	-108.4	3.2	-62.4	-111.1	-12.9
	FUC4	-59.4	-92.6	37.8	-58.2	-94.8	25.8	-62.7	-91.4	12.7	-64.1	-91.6	1.2
	FUC8	-58.4	-90.8	38.1	-57.1	-93.0	27.0	-62.4	-88.8	14.3	-64.0	-88.7	3.6
$\sigma_{e1}^2 = 1.5\sigma_{e0}^2$	True	2.3	0.2	1.5	1.4	0.9	0.1	0.8	1.2	-1.3	-1.4	2.7	-1.8
	Naive	-186.9	-428.3	152.8	-183.2	-437.0	137.5	-181.4	-443.1	124.0	-178.9	-450.8	109.8
	MVC	-53.3	-137.0	25.4	-51.0	-142.4	7.3	-56.1	-140.2	-11.5	-58.4	-141.8	-30.1
	FUC4	-46.4	-123.3	37.3	-47.0	-125.2	24.0	-53.9	-120.8	9.3	-58.3	-119.5	-4.2
	FUC8	-45.2	-121.1	39.2	-45.7	-123.0	27.0	-53.5	-117.6	12.7	-58.0	-116.0	0.1
$\sigma_{e1}^2 = 2\sigma_{e0}^2$	True	2.3	0.2	1.5	1.4	0.9	0.1	0.8	1.2	-1.3	-1.4	2.7	-1.8
	Nave	-125.6	-536.6	188.8	-120.4	-548.1	176.3	-117.4	-556.2	164.7	-113.9	-566.0	152.2
	MVC	-49.5	-150.3	19.3	-48.2	-155.5	0.4	-54.8	-152.7	-19.5	-58.8	-153.5	-39.4
	FUC4	-41.6	-136.0	36.4	-43.2	-137.6	22.5	-51.3	-132.4	7.1	-57.3	-130.2	-7.5
	FUC8	-40.1	-133.7	39.2	-41.6	-135.4	26.4	-50.7	-129.0	11.3	-56.8	-126.5	-2.3

Table 2.5: Summary of estimated biases ($\times 10^3$) in parameter estimates ($\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_Z$) with 95% censoring at the end of study, with the uncorrelated measurement errors. Measurement error variances differ in the two treatment arms, and true β_Z value changes.

2.6.2 Simulation studies with Distribution Parameters (ρ_0, ρ_1) Estimated

To investigate the robustness of the results to the distribution parameters specification, a simulation study with distribution parameters (ρ_0, ρ_1) estimated from an external data is conducted. The main part of this simulation is similar as described before. The only difference is that now we do not plug in the true distribution parameters (ρ_0, ρ_1) into the calibration procedure. Instead, we simulate the external data as described in Table 2.1 to estimate ρ_0 and ρ_1 . The calibration procedure uses the estimated $(\hat{\rho}_0, \hat{\rho}_1)$. With this simulation, we take into account variabilities from both the calibration procedure and the distribution parameter estimation procedure. We let $\sigma_1^2 = 1$, $\sigma_e^2 = 0.2$ and $\beta_Z = \log(1.5)$. The true values of (ρ_0, ρ_1) are $(0.95, 0.9)$. Biases and SEs are summarized in Table 2.6. The first column shows biases and SEs when the (ρ_0, ρ_1) are known. The second and third columns correspond to (ρ_0, ρ_1) estimated from external data sets with sample size 1,000 and 500. The means and SEs of the estimated ρ_0 are 0.946 (0.014), 0.943 (0.020) and the corresponding means and SEs of ρ_1 are 0.896 (0.021), 0.892 (0.029). As expected, the biases and SEs in the estimations of (ρ_0, ρ_1) increase as the sample size of the external data set decreases.

Compare to when (ρ_0, ρ_1) are known, both the biases and the SEs of β estimates are slightly larger when distribution parameters (ρ_0, ρ_1) are estimated. We still observe that both MVC and FUC reduces biases of all the three parameters in most scenarios, and biases decrease as the number of intervals increases. Hence, both methods are robust to distribution parameters specification, as long as (ρ_0, ρ_1) are reasonably specified.

When $(\hat{\rho}_0, \hat{\rho}_1)$ are used in the calibration procedure, biases of all the three parameters tend to be larger when $(\hat{\rho}_0, \hat{\rho}_1)$ are less precise. Standard errors have a small increasing trend as well. In Chapter 4, we provide the form of the asymptotic variance. Generally speaking, the biases and variability of β estimates are from two parts: one from estimating $(\hat{\rho}_0, \hat{\rho}_1)$, and another from the calibration procedure. Sample size of the external data set only changes the first part, and the contribution is generally small with the uncorrelated measurement error. No matter how accurate (ρ_0, ρ_1) estimates are, the abilities of MVC and FUC in reducing biases are preserved. We can expect conclusions from the previous

		(ρ_0, ρ_1) known			$(\hat{\rho}_0, \hat{\rho}_1), n_E = 1000$			$(\hat{\rho}_0, \hat{\rho}_1), n_E = 500$		
meth.		$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$
bias($\times 10^3$)	True	0.9	0.1	0.1	0.9	0.1	0.1	0.9	0.1	0.1
	Naive	-136.1	-136.8	20.7	-136.1	-136.8	20.7	-136.1	-136.8	20.7
	MVC	-44.0	-48.2	-16.0	-50.1	-46.6	-16.4	-55.9	-44.7	-16.9
	FUC4	-38.5	-44.9	-6.8	-41.3	-47.0	-6.2	-46.3	-46.3	-6.8
	FUC8	-37.4	-44.4	-5.1	-40.1	-46.7	-4.5	-44.8	-46.4	-5.1
SE	True	0.109	0.111	0.106	0.109	0.111	0.106	0.109	0.111	0.106
	Naive	0.065	0.065	0.099	0.065	0.065	0.099	0.065	0.065	0.099
	MVC	0.202	0.203	0.136	0.204	0.207	0.136	0.208	0.213	0.138
	FUC4	0.200	0.200	0.137	0.199	0.200	0.136	0.200	0.202	0.137
	FUC8	0.200	0.200	0.137	0.199	0.200	0.136	0.200	0.201	0.137

Table 2.6: Summary of estimated biases ($\times 10^3$) and SEs in parameter estimate $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_Z)$ with 95% censoring at the end of study, with the uncorrelated measurement errors. Distribution parameters (ρ_0, ρ_1) are known, estimated from external data sets with sample size 1,000 and 500.

simulation studies with (ρ_0, ρ_1) known to hold when these two parameters are estimated from either internal or external data sets.

2.6.3 Simulation Studies of Sensitivity on Censoring Mechanism

Now we investigate how censoring mechanism and censoring probability influence the performance of our proposed methods. We look at three different censoring mechanisms. The first is the censoring mechanism that we have used so far: all subjects are censored at a fixed time point C_{end} when study ends (Censor I). The second mechanism is that censoring time $C_i, i = 1, 2, \dots, n$, follows an exponential distribution with a fixed rate λ_C (Censor II). This censoring mechanism may have longer time span than the first censoring mechanism. The last is that the censoring time follows an exponential distribution within each treatment group, but the censoring rates are different between the two groups (Censor III). Now we let $\sigma_0^2 = 1, \sigma_0^2 = 1.2, \sigma_{e0}^2 = \sigma_{e1}^2 = \sigma_e^2 = 0.5$, and $\beta_Z = \log(1.5)$. (ρ_0, ρ_1) is assumed to be known. We vary the overall censoring probability from 50% to 99%. For the third censoring mechanism, we let the censoring probability in the treatment group to be 5% higher than

that in the control group. For example, when the overall censoring probability is 95%, we let the censoring probability in the control group to be 92.5% and that in the treatment group to be 97.5%. Since the study subjects are randomized to the two treatment arms with equal probability, the overall censoring probability is 95%. We do not use this censoring mechanism when the overall censoring probability is 99%, since it is not possible to have a 5% difference in the two censoring probabilities. This censoring mechanism may happen when the treatment causes some side effects which may result in treated subjects being more likely to quit the study early. In this setting, calibration intervals in the follow-up time calibration needs to be defined carefully. Instead of using the quantiles of all failure times, we now define intervals in terms of the quantiles of failure times in the treatment group. That is, we define intervals as $[0, Q_{T,1/L,Z=1}), [Q_{T,1/L,Z=1}, Q_{T,2/L,Z=1}), \dots, [Q_{T,(L-1)/L,Z=1}, \infty)$, where $Q_{T,k/L,Z=1}$ is the k^{th} L-quantile of failure times in the treatment group. We use this definition to ensure that there are enough treated subjects within each calibration. This definition of intervals may change the results slightly. For all the simulations, we fix the total expected number of failures to be 500 and vary cohort size with different censoring probabilities. Results are summarized in Table 2.7.

Biases of β_0 and β_1 are similar among the three censoring mechanisms. Naive biases of β_Z with the third censoring mechanism is slightly larger than those of the other two mechanisms, especially when censoring probability is large. Differences might be due to censoring is differential for the two treatment groups, as well as due to different interval definitions. As censoring probability increases, biases of β_0 and β_1 are decreasing with all the methods.

For β_Z , biases with naive approach are increasing as censoring probability increases, while for MVC and FUC, biases of β_Z decrease. When censoring probability is low, MVC does not have the ability to recover the true β_Z . It provides a even bigger bias than the naive approach with censoring probability below 95%. Using FUC largely reduces the bias in β_Z , and over-correction does not present. When censoring probability is relatively high, MVC starts to show the ability to reduce β_Z bias. Hence, we do not recommend using MVC when censoring probability is below 95%, since the performance of this method highly depends on the rare disease assumption. Using FUC with a few calibrations (such as $L = 4$) has

more robust performance to the censoring probability, while the computational time is not too long.

2.6.4 *Simulation Studies of Violation to Normality Assumption*

Up to now, we assume both the true biomarker values \mathbf{X} and the measurement errors \mathbf{U} are normally distributed. Our proposed MVC and FUC make use of this assumption in calibrating the conditional means and variances, and in approximating the induced hazard. In this subsection, we investigate the robustness of MVC and FUC to non-normality. We assume \mathbf{X} is still normally distributed, while \mathbf{W} are non-normal. This corresponds to the situation that in practice, we may be able to find a suitable transformation for \mathbf{X} to be normal, but the corresponding \mathbf{U} may be non-normal under this transformation. To generate non-normal multivariate distributions, we use methods described in Vale and Maurelli (1983). Basically, this method uses linear, quadratic and cubic forms of a multivariate normal random variable to generate non-normal distributions with specific correlation structure and marginal mean, variance, skewness and kurtosis. In some extreme combinations of these four moments, this method may fail.

In this simulation, we investigate four skewness and kurtosis combinations: $(0, 0)$, $(0, 6/5)$, $(1, 3)$, $(2, 7)$. Here, $(0, 0)$ corresponds to the normal case. $(0, 6/5)$ is chosen to resemble a logistic distribution, which is close to a normal distribution but has heavier tails. There are several versions of multivariate logistic distributions. But all of them have some specific correlation structures, which make them not flexible to use in our simulation. The combination of $(1, 3)$ and $(2, 7)$ are chosen to allow the distributions to be both skewed and with heavy tails. We consider $(1, 3)$ to be moderate non-normality and $(2, 7)$ to be severely non-normality. We focus on the first rows of Table 2.3 and 2.4 that $\sigma_1^2 = 0.8$ and $\sigma_e^2 = 0.2$. Biases are summarized in Table 2.8.

As the measurement error distribution becomes further away from a normal distribution, the naive approach tends to have larger biases in all the three parameter estimates, especially in β_0 and β_1 . In all the scenarios, both MVC and FUC show the ability to reduce naive biases. Even with severe violation of normality, MVC reduces the biases of naive approach

bias($\times 10^3$)		Censor I			Censor II			Censor III		
P(censor)	method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$
0.5	True	3.5	1.2	-0.6	5.4	-0.4	0.4	-0.7	6.6	-1.6
	Naive	-387.3	-359.3	-9.8	-388.7	-356.3	-11.1	-391.5	-364.6	-12.8
	MVC	-204.2	-223.8	-80.0	-219.2	-205.8	-89.9	-232.7	-207.9	-98.2
	FUC4	-157.4	-174.9	-25.1	-163.6	-165.4	-30.8	-181.1	-162.0	-42.0
	FUC8	-150.5	-162.3	-13.9	-157.6	-151.3	-19.0	-170.8	-151.1	-29.2
0.8	True	3.8	1.6	-1.0	1.3	4.8	-6.4	-3.0	8.9	-4.2
	Naive	-339.3	-304.5	28.9	-339.1	-304.3	26.8	-346.3	-314.3	32.0
	MVC	-130.7	-172.5	-50.5	-144.3	-158.0	-61.2	-211.0	-115.4	-90.0
	FUC4	-104.1	-148.2	-9.4	-115.6	-136.5	-17.3	-177.2	-100.0	-44.7
	FUC8	-102.0	-140.7	-3.2	-110.9	-130.7	-8.0	-174.2	-91.8	-37.6
0.9	True	0.2	2.4	0.5	6.5	-2.2	4.6	-4.1	7.7	-6.5
	Naive	-312.5	-276.1	58.3	-309.7	-275.5	62.3	-312.9	-293.8	63.8
	MVC	-89.0	-147.0	-26.5	-84.5	-146.6	-22.6	-147.5	-111.4	-58.3
	FUC4	-77.3	-128.3	-0.4	-72.2	-129.2	3.9	-130.3	-100.8	-27.6
	FUC8	-75.5	-124.4	4.1	-71.2	-124.2	8.7	-126.5	-97.3	-21.7
0.95	True	-1.4	2.7	-1.8	-1.7	3.5	-0.2	4.5	-0.8	4.4
	Naive	-290.2	-251.5	75.8	-291.0	-249.9	76.6	-292.9	-273.4	106.3
	MVC	-60.8	-118.1	-17.8	-72.4	-106.6	-24.6	-157.9	-48.6	-53.0
	FUC4	-51.8	-108.6	2.0	-64.0	-97.2	-5.5	-155.2	-39.2	-38.9
	FUC8	-50.6	-106.5	5.4	-63.7	-93.6	-2.2	-154.4	-36.2	-36.4
0.99	True	-1.0	2.3	5.4	-3.2	3.7	1.1	—	—	—
	Naive	-255.2	-209.0	113.1	-254.7	-207.4	109.7	—	—	—
	MVC	-29.7	-56.2	-2.5	-38.0	-45.7	-12.9	—	—	—
	FUC4	-27.8	-52.7	4.4	-37.3	-41.8	-7.5	—	—	—
	FUC8	-27.9	-51.7	5.3	-36.9	-41.0	-6.1	—	—	—

Table 2.7: Summary of estimated biases ($\times 10^3$) and standard errors in parameter estimate ($\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_Z$) with different censoring mechanisms and censoring probabilities, with the uncorrelated measurement errors.

bias($\times 10^3$)		$\beta_Z = \log(0.8)$			$\beta_Z = 0$			$\beta_Z = \log(1.2)$			$\beta_Z = \log(1.5)$		
(sk, ku)	method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$
(0, 0)	True	2.6	0.3	1.0	4.2	-1.2	1.7	3.8	-1.1	1.6	2.4	-0.2	0.8
	Naive	-118.9	-157.5	32.1	-117.4	-160.0	23.3	-117.0	-161.5	15.5	-117.4	-162.9	5.7
	MVC	-36.7	-49.1	15.6	-37.0	-48.6	6.1	-39.1	-46.5	-2.6	-43.8	-42.4	-13.8
	FUC4	-32.1	-45.1	14.2	-31.9	-45.0	8.0	-33.5	-43.4	1.9	-37.6	-39.7	-6.1
	FUC8	-31.4	-44.3	14.1	-30.9	-44.5	8.4	-32.6	-42.8	2.8	-36.4	-39.4	-4.7
(0, 6/5)	True	2.6	0.3	1.0	4.2	-1.2	1.7	3.8	-1.1	1.6	2.4	-0.2	0.8
	Naive	-122.9	-159.8	32.8	-121.7	-162.0	23.7	-121.6	-163.3	15.8	-122.1	-164.8	5.8
	MVC	-43.8	-48.9	15.8	-45.6	-46.7	5.5	-48.3	-44.2	-3.5	-53.2	-40.1	-14.9
	FUC4	-39.1	-45.0	14.6	-40.3	-43.4	7.6	-42.5	-41.3	1.3	-46.6	-37.8	-6.9
	FUC8	-38.1	-44.5	14.6	-39.1	-43.1	8.1	-41.2	-41.0	2.2	-45.2	-37.7	-5.4
(1, 3)	True	2.6	0.3	1.0	4.2	-1.2	1.7	3.8	-1.1	1.6	2.4	-0.2	0.8
	Naive	-152.3	-183.6	37.3	-151.3	-185.5	26.7	-151.2	-186.9	17.5	-151.8	-188.5	6.0
	MVC	-82.2	-69.0	18.8	-83.9	-67.0	7.0	-86.3	-64.9	-3.1	-90.6	-61.4	-15.9
	FUC4	-77.4	-65.3	17.9	-78.7	-63.6	9.3	-80.6	-61.9	1.7	-84.3	-59.0	-8.0
	FUC8	-76.5	-64.7	17.9	-77.7	-63.2	9.8	-79.5	-61.5	2.7	-83.1	-58.7	-6.5
(2, 7)	True	2.6	0.3	1.0	4.2	-1.2	1.7	3.8	-1.1	1.6	2.4	-0.2	0.8
	Naive	-184.6	-210.8	42.9	-183.8	-212.5	30.6	-183.7	-213.9	20.2	-184.3	-215.6	7.1
	MVC	-122.7	-94.4	23.1	-124.3	-92.5	9.8	-126.3	-90.7	-1.5	-130.1	-87.9	-15.7
	FUC4	-117.9	-90.7	22.6	-119.2	-89.2	12.2	-120.7	-87.8	3.4	-124.1	-85.2	-7.9
	FUC8	-117.1	-90.1	22.6	-118.2	-88.7	12.8	-119.7	-87.3	4.3	-123.0	-84.8	-6.5

Table 2.8: Summary of estimated biases ($\times 10^3$) and standard errors in parameter estimate ($\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_Z$) with violation of normality, with the uncorrelated measurement errors.

by at least 30%, and even more with FUC. Hence our proposed MVC and FUC are robust to the normality assumption.

2.6.5 Summary of Simulation Studies

Through simulation studies, we find that both the mean-variance regression calibration (MVC) and the follow-up time calibration (FUC) have good performance in terms of reducing biases introduced by the uncorrelated measurement errors. FUC generally has smaller biases than MVC, with slightly larger standard errors, and it is less sensitive to the rare disease assumption. We recommend using this method when sample size is relatively large. However, with large L , there is a possibility of over-correction. This can be due to unstable

calibrations at later time intervals with small sample sizes. To avoid over-correction, we recommend using FUC with $L \leq 4$. Both MVC and FUC work well when (ρ_0, ρ_1) are estimated. The abilities of MVC and FUC to reduce biases are robust to censoring mechanisms and violation of normality.

Chapter 3

INDUCED HAZARD APPROACH— CORRELATED
MEASUREMENT ERRORS**3.1 Overview**

In Chapter 2, we discussed the influence of the uncorrelated measurement error (i.e., technical measurement error) on mediation analysis, and proposed the mean-variance regression calibration (MVC) and the follow-up time calibration (FUC) based on the induced hazard to correct for biases. In practice, the correlated measurement error described in Section 1.5 may be more important to correct for in order to understand the mediation effects. As discussed in Section 1.5, the correlated measurement error is preferred if the scientific question is to determine if the long-term biomarker level is a potential mediator. In the motivating example, it is possible that the long-term average IGFBP4 level of a subject is mediating the effect between hormone therapy and stroke. The measure of IGFBP4 at a specific clinic visit is an estimate of the long-term average level. But the temporal variation and the technical measurement error associated with the IGFBP4 measure may introduce extra variation and result in biases in the hazard ratio estimates. In this chapter, we continue to use the induced hazard function and discuss the performance of MVC and FUC with the correlated measurement error. The underlying mediation pathway is the same as in Figure 2.1.

Similar as before, we denote the observed event time and censoring indicator as (T, δ) , treatment assignment as Z , the true biomarker values as $\mathbf{X} = (X_0, X_1)$, the observed biomarker values as $\mathbf{W} = (W_0, W_1)$, and $\mathbf{W} = \mathbf{X} + \mathbf{U}$, where \mathbf{U} is the correlated measurement error independent of \mathbf{X} given Z . Note that the definition of \mathbf{X} now becomes the long-term average of the biomarker level instead of the true biomarker value at a specific

time point. The joint distribution of \mathbf{X} and \mathbf{U} given Z is as in (1.13), (1.14):

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{U} \end{pmatrix} | Z \sim N\left(\begin{pmatrix} \mathbf{M}_Z \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_Z & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}_Z \end{pmatrix}\right), \quad (3.1)$$

where

$$\begin{aligned} \mathbf{M}_0 &= \begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix}, & \boldsymbol{\Sigma}_0 &= \begin{pmatrix} \sigma_0^2 & \sigma_0^2 \\ \sigma_0^2 & \sigma_0^2 \end{pmatrix}, & \boldsymbol{\Delta}_0 &= \begin{pmatrix} \sigma_{e0}^2 & r_0 \sigma_{e0}^2 \\ r_0 \sigma_{e0}^2 & \sigma_{e0}^2 \end{pmatrix}, \\ \mathbf{M}_1 &= \begin{pmatrix} \mu_0 \\ \mu_1 + d \end{pmatrix}, & \boldsymbol{\Sigma}_1 &= \begin{pmatrix} \sigma_0^2 & \rho \sigma_0 \sigma_1 \\ \rho \sigma_0 \sigma_1 & \sigma_1^2 \end{pmatrix}, & \boldsymbol{\Delta}_1 &= \begin{pmatrix} \sigma_{e0}^2 & r_1 \sigma_{e0} \sigma_{e1} \\ r_1 \sigma_{e0} \sigma_{e1} & \sigma_{e1}^2 \end{pmatrix}. \end{aligned} \quad (3.2)$$

The variance of temporal variation may vary with treatment Z , hence we do not restrict σ_{e0}^2 and σ_{e1}^2 to be the same for the correlated measurement error case. Distribution parameters for this joint distribution are $\mathcal{A} = (\mu_0, \mu_1, d, \sigma_0^2, \sigma_1^2, \rho, \sigma_{e0}^2, \sigma_{e1}^2, r_0, r_1)$. We restrict our discussion to cohort study with one biomarker. Other study designs and multiple biomarker scenarios are discussed in Chapter 5.

3.2 Naive Approach

To analyze the potential mediation effect of a biomarker's long-term average, we compare the following two models:

$$\lambda(t; X_0, Z) = \lambda_0(t) \exp(\alpha_0 X_0 + \alpha_Z Z) \quad (3.3)$$

$$\lambda(t; X_0, X_1, Z) = \lambda_1(t) \exp(\beta_0 X_0 + \beta_1 X_1 + \beta_Z Z) \quad (3.4)$$

These two models are the same as discussed in Chapter 2 with the uncorrelated measurement error, but \mathbf{X} is defined differently. If β_Z moves substantially towards null compared to α_Z , we conclude that there is evidence that X_1 mediates the relationship between T and Z . Again, with the definition of the correlated measurement error, baseline X_0 and U_0 distributions do not differ in the two treatment groups, and α_Z estimate is not seriously biased. But X_1 and U_1 distributions depend on Z , thus can result in a biased estimate of β_Z , and potentially lead to a wrong conclusion about mediation effect. We expect the biases due to the correlated measurement error to be larger than those due to the uncorrelated

measurement error, since temporal variation introduces extra variability into the estimation procedure.

For the naive approach, we again fit the following two models:

$$\lambda(t; W_0, Z) = \lambda_2(t) \exp(a_0 W_0 + a_Z Z)$$

$$\lambda(t; W_0, W_1, Z) = \lambda_3(t) \exp(b_0 W_0 + b_1 W_1 + b_Z Z).$$

Parameter (b_0, b_1, b_Z) can be estimated similarly as in Appendix A with the uncorrelated measurement error through partial likelihood approach. Numeric solutions are needed. We skip the theoretical formula but provide a simulation study to understand the performance of the naive approach.

We again consider a cohort study of sample size 10,000 with half of the study participants randomized to the treatment group, and all subjects are censored at a fixed end of study C_{end} (administrative censoring). Long-term average biomarker level X_0, X_1 are jointly normally distributed within each treatment group, with the following distribution parameters: $\mu_0 = \mu_1 = 0$, $d = 0.5$, $\sigma_0^2 = 1$, $\rho = 0.9$. Suppose that the baseline hazard $\lambda_1(t)$ is constantly 1 and $\beta_0 = \beta_1 = 1$. We investigate how the bias of each parameter varies with σ_1^2 , σ_{e0}^2 , σ_{e1}^2 , β_Z and censoring probability. Results are summarized in Figure 3.1.

In Figure 3.1 (a)-(c), we constrain measurement error variances to be the same (i.e. $\sigma_{e0}^2 = \sigma_{e1}^2$), and vary σ_1^2 , β_Z and censoring probability. We let $\sigma_1^2 = 1.5$, $\beta_Z = \log(1.5)$ and censoring probability 0.95, if not otherwise stated. For all the scenarios, biases of all the three parameters get larger as measurement error variances increase. Comparing with Figure 2.2, naive biases of all the three parameters have similar directions as with the uncorrelated measurement error. The magnitude of bias for β_0 is larger with the correlated measurement error. Although not obvious from the figure, the biases of β_1 and β_Z are also larger with the correlated measurement error. As σ_1^2 and censoring probability change, biases of all the three parameters change. When β_Z changes, only the bias of β_Z is influenced.

In Figure 3.1 (d), we allow measurement error variances to differ between the two treatment arms. As observed in the uncorrelated measurement error case, biases of β_1 and β_Z are bigger when the relative difference between σ_{e0}^2 and σ_{e1}^2 gets larger.

Due to the fact that variances of the correlated measurement error are usually larger

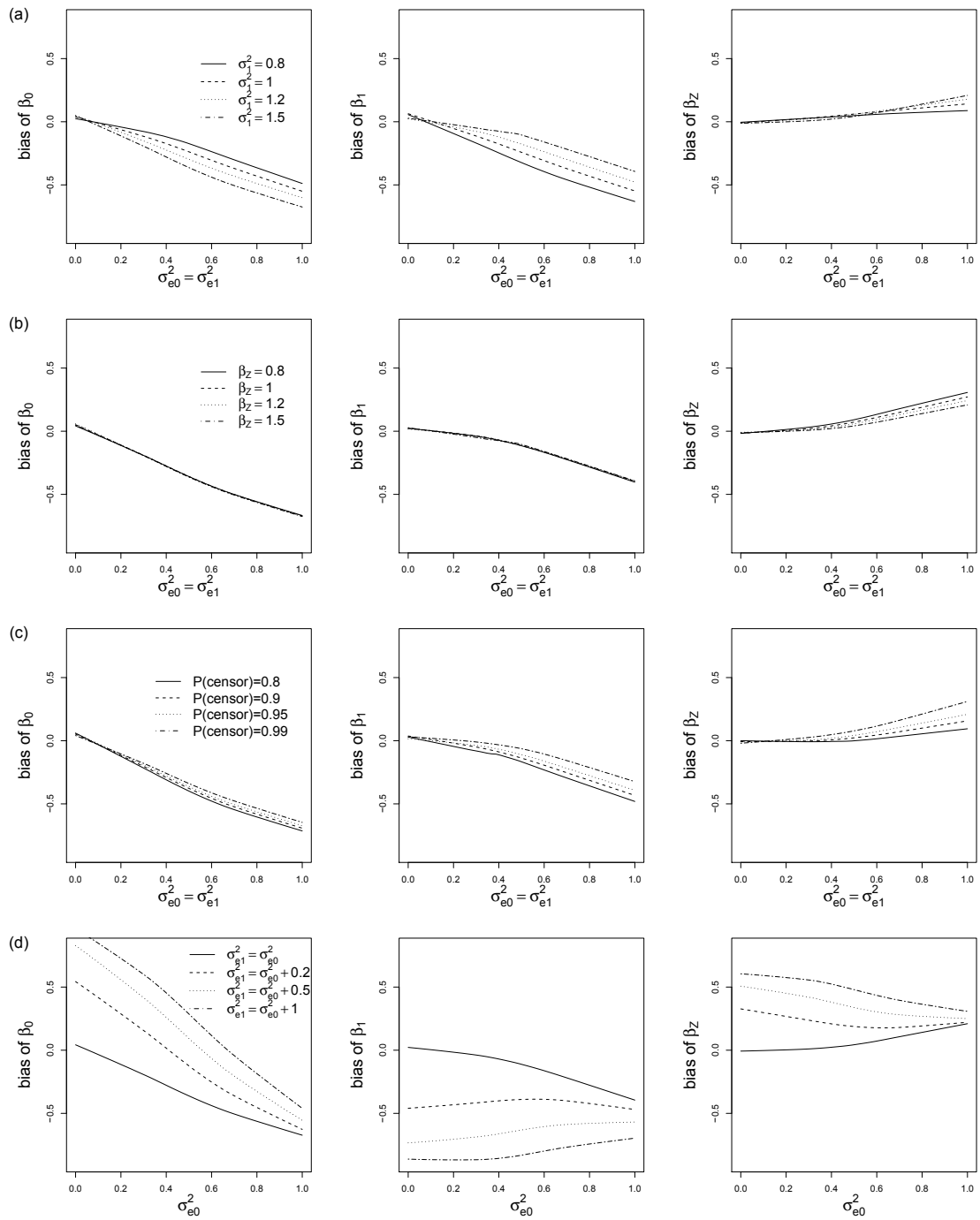


Figure 3.1: Summary of estimated biases of naive approach with correlated measurement error, for different configurations of parameters: (a) different σ_1^2 , (b) different β_Z , (c) different censoring probabilities, (d) different $\sigma_{e0}^2, \sigma_{e1}^2$ relationship.

than those of the uncorrelated measurement error and are differential in the two treatment groups, this simulation study suggests that the naive approach that ignores the correlated measurement error may result in more serious biases in parameter estimates, compared to the uncorrelated measurement error situation. Biases depend on several parameters in a complicated fashion. Correction methods are needed to provide a less biased treatment effect estimate (β_Z) while biases of other parameters (β_0, β_1) remain reasonably small.

3.3 Mean-Variance Regression Calibration and Follow-up Time Calibration

The mean-variance regression calibration and the follow-up time calibration can be applied to the correlated measurement error scenario with modification.

Similar as before, under the rare disease assumption $Pr(\tilde{T} \geq t | \mathbf{X}, Z) \approx 1$ and when the conditional distribution of $(\mathbf{X} | \mathbf{W}, Z)$ is normal, we can approximate the induced hazard $\lambda(t; \mathbf{W}, Z)$ as

$$\begin{aligned} \lambda(t; \mathbf{W}, Z) &\approx \lambda_1(t) E[\exp(\boldsymbol{\beta}_X^T \mathbf{X} + \beta_Z Z) | \mathbf{W}, Z] \\ &\approx \lambda_1(t) \exp[\beta_Z Z + \boldsymbol{\beta}_X^T \mathbf{E}(\mathbf{X} | \mathbf{W}, Z) + \frac{1}{2} \boldsymbol{\beta}_X^T \mathbf{V}(\mathbf{X} | \mathbf{W}, Z) \boldsymbol{\beta}_X]. \end{aligned}$$

Conditional mean and variance $\mathbf{E}(\mathbf{X} | \mathbf{W}, Z)$, $\mathbf{V}(\mathbf{X} | \mathbf{W}, Z)$ are defined as

$$\begin{aligned} \mathbf{E}(\mathbf{X} | \mathbf{W}, Z = 0) &= \frac{1}{k_0} \begin{pmatrix} a_1 W_0 + a_1 W_1 + a_2 \\ a_1 W_0 + a_1 W_1 + a_3 \end{pmatrix} \\ \mathbf{E}(\mathbf{X} | \mathbf{W}, Z = 1) &= \frac{1}{k_1} \begin{pmatrix} a_4 W_0 + a_5 W_1 + a_6 \\ a_7 W_0 + a_8 W_1 + a_9 \end{pmatrix} \\ \mathbf{V}(\mathbf{X} | \mathbf{W}, Z = 0) &= \frac{1}{k_0} \begin{pmatrix} v_1 & v_1 \\ v_1 & v_1 \end{pmatrix} \\ \mathbf{V}(\mathbf{X} | \mathbf{W}, Z = 1) &= \frac{1}{k_1} \begin{pmatrix} v_2 & v_3 \\ v_3 & v_4 \end{pmatrix}, \end{aligned}$$

where

$$\begin{aligned}
k_0 &= 2\sigma_0^2 + (1+r_0)\sigma_{e0}^2, \quad k_1 = (\sigma_0^2 + \sigma_{e0}^2)(\sigma_1^2 + \sigma_{e1}^2) - (\rho\sigma_0\sigma_1 + r_1\sigma_{e0}\sigma_{e1})^2, \\
a_1 &= \sigma_0^2, \quad a_2 = [\sigma_0^2 + (1+r_0)\sigma_{e0}^2]\mu_0 - \sigma_0^2\mu_1, \quad a_3 = -\sigma_0^2\mu_0 + [\sigma_0^2 + (1+r_0)\sigma_{e0}^2]\mu_1, \\
a_4 &= \sigma_0^2(\sigma_1^2 + \sigma_{e1}^2) - \rho\sigma_0\sigma_1(\rho\sigma_0\sigma_1 + r_1\sigma_{e0}\sigma_{e1}), \quad a_5 = \sigma_0\sigma_{e0}(\rho\sigma_1\sigma_{e0} - r_1\sigma_0\sigma_{e1}) \\
a_6 &= [\sigma_{e0}^2(\sigma_1^2 + \sigma_{e1}^2) - r_1\sigma_{e0}\sigma_{e1}(\rho\sigma_0\sigma_1 + r_1\sigma_{e0}\sigma_{e1})]\mu_0 + \sigma_0\sigma_{e0}(r_1\sigma_0\sigma_{e1} - \rho\sigma_1\sigma_{e0}) \\
a_7 &= \sigma_1\sigma_{e1}(\rho\sigma_0\sigma_{e1} - r_1\sigma_1\sigma_{e0}), \quad a_8 = \sigma_1^2(\sigma_0^2 + \sigma_{e0}^2) - \rho\sigma_0\sigma_1(\rho\sigma_0\sigma_1 + r_1\sigma_{e0}\sigma_{e1}) \\
a_9 &= -\sigma_1\sigma_{e1}(\rho\sigma_0\sigma_{e1} - r_1\sigma_1\sigma_{e0})\mu_0 + [\sigma_{e1}^2(\sigma_0^2 + \sigma_{e0}^2) - r_1\sigma_{e0}\sigma_{e1}(\rho\sigma_0\sigma_1 + r_1\sigma_{e0}\sigma_{e1})](\mu_1 + d) \\
v_1 &= (1+r_0)\sigma_0^2\sigma_{e0}^2, \quad v_2 = \sigma_0^2\sigma_{e0}^2[(1-\rho^2)\sigma_1^2 + (1-r_1^2)\sigma_{e1}^2] \\
v_3 &= \sigma_0\sigma_1\sigma_{e0}\sigma_{e1}[r_1(1-\rho^2)\sigma_0\sigma_1 + \rho(1-r_1^2)\sigma_{e0}\sigma_{e1}], \quad v_4 = \sigma_1^2\sigma_{e1}^2[(1-\rho^2)\sigma_0^2 + (1-r_1^2)\sigma_{e0}^2].
\end{aligned}$$

Plugging these parameters into the induced hazard function, $\lambda(t; \mathbf{W}, Z)$ has the following form:

$$\lambda(t; \mathbf{W}, Z) = \lambda_2(t) \exp(b_0 W'_0 + b_2 Z + b_3 W'_0 Z + b_4 W_1 Z),$$

where W'_0 is defined as

$$W'_0 = \begin{cases} \frac{W_0 + W_1}{2} & \text{if } Z = 0 \\ W_0 & \text{if } Z = 1. \end{cases}$$

Notice that this is an interaction model but the main effect of W_1 drops from the model due to the collinearity between X_0 and X_1 in the control group. W'_0 provides a better estimate of X_0 for the control group. The corresponding partial likelihood with $\mathbf{b} = (b_0, b_2, b_3, b_4)$ is

$$PL(\mathbf{b}) = \prod_{i=1}^k \frac{\exp[b_0 W'_{i0} + b_2 Z_i + b_3 W'_{i0} Z_i + b_4 W_{i1} Z_i]}{\sum_{j \in R(t_i)} \exp[b_0 W'_{0j} + b_2 Z_j + b_3 W'_{0j} Z_j + b_4 W_{1j} Z_j]}.$$

After some calculation, we get the relationship between \mathbf{b} and β :

$$\begin{cases} b_0 = \frac{2}{k_0}(a_1\beta_0 + a_2\beta_1) \\ \frac{b_0}{2} + b_3 = \frac{1}{k_1}(a_4\beta_0 + a_7\beta_1) \\ b_4 = \frac{1}{k_1}(a_5\beta_0 + a_8\beta_1) \\ b_2 = \beta_Z + \frac{1}{k_1}(a_6\beta_0 + a_9\beta_1) + \frac{1}{2k_1}(v_2\beta_0^2 + 2v_3\beta_0\beta_1 + v_4\beta_1^2) \\ \quad - \frac{1}{k_0}(a_2\beta_0 + a_3\beta_1) - \frac{v_1}{2k_0}(\beta_0^2 + 2\beta_0\beta_1 + \beta_1^2) \end{cases}$$

Plugging this relationship into the partial likelihood, we have

$$PL(\boldsymbol{\beta}) = \prod_{i=1}^k \frac{\exp[b_0(\boldsymbol{\beta})W'_{i0} + b_2(\boldsymbol{\beta})Z_i + b_3(\boldsymbol{\beta})W'_{i0}Z_i + b_4(\boldsymbol{\beta})W_{i1}Z_i]}{\sum_{j \in R(t_i)} \exp[b_0(\boldsymbol{\beta})W'_{0j} + b_2(\boldsymbol{\beta})Z_j + b_3(\boldsymbol{\beta})W'_{0j}Z_j + b_4(\boldsymbol{\beta})W_{1j}Z_j]}.$$

Maximizing $PL(\boldsymbol{\beta})$ through Newton-Raphson method will give a set of $\boldsymbol{\beta}$ estimates.

To reduce sensitivity on the rare disease assumption, the follow-up time calibration can be used. As described before, we first divide time into L intervals: $[I_1, I_2), [I_2, I_3), \dots, [I_L, I_{L+1})$, where $I_1 = 0, I_{L+1} = \infty$. And we approximate the partial likelihood by

$$\begin{aligned} PL(\boldsymbol{\beta}) &\approx \prod_{l=1}^L \left\{ \prod_{i: t_i \in [I_l, I_{l+1})} \frac{E[\exp(\boldsymbol{\beta}_X^T \mathbf{X}_i + \beta_Z Z_i) | \tilde{T}_i \geq I_l, \mathbf{W}_i, Z_i]}{\sum_{j \in R(t_i)} E[\exp(\boldsymbol{\beta}_X^T \mathbf{X}_j + \beta_Z Z_j) | \tilde{T}_j \geq I_l, \mathbf{W}_j, Z_j]} \right\} \\ &\approx \prod_{l=1}^L \prod_{i: t_i \in [I_l, I_{l+1})} \frac{\exp\{\beta_Z Z_i + \boldsymbol{\beta}_X^T \mathbf{E}(\mathbf{X}_i | I_l, \mathbf{W}_i, Z_i) + \frac{1}{2} \boldsymbol{\beta}_X^T \mathbf{V}(\mathbf{X}_i | I_l, \mathbf{W}_i, Z_i) \boldsymbol{\beta}_X\}}{\sum_{j \in R(t_i)} \exp\{\beta_Z Z_j + \boldsymbol{\beta}_X^T \mathbf{E}(\mathbf{X}_j | I_l, \mathbf{W}_j, Z_j) + \frac{1}{2} \boldsymbol{\beta}_X^T \mathbf{V}(\mathbf{X}_j | I_l, \mathbf{W}_j, Z_j) \boldsymbol{\beta}_X\}}. \end{aligned}$$

The last “ \approx ” is exact if the conditional distribution of $(\mathbf{X} | \tilde{T} \geq I_l, \mathbf{W}, Z)$ is normal for each $l = 1, 2, \dots, L$. Otherwise, it can be viewed as a second-order Taylor approximation. Again, Newton-Raphson method can be used to solve for $\boldsymbol{\beta}$ once $\mathbf{E}(\mathbf{X} | I_l, \mathbf{W}, Z)$ and $\mathbf{V}(\mathbf{X} | I_l, \mathbf{W}, Z)$ are known. $\mathbf{E}(\mathbf{X} | I_l, \mathbf{W}, Z)$ and $\mathbf{V}(\mathbf{X} | I_l, \mathbf{W}, Z)$ can be estimated when distribution parameters $\mathcal{A}(I_l)$ are available.

Similar methods as discussed in Section 2.5 can be used to estimate distribution parameters $\mathcal{A} = (\mu_0, \mu_1, d, \sigma_0^2, \sigma_1^2, \rho, \sigma_{e0}^2, \sigma_{e1}^2, r_0, r_1)$ for the correlated measurement error scenario. With the biomarker process (1.9)

$$W_{ij} = \mu(Z_i, t_j) + b_i(Z_i, t_j) + S_i(Z_i, t_j) + \epsilon_{ij},$$

and the distribution specified in Table 2.1, we can compute the distribution parameters as

$$\begin{aligned} \sigma_0^2 &= \sigma_{b0}^2 & \sigma_1^2 &= \sigma_{b1}^2 & \rho &= \rho_b \\ \sigma_{e0}^2 &= \sigma_{s0}^2 + \sigma_\epsilon^2 & \sigma_{e1}^2 &= \sigma_{s1}^2 + \sigma_\epsilon^2 & r_0 &= \frac{\rho_{s0} \sigma_{s0}^2}{\sigma_{s0}^2 + \sigma_\epsilon^2} & r_1 &= \frac{\rho_{s1} \sigma_{s0} \sigma_{s1}}{\sqrt{(\sigma_{s0}^2 + \sigma_\epsilon^2)(\sigma_{s1}^2 + \sigma_\epsilon^2)}}. \end{aligned}$$

Similar as before, if these estimates are based on an internal data set, we can use all the distribution parameter estimates. If these estimates are based on an external data set, we will only use the most robust parameters across different populations, such as (ρ, r_0, r_1) . Other parameters can be estimated similarly as before with method of moments. Here

the variance structure is more complicated, and solving $(\mu_0, \mu_1, d, \sigma_0^2, \sigma_1^2, \sigma_{e0}^2, \sigma_{e1}^2)$ involves a quadratic equation. When $\rho\sigma_0\sigma_{e1} \neq r_1\sigma_1\sigma_{e0}$, there is a unique set of solutions to these parameters:

$$\begin{aligned}\hat{\mu}_0 &= \frac{1}{n} \sum_{i=1}^n W_{i0}, \quad \hat{\mu}_1 = \frac{1}{n_0} \sum_{i:Z_i=0} W_{i1}, \quad \hat{d} = \frac{1}{n_1} \sum_{i:Z_i=1} W_{i1} - \hat{\mu}_1 \\ \hat{\sigma}_0^2 &= \frac{C - r_0 A}{1 - r_0}, \quad \hat{\sigma}_{e0}^2 = \frac{A - C}{1 - r_0}, \\ \hat{\sigma}_1^2 &= \begin{cases} \left[\frac{\rho\hat{\sigma}_0 D - r_1\hat{\sigma}_{e0} \sqrt{(\rho^2\hat{\sigma}_0^2 + r_1^2\hat{\sigma}_{e0}^2)B - D^2}}{\rho^2\hat{\sigma}_0^2 + r_1^2\hat{\sigma}_{e0}^2} \right]^2, & \text{if } \rho\sigma_0\sigma_{e1} > r_1\sigma_1\sigma_{e0} \\ \left[\frac{\rho\hat{\sigma}_0 D + r_1\hat{\sigma}_{e0} \sqrt{(\rho^2\hat{\sigma}_0^2 + r_1^2\hat{\sigma}_{e0}^2)B - D^2}}{\rho^2\hat{\sigma}_0^2 + r_1^2\hat{\sigma}_{e0}^2} \right]^2, & \text{if } \rho\sigma_0\sigma_{e1} \leq r_1\sigma_1\sigma_{e0} \end{cases} \\ \hat{\sigma}_{e1}^2 &= \begin{cases} \left[\frac{r_1\hat{\sigma}_{e0} D + \rho\hat{\sigma}_0 \sqrt{(\rho^2\hat{\sigma}_0^2 + r_1^2\hat{\sigma}_{e0}^2)B - D^2}}{\rho^2\hat{\sigma}_0^2 + r_1^2\hat{\sigma}_{e0}^2} \right]^2, & \text{if } \rho\sigma_0\sigma_{e1} > r_1\sigma_1\sigma_{e0} \\ \left[\frac{r_1\hat{\sigma}_{e0} D - \rho\hat{\sigma}_0 \sqrt{(\rho^2\hat{\sigma}_0^2 + r_1^2\hat{\sigma}_{e0}^2)B - D^2}}{\rho^2\hat{\sigma}_0^2 + r_1^2\hat{\sigma}_{e0}^2} \right]^2, & \text{if } \rho\sigma_0\sigma_{e1} \leq r_1\sigma_1\sigma_{e0} \end{cases}\end{aligned}$$

where

$$\begin{aligned}A &= \frac{n}{n+n_0} \left\{ \frac{1}{n-1} \sum_{i=1}^n (W_{i0} - \hat{\mu}_0)^2 \right\} + \frac{n_0}{n+n_0} \left\{ \frac{1}{n_0-1} \sum_{i:Z_i=0} (W_{i1} - \hat{\mu}_1)^2 \right\}, \\ B &= \frac{1}{n_1-1} \sum_{i:Z_i=1} (W_{i0} - \widehat{\mu_0 + d})^2, \\ C &= \frac{1}{n_0-1} \sum_{i:Z_i=0} (W_{i0} - \hat{\mu}_0)(W_{i1} - \hat{\mu}_1), \quad D = \frac{1}{n_1-1} \sum_{i:Z_i=1} (W_{i0} - \hat{\mu}_0)(W_{i1} - \hat{\mu}_1 - \hat{d}).\end{aligned}$$

Another more straightforward approach is to use maximum likelihood estimates (MLEs).

That is, we plug the $(\hat{\rho}, \hat{r}_0, \hat{r}_1)$ into the distribution of \mathbf{W} :

$$\begin{aligned}\begin{pmatrix} W_0 \\ W_1 \end{pmatrix} | Z=0 &\sim N\left(\begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 + \sigma_{e0}^2 & \sigma_0^2 + \hat{r}_0\sigma_{e0}^2 \\ \sigma_0^2 + \hat{r}_0\sigma_{e0}^2 & \sigma_0^2 + \sigma_{e0}^2 \end{pmatrix} \right) \\ \begin{pmatrix} W_0 \\ W_1 \end{pmatrix} | Z=1 &\sim N\left(\begin{pmatrix} \mu_0 \\ \mu_1 + d \end{pmatrix}, \begin{pmatrix} \sigma_0^2 + \sigma_{e0}^2 & \hat{\rho}\sigma_0\sigma_1 + \hat{r}_1\sigma_{e0}\sigma_{e1} \\ \hat{\rho}\sigma_0\sigma_1 + \hat{r}_1\sigma_{e0}\sigma_{e1} & \sigma_1^2 + \sigma_{e1}^2 \end{pmatrix} \right),\end{aligned}$$

and solve for the MLE of $(\mu_0, \mu_1, d, \sigma_0^2, \sigma_1^2, \sigma_{e0}^2, \sigma_{e1}^2)$. In practice, the MLE approach is more stable to use. First, we may not know beforehand the relationship between $\rho\sigma_0\sigma_{e1}$ and $r_1\sigma_1\sigma_{e0}$, thus do not know which set of solutions to use for $(\sigma_1^2, \sigma_{e1}^2)$. Second, with random samples we may encounter situations with $(\rho^2\hat{\sigma}_0^2 + r_1^2\hat{\sigma}_{e0}^2)B - D^2 < 0$, which means there is no solution for σ_1^2 and σ_{e1}^2 . This should not cause serious problem when the true values

of $\rho\sigma_0\sigma_{e1}$ and $r_1\sigma_1\sigma_{e0}$ are far away from each other, but again we would not know the true values of these parameters in practice. Thus, we would recommend using MLE for MVC and for the first calibration of FUC. The MLEs are expected to have similar point estimates as the method of moments estimates, with slightly smaller standard errors when the sample size is large.

For the follow-up time calibration, we use exactly the same methods as mentioned above to estimate the parameters at the beginning of the study, and denote them as $\hat{\mathcal{A}}(0)$. For the following intervals, we keep the distribution of measurement error to be constant. That is, $(\hat{\sigma}_{e0}^2, \hat{\sigma}_{e1}^2, \hat{r}_0, \hat{r}_1)$ are constant. The distribution parameters at t are

$$\mathcal{A}(t) = (\sigma_{e0}^2, \sigma_{e1}^2, r_0, r_1, \mu_0(0, t), \mu_1(0, t), \sigma_0^2(0, t), \\ \mu_0(1, t), (\mu_1 + d)(1, t), \sigma_0^2(1, t), \sigma_1^2(1, t), \rho(1, t)).$$

When given $(\hat{\sigma}_{e0}^2, \hat{\sigma}_{e1}^2, \hat{r}_0, \hat{r}_1)$, other parameters $(\mu_0(0, I_l), \mu_1(0, I_l), \mu_0(1, I_l), (\mu_1 + d)(1, I_l), \sigma_0^2(0, I_l), \sigma_0^2(1, I_l), \sigma_1^2(1, I_l), \rho(1, I_l))$ are estimated through the method of moments:

$$\begin{aligned} \hat{\mu}_0(0, I_l) &= \frac{1}{n_{0l}} \sum_{i \in R(I_l), Z_i=0} W_{i0}, & \hat{\mu}_1(0, I_l) &= \frac{1}{n_{0l}} \sum_{i \in R(I_l), Z_i=0} W_{i1}, \\ \hat{\sigma}_0^2(0, I_l) &= \frac{1}{3(n_{0l}-1)} \sum_{i \in R(I_l), Z_i=0} \{ [W_{i0} - \hat{\mu}_0(0, I_l)]^2 + [W_{i1} - \hat{\mu}_1(0, I_l)]^2 + [W_{i0} - \hat{\mu}_0(0, I_l)][W_{i1} - \hat{\mu}_1(0, I_l)] \} \\ &\quad - \frac{2 + \hat{r}_0}{3} \hat{\sigma}_e^2 \\ \hat{\mu}_0(1, I_l) &= \frac{1}{n_{1l}} \sum_{i \in R(I_l), Z_i=1} W_{i0}, & \widehat{\mu_1 + d}(1, I_l) &= \frac{1}{n_{1l}} \sum_{i \in R(I_l), Z_i=1} W_{i1}, \\ \hat{\sigma}_0^2(1, I_l) &= \frac{1}{n_{1l}-1} \sum_{i \in R(I_l), Z_i=1} [W_{i0} - \hat{\mu}_0(1, I_l)]^2 - \hat{\sigma}_{e0}^2 \\ \hat{\sigma}_1^2(1, I_l) &= \frac{1}{n_{1l}-1} \sum_{i \in R(I_l), Z_i=1} [W_{i1} - \widehat{\mu_1 + d}(1, I_l)]^2 - \hat{\sigma}_{e1}^2 \\ \hat{\rho}(1, I_l) &= \frac{\frac{1}{n_{1l}-1} \sum_{i \in R(I_l), Z_i=1} [W_{i0} - \hat{\mu}_0(1, I_l)][W_{i1} - \widehat{\mu_1 + d}(1, I_l)] - \hat{r}_1 \hat{\sigma}_{e0} \hat{\sigma}_{e1}}{\sqrt{\frac{1}{n_{1l}-1} \sum_{i \in R(I_l), Z_i=1} [W_{i0} - \hat{\mu}_0(1, I_l)]^2 - \hat{\sigma}_{e0}^2} \sqrt{\frac{1}{n_{1l}-1} \sum_{i \in R(I_l), Z_i=1} [W_{i1} - \widehat{\mu_1 + d}(1, I_l)]^2 - \hat{\sigma}_{e1}^2}} \end{aligned}$$

where n_{0l} and n_{1l} are the number of subjects who survived to I_l in the control and treatment groups. We do not recommend using the MLE in the follow-up time calibration other than

at $t = 0$. This is because that the MLE greatly increases the computation time, and when $(\sigma_{e0}^2, \sigma_{e1}^2, r_0, r_1)$ are known, other parameter estimates no longer involve quadratic equations, thus can be easily estimated through methods of moments with similar accuracy. Conditional means and variances can be computed given all the distribution parameters.

3.4 Simulation Studies

The mean-variance regression calibration (MVC) and the follow-up time calibration (FUC) can be applied to correct for the correlated measurement error. In this section, we will study the performance of both approaches. Similar to Section 2.6, we investigate situations with distribution parameters (ρ, r_0, r_1) known or estimated. Simulation studies are performed to compare the performance of MVC and FUC with the naive approach, and to understand the sensitivity of the results to the rare disease and normality assumptions and to (ρ, r_0, r_1) specification.

3.4.1 Simulation Studies with Distribution Parameters (ρ, r_0, r_1) Known

In this subsection, we assume that the values of the distribution parameters (ρ, r_0, r_1) to be known. We fix $\mu_0 = \mu_1 = 0$, $d = 0.5$, $\rho = 0.9$, $r_0 = 0.7$, $r_1 = 0.5$, and the underlying $\lambda_0 = 1$, $\beta_0 = \beta_1 = 1$. We assume half of the cohort members are randomized to the treatment group, and the remaining are in the control group. We first investigate a simple censoring mechanism that all subjects are censored at the end of the study (C_{end}) and the censoring probability is 95%. Table 3.1 and Table 3.2 summarize the estimated biases and standard errors with different $\sigma_1^2, \sigma_{e0}^2, \sigma_{e1}^2$ and β_Z values. Sample size is 10,000 and results are based on 1,000 simulation replicates. Five approaches are compared, including models with the true \mathbf{X} (True), with the observed \mathbf{W} (Naive), the mean-variance regression calibration (MVC), and the follow-up time calibrations with 4 and 8 intervals (FUC4, FUC8). Intervals are defined 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.

Biases with the naive approach are considerably larger than those with the uncorrelated measurement error, so the correlated measurement error is more likely to result in wrong conclusions about mediation effects. For example, with $\sigma_1^2 = 1.2, \sigma_{e0}^2 = 0.5, \sigma_{e1}^2 = 1$ and

when the true hazard ratio of treatment is 1.5 ($\beta_Z = \log(1.5)$), we estimate the HR as 1.83 with the naive approach. With MVC and FUC, biases of the three parameters all reduce significantly. For this combination of parameters, we estimate the hazard ratio of active treatment to be around 1.43 with both MVC and FUC.

Comparing among MVC, FUC4 and FUC8, more calibration generally results in smaller biases. The improvement from MVC to FUC4 is much larger than that from FUC4 to FUC8, especially for β_0 and β_1 . With FUC, over-correction in β_Z is observed with some combinations of parameters. However, even with over-correction, biases of β_Z with FUC4 are reasonably small. Considering that FUC performs better than MVC in estimating β_0 and β_1 , and over-correction is usually quite mild with small L , we recommend using FUC with $L \leq 4$. Standard errors are bigger with MVC and FUC compared to the naive approach, and they increase as the number of calibrations increases, but the increment is small.

When we fix $\sigma_1^2 = 1.2$, $\sigma_{e_0}^2 = 0.5$ and $\sigma_{e_1}^2 = 1$, and let d vary from 0 to 1, biases are summarized in Table 3.3. As d increases, the naive bias of β_Z increases dramatically, while the changes in β_0 and β_1 biases are relatively small. However, no matter what the true value of d is, MVC and FUC both result in much smaller biases in β_0 and β_1 estimates compared to those of the naive approach. When d is small (e.g., $d = 0$), naive biases in β_Z are quite small. In this setting, both MVC and FUC are associated with reasonably small β_Z estimates, but they do not show the ability to further reduce the bias in the naive β_Z estimate. As d increases, the bias of the naive β_Z estimate increases significantly. In this situation, MVC and FUC both have the ability to recover the true β_Z , and over-correction is less frequent with large d . d is defined as the mean change of biomarker value due to treatment. In practice, we would only consider a biomarker to be a potential mediator if its pre- and post-treatment values are significantly different. Hence $d \neq 0$ is of more practical importance, and both MVC and FUC perform well under this scenario.

bias($\times 10^3$)			$\beta_Z = \log(0.8)$			$\beta_Z = 0$			$\beta_Z = \log(1.2)$			$\beta_Z = \log(1.5)$		
σ_1^2	$\sigma_{e0}^2, \sigma_{e1}^2$	meth.	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$
0.8	0.5, 0.75	True	1.5	1.3	-0.1	2.9	0.0	0.5	2.2	0.9	0.1	1.4	0.8	-0.3
		Naive	-207.4	-588.8	132.3	-201.0	-598.4	112.2	-197.2	-605.2	95.0	-193.8	-613.1	73.5
		MVC	-69.0	-147.9	45.3	-81.0	-134.1	17.6	-91.4	-122.4	-5.4	-104.3	-109.4	-33.2
		FUC4	-64.7	-110.9	4.4	-72.9	-101.8	-12.7	-80.4	-93.7	-27.2	-90.1	-84.3	-45.1
		FUC8	-65.2	-103.2	-3.4	-72.4	-95.3	-18.6	-79.2	-88.0	-31.8	-88.2	-79.5	-47.9
	0.5, 1	True	1.5	1.3	-0.1	2.9	0.0	0.5	2.2	0.9	0.1	1.4	0.8	-0.3
		Naive	-128.9	-694.4	151.3	-121.4	-706.7	132.6	-116.8	-715.4	116.2	-112.8	-725.2	95.5
		MVC	-72.1	-147.7	39.8	-87.5	-130.3	10.6	-100.4	-116.0	-13.4	-116.1	-100.0	-42.4
		FUC4	-70.1	-107.4	-0.1	-80.8	-95.6	-18.2	-90.1	-85.5	-33.5	-101.9	-73.9	-52.2
		FUC8	-71.3	-98.8	-7.9	-80.8	-88.4	-24.1	-89.2	-79.4	-37.8	-100.0	-68.9	-54.6
1	0.5, 0.75	True	1.4	1.4	0.4	0.3	2.6	-0.5	-0.3	3.1	-1.6	-0.3	1.8	-1.4
		Naive	-247.9	-530.9	186.6	-243.8	-537.3	164.5	-241.0	-542.3	145.8	-239.0	-547.6	123.6
		MVC	-95.6	-129.1	29.3	-97.8	-128.0	3.7	-99.8	-127.6	-17.5	-103.5	-126.4	-42.6
		FUC4	-79.0	-106.5	-1.4	-80.7	-106.3	-17.2	-82.4	-106.3	-30.7	-85.6	-105.7	-46.9
		FUC8	-76.1	-102.7	-7.0	-77.7	-102.8	-21.2	-79.2	-103.1	-33.4	-82.3	-102.4	-48.0
	0.5, 1	True	1.4	1.4	0.4	0.3	2.6	-0.5	-0.3	3.1	-1.6	-0.3	1.8	-1.4
		Naive	-162.1	-643.7	216.8	-156.5	-653.1	196.2	-152.8	-660.2	178.5	-149.9	-667.6	157.1
		MVC	-102.1	-127.2	19.9	-106.3	-124.5	-6.9	-109.7	-123.0	-28.9	-115.3	-120.3	-55.2
		FUC4	-87.6	-101.4	-9.1	-90.6	-100.1	-25.6	-93.2	-99.5	-39.6	-97.7	-97.6	-56.5
		FUC8	-85.0	-97.0	-14.5	-87.7	-96.2	-29.2	-90.1	-95.9	-41.8	-94.5	-94.2	-57.0
1.2	0.5, 0.75	True	-0.6	3.1	-0.6	-0.3	2.2	-1.1	-1.1	3.1	-3.1	-2.1	3.1	-3.0
		Naive	-288.4	-478.8	223.4	-285.4	-483.6	199.6	-284.4	-486.3	179.0	-282.9	-490.1	156.6
		MVC	-91.8	-143.9	29.3	-87.0	-150.9	7.1	-87.4	-153.1	-13.6	-87.3	-156.2	-36.3
		FUC4	-69.2	-127.9	7.7	-66.3	-133.5	-6.3	-67.7	-134.6	-20.3	-68.5	-136.8	-35.0
		FUC8	-64.8	-125.8	3.9	-62.1	-131.3	-8.7	-63.8	-132.1	-21.6	-64.9	-134.0	-35.0
	0.5, 1	True	-0.6	3.1	-0.6	-0.3	2.2	-1.1	-1.1	3.1	-3.1	-2.1	3.1	-3.0
		Naive	-197.2	-596.6	264.8	-192.6	-604.2	242.5	-190.5	-609.2	223.0	-187.9	-615.2	201.5
		MVC	-94.8	-147.8	20.5	-91.2	-154.5	-2.5	-92.2	-156.5	-23.7	-93.6	-159.0	-47.5
		FUC4	-74.4	-128.4	1.1	-72.3	-133.7	-13.3	-74.2	-134.8	-27.6	-76.2	-136.2	-43.0
		FUC8	-70.2	-125.8	-2.3	-68.3	-131.1	-15.2	-70.5	-131.8	-28.3	-72.7	-133.1	-42.4

Table 3.1: Summary of estimated biases ($\times 10^3$) in parameter estimates ($\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_Z$) with 95% censoring at the end of study, with the correlated measurement errors; with σ_1^2, σ_e^2 and true β_Z value changing.

σ_1^2	SE		meth.	$\beta_Z = \log(0.8)$			$\beta_Z = 0$			$\beta_Z = \log(1.2)$			$\beta_Z = \log(1.5)$		
	$\sigma_{e0}^2, \sigma_{e1}^2$			$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$
0.8	0.5, 0.75	True	0.139	0.149	0.102	0.134	0.144	0.100	0.131	0.140	0.101	0.125	0.135	0.100	
		Naive	0.063	0.061	0.100	0.062	0.061	0.099	0.061	0.060	0.099	0.059	0.058	0.099	
		MVC	0.294	0.316	0.137	0.289	0.314	0.135	0.285	0.312	0.135	0.277	0.307	0.133	
		FUC4	0.300	0.324	0.146	0.296	0.322	0.144	0.292	0.320	0.144	0.284	0.315	0.141	
		FUC8	0.301	0.325	0.148	0.297	0.324	0.146	0.293	0.322	0.146	0.285	0.316	0.143	
	0.5, 1	True	0.139	0.149	0.102	0.134	0.144	0.100	0.131	0.140	0.101	0.125	0.135	0.100	
		Naive	0.061	0.057	0.101	0.060	0.056	0.100	0.059	0.055	0.100	0.058	0.054	0.100	
		MVC	0.304	0.330	0.142	0.299	0.328	0.141	0.295	0.325	0.140	0.287	0.320	0.138	
		FUC4	0.310	0.338	0.151	0.306	0.336	0.150	0.302	0.334	0.149	0.293	0.328	0.146	
		FUC8	0.311	0.339	0.154	0.307	0.337	0.152	0.303	0.335	0.151	0.294	0.329	0.149	
1	0.5, 0.75	True	0.134	0.137	0.111	0.131	0.132	0.110	0.125	0.126	0.109	0.121	0.122	0.109	
		Naive	0.063	0.061	0.101	0.062	0.060	0.101	0.061	0.058	0.101	0.061	0.058	0.102	
		MVC	0.311	0.318	0.173	0.304	0.313	0.172	0.296	0.307	0.170	0.293	0.304	0.169	
		FUC4	0.318	0.326	0.179	0.311	0.322	0.178	0.303	0.315	0.176	0.300	0.312	0.174	
		FUC8	0.319	0.328	0.180	0.312	0.323	0.180	0.304	0.316	0.177	0.301	0.313	0.176	
	0.5, 1	True	0.134	0.137	0.111	0.131	0.132	0.110	0.125	0.126	0.109	0.121	0.122	0.109	
		Naive	0.061	0.057	0.102	0.060	0.055	0.102	0.059	0.054	0.102	0.059	0.053	0.102	
		MVC	0.323	0.333	0.181	0.315	0.328	0.181	0.307	0.320	0.178	0.302	0.316	0.176	
		FUC4	0.330	0.342	0.187	0.323	0.337	0.187	0.314	0.329	0.183	0.310	0.325	0.181	
		FUC8	0.332	0.343	0.189	0.324	0.338	0.188	0.315	0.330	0.184	0.311	0.327	0.182	
1.2	0.5, 0.75	True	0.129	0.126	0.121	0.124	0.120	0.118	0.123	0.117	0.117	0.120	0.112	0.118	
		Naive	0.063	0.059	0.103	0.062	0.058	0.103	0.062	0.057	0.103	0.062	0.057	0.105	
		MVC	0.307	0.301	0.202	0.301	0.295	0.199	0.295	0.289	0.197	0.294	0.288	0.198	
		FUC4	0.314	0.308	0.206	0.308	0.303	0.203	0.302	0.297	0.201	0.301	0.296	0.202	
		FUC8	0.316	0.310	0.207	0.309	0.304	0.204	0.304	0.299	0.202	0.302	0.298	0.203	
	0.5, 1	True	0.129	0.126	0.121	0.124	0.120	0.118	0.123	0.117	0.117	0.120	0.112	0.118	
		Naive	0.061	0.055	0.104	0.060	0.054	0.104	0.060	0.053	0.103	0.061	0.053	0.105	
		MVC	0.316	0.311	0.211	0.308	0.304	0.207	0.302	0.298	0.204	0.300	0.297	0.205	
		FUC4	0.324	0.320	0.214	0.316	0.313	0.211	0.310	0.307	0.208	0.308	0.306	0.209	
		FUC8	0.325	0.321	0.215	0.317	0.314	0.212	0.311	0.309	0.209	0.310	0.307	0.210	

Table 3.2: Summary of standard errors in parameter estimates $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_Z)$ with 95% censoring at the end of study, with the correlated measurement errors; with σ_1^2 , σ_e^2 and true β_Z value changing.

bias($\times 10^3$)		$\beta_Z = \log(0.8)$			$\beta_Z = 0$			$\beta_Z = \log(1.2)$			$\beta_Z = \log(1.5)$		
d	meth.	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$
0	True	-0.1	2.5	0.5	-1.1	2.6	0.6	-0.5	2.8	1.7	0.4	1.6	0.4
	Naive	-211.8	-577.4	21.7	-204.3	-586.7	-2.8	-199.1	-593.5	-22.1	-193.8	-601.5	-48.5
	MVC	-105.1	-134.5	2.6	-98.1	-142.5	-22.0	-94.9	-146.7	-41.7	-90.0	-154.1	-68.3
	FUC4	-80.5	-118.0	-28.5	-75.7	-124.6	-43.5	-73.9	-127.7	-55.3	-70.7	-133.8	-72.5
	FUC8	-75.2	-116.2	-34.2	-71.0	-122.4	-47.6	-69.5	-125.3	-58.1	-66.5	-131.2	-73.8
0.5	True	-0.6	3.1	-0.6	-0.3	2.2	-1.1	-1.1	3.1	-3.1	-2.1	3.1	-3.0
	Naive	-197.2	-596.6	264.8	-192.6	-604.2	242.5	-190.5	-609.2	223.0	-187.9	-615.2	201.5
	MVC	-94.8	-147.8	20.5	-91.2	-154.5	-2.5	-92.2	-156.5	-23.7	-93.6	-159.0	-47.5
	FUC4	-74.4	-128.4	1.1	-72.3	-133.7	-13.3	-74.2	-134.8	-27.6	-76.2	-136.2	-43.0
	FUC8	-70.2	-125.8	-2.3	-68.3	-131.1	-15.2	-70.5	-131.8	-28.3	-72.7	-133.1	-42.4
1	True	-1.4	3.2	-4.0	-2.2	2.8	-4.1	-2.3	2.7	-3.0	-0.5	2.0	-1.7
	Naive	-189.3	-611.6	520.2	-187.2	-617.8	501.5	-186.0	-622.3	487.4	-184.9	-626.9	468.6
	MVC	-93.0	-157.0	45.0	-95.0	-160.0	22.7	-97.2	-161.9	5.7	-99.1	-164.2	-15.6
	FUC4	-75.3	-134.9	33.7	-77.7	-136.9	18.9	-80.2	-138.0	7.9	-82.3	-139.4	-6.4
	FUC8	-71.6	-131.9	31.9	-74.3	-133.6	18.4	-76.9	-134.6	8.4	-79.1	-135.8	-4.8

Table 3.3: Summary of estimated biases ($\times 10^3$) in parameter estimates ($\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_Z$) with 95% censoring at the end of study, with the correlated measurement errors; with d and true β_Z value changing.

3.4.2 Simulation Studies with Distribution Parameters (ρ, r_0, r_1) Estimated

When (ρ, r_0, r_1) are estimated from an external data set, the performance of MVC and FUC are studied in this section. Here $\sigma_1^2 = 1, \sigma_{e_0}^2 = 0.5, \sigma_{e_1}^2 = 0.75$, and $\beta_Z = \log(1.5)$. We generate external data sets with sample size 1,000 and 500 to estimate (ρ, r_0, r_1) . The true values of (ρ, r_0, r_1) in this simulation are $(0.9, 0.7, 0.5)$. In the estimating procedure with an external data set, since the true ρ value is 0.9, it is possible that for some simulated data set, the estimated ρ s are quite close to 1, which will cause collinearity problem in fitting the Cox model. In practice, if we found the biomarker change is exactly collinear with the treatment assignment, we can immediately conclude this biomarker is a mediator without fitting models. Hence we exclude the simulated data set with $\hat{\rho}$ close to 1. With an external data set of sample size 1,000 and 500, the excluding percentages are about 30% and 50%. Excluding these external data sets makes the estimates of (ρ, r_0, r_1) biased. The mean (SE) for the three parameters are $0.86(0.06), 0.67(0.06), 0.53(0.07)$ for using an external data set of sample size 1000, and $0.85(0.08), 0.65(0.08), 0.53(0.09)$ for using an external data set of sample size 500. Biases and standard errors both increase as the sample size of the external data set gets smaller.

With the estimated (ρ, r_0, r_1) , we conduct simulation studies to evaluate the performance of MVC and FUC. Results are summarized in Table 3.4. As (ρ, r_0, r_1) estimates get less stable, biases in β_Z change from negative to positive, and the biases are increasing. β_0 and β_1 biases are changing in opposite directions. Standard errors increase significantly as $(\hat{\rho}, \hat{r}_0, \hat{r}_1)$ are plugged in. This indicates that the $(\hat{\rho}, \hat{r}_0, \hat{r}_1)$ contributes a big portion of the total variability. When $(\hat{\rho}, \hat{r}_0, \hat{r}_1)$ are relatively precise, MVC and FUC still provide much smaller biases compared to the naive approach. With smaller external data sets, MVC and FUC start to lose the ability to recover β_Z estimate. Compared to the previous results in the uncorrelated measurement error scenario, we find that the performance of MVC and FUC are more sensitive to the specification of distribution parameters when the correlated measurement error is present. We recommend the sample size of the external data set to be at least comparable to the total number of failures in the main study, in order to achieve stable performance of MVC and FUC. If such an external data set is not available, one

		(ρ, r_0, r_1)			$(\hat{\rho}, \hat{r}_0, \hat{r}_1), n_E = 1000$			$(\hat{\rho}, \hat{r}_0, \hat{r}_1), n_E = 500$		
meth.		$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$
bias($\times 10^3$)	True	-0.3	1.8	-1.4	-0.3	1.8	-1.4	-0.3	1.8	-1.4
	Naive	-239.0	-547.6	123.6	-239.0	-547.6	123.6	-239.0	-547.6	123.6
	MVC	-103.5	-126.4	-42.6	-139.3	-112.5	5.2	-60.0	-182.7	55.2
	FUC4	-85.6	-105.7	-46.9	-113.2	-102.7	12.8	-24.0	-185.8	66.9
	FUC8	-82.3	-102.4	-48.0	-107.9	-102.0	14.0	-17.4	-186.8	68.7
sd	True	0.121	0.122	0.109	0.121	0.122	0.109	0.121	0.122	0.109
	Naive	0.061	0.058	0.102	0.061	0.058	0.102	0.061	0.058	0.102
	MVC	0.293	0.304	0.169	0.794	0.875	0.346	0.879	0.934	0.384
	FUC4	0.300	0.312	0.174	0.808	0.896	0.350	0.908	0.964	0.393
	FUC8	0.301	0.313	0.176	0.811	0.900	0.351	0.915	0.971	0.396

Table 3.4: Summary of estimated biases ($\times 10^3$) and SEs in parameter estimate $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_Z)$ with 95% censoring at the end of study, with the correlated measurement errors. Distribution parameters (ρ, r_0, r_1) are known, estimated from external data sets with sample size 1,000 and 500.

should either reduce the complexity of the biomarker process model, or do a sensitivity study covering a range of potential (ρ, r_0, r_1) values.

3.4.3 Simulation Studies of Sensitivity on Censoring Mechanism

Now we investigate the performance of MVC and FUC under different censoring mechanisms and censoring probabilities. Again, we are interested in three different censoring mechanisms: administrative censoring at a pre-specified time C_{end} (Censor I), exponential censoring with a constant rate λ_C (Censor II), and exponential censoring with rates differ in the two treatment arms (Censor III). We vary the overall censoring probability from 50% to 99%. For the third censoring mechanism, we again let the censoring probability of the treatment group be 5% higher than that of the control group. In this situation, intervals for FUC are defined similarly as in the uncorrelated measurement error situation. That is, intervals are $[0, Q_{T,1/L,Z=1}), [Q_{T,1/L,Z=1}, Q_{T,2/L,Z=1}), \dots, [Q_{T,(L-1)/L,Z=1}, \infty)$, where $Q_{T,k/L,Z=1}$ is the k^{th} L-quantile of the failure times in the treatment group. We let $\sigma_0^2 = 1$, $\sigma_1^2 = 1.2$, $\sigma_{e0}^2 = 0.5$, $\sigma_{e1}^2 = 1$, and $\beta_Z = \log(1.5)$. The expected number of events is fixed at 500 and

the cohort size differs with censoring probabilities. Results are summarized in Table 3.5.

As the censoring probability gets larger, the naive bias of β_Z increases, while biases of the other two parameters also change. Naive biases are quite similar for the three censoring mechanisms. MVC performs well with all the censoring mechanisms and tends to result in reduced β_0 and β_1 biases. However, when censoring probability is low, MVC does not show the ability to reduce β_Z biases. As censoring probability increases, it starts to reduce naive bias in β_Z . For administrative censoring, the censoring probability has to be at least 80% for MVC to show benefits. For exponential censoring and exponential censoring with differential rates, the censoring probability has to be 90% to show MVC benefits. In contrast, FUC almost always provides good correction to the bias in β_Z , due to the fact that FUC allows differential distributions for the two treatment groups over time. Again, the improvement from MVC to FUC4 is much larger than that from FUC4 to FUC8. From this point of view, we recommend choosing FUC4 over MVC when censoring probability is low, especially if the total time span is long or censoring is dependent on treatment assignment.

3.4.4 *Simulation Studies of Violation to Normality Assumption*

In this section, we evaluate the performance of MVC and FUC when the normality assumption of the measurement error is violated. We use similar techniques as discussed in Chapter 2 to generate non-normal distributions with specific skewness and kurtosis. Here we let $\sigma_1^2 = 0.8$, $\sigma_{e0}^2 = 0.5$ and $\sigma_{e1}^2 = 0.75$. We let the skewness and kurtosis be $(0, 0)$, $(0, 6/5)$, $(1, 3)$ and $(2, 7)$, which correspond to normal, symmetric with heavy tail, moderate non-normal and severe non-normal distributions. Results are summarized in Table 3.6.

The naive approach has larger biases as the violation of normality gets more severe. MVC and FUC biases are also increasing, but they are still considerably smaller than those from the naive approach. Even with severe violation of normality, MVC and FUC have the ability to correct for at least 50% of the naive biases in β_Z . The performance in correcting β_0 and β_1 are not very robust to non-normality. With severe non-normality, the remaining biases in β_0 and β_1 can be quite large. However, since the primary interest in this dissertation is to recover treatment effect, MVC and FUC are performing well in recovering β_Z . Another

bias($\times 10^3$)		Censor I			Censor II			Censor III		
P(censor)	method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$
0.5	True	2.0	2.6	-1.0	-0.1	8.5	0.5	-1.6	7.5	-1.8
	Naive	-352.8	-676.0	52.4	-359.7	-677.2	49.2	-371.9	-661.7	39.4
	MVC	-252.3	-320.3	-146.9	-261.5	-319.9	-156.0	-267.4	-314.1	-160.3
	FUC4	-143.0	-261.6	-38.3	-129.6	-281.3	-35.7	-142.7	-265.0	-49.5
	FUC8	-125.1	-246.4	-19.4	-100.6	-285.2	-9.3	-102.0	-273.1	-17.7
0.8	True	5.9	-0.1	-0.2	3.7	-0.3	2.6	-8.5	14.1	-7.3
	Naive	-273.6	-646.0	121.0	-288.7	-648.8	111.0	-312.8	-615.2	97.9
	MVC	-159.9	-257.7	-92.0	-215.1	-230.1	-122.3	-233.3	-206.9	-129.6
	FUC4	-112.6	-207.8	-39.1	-164.2	-177.2	-65.5	-173.5	-170.8	-71.6
	FUC8	-105.1	-198.3	-30.8	-148.1	-173.2	-51.1	-156.5	-169.2	-55.6
0.9	True	3.2	-0.4	1.8	2.1	5.0	0.6	-4.9	8.4	-7.6
	Naive	-226.6	-630.3	169.0	-239.0	-634.4	154.2	-276.6	-586.0	134.0
	MVC	-148.9	-183.0	-76.1	-165.6	-188.6	-91.4	-174.2	-177.0	-90.0
	FUC4	-125.2	-144.6	-54.8	-134.3	-149.0	-58.9	-137.2	-152.4	-63.0
	FUC8	-120.1	-139.3	-51.3	-123.8	-145.8	-49.8	-125.0	-152.4	-53.1
0.95	True	-2.1	3.1	-3.0	-6.2	9.6	-12.0	9.1	-5.5	5.5
	Naive	-187.9	-615.2	201.5	-206.8	-617.4	180.1	-280.6	-528.0	158.0
	MVC	-93.6	-159.0	-47.5	-123.3	-161.1	-70.8	-179.0	-100.2	-73.3
	FUC4	-76.2	-136.2	-43.0	-104.1	-133.3	-61.9	-150.1	-89.0	-76.4
	FUC8	-72.7	-133.1	-42.4	-98.7	-129.5	-58.3	-142.5	-89.8	-72.1
0.99	True	0.0	1.6	5.6	-1.1	3.1	7.0	—	—	—
	Naive	-125.3	-586.7	264.5	-138.0	-591.0	255.1	—	—	—
	MVC	-53.1	-69.9	-19.4	-65.7	-82.0	-24.8	—	—	—
	FUC4	-43.3	-63.6	-41.5	-54.4	-74.3	-42.4	—	—	—
	FUC8	-41.5	-62.8	-45.3	-51.7	-73.2	-44.4	—	—	—

Table 3.5: Summary of estimated biases ($\times 10^3$) and standard errors in parameter estimate ($\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_Z$) with different censoring mechanisms and censoring probabilities, with the correlated measurement errors.

bias($\times 10^3$)		$\beta_Z = \log(0.8)$			$\beta_Z = 0$			$\beta_Z = \log(1.2)$			$\beta_Z = \log(1.5)$		
(sk, ku)	method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$
(0, 0)	True	1.5	1.3	-0.1	2.9	0.0	0.5	2.2	0.9	0.1	1.4	0.8	-0.3
	Naive	-208.5	-587.5	132.9	-202.2	-596.9	112.7	-198.6	-603.5	95.4	-195.3	-611.5	73.9
	MVC	-79.7	-137.1	43.8	-94.7	-119.6	15.0	-106.0	-107.0	-8.2	-118.6	-94.3	-35.7
	FUC4	-75.4	-100.1	3.8	-86.5	-87.5	-14.5	-95.1	-78.2	-29.3	-104.4	-69.2	-46.9
	FUC8	-75.8	-92.5	-3.9	-85.9	-81.0	-20.3	-93.7	-72.8	-33.6	-102.4	-64.5	-49.4
(0, 6/5)	True	1.5	1.3	-0.1	2.9	0.0	0.5	2.2	0.9	0.1	1.4	0.8	-0.3
	Naive	-245.2	-592.8	134.6	-239.1	-601.7	113.4	-235.4	-608.0	95.4	-231.8	-615.7	73.2
	MVC	-130.0	-147.6	44.1	-145.0	-129.7	14.3	-156.2	-116.7	-9.8	-168.8	-103.4	-38.4
	FUC4	-126.3	-111.3	5.6	-137.2	-98.3	-14.0	-145.6	-88.7	-29.9	-154.8	-79.2	-48.9
	FUC8	-126.6	-104.0	-1.8	-136.6	-92.1	-19.5	-144.2	-83.5	-34.0	-152.9	-74.7	-51.3
(1, 3)	True	1.5	1.3	-0.1	2.9	0.0	0.5	2.2	0.9	0.1	1.4	0.8	-0.3
	Naive	-366.0	-636.7	147.8	-360.5	-643.9	123.2	-356.9	-649.3	102.6	-353.4	-655.8	77.2
	MVC	-279.4	-240.3	62.9	-292.6	-223.4	30.0	-302.2	-211.5	3.7	-313.4	-198.5	-28.0
	FUC4	-276.5	-208.2	30.5	-286.1	-195.8	6.4	-293.0	-187.0	-12.9	-301.2	-177.5	-36.3
	FUC8	-276.5	-202.2	24.4	-285.4	-190.7	1.9	-291.7	-182.7	-16.2	-299.3	-173.8	-38.2
(2, 7)	True	1.5	1.3	-0.1	2.9	0.0	0.5	2.2	0.9	0.1	1.4	0.8	-0.3
	Naive	-455.2	-681.3	162.4	-450.4	-687.3	135.3	-447.0	-691.9	112.9	-443.7	-697.4	85.2
	MVC	-369.6	-349.1	90.5	-381.3	-333.7	55.9	-389.5	-322.9	28.2	-399.5	-310.7	-5.4
	FUC4	-367.3	-320.8	63.3	-375.8	-309.5	36.3	-381.5	-301.7	14.7	-388.7	-292.6	-11.8
	FUC8	-367.3	-315.6	58.4	-375.1	-305.1	32.6	-380.3	-298.0	12.0	-387.0	-289.5	-13.2

Table 3.6: Summary of estimated biases ($\times 10^3$) and standard errors in parameter estimate ($\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_Z$) with violation of normality, with the correlated measurement errors.

interesting observation is that over-correction seems to be less frequent with violation of the normality assumption.

3.4.5 Summary of Simulation Studies

According to the simulation study results, ignoring the correlated measurement errors is very likely to result in wrong conclusions about mediation effects. Both the mean-variance regression calibration and the follow-up time calibration have the ability to reduce biases associated with the naive approach and to recover the true treatment effect. Simulation studies suggest that compared to the uncorrelated measurement error case, both MVC and FUC are more sensitive to the specification of distribution parameters in the correlated

measurement error case. Both MVC and FUC are robust to censoring mechanisms, and FUC performs better when censoring probability is low. Both methods work well when there is violation of normality assumptions. Generally speaking, FUC has more stable performance than MVC. However, as the number of calibrations gets larger, there can be over-correction. We do not recommend FUC with more than 4 calibration intervals.

Chapter 4

DISTRIBUTION THEORY**4.1 Introduction**

In this chapter, we develop distribution theory for the mean-variance regression calibration and the follow-up time calibration, including the consistency of $\hat{\beta}$ from the follow-up time calibration and its asymptotic distribution. We consider the mean-variance regression calibration as a special case of the follow-up time calibration, in which we only calibrate once with all subjects at risk at the beginning of the study. The basic idea of proving both consistency and asymptotic distribution is similar to that in Andersen and Gill (1982). If we view the conditional means and variances as covariates, our proposed method has similar model structure as in a regular Cox proportional hazards model. The difference is that our model involves a quadratic term of parameters in addition to the linear term. Also, the conditional means and variances are estimated from samples, which introduces additional variability into the parameter estimates. These issues are discussed in this chapter. We introduce notations and assumptions in Section 4.2, and have a general discussion of consistency and asymptotic normality in Sections 4.3 and 4.4. The special cases of the uncorrelated and correlated measurement error are discussed in Section 4.5.

4.2 Notation and Assumptions*4.2.1 Formulation of the Model*

In this chapter, we formulate the Cox proportional hazards model and our proposed estimating equations in the framework of multivariate counting processes. For simplicity and without loss of generality, we will work on the time interval $(0, 1]$.

Suppose that we have n subjects. We denote T_i as the observed failure or censoring time random variable, and δ_i as the censoring indicator. Let Z_i be the treatment assignment, $\mathbf{X}_i = \{X_{i0}, X_{i1}\}$ be the true pre- and post-randomization biomarker values, and

$\mathbf{W}_i = \{W_{i0}, W_{i1}\}$ be the observed values. We assume independent censoring, that is, failure time and censoring time are independent given (\mathbf{X}_i, Z_i) . Let $N = (N_1, N_2, \dots, N_n)$ be a multivariate counting process in which $N_i(t)$ records the number of failures, 0 or 1, in time $[0, t]$ for the i^{th} subject. Thus, we can write

$$N_i(t) = I(T_i \leq t, \delta_i = 1).$$

Suppose that $N_i(t)$ is right continuous and no two components of $N(t)$ have simultaneous jumps. We also define

$$Y_i(t) = I(T_i \geq t)$$

as the at risk process. $Y_i(t)$ is left-continuous. Both $N_i(t)$ and $Y_i(t)$ are bounded.

With the follow-up time calibration, we first divide time $(0, 1]$ into L fixed intervals: $(I_1, I_2], (I_2, I_3], \dots, (I_L, I_{L+1}]$, where $I_1 = 0, I_{L+1} = 1$, and $L \ll n$. Then we approximate the induced hazard

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

by

$$\begin{aligned} \lambda(t; \mathbf{W}, Z) &\approx \lambda_1(t) E[\exp(\beta_Z Z + \beta_X^T \mathbf{X}) | T > I_l, \mathbf{W}, Z] \\ &\approx \lambda_1(t) \exp\{\beta_Z Z + \beta_X^T E(\mathbf{X} | T > I_l, \mathbf{W}, Z) + \frac{1}{2} \beta_X^T V(\mathbf{X} | T > I_l, \mathbf{W}, Z) \beta_X\} \end{aligned} \quad (4.1)$$

for $t \in (I_l, I_{l+1}]$. The first “ \approx ” is due to replacing the condition $\{T \geq t\}$ by the closest interval $\{T > I_l\}$, and the second “ \approx ” is by Taylor expansion and is exact if the conditional distribution $(\mathbf{X} | T > I_l, \mathbf{W}, Z)$ is normal. We now define two more random processes $\mathbf{E}(t)$ and $\mathbf{V}(t)$: for $t \in (I_l, I_{l+1}]$,

$$\begin{aligned} \mathbf{E}(t) &= E(\mathbf{X} | T > I_l, \mathbf{W}, Z) = E(\mathbf{X} | \mathbf{W}, Z, \mathcal{A}(I_l)) \\ \mathbf{V}(t) &= V(\mathbf{X} | T > I_l, \mathbf{W}, Z) = V(\mathbf{X} | \mathbf{W}, Z, \mathcal{A}(I_l)), \end{aligned}$$

where $\mathcal{A}(I_l)$ is the set of distribution parameters for subjects who survive longer than time I_l . These two random processes represent the conditional mean and variance of the

biomarker value \mathbf{X} given (\mathbf{W}, Z) and that subject survives longer than I_l . Notice that $\mathbf{E}(t)$ and $\mathbf{V}(t)$ are step functions with constant values within each interval, and they are both left-continuous. For a general situation, we can allow $\mathbf{E}(t)$ and $\mathbf{V}(t)$ to be arbitrary left-continuous time-dependent random variables. In the following discussion, despite the fact that $\mathbf{E}(t)$, $\mathbf{V}(t)$ are constant within each interval, we will keep the general notation as much as possible for the ease of future extensions. The condition of subjects surviving to time I_l can be replaced by $\mathcal{A}(I_l)$, which is defined as the distribution parameters in the conditional distribution of $(\mathbf{X}|T > I_l, \mathbf{W}, Z)$. For example, in the uncorrelated measurement error setting with $\sigma_{e0}^2 = \sigma_{e1}^2 = \sigma_e^2$,

$$\begin{aligned} \mathcal{A}(t) = & (\sigma_e^2, \mu_0(0, t), \mu_1(0, t), \sigma_0^2(0, t), \rho_0(0, t), \\ & \mu_0(1, t), (\mu_1 + d)(1, t), \sigma_0^2(1, t), \sigma_1^2(1, t), \rho_1(1, t)) \end{aligned}$$

and in the correlated measurement error setting,

$$\begin{aligned} \mathcal{A}(t) = & (\sigma_{e0}^2, \sigma_{e1}^2, r_0, r_1, \mu_0(0, t), \mu_1(0, t), \sigma_0^2(0, t), \\ & \mu_0(1, t), (\mu_1 + d)(1, t), \sigma_0^2(1, t), \sigma_1^2(1, t), \rho(1, t)). \end{aligned}$$

In practice, we need to estimate $\mathcal{A}(I_l)$. Once $\hat{\mathcal{A}}(I_l)$ is obtained, $(\hat{\mathbf{E}}(t), \hat{\mathbf{V}}(t))$ is determined. The induced hazard we use is

$$\lambda(t; Z_i, \mathbf{W}_i) \approx \lambda_0(t) \exp\{\beta_Z Z_i + \beta_X^T \hat{\mathbf{E}}_i(I_l) + \frac{1}{2} \beta_X^T \hat{\mathbf{V}}_i(I_l) \beta_X\}$$

for $t \in (I_l, I_{l+1}]$, rather than (4.1). The corresponding log partial likelihood $LPL(\boldsymbol{\beta}; \hat{\mathcal{A}})$, score equation $U(\boldsymbol{\beta}; \hat{\mathcal{A}})$ and information matrix $\mathcal{I}(\boldsymbol{\beta}; \hat{\mathcal{A}})$ in counting process notation are

$$\begin{aligned} LPL_{n,l}(\boldsymbol{\beta}; \hat{\mathcal{A}}) = & n^{-1} \left\{ \sum_{i=1}^n \int_{I_l}^{I_{l+1}} [\beta_Z Z_i + \beta_X^T \hat{\mathbf{E}}_i(u) + \frac{1}{2} \beta_X^T \hat{\mathbf{V}}_i(u) \beta_X] dN_i(u) \right. \\ & \left. - \int_{I_l}^{I_{l+1}} \log[n\hat{S}^{(0)}(\boldsymbol{\beta}, u)] d\bar{N}(u) \right\} \end{aligned} \quad (4.2)$$

$$LPL_n(\boldsymbol{\beta}; \hat{\mathcal{A}}) = \sum_{l=1}^L LPL_{n,l}(\boldsymbol{\beta}; \hat{\mathcal{A}}) \quad (4.3)$$

$$U_{n,l}(\boldsymbol{\beta}; \hat{\mathbf{A}}) = n^{-1} \left\{ \sum_{i=1}^n \int_{I_l}^{I_{l+1}} \begin{pmatrix} Z_i \\ \hat{\mathbf{E}}_i(u) + \hat{\mathbf{V}}_i(u) \boldsymbol{\beta}_X \end{pmatrix} dN_i(u) - \int_{I_l}^{I_{l+1}} \frac{\hat{S}^{(1)}(\boldsymbol{\beta}, u)}{\hat{S}^{(0)}(\boldsymbol{\beta}, u)} d\bar{N}(u) \right\} \quad (4.4)$$

$$U_n(\boldsymbol{\beta}; \hat{\mathbf{A}}) = \sum_{l=1}^L U_{n,l}(\boldsymbol{\beta}; \hat{\mathbf{A}}) \quad (4.5)$$

$$\mathcal{I}_{n,l}(\boldsymbol{\beta}; \hat{\mathbf{A}}) = n^{-1} \left\{ \sum_{i=1}^n \int_{I_l}^{I_{l+1}} \begin{pmatrix} 0 & 0 \\ 0 & \hat{\mathbf{V}}_i(u) \end{pmatrix} dN_i(u) - \int_{I_l}^{I_{l+1}} \left[\frac{\hat{S}^{(2)}(\boldsymbol{\beta}, u)}{\hat{S}^{(0)}(\boldsymbol{\beta}, u)} - \left(\frac{\hat{S}^{(1)}(\boldsymbol{\beta}, u)}{\hat{S}^{(0)}(\boldsymbol{\beta}, u)} \right)^{\otimes 2} \right] d\bar{N}(u) \right\} \quad (4.6)$$

$$\mathcal{I}_n(\boldsymbol{\beta}; \hat{\mathbf{A}}) = \sum_{l=1}^L \mathcal{I}_{n,l}(\boldsymbol{\beta}; \hat{\mathbf{A}}) \quad (4.7)$$

where $\bar{N}(u) = \sum_{i=1}^n N_i(u)$, $a^{\otimes 2} = aa^T$ for any vector a , and

$$\begin{aligned} \hat{S}^{(0)}(\boldsymbol{\beta}, t) &= n^{-1} \sum_{i=1}^n Y_i(t) \exp[\beta_Z Z_i + \boldsymbol{\beta}_X^T \hat{\mathbf{E}}_i(t) + \frac{1}{2} \boldsymbol{\beta}_X^T \hat{\mathbf{V}}_i(t) \boldsymbol{\beta}_X] \\ \hat{S}_Z^{(1)}(\boldsymbol{\beta}, t) &= n^{-1} \sum_{i=1}^n Y_i(t) Z_i \exp[\beta_Z Z_i + \boldsymbol{\beta}_X^T \hat{\mathbf{E}}_i(t) + \frac{1}{2} \boldsymbol{\beta}_X^T \hat{\mathbf{V}}_i(t) \boldsymbol{\beta}_X] \\ \hat{S}_E^{(1)}(\boldsymbol{\beta}, t) &= n^{-1} \sum_{i=1}^n Y_i(t) \hat{\mathbf{E}}_i(t) \exp[\beta_Z Z_i + \boldsymbol{\beta}_X^T \hat{\mathbf{E}}_i(t) + \frac{1}{2} \boldsymbol{\beta}_X^T \hat{\mathbf{V}}_i(t) \boldsymbol{\beta}_X] \\ \hat{S}_V^{(1)}(\boldsymbol{\beta}, t) &= n^{-1} \sum_{i=1}^n Y_i(t) \hat{\mathbf{V}}_i(t) \exp[\beta_Z Z_i + \boldsymbol{\beta}_X^T \hat{\mathbf{E}}_i(t) + \frac{1}{2} \boldsymbol{\beta}_X^T \hat{\mathbf{V}}_i(t) \boldsymbol{\beta}_X] \\ \hat{S}^{(1)}(\boldsymbol{\beta}, t) &= \begin{pmatrix} \hat{S}_Z^{(1)}(\boldsymbol{\beta}, t) \\ \hat{S}_E^{(1)}(\boldsymbol{\beta}, t) + \hat{S}_V^{(1)}(\boldsymbol{\beta}, t) \boldsymbol{\beta}_X \end{pmatrix}. \end{aligned}$$

The corresponding definitions in terms of \mathbf{E}_i and \mathbf{V}_i are

$$\begin{aligned} S^{(0)}(\boldsymbol{\beta}, t) &= n^{-1} \sum_{i=1}^n Y_i(t) \exp[\beta_Z Z_i + \boldsymbol{\beta}_X^T \mathbf{E}_i(t) + \frac{1}{2} \boldsymbol{\beta}_X^T \mathbf{V}_i(t) \boldsymbol{\beta}_X] \\ S_Z^{(1)}(\boldsymbol{\beta}, t) &= n^{-1} \sum_{i=1}^n Y_i(t) Z_i \exp[\beta_Z Z_i + \boldsymbol{\beta}_X^T \mathbf{E}_i(t) + \frac{1}{2} \boldsymbol{\beta}_X^T \mathbf{V}_i(t) \boldsymbol{\beta}_X] \\ S_E^{(1)}(\boldsymbol{\beta}, t) &= n^{-1} \sum_{i=1}^n Y_i(t) \mathbf{E}_i(t) \exp[\beta_Z Z_i + \boldsymbol{\beta}_X^T \mathbf{E}_i(t) + \frac{1}{2} \boldsymbol{\beta}_X^T \mathbf{V}_i(t) \boldsymbol{\beta}_X] \\ S_V^{(1)}(\boldsymbol{\beta}, t) &= n^{-1} \sum_{i=1}^n Y_i(t) \mathbf{V}_i(t) \exp[\beta_Z Z_i + \boldsymbol{\beta}_X^T \mathbf{E}_i(t) + \frac{1}{2} \boldsymbol{\beta}_X^T \mathbf{V}_i(t) \boldsymbol{\beta}_X] \end{aligned}$$

$$S^{(1)}(\boldsymbol{\beta}, t) = \begin{pmatrix} S_Z^{(1)}(\boldsymbol{\beta}, t) \\ S_E^{(1)}(\boldsymbol{\beta}, t) + S_V^{(1)}(\boldsymbol{\beta}, t)\boldsymbol{\beta}_X \end{pmatrix}$$

Some further definitions are:

$$\begin{aligned} \hat{S}_Z^{(2)}(\boldsymbol{\beta}, t) &= n^{-1} \sum_{i=1}^n Y_i(t) Z_i^2 \exp[\beta_Z Z_i + \boldsymbol{\beta}_X^T \hat{\mathbf{E}}_i(t) + \frac{1}{2} \boldsymbol{\beta}_X^T \hat{\mathbf{V}}_i(t) \boldsymbol{\beta}_X] \\ \hat{S}_E^{(2)}(\boldsymbol{\beta}, t) &= n^{-1} \sum_{i=1}^n Y_i(t) Z_i [\hat{\mathbf{E}}_i(t) + \hat{\mathbf{V}}_i(t) \boldsymbol{\beta}_X] \exp[\beta_Z Z_i + \boldsymbol{\beta}_X^T \hat{\mathbf{E}}_i(t) + \frac{1}{2} \boldsymbol{\beta}_X^T \hat{\mathbf{V}}_i(t) \boldsymbol{\beta}_X] \\ \hat{S}_V^{(2)}(\boldsymbol{\beta}, t) &= n^{-1} \sum_{i=1}^n Y_i(t) [\hat{\mathbf{E}}_i(t) + \hat{\mathbf{V}}_i(t) \boldsymbol{\beta}_X]^{\otimes 2} \exp[\beta_Z Z_i + \boldsymbol{\beta}_X^T \hat{\mathbf{E}}_i(t) + \frac{1}{2} \boldsymbol{\beta}_X^T \hat{\mathbf{V}}_i(t) \boldsymbol{\beta}_X] \\ \hat{S}^{(2)}(\boldsymbol{\beta}, t) &= \begin{pmatrix} \hat{S}_Z^{(2)}(\boldsymbol{\beta}, t) & \hat{S}_E^{(2)}(\boldsymbol{\beta}, t)^T \\ \hat{S}_E^{(2)}(\boldsymbol{\beta}, t) & \hat{S}_V^{(1)}(\boldsymbol{\beta}, t) + \hat{S}_V^{(2)}(\boldsymbol{\beta}, t) \end{pmatrix} \end{aligned}$$

and

$$\begin{aligned} S_Z^{(2)}(\boldsymbol{\beta}, t) &= n^{-1} \sum_{i=1}^n Y_i(t) Z_i^2 \exp[\beta_Z Z_i + \boldsymbol{\beta}_X^T \mathbf{E}_i(t) + \frac{1}{2} \boldsymbol{\beta}_X^T \mathbf{V}_i(t) \boldsymbol{\beta}_X] \\ S_E^{(2)}(\boldsymbol{\beta}, t) &= n^{-1} \sum_{i=1}^n Y_i(t) Z_i [\mathbf{E}_i(t) + \mathbf{V}_i(t) \boldsymbol{\beta}_X] \exp[\beta_Z Z_i + \boldsymbol{\beta}_X^T \mathbf{E}_i(t) + \frac{1}{2} \boldsymbol{\beta}_X^T \mathbf{V}_i(t) \boldsymbol{\beta}_X] \\ S_V^{(2)}(\boldsymbol{\beta}, t) &= n^{-1} \sum_{i=1}^n Y_i(t) [\mathbf{E}_i(t) + \mathbf{V}_i(t) \boldsymbol{\beta}_X]^{\otimes 2} \exp[\beta_Z Z_i + \boldsymbol{\beta}_X^T \mathbf{E}_i(t) + \frac{1}{2} \boldsymbol{\beta}_X^T \mathbf{V}_i(t) \boldsymbol{\beta}_X] \\ S^{(2)}(\boldsymbol{\beta}, t) &= \begin{pmatrix} S_Z^{(2)}(\boldsymbol{\beta}, t) & S_E^{(2)}(\boldsymbol{\beta}, t)^T \\ S_E^{(2)}(\boldsymbol{\beta}, t) & S_V^{(1)}(\boldsymbol{\beta}, t) + S_V^{(2)}(\boldsymbol{\beta}, t) \end{pmatrix}, \end{aligned}$$

and

$$\begin{aligned} s^{(0)}(\boldsymbol{\beta}, t) &= E\{Y_1(t) \exp(\beta_Z Z_1 + \boldsymbol{\beta}_X^T \mathbf{E}_1(t) + \frac{1}{2} \boldsymbol{\beta}_X^T \mathbf{V}_1(t) \boldsymbol{\beta}_X)\} \\ s_Z^{(1)}(\boldsymbol{\beta}, t) &= E\{Y_1(t) Z_1 \exp(\beta_Z Z_1 + \boldsymbol{\beta}_X^T \mathbf{E}_1(t) + \frac{1}{2} \boldsymbol{\beta}_X^T \mathbf{V}_1(t) \boldsymbol{\beta}_X)\} \\ s_E^{(1)}(\boldsymbol{\beta}, t) &= E\{Y_1(t) \mathbf{E}_1(t) \exp(\beta_Z Z_1 + \boldsymbol{\beta}_X^T \mathbf{E}_1(t) + \frac{1}{2} \boldsymbol{\beta}_X^T \mathbf{V}_1(t) \boldsymbol{\beta}_X)\} \\ s_V^{(1)}(\boldsymbol{\beta}, t) &= E\{Y_1(t) \mathbf{V}_1(t) \exp(\beta_Z Z_1 + \boldsymbol{\beta}_X^T \mathbf{E}_1(t) + \frac{1}{2} \boldsymbol{\beta}_X^T \mathbf{V}_1(t) \boldsymbol{\beta}_X)\} \\ s_Z^{(2)}(\boldsymbol{\beta}, t) &= E\{Y_1(t) Z_1^2 \exp[\beta_Z Z_1 + \boldsymbol{\beta}_X^T \mathbf{E}_1(t) + \frac{1}{2} \boldsymbol{\beta}_X^T \mathbf{V}_1(t) \boldsymbol{\beta}_X]\} \\ s_E^{(2)}(\boldsymbol{\beta}, t) &= E\{Y_1(t) Z_1 [\mathbf{E}_1(t) + \mathbf{V}_1(t) \boldsymbol{\beta}_X] \exp[\beta_Z Z_1 + \boldsymbol{\beta}_X^T \mathbf{E}_1(t) + \frac{1}{2} \boldsymbol{\beta}_X^T \mathbf{V}_1(t) \boldsymbol{\beta}_X]\} \\ s_V^{(2)}(\boldsymbol{\beta}, t) &= E\{Y_1(t) [\mathbf{E}_1(t) + \mathbf{V}_1(t) \boldsymbol{\beta}_X]^{\otimes 2} \exp[\beta_Z Z_1 + \boldsymbol{\beta}_X^T \mathbf{E}_1(t) + \frac{1}{2} \boldsymbol{\beta}_X^T \mathbf{V}_1(t) \boldsymbol{\beta}_X]\} \end{aligned}$$

$$s^{(1)}(\boldsymbol{\beta}, t) = \begin{pmatrix} s_Z^{(1)}(\boldsymbol{\beta}, t) \\ s_E^{(1)}(\boldsymbol{\beta}, t) + s_V^{(1)}(\boldsymbol{\beta}, t)\boldsymbol{\beta}_X \end{pmatrix}$$

$$s^{(2)}(\boldsymbol{\beta}, t) = \begin{pmatrix} s_Z^{(2)}(\boldsymbol{\beta}, t) & s_E^{(2)}(\boldsymbol{\beta}, t)^T \\ s_E^{(2)}(\boldsymbol{\beta}, t) & s_V^{(1)}(\boldsymbol{\beta}, t) + s_V^{(2)}(\boldsymbol{\beta}, t) \end{pmatrix}.$$

4.2.2 Regularity Conditions

Here we provide a set of regularity conditions for the consistency to a suitable (generally biased) target $\boldsymbol{\beta}^*$ and the asymptotic normality in general. We will discuss how to relax these conditions in the special cases of the uncorrelated and correlated measurement errors in Section 4.5.

Regularity conditions:

1. $\int_0^1 \lambda_0(u) du < \infty$.

2. As $n \rightarrow \infty$,

$$\sup_{l=1, \dots, L} \|\hat{\mathcal{A}}(I_l) - \mathcal{A}(I_l)\| \xrightarrow{P} 0.$$

3. $(\mathbf{E}_i, \mathbf{V}_i)$ are predictable and locally bounded.

4. For $S^{(k)}$, $k = 0, 1, 2$, there exists a neighborhood \mathcal{B} of $\boldsymbol{\beta}^*$ (define later) and, respectively, $s^{(k)}$, $k = 0, 1, 2$ defined on $\mathcal{B} \times [0, 1]$ such that,

$$\sup_{u \in [0, 1], \boldsymbol{\beta} \in \mathcal{B}} \|S^{(k)}(\boldsymbol{\beta}, u) - s^{(k)}(\boldsymbol{\beta}, u)\| \rightarrow 0$$

in probability as $n \rightarrow \infty$.

5. The function $s^{(k)}$, $k = 0, 1, 2$ are bounded, and $s^{(0)}$ is bounded away from 0 on $\mathcal{B} \times [0, 1]$.

6. The matrix

$$\Sigma(\boldsymbol{\beta}) = \int_0^1 \left\{ \begin{pmatrix} 0 & 0 \\ 0 & s_V^{(1)}(\boldsymbol{\beta}^*, u) \end{pmatrix} - \frac{s^{(2)}(\boldsymbol{\beta}, u)s^{(0)}(\boldsymbol{\beta}, u) - [s^{(1)}(\boldsymbol{\beta}, u)]^{\otimes 2}}{[s^{(0)}(\boldsymbol{\beta}, u)]^2} s^{(0)}(\boldsymbol{\beta}^*, u) \right\} \lambda_0(u) du$$

is negative semidefinite for all $\boldsymbol{\beta} \in \mathcal{B}$, and is negative definite when $\boldsymbol{\beta} = \boldsymbol{\beta}^*$.

7. (Lindeberg Condition) Let

$$\mathcal{V}_i = \frac{1}{\sqrt{n}} \left\{ \int_0^1 \left[\begin{array}{c} Z_i \\ \mathbf{E}_i(u) + \mathbf{V}_i(u) \boldsymbol{\beta}_X^* \end{array} \right] \frac{s^{(1)}(\boldsymbol{\beta}^*, u)}{s^{(0)}(\boldsymbol{\beta}^*, u)} dM_i^*(u) - \sum_{l=1}^L D_l(\boldsymbol{\beta}^*) F_l(\mathbf{W}_i, Z_i, \mathcal{A}(I_l)) Y_i(I_l) \right\},$$

where $D_l(\cdot)$ and $F_l(\cdot)$ are defined later. Let $\{e_1, e_2, e_3\}$ be any basis of \mathbb{R}^3 , then

$$\sum_{i=1}^n E[(e_k^T \mathcal{V}_i)^2 I\{|e_k^T \mathcal{V}_i| > \epsilon\} | \mathcal{F}_{n,i-1}] \xrightarrow{P} 0$$

for every $\epsilon > 0$ and $k = 1, 2, 3$, and $\mathcal{F}_{n,i}$ is the filtration up to time I_i .

4.2.3 Some Basic Tools

Lemma 1 is about the convergence of a sequence of convex functions and their corresponding maximum points.

Lemma 1 (*Fleming and Harrington (1991) Lemma 8.3.1*)

Let E be an open convex subset of \mathbb{R}^p , and let F_1, F_2, \dots , be a sequence of random concave functions on E and f a real-valued function on E such that, for all $x \in E$,

$$\lim_{n \rightarrow \infty} F_n(x) = f(x)$$

in probability. Then:

1. The function f is concave.
2. For all compact subsets A of E ,

$$\sup_{x \in A} |F_n(x) - f(x)| \rightarrow 0$$

in probability, as $n \rightarrow \infty$.

3. If F_n has a unique maximum at X_n and f has one at x , then $X_n \rightarrow x$ in probability as $n \rightarrow \infty$.

Theorem 1 is the strong law of large numbers for a sequence of stochastic processes.

Theorem 1 (SLLN for $D[0, 1]$ from Rao (1963))

Let $X; X_1, X_2, \dots$ be i.i.d random elements of $D_E[0, 1]$ (endowed with the Skorohod topology) where the elements of $D_E[0, 1]$ are right continuous functions on $[0, 1]$ with left hand limits taking values in a separable Banach space E . Suppose that

$$E\|X\| = E \sup_{t \in [0, 1]} \|X(t)\| < \infty,$$

then

$$\left\| \frac{1}{n} \sum_{i=1}^n X_i - EX \right\| \rightarrow 0 \text{ almost surely as } n \rightarrow \infty.$$

Theorem 2 and Corollary 1 state the Lengart's inequality and its two applications, which will be used in the proof of asymptotic normality.

Theorem 2 (Lengart's Inequality)

Let X be a right-continuous adapted process, and Y a nondecreasing predictable process with $Y(0) = 0$. Suppose, for all bounded stopping time T ,

$$E\{|X(t)|\} \leq E\{Y(t)\}.$$

Then for any stopping time T , and any $\epsilon, \eta > 0$,

$$\Pr(\sup_{t \leq T} |X(t)| \geq \epsilon) \leq \frac{\eta}{\epsilon} + \Pr\{Y(t) \geq \eta\}.$$

Corollary 1 (Two applications of the Lengart's Inequality from Andersen and Gill (1982))

Suppose all stochastic processes are defined on the time interval $(0, 1]$.

1. Let N be a univariate counting process with intensity process λ . Then for all $\epsilon, \eta > 0$,

$$\Pr(N(1) > \eta) \leq \frac{\epsilon}{\eta} + \Pr\left(\int_0^1 \lambda(t) dt > \epsilon\right).$$

2. Let W be a local square integrable martingale. Then for all $\epsilon, \eta > 0$,

$$\Pr\left(\sup_{t \in (0, 1]} |W(t)| > \eta\right) \leq \frac{\epsilon}{\eta^2} + \Pr\langle W, W \rangle(1) > \epsilon.$$

Theorem 3 is the multivariate version of Lindeberg Central Limit Theorem (Kundua et al., 2000). This can be used to prove asymptotic normality of the sum of independently but not identically distributed random variables.

Theorem 3 (*Multivariate Lindeberg Central Limit Theorem*)

Let $\{X_{nj}\}$ be a \mathbb{R}^m -valued martingale difference array with respect to $\{\mathcal{F}_{nj}\}$ such that $E(\|X_{nj}\|^2) < \infty$ for every $1 \leq j \leq n$, $n \geq 1$. Let $\{e_1, \dots, e_m\}$ be any basis of \mathbb{R}^m . Assume that the following conditions hold.

1. For every $b \in \mathbb{R}^m$, $\sum_{j=1}^n E[(b^T X_{nj})^2 | \mathcal{F}_{n,j-1}] \xrightarrow{P} \sigma_b^2$ for some σ_b .
2. $L_n(\epsilon, e_k) \xrightarrow{P} 0$ for every $\epsilon > 0$ and every $1 \leq k \leq m$, where for $b \in \mathbb{R}^m$

$$L_n(\epsilon, b) := \sum_{j=1}^n E[(b^T X_{nj})^2 I\{|b^T X_{nj}| > \epsilon\} | \mathcal{F}_{n,j-1}].$$

Then

$$S_n \xrightarrow{D} N_m(0, \Sigma),$$

where the covariance matrix Σ is characterized by $b^T \Sigma b = \sigma_b^2$, for all $b \in \mathbb{R}^m$.

Lemma 2 proves that when distribution parameters $\mathcal{A}(I_l)$ are consistent, then estimates of conditional means $\hat{\mathbf{E}}(t)$ and variances $\hat{\mathbf{V}}(t)$, and $\hat{S}^{(k)}$, $k = 0, 1, 2$, are also consistent.

Lemma 2 *If $\sup_{l=1, \dots, L} \|\hat{\mathcal{A}}(I_l) - \mathcal{A}(I_l)\| \xrightarrow{P} 0$, then*

$$\sup_{t \in (0,1], i=1, \dots, n} \|\hat{\mathbf{E}}_i(t) - \mathbf{E}_i(t)\| \xrightarrow{P} 0$$

$$\sup_{t \in (0,1], i=1, \dots, n} \|\hat{\mathbf{V}}_i(t) - \mathbf{V}_i(t)\| \xrightarrow{P} 0$$

$$\text{and } \sup_{t \in (0,1]} \|\hat{S}^{(k)}(t) - S^{(k)}(t)\| \xrightarrow{P} 0, \quad k = 1, 2, 3.$$

Proof. When $\sup_{l=1, \dots, L} \|\hat{\mathcal{A}}(I_l) - \mathcal{A}(I_l)\| \xrightarrow{P} 0$, it is obvious to prove the convergence of the mean and variance matrix in the joint distribution of $(\mathbf{X}, \mathbf{W} | Z)$. $\hat{\mathbf{E}}(t)$ and $\hat{\mathbf{V}}(t)$ are both continuous functions of the mean and variance matrices. So we have

$$\sup_{t \in (0,1], i=1, \dots, n} \|\hat{\mathbf{E}}_i(t) - \mathbf{E}_i(t)\| \xrightarrow{P} 0$$

$$\sup_{t \in (0,1], i=1, \dots, n} \|\hat{\mathbf{V}}_i(t) - \mathbf{V}_i(t)\| \xrightarrow{P} 0.$$

Again, $S^{(k)}(t)$, $k = 0, 1, 2$ are also continuous functions of $\hat{\mathbf{E}}(t)$ and $\hat{\mathbf{V}}(t)$. Then

$$\sup_{t \in (0,1]} \|\hat{S}^{(k)}(t) - S^{(k)}(t)\| \xrightarrow{P} 0.$$

■

4.3 Consistency to the Target β^*

With the log partial likelihood defined in (4.2) and (4.3), let $\hat{\beta} = (\hat{\beta}_Z, \hat{\beta}_X^T)^T$ denote the solution to the partial likelihood.

Theorem 4 *Under regularity conditions, $\hat{\beta} \xrightarrow{P} \beta^*$, where β^* is the true value of β in the approximated induced hazard (4.1). That is, β converges in probability to the solution of $E[U_n(\beta, \mathcal{A})] = 0$.*

Proof. We start with a discussion of the properties of $LPL_{n,l}(\beta; \hat{\mathcal{A}})$ for each $l = 1, 2, \dots, L$.

First, we prove that the observed log partial likelihood $LPL_{n,l}(\beta; \hat{\mathcal{A}})$ is equivalent to the exact partial likelihood $LPL_{n,l}(\beta; \mathcal{A})$.

$$\begin{aligned} & LPL_{n,l}(\beta; \hat{\mathcal{A}}) - LPL_{n,l}(\beta; \mathcal{A}) \\ &= n^{-1} \sum_{i=1}^n \int_{I_l}^{I_{l+1}} [\beta_X(\hat{\mathbf{E}}_i(u) - \mathbf{E}_i(u)) + \frac{1}{2} \beta_X^T(\hat{\mathbf{V}}_i(u) - \mathbf{V}_i(u))\beta_X] dN_i(u) \end{aligned} \quad (4.8)$$

$$- n^{-1} \sum_{i=1}^n \int_{I_l}^{I_{l+1}} [\log \hat{S}_0(\beta, u) - \log S_0(\beta, u)] d\bar{N}(u) \quad (4.9)$$

where $\bar{N}(u) = \sum_{i=1}^n N_i(u)$.

For expression (4.8), it holds that

$$\begin{aligned} & n^{-1} \sum_{i=1}^n \int_{I_l}^{I_{l+1}} [\beta_X(\hat{\mathbf{E}}_i(u) - \mathbf{E}_i(u)) + \frac{1}{2} \beta_X^T(\hat{\mathbf{V}}_i(u) - \mathbf{V}_i(u))\beta_X] dN_i(u) \\ & \leq n^{-1} \sum_{i=1}^n \int_{I_l}^{I_{l+1}} [|\beta_X(\hat{\mathbf{E}}_i(u) - \mathbf{E}_i(u))| + \frac{1}{2} |\beta_X^T(\hat{\mathbf{V}}_i(u) - \mathbf{V}_i(u))\beta_X|] dN_i(u) \\ & \leq n^{-1} \sum_{i=1}^n \int_{I_l}^{I_{l+1}} [2 \max |\beta_X| \sup_{\substack{u \in (0,1], \\ i=1, \dots, n}} \|\hat{\mathbf{E}}_i(u) - \mathbf{E}_i(u)\| + 2(\max |\beta_X|)^2 \sup_{\substack{u \in (0,1], \\ i=1, \dots, n}} \|\hat{\mathbf{V}}_i(u) - \mathbf{V}_i(u)\|] dN_i(u) \\ & = \frac{d_l}{n} [2 \max |\beta_X| \sup_{\substack{u \in (0,1], \\ i=1, \dots, n}} \|\hat{\mathbf{E}}_i(u) - \mathbf{E}_i(u)\| + 2(\max |\beta_X|)^2 \sup_{\substack{u \in (0,1], \\ i=1, \dots, n}} \|\hat{\mathbf{V}}_i(u) - \mathbf{V}_i(u)\|], \end{aligned}$$

where d_l is the total number of failures in $(I_l, I_{l+1}]$. Since $0 \leq d_l/n < 1$, and as $n \rightarrow \infty$, the rest converges to 0 due to Lemma 2. So expression (4.8) = $o_p(1)$.

For expression (4.9), it follows that

$$\begin{aligned}
& n^{-1} \int_{I_l}^{I_{l+1}} \{\log \hat{S}_0(\boldsymbol{\beta}, u) - \log S_0(\boldsymbol{\beta}, u)\} d\bar{N}(u) \\
& \leq n^{-1} \int_{I_l}^{I_{l+1}} \sup_{u \in (0,1]} \|\log \hat{S}_0(\boldsymbol{\beta}, u) - \log S_0(\boldsymbol{\beta}, u)\| d\bar{N}(u) \\
& = \frac{d_l}{n} \sup_{u \in (0,1]} \|\log \hat{S}_0(\boldsymbol{\beta}, u) - \log S_0(\boldsymbol{\beta}, u)\|
\end{aligned}$$

By the uniform convergence of $\hat{S}(\boldsymbol{\beta}, u)$ in Lemma 2 and the continuity of the log function, with similar arguments as used for expression (4.8), we can prove that

$$\sup_{u \in (0,1]} \|\log \hat{S}_0(\boldsymbol{\beta}, u) - \log S_0(\boldsymbol{\beta}, u)\| \xrightarrow{P} 0,$$

and expression (4.9) converges to 0 in probability, as $n \rightarrow \infty$.

Hence

$$LPL_{n,l}(\boldsymbol{\beta}, \hat{\mathcal{A}}) = LPL_{n,l}(\boldsymbol{\beta}, \mathcal{A}) + o_p(1),$$

and when adding across the fixed number of intervals L , we have

$$LPL_n(\boldsymbol{\beta}, \hat{\mathcal{A}}) = LPL_n(\boldsymbol{\beta}, \mathcal{A}) + o_p(1).$$

In the following proof, we replace $LPL_n(\boldsymbol{\beta}, \hat{\mathcal{A}})$ by its equivalent $LPL_n(\boldsymbol{\beta}, \mathcal{A})$.

Now we want to prove convergence of $LPL_n(\boldsymbol{\beta}, \mathcal{A})$ by expressing a closely related term as a martingale.

$$\begin{aligned}
X_{n,l}(\boldsymbol{\beta}, t) &= LPL_{n,l}(\boldsymbol{\beta}, \mathcal{A}, t) - LPL_{n,l}(\boldsymbol{\beta}^*, \mathcal{A}, t) \\
&= n^{-1} \left\{ \sum_{i=1}^n \int_{I_l}^t [(\beta_Z - \beta_Z^*) Z_i + (\boldsymbol{\beta}_X - \boldsymbol{\beta}_X^*)^T \mathbf{E}_i(u) + \frac{1}{2} (\boldsymbol{\beta}_X^T \mathbf{V}_i(u) \boldsymbol{\beta}_X - \boldsymbol{\beta}_X^{*T} \mathbf{V}_i(u) \boldsymbol{\beta}_X^*)] dN_i(u) \right. \\
&\quad \left. - \int_{I_l}^t \log \frac{S^{(0)}(\boldsymbol{\beta}, u)}{S^{(0)}(\boldsymbol{\beta}^*, u)} d\bar{N}(u) \right\}
\end{aligned}$$

Let

$$\begin{aligned}
& A_{n,l}(\boldsymbol{\beta}, t) \\
&= n^{-1} \sum_{i=1}^n \left\{ \int_{I_i}^t [(\beta_Z - \beta_Z^*)Z_i + (\boldsymbol{\beta}_X - \boldsymbol{\beta}_X^*)^T \mathbf{E}_i(u) + \frac{1}{2}(\boldsymbol{\beta}_X^T \mathbf{V}_i(u)\boldsymbol{\beta}_X - \boldsymbol{\beta}_X^{*T} \mathbf{V}_i(u)\boldsymbol{\beta}_X^*)] \right. \\
&\quad \times \exp(\beta_Z^* Z_i + \boldsymbol{\beta}_X^* \mathbf{E}_i(u) + \frac{1}{2}\boldsymbol{\beta}_X^* \mathbf{V}_i(u)\boldsymbol{\beta}_X^*) Y_i(u) \lambda_0(u) du \\
&\quad \left. - \int_{I_i}^t \log \frac{S^{(0)}(\boldsymbol{\beta}, u)}{S^{(0)}(\boldsymbol{\beta}^*, u)} \exp(\beta_Z^* Z_i + \boldsymbol{\beta}_X^* \mathbf{E}_i(u) + \frac{1}{2}\boldsymbol{\beta}_X^* \mathbf{V}_i(u)\boldsymbol{\beta}_X^*) Y_i(u) \lambda_0(u) du \right\} \\
&= n^{-1} \int_{I_i}^t \{ [(\beta_Z - \beta_Z^*)S_Z^{(1)}(\boldsymbol{\beta}^*, u) + (\boldsymbol{\beta}_X - \boldsymbol{\beta}_X^*)^T S_E^{(1)}(\boldsymbol{\beta}^*, u) \\
&\quad + \frac{1}{2}(\boldsymbol{\beta}_X^T S_V^{(1)}(\boldsymbol{\beta}^*, u)\boldsymbol{\beta}_X - \boldsymbol{\beta}_X^{*T} S_V^{(1)}(\boldsymbol{\beta}^*, u)\boldsymbol{\beta}_X^*) - \log \frac{S^{(0)}(\boldsymbol{\beta}, u)}{S^{(0)}(\boldsymbol{\beta}^*, u)} S^{(0)}(\boldsymbol{\beta}^*, u)] \lambda_0(u) \} du
\end{aligned}$$

Hence,

$$\begin{aligned}
& X_{n,l}(\boldsymbol{\beta}, t) - A_{n,l}(\boldsymbol{\beta}, t) \\
&= n^{-1} \sum_{i=1}^n \int_{I_i}^t \left\{ (\beta_Z - \beta_Z^*)Z_i + (\boldsymbol{\beta}_X - \boldsymbol{\beta}_X^*)^T \mathbf{E}_i(u) + \frac{1}{2}[\boldsymbol{\beta}_X^T \mathbf{V}_i(u)\boldsymbol{\beta}_X - \boldsymbol{\beta}_X^{*T} \mathbf{V}_i(u)\boldsymbol{\beta}_X^*] \right. \\
&\quad \left. - \log \frac{S^{(0)}(\boldsymbol{\beta}, u)}{S^{(0)}(\boldsymbol{\beta}^*, u)} \right\} dM_i(u)
\end{aligned}$$

Since $\mathbf{E}_i, \mathbf{V}_i$ are locally bounded and predictable (Condition 3), the integrand is also locally bounded and predictable. Hence $X_{n,l}(\boldsymbol{\beta}, t) - A_{n,l}(\boldsymbol{\beta}, t)$ is a locally square integrable martingale. The predictable variation process is

$$\begin{aligned}
& n \langle X_{n,l} - A_{n,l}, X_{n,l} - A_{n,l} \rangle (\boldsymbol{\beta}, t) \\
&= n^{-1} \sum_{i=1}^n \int_{I_i}^t \left\{ (\beta_Z - \beta_Z^*)Z_i + (\boldsymbol{\beta}_X - \boldsymbol{\beta}_X^*)^T \mathbf{E}_i(u) + \frac{1}{2}[\boldsymbol{\beta}_X^T \mathbf{V}_i(u)\boldsymbol{\beta}_X - \boldsymbol{\beta}_X^{*T} \mathbf{V}_i(u)\boldsymbol{\beta}_X^*] \right. \\
&\quad \left. - \log \frac{S^{(0)}(\boldsymbol{\beta}, u)}{S^{(0)}(\boldsymbol{\beta}^*, u)} \right\}^2 Y_i(u) \lambda_0(u) \exp(\beta_Z^* Z_i + \boldsymbol{\beta}_X^* \mathbf{E}_i(u) + \frac{1}{2}\boldsymbol{\beta}_X^* \mathbf{V}_i(u)\boldsymbol{\beta}_X^*) du \\
&= \frac{1}{n} \sum_{i=1}^n \int_{I_i}^t \left[(\beta_Z - \beta_Z^*)Z_i + (\boldsymbol{\beta}_X - \boldsymbol{\beta}_X^*)^T \mathbf{E}_i(u) + \frac{1}{2}(\boldsymbol{\beta}_X^T \mathbf{V}_i(u)\boldsymbol{\beta}_X - \boldsymbol{\beta}_X^{*T} \mathbf{V}_i(u)\boldsymbol{\beta}_X^*) \right]^2 \\
&\quad \times Y_i(u) \lambda_0(u) \exp(\beta_Z^* Z_i + \boldsymbol{\beta}_X^* \mathbf{E}_i(u) + \frac{1}{2}\boldsymbol{\beta}_X^* \mathbf{V}_i(u)\boldsymbol{\beta}_X^*) du \quad (4.10) \\
&- \frac{2}{n} \sum_{i=1}^n \int_0^t \left[(\beta_Z - \beta_Z^*)Z_i + (\boldsymbol{\beta}_X - \boldsymbol{\beta}_X^*)^T \mathbf{E}_i(u) + \frac{1}{2}(\boldsymbol{\beta}_X^T \mathbf{V}_i(u)\boldsymbol{\beta}_X - \boldsymbol{\beta}_X^{*T} \mathbf{V}_i(u)\boldsymbol{\beta}_X^*) \right] \log \frac{S^{(0)}(\boldsymbol{\beta}, u)}{S^{(0)}(\boldsymbol{\beta}^*, u)}
\end{aligned}$$

$$\times Y_i(u)\lambda_0(u) \exp(\beta_Z^* Z_i + \beta_X^* \mathbf{E}_i(u) + \frac{1}{2}\beta_X^* \mathbf{V}_i(u)\beta_X^*) du \quad (4.11)$$

$$+ \frac{1}{n} \sum_{i=1}^n \int_0^t \left[\log \frac{S^{(0)}(\boldsymbol{\beta}, u)}{S^{(0)}(\boldsymbol{\beta}^*, u)} \right]^2 \times Y_i(u)\lambda_0(u) \exp(\beta_Z^* Z_i + \beta_X^* \mathbf{E}_i(u) + \frac{1}{2}\beta_X^* \mathbf{V}_i(u)\beta_X^*) du. \quad (4.12)$$

Expression (4.10)

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \int_{I_l}^t \left[(\beta_Z - \beta_Z^*) Z_i + (\beta_X - \beta_X^*)^T \mathbf{E}_i(u) + \frac{1}{2}(\beta_X^T \mathbf{V}_i(u)\beta_X - \beta_X^{*T} \mathbf{V}_i(u)\beta_X^*) \right]^2 \\ & \times Y_i(u)\lambda_0(u) \exp(\beta_Z^* Z_i + \beta_X^* \mathbf{E}_i(u) + \frac{1}{2}\beta_X^* \mathbf{V}_i(u)\beta_X^*) du \end{aligned}$$

has finite limit, because by Condition 1 and 3, each component of the integrand is bounded.

Expression (4.11)

$$\begin{aligned} & - \frac{2}{n} \sum_{i=1}^n \int_{I_l}^t \left[(\beta_Z - \beta_Z^*) Z_i + (\beta_X - \beta_X^*)^T \mathbf{E}_i(u) + \frac{1}{2}(\beta_X^T \mathbf{V}_i(u)\beta_X - \beta_X^{*T} \mathbf{V}_i(u)\beta_X^*) \right] \log \frac{S^{(0)}(\boldsymbol{\beta}, u)}{S^{(0)}(\boldsymbol{\beta}^*, u)} \\ & \times Y_i(u)\lambda_0(u) \exp(\beta_Z^* Z_i + \beta_X^* \mathbf{E}_i(u) + \frac{1}{2}\beta_X^* \mathbf{V}_i(u)\beta_X^*) du \\ = & - \frac{2}{n} \int_{I_l}^t \left[(\beta_Z - \beta_Z^*) S_Z^{(1)}(\boldsymbol{\beta}^*, u) + (\beta_X - \beta_X^*)^T S_E^{(1)}(\boldsymbol{\beta}^*, u) \right. \\ & \left. + \frac{1}{2}(\beta_X^T S_V^{(1)}(\boldsymbol{\beta}^*, u)\beta_X - \beta_X^{*T} S_V^{(1)}(\boldsymbol{\beta}^*, u)\beta_X^*) \log \frac{S^{(0)}(\boldsymbol{\beta}, u)}{S^{(0)}(\boldsymbol{\beta}^*, u)} \right] \lambda_0(u) du \\ \rightarrow & - \frac{2}{n} \int_{I_l}^t \left[(\beta_Z - \beta_Z^*) s_Z^{(1)}(\boldsymbol{\beta}^*, u) + (\beta_X - \beta_X^*)^T s_E^{(1)}(\boldsymbol{\beta}^*, u) \right. \\ & \left. + \frac{1}{2}(\beta_X^T s_V^{(1)}(\boldsymbol{\beta}^*, u)\beta_X - \beta_X^{*T} s_V^{(1)}(\boldsymbol{\beta}^*, u)\beta_X^*) \log \frac{s^{(0)}(\boldsymbol{\beta}, u)}{s^{(0)}(\boldsymbol{\beta}^*, u)} \right] \lambda_0(u) du \end{aligned}$$

by Condition 4. Similarly, expression (4.12)

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \int_{I_l}^t \left[\log \frac{S^{(0)}(\boldsymbol{\beta}, u)}{S^{(0)}(\boldsymbol{\beta}^*, u)} \right]^2 Y_i(u)\lambda_0(u) \exp(\beta_Z^* Z_i + \beta_X^* \mathbf{E}_i(u) + \frac{1}{2}\beta_X^* \mathbf{V}_i(u)\beta_X^*) du \\ = & \int_{I_l}^t \left[\log \frac{S^{(0)}(\boldsymbol{\beta}, u)}{S^{(0)}(\boldsymbol{\beta}^*, u)} \right]^2 S^{(0)}(\boldsymbol{\beta}^*, u)\lambda_0(u) du \\ \rightarrow & \int_0^t \left[\log \frac{s^{(0)}(\boldsymbol{\beta}, u)}{s^{(0)}(\boldsymbol{\beta}^*, u)} \right]^2 s^{(0)}(\boldsymbol{\beta}^*, u)\lambda_0(u) du. \end{aligned}$$

By the boundedness of $s^{(k)}$, $k = 0, 1$ in Condition 5, the limits of expression (4.11) and (4.12) are finite. Hence $n < X_{n,l} - A_{n,l}, X_{n,l} - A_{n,l} >$ converge to a finite limit. This indicates that

$$\lim_{n \rightarrow \infty} [X_{n,l}(\boldsymbol{\beta}, I_{l+1}) - A_{n,l}(\boldsymbol{\beta}, I_{l+1})] = 0$$

in probability. By Condition 1 and the uniform convergence of $S^{(k)}(\boldsymbol{\beta}, u)$, $k = 0, 1$ in Condition 4, $A_{n,l}(\boldsymbol{\beta}, I_{l+1})$ converges to

$$A_{.,l}(\boldsymbol{\beta}, I_{l+1}) = \int_{I_l}^{I_{l+1}} \{(\boldsymbol{\beta}_Z - \boldsymbol{\beta}_Z^*)^T s_Z^{(1)}(\boldsymbol{\beta}^*, u) + (\boldsymbol{\beta}_X - \boldsymbol{\beta}_X^*)^T s_E^{(1)}(\boldsymbol{\beta}^*, u) \\ + \frac{1}{2}[\boldsymbol{\beta}_X^T s_V^{(1)}(\boldsymbol{\beta}^*, u)\boldsymbol{\beta}_X - \boldsymbol{\beta}_X^{*T} s_V^{(1)}(\boldsymbol{\beta}^*, u)\boldsymbol{\beta}_X^*] - \log \frac{s^{(0)}(\boldsymbol{\beta}, u)}{s^{(0)}(\boldsymbol{\beta}^*, u)} s^{(0)}(\boldsymbol{\beta}^*, u)\} \lambda_0(u) du$$

then $X_{n,l}(\boldsymbol{\beta}, I_{l+1})$ converges to the same limit, as long as $\boldsymbol{\beta} \in \mathcal{B}$. Adding $X_{n,l}(\boldsymbol{\beta}, I_{l+1})$ across time intervals, then $X_n(\boldsymbol{\beta}) = \sum_{l=1}^L X_{n,l}(\boldsymbol{\beta}, I_{l+1})$ converges to the same limit as $A_n(\boldsymbol{\beta}) = \sum_{l=1}^L A_{n,l}(\boldsymbol{\beta}, I_{l+1})$, and the limit is

$$A(\boldsymbol{\beta}) = \sum_{l=1}^L A_{.,l}(\boldsymbol{\beta}, I_{l+1}) \\ = \int_0^1 \{(\boldsymbol{\beta}_Z - \boldsymbol{\beta}_Z^*)^T s_Z^{(1)}(\boldsymbol{\beta}^*, u) + (\boldsymbol{\beta}_X - \boldsymbol{\beta}_X^*)^T s_E^{(1)}(\boldsymbol{\beta}^*, u) \\ + \frac{1}{2}[\boldsymbol{\beta}_X^T s_V^{(1)}(\boldsymbol{\beta}^*, u)\boldsymbol{\beta}_X - \boldsymbol{\beta}_X^{*T} s_V^{(1)}(\boldsymbol{\beta}^*, u)\boldsymbol{\beta}_X^*] - \log \frac{s^{(0)}(\boldsymbol{\beta}, u)}{s^{(0)}(\boldsymbol{\beta}^*, u)} s^{(0)}(\boldsymbol{\beta}^*, u)\} \lambda_0(u) du$$

The first derivative of $A(\boldsymbol{\beta})$ is

$$A^{(1)}(\boldsymbol{\beta}) = \int_0^1 \left\{ \begin{pmatrix} s_Z^{(1)}(\boldsymbol{\beta}^*, u) \\ s_E^{(1)}(\boldsymbol{\beta}^*, u) + s_V^{(1)}(\boldsymbol{\beta}^*, u)\boldsymbol{\beta}_X \end{pmatrix} - \frac{s^{(1)}(\boldsymbol{\beta}, u)}{s^{(0)}(\boldsymbol{\beta}, u)} s^{(0)}(\boldsymbol{\beta}^*, u) \right\} \lambda_0(u) du$$

$A^{(1)}(\boldsymbol{\beta})$ is equal to 0 when $\boldsymbol{\beta} = \boldsymbol{\beta}^*$. The second derivative of $A(\boldsymbol{\beta})$ is

$$A^{(2)}(\boldsymbol{\beta}) = \int_0^1 \left\{ \begin{pmatrix} 0 & 0 \\ 0 & s_V^{(1)}(\boldsymbol{\beta}^*, u) \end{pmatrix} - \frac{s^{(2)}(\boldsymbol{\beta}, u)s^{(0)}(\boldsymbol{\beta}, u) - [s^{(1)}(\boldsymbol{\beta}, u)]^{\otimes 2}}{[s^{(0)}(\boldsymbol{\beta}, u)]^2} s^{(0)}(\boldsymbol{\beta}^*, u) \right\} \lambda_0(u) du \\ = \Sigma(\boldsymbol{\beta}).$$

According to Condition 6, $\Sigma(\boldsymbol{\beta})$ is negative semidefinite for $\boldsymbol{\beta} \in \mathcal{B}$, and is negative definite at $\boldsymbol{\beta} = \boldsymbol{\beta}^*$. Hence $A(\boldsymbol{\beta})$ is a concave function with a unique maximum at $\boldsymbol{\beta}^*$. Since $X_n(\boldsymbol{\beta})$ is also a concave function with a unique maximum at $\hat{\boldsymbol{\beta}}$, then by Lemma 1, $\hat{\boldsymbol{\beta}} \xrightarrow{P} \boldsymbol{\beta}^*$. ■

This completes our proof of the consistency of $\hat{\boldsymbol{\beta}}$ to $\boldsymbol{\beta}^*$, which is the solution to $E[U_n(\boldsymbol{\beta}, \mathcal{A})] = 0$. $\boldsymbol{\beta}^*$ is generally biased. However, the biases are usually small according to simulation studies.

4.4 Asymptotic Normality

Theorem 5 *Under regularity conditions,*

$$n^{1/2}(\hat{\beta} - \beta^*) \xrightarrow{D} N(0, \Sigma(\beta^*)^{-1}[B(\beta^*) + D(\beta^*)]\Sigma(\beta^*)^{-1}),$$

where

$$\begin{aligned} \Sigma(\beta) &= \int_0^1 \left\{ \begin{pmatrix} 0 & 0 \\ 0 & s_V^{(1)}(\beta^*, u) \end{pmatrix} - \frac{s^{(2)}(\beta, u)s^{(0)}(\beta, u) - [s^{(1)}(\beta, u)]^{\otimes 2}}{[s^{(0)}(\beta, u)]^2} s^{(0)}(\beta^*, u) \right\} \lambda_0(u) du, \\ B(\beta) &= E \left\{ \int_0^1 \left[\begin{pmatrix} Z_1 \\ \mathbf{E}_1(u) + \mathbf{V}_1(u)\beta_X \end{pmatrix} - \frac{s^{(1)}(\beta, u)}{s^{(0)}(\beta, u)} \right] dN_1(u) - \sum_{l=1}^L F_l(W_1, Z_1, \mathcal{A}(I_l)) D_l(\beta) \right\}^{\otimes 2}, \\ D(\beta) &= \left[\sum_{l=1}^L D_l(\beta) C(I_l) \right] V_q \left[\sum_{l=1}^L D_l(\beta) C(I_l) \right]^T, \end{aligned}$$

and $D_l(\cdot), F_l(\cdot), C(\cdot), V_q$ are defined later.

Proof. By Taylor expansion,

$$U_n(\beta, \hat{\mathcal{A}}) - U_n(\beta^*, \hat{\mathcal{A}}) = \mathcal{I}_n(\beta^{**}, \hat{\mathcal{A}})(\beta - \beta^*), \quad (4.13)$$

where β^{**} is on the line segment between β and β^* . Inserting $\hat{\beta}$ into (4.13), we get

$$n^{1/2}U_n(\beta^*, \hat{\mathcal{A}}) = -\mathcal{I}_n(\beta^{**}, \hat{\mathcal{A}})[n^{1/2}(\hat{\beta} - \beta^*)], \quad (4.14)$$

since $U_n(\hat{\beta}, \hat{\mathcal{A}}) = 0$.

In order to find the asymptotic normality of $n^{1/2}(\hat{\beta} - \beta^*)$, we need to prove weak convergence to a Gaussian process of the local martingale $n^{1/2}U_n(\beta^*, \hat{\mathcal{A}})$ and to prove convergence in probability to a non-singular matrix of $\mathcal{I}_n(\beta^{**}, \hat{\mathcal{A}})$.

Claim 1

$$\mathcal{I}_n(\beta^{**}, \hat{\mathcal{A}}) \xrightarrow{P} \Sigma(\beta^*).$$

as $n \rightarrow \infty$.

We now prove Claim 1 by first decomposing $\|\mathcal{I}_n(\beta^{**}, \hat{\mathcal{A}}) - \Sigma(\beta^*)\|$ into two parts:

$$\|\mathcal{I}_n(\beta^{**}, \hat{\mathcal{A}}) - \Sigma(\beta^*)\| \leq \|\mathcal{I}_n(\beta^{**}, \hat{\mathcal{A}}) - \mathcal{I}_n(\beta^{**}, \mathcal{A})\| + \|\mathcal{I}_n(\beta^{**}, \mathcal{A}) - \Sigma(\beta^*)\|. \quad (4.15)$$

For interval $(I_l, I_{l+1}]$,

$$\begin{aligned}
& \|\mathcal{I}_{n,l}(\boldsymbol{\beta}^{**}, \hat{\mathcal{A}}) - \mathcal{I}_{n,l}(\boldsymbol{\beta}^{**}, \mathcal{A})\| \\
& \leq \left\| \frac{1}{n} \sum_{i=1}^n \int_{I_l}^{I_{l+1}} \begin{pmatrix} 0 & 0 \\ 0 & \hat{\mathbf{V}}_i(u) - \mathbf{V}_i(u) \end{pmatrix} dN_i(u) \right\| \\
& + \left\| \frac{1}{n} \int_{I_l}^{I_{l+1}} \left[\frac{\hat{S}^{(2)}(\boldsymbol{\beta}^{**}, u)}{\hat{S}^{(0)}(\boldsymbol{\beta}^{**}, u)} - \frac{S^{(2)}(\boldsymbol{\beta}^{**}, u)}{S^{(0)}(\boldsymbol{\beta}^{**}, u)} \right] d\bar{N}(u) \right\| \\
& + \left\| \frac{1}{n} \int_{I_l}^{I_{l+1}} \left[\left(\frac{\hat{S}^{(1)}(\boldsymbol{\beta}^{**}, u)}{\hat{S}^{(0)}(\boldsymbol{\beta}^{**}, u)} \right)^{\otimes 2} - \left(\frac{S^{(1)}(\boldsymbol{\beta}^{**}, u)}{S^{(0)}(\boldsymbol{\beta}^{**}, u)} \right)^{\otimes 2} \right] d\bar{N}(u) \right\|.
\end{aligned}$$

Each part can be proven to be $o_p(1)$ due to Lemma 2. Thus, summing over the finite number of intervals, we have

$$\mathcal{I}_n(\boldsymbol{\beta}^{**}, \hat{\mathcal{A}}) = \mathcal{I}_n(\boldsymbol{\beta}^{**}, \mathcal{A}) + o_p(1).$$

The second part of the right side of (4.15) is

$$\|\mathcal{I}_n(\boldsymbol{\beta}^{**}, \mathcal{A}) - \Sigma(\boldsymbol{\beta}^*)\| \leq \sum_{l=1}^L \|\mathcal{I}_{n,l}(\boldsymbol{\beta}^{**}, \mathcal{A}) - \Sigma_l(\boldsymbol{\beta}^*)\|$$

where

$$\Sigma_l(\boldsymbol{\beta}^*) = \int_{I_l}^{I_{l+1}} \left\{ \begin{pmatrix} 0 & 0 \\ 0 & s_V^{(1)}(\boldsymbol{\beta}^*, u) \end{pmatrix} - \frac{s^{(2)}(\boldsymbol{\beta}^*, u)s^{(0)}(\boldsymbol{\beta}^*, u) - [s^{(1)}(\boldsymbol{\beta}^*, u)]^{\otimes 2}}{s^{(0)}(\boldsymbol{\beta}^*, u)} \right\} \lambda_0(u) du$$

For each interval,

$$\begin{aligned}
& \|\mathcal{I}_{n,l}(\boldsymbol{\beta}^{**}, \mathcal{A}) - \Sigma_l(\boldsymbol{\beta}^*)\| \\
& \leq \left\| \frac{1}{n} \sum_{i=1}^n \int_{I_l}^{I_{l+1}} \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{V}_i(u) \end{pmatrix} dN_i(u) - \int_{I_l}^{I_{l+1}} \begin{pmatrix} 0 & 0 \\ 0 & s_V^{(1)}(\boldsymbol{\beta}^*, u) \end{pmatrix} \lambda_0(u) du \right\| \quad (4.16)
\end{aligned}$$

$$\begin{aligned}
& + \left\| \int_{I_l}^{I_{l+1}} \frac{S^{(2)}(\boldsymbol{\beta}^{**}, u)S^{(0)}(\boldsymbol{\beta}^{**}, u) - [S^{(1)}(\boldsymbol{\beta}^{**}, u)]^{\otimes 2}}{[S^{(0)}(\boldsymbol{\beta}^{**}, u)]^2} \frac{d\bar{N}(u)}{n} \right. \\
& \quad \left. - \int_{I_l}^{I_{l+1}} \frac{s^{(2)}(\boldsymbol{\beta}^*, u)s^{(0)}(\boldsymbol{\beta}^*, u) - [s^{(1)}(\boldsymbol{\beta}^*, u)]^{\otimes 2}}{s^{(0)}(\boldsymbol{\beta}^*, u)} \lambda_0(u) du \right\| \quad (4.17)
\end{aligned}$$

First,

$$\begin{aligned}
& \left\| \frac{1}{n} \sum_{i=1}^n \int_{I_i}^{I_{i+1}} \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{V}_i(I_i) \end{pmatrix} dN_i(u) - \int_{I_i}^{I_{i+1}} \begin{pmatrix} 0 & 0 \\ 0 & s_V^{(1)}(\boldsymbol{\beta}^*, u) \end{pmatrix} \lambda_0(u) du \right\| \\
&= \left\| \frac{1}{n} \sum_{i=1}^n \int_{I_i}^{I_{i+1}} \mathbf{V}_i(u) dN_i(u) - \int_{I_i}^{I_{i+1}} s_V^{(1)}(\boldsymbol{\beta}^*, u) \lambda_0(u) du \right\| \\
&\leq \left\| \frac{1}{n} \sum_{i=1}^n \int_{I_i}^{I_{i+1}} \mathbf{V}_i(u) [dN_i(u) - \lambda_i(u) du] \right\| + \left\| \frac{1}{n} \sum_{i=1}^n \int_{I_i}^{I_{i+1}} \mathbf{V}_i(u) \lambda_i(u) du - \int_{I_i}^{I_{i+1}} s_V^{(1)}(\boldsymbol{\beta}^*, u) \lambda_0(u) du \right\| \\
&= \left\| \frac{1}{n} \sum_{i=1}^n \int_{I_i}^{I_{i+1}} \mathbf{V}_i(u) dM_i^*(u) \right\| + \left\| \int_{I_i}^{I_{i+1}} [S_V^{(1)}(\boldsymbol{\beta}^*, u) - s_V^{(1)}(\boldsymbol{\beta}^*, u)] \lambda_0(u) du \right\|
\end{aligned}$$

where

$$dM_i^*(u) = d[N_i(u) - \lambda_0(u) \exp(\beta_Z^* Z_i + \boldsymbol{\beta}_X^{*T} \mathbf{E}_i(u) + \frac{1}{2} \boldsymbol{\beta}_X^{*T} \mathbf{V}_i(u) \boldsymbol{\beta}_X^*)].$$

Here,

$$\left\| \frac{1}{n} \sum_{i=1}^n \int_{I_i}^{I_{i+1}} \mathbf{V}_i(u) dM_i^*(u) \right\| = o_p(1)$$

since $\int_{I_i}^{I_{i+1}} \mathbf{V}_i(u) dM_i^*(u)$ is a martingale with expectation 0. By the SLLN in Theorem 1, this part is $o_p(1)$. And

$$\left\| \int_{I_i}^{I_{i+1}} [S_V^{(1)}(\boldsymbol{\beta}^*, u) - s_V^{(1)}(\boldsymbol{\beta}^*, u)] \lambda_0(u) du \right\|$$

is also $o_p(1)$, due to Conditions 1 and 4. Thus, (4.16) is $o_p(1)$. Now denote

$$\begin{aligned}
K_s(\boldsymbol{\beta}, u) &= \frac{S^{(2)}(\boldsymbol{\beta}, u) S^{(0)}(\boldsymbol{\beta}, u) - [S^{(1)}(\boldsymbol{\beta}, u)]^{\otimes 2}}{[S^{(0)}(\boldsymbol{\beta}, u)]^2} \\
k_s(\boldsymbol{\beta}, u) &= \frac{s^{(2)}(\boldsymbol{\beta}, u) s^{(0)}(\boldsymbol{\beta}, u) - [s^{(1)}(\boldsymbol{\beta}, u)]^{\otimes 2}}{[s^{(0)}(\boldsymbol{\beta}, u)]^2},
\end{aligned}$$

then $\sup_{u \in (0,1]} \|K_s(\boldsymbol{\beta}, u) - k_s(\boldsymbol{\beta}, u)\| \xrightarrow{P} 0$ by Condition 4. Then Equation (4.17) can be decomposed into four parts:

$$\begin{aligned}
& \left\| \int_{I_i}^{I_{i+1}} K_s(\boldsymbol{\beta}^{**}, u) \frac{d\bar{N}}{n}(u) - \int_{I_i}^{I_{i+1}} k_s(\boldsymbol{\beta}^*, u) \lambda(u) du \right\| \\
&\leq \left\| \int_{I_i}^{I_{i+1}} \{K_s(\boldsymbol{\beta}^{**}, u) - k_s(\boldsymbol{\beta}^{**}, u)\} \frac{d\bar{N}}{n}(u) \right\| + \left\| \int_{I_i}^{I_{i+1}} \{k_s(\boldsymbol{\beta}^{**}, u) - k_s(\boldsymbol{\beta}^*, u)\} \frac{d\bar{N}}{n}(u) \right\| \\
&+ \left\| \int_{I_i}^{I_{i+1}} k_s(\boldsymbol{\beta}^*, u) \left\{ \frac{d\bar{N}}{n}(u) - \frac{\bar{\lambda}(u)}{n} du \right\} \right\| + \left\| \int_{I_i}^{I_{i+1}} k_s(\boldsymbol{\beta}^*, u) \{S^{(0)}(\boldsymbol{\beta}^*, u) - s^{(0)}(\boldsymbol{\beta}^*, u)\} \lambda_0(u) du \right\|,
\end{aligned}$$

where $\bar{\lambda}(u) = \sum_{i=1}^n \lambda_i(u)$. By Corollary 1(1),

$$\Pr\left(\frac{\bar{N}(I_{l+1})}{n} > \eta\right) \leq \frac{\delta}{\eta} + \Pr\left(\int_{I_l}^{I_{l+1}} S^{(0)}(\boldsymbol{\beta}^*, u) \lambda_0(u) du > \delta\right).$$

For $\delta > \int_{I_l}^{I_{l+1}} s^{(0)}(\boldsymbol{\beta}^*, u) \lambda_0(u) du$, the latter probability tends to 0 as $n \rightarrow \infty$. Hence as $\eta \uparrow \infty$, we have

$$\lim_{\eta \uparrow \infty} \lim_{n \rightarrow \infty} \Pr\left(\frac{\bar{N}(I_{l+1})}{n} > \eta\right) = 0.$$

Together with $\sup_{u \in (0,1]} \|K_s(\boldsymbol{\beta}, u) - k_s(\boldsymbol{\beta}, u)\| \xrightarrow{P} 0$ and $\boldsymbol{\beta}^{**} \xrightarrow{P} \boldsymbol{\beta}^*$, we can conclude the first term of the above decomposition is $o_p(1)$. Also, the continuity of $k_s(\boldsymbol{\beta}, u)$ in $\boldsymbol{\beta}$, uniformly in u and $\boldsymbol{\beta}^{**} \xrightarrow{P} \boldsymbol{\beta}^*$ imply that the second term is $o_p(1)$. By Corollary 1(2), for each component of $k_s(\boldsymbol{\beta}^*, u)$, it follows that

$$\Pr\left(\left|\int_{I_l}^{I_{l+1}} k_s(\boldsymbol{\beta}^*, u) \frac{d\bar{M}^*}{n}(u)\right| > \delta\right) \leq \frac{\eta}{\delta^2} + \Pr\left(\frac{1}{n} \int_{I_l}^{I_{l+1}} \{k_s(\boldsymbol{\beta}^*, u)\}^2 S^{(0)}(\boldsymbol{\beta}^*, u) \lambda_0(u) du > \eta\right)$$

When $\eta < \frac{1}{n} \int_{I_l}^{I_{l+1}} \{k_s(\boldsymbol{\beta}^*, u)\}^2 s^{(0)}(\boldsymbol{\beta}^*, u) \lambda_0(u) du$, the probability on the right side tends to 0 as $n \rightarrow \infty$. As $n \rightarrow \infty$, η can get closer to 0, which implies that η/δ^2 can converge to 0. Hence the third part is $o_p(1)$. By Condition 1, 4, 5, the last part is $o_p(1)$. So Equation (4.17) is also $o_p(1)$, and

$$\|\mathcal{I}_{n,l}(\boldsymbol{\beta}^{**}, \mathcal{A}) - \Sigma_l(\boldsymbol{\beta}^*)\| = o_p(1).$$

Adding over finite intervals, we have

$$\|\mathcal{I}_n(\boldsymbol{\beta}^{**}, \mathcal{A}) - \Sigma(\boldsymbol{\beta}^*)\| = o_p(1).$$

This completes the proof of Claim 1.

Claim 2

$$n^{1/2} U_n(\boldsymbol{\beta}^*, \hat{\mathcal{A}}) \xrightarrow{D} N(0, B(\boldsymbol{\beta}^*) + D(\boldsymbol{\beta}^*))$$

as $n \rightarrow \infty$.

In this proof, we assume that subjects are sorted by their event time t , and the number of subjects having events in interval $(I_l, I_{l+1}]$ is $n_l - n_{l+1}$, where $n_1 = n$ and $n_{L+1} = 0$. This implies that

$$0 = I_1 < t_1 < \dots \leq I_l < t_{n_1 - n_{l+1}} < \dots < t_{n_1 - n_{l+1}} \leq I_{l+1} < \dots < t_n < I_{L+1} = 1$$

By the definition of $U_n(\boldsymbol{\beta}, \hat{\mathcal{A}})$, we have

$$n^{1/2}[U_n(\boldsymbol{\beta}^*, \hat{\mathcal{A}}) - U_n(\boldsymbol{\beta}^*, \mathcal{A})] = n^{1/2} \sum_{l=1}^L [U_{n,l}(\boldsymbol{\beta}^*, \hat{\mathcal{A}}) - U_{n,l}(\boldsymbol{\beta}^*, \mathcal{A})],$$

By similar arguments as in Wang et al. (1997),

$$\begin{aligned} & n^{1/2}[U_{n,l}(\boldsymbol{\beta}^*, \hat{\mathcal{A}}) - U_{n,l}(\boldsymbol{\beta}^*, \mathcal{A})] \\ &= n^{-1/2} \sum_{i=n_1-n_l+1}^{n_1-n_l+1} \delta_i \left\{ \begin{pmatrix} 0 \\ [\hat{\mathbf{E}}_i(T_i) - \mathbf{E}_i(T_i)] + [\hat{\mathbf{V}}_i(T_i) - \mathbf{V}_i(T_i)]\boldsymbol{\beta}_X^* \end{pmatrix} - \left[\frac{\hat{S}^{(1)}(\boldsymbol{\beta}^*, T_i)}{\hat{S}^{(0)}(\boldsymbol{\beta}^*, T_i)} - \frac{S^{(1)}(\boldsymbol{\beta}^*, T_i)}{S^{(0)}(\boldsymbol{\beta}^*, T_i)} \right] \right\} \\ &= n^{-1/2} \sum_{i=n_1-n_l+1}^{n_1-n_l+1} \delta_i \left\{ \begin{pmatrix} 0 \\ [\hat{\mathbf{E}}_i(T_i) - \mathbf{E}_i(T_i)] + [\hat{\mathbf{V}}_i(T_i) - \mathbf{V}_i(T_i)]\boldsymbol{\beta}_X^* \end{pmatrix} - \frac{\hat{S}^{(1)}(\boldsymbol{\beta}^*, T_i) - S^{(1)}(\boldsymbol{\beta}^*, T_i)}{S^{(0)}(\boldsymbol{\beta}^*, T_i)} \right. \\ &\quad \left. + \frac{S^{(1)}(\boldsymbol{\beta}^*, T_i)[\hat{S}^{(0)}(\boldsymbol{\beta}^*, T_i) - S^{(0)}(\boldsymbol{\beta}^*, T_i)]}{[S^{(0)}(\boldsymbol{\beta}^*, T_i)]^2} \right\} + o_p(1) \end{aligned} \quad (4.18)$$

by Lemma 2. Here by Taylor expansion,

$$\begin{aligned} & \hat{S}^{(0)}(\boldsymbol{\beta}^*, T_i) - S^{(0)}(\boldsymbol{\beta}^*, T_i) \\ &= \frac{1}{n} \sum_{j=1}^n Y_j(T_i) \exp(\boldsymbol{\beta}_Z^* Z_j) \{ \exp[\boldsymbol{\beta}_X^{*T} \hat{\mathbf{E}}_j(T_i) + \frac{1}{2} \boldsymbol{\beta}_X^{*T} \hat{\mathbf{V}}_j(T_i) \boldsymbol{\beta}_X^*] - \exp[\boldsymbol{\beta}_X^{*T} \mathbf{E}_j(T_i) + \frac{1}{2} \boldsymbol{\beta}_X^{*T} \mathbf{V}_j(T_i) \boldsymbol{\beta}_X^*] \} \\ &= \frac{1}{n} \sum_{j=1}^n Y_j(T_i) \exp[\boldsymbol{\beta}_Z^* Z_j + \boldsymbol{\beta}_X^{*T} \mathbf{E}_j(T_i) + \frac{1}{2} \boldsymbol{\beta}_X^{*T} \mathbf{V}_j(T_i) \boldsymbol{\beta}_X^*] \\ &\quad \times \{ \boldsymbol{\beta}_X^{*T} [\hat{\mathbf{E}}_j(T_i) - \mathbf{E}_j(T_i)] + \frac{1}{2} \boldsymbol{\beta}_X^{*T} [\hat{\mathbf{V}}_j(T_i) - \mathbf{V}_j(T_i)] \boldsymbol{\beta}_X^* \} + o_p(1) \end{aligned}$$

and

$$\hat{S}^{(1)}(\boldsymbol{\beta}^*, T_i) - S^{(1)}(\boldsymbol{\beta}^*, T_i) = \begin{pmatrix} \hat{S}_Z^{(1)}(\boldsymbol{\beta}^*, T_i) - S_Z^{(1)}(\boldsymbol{\beta}^*, T_i) \\ [\hat{S}_E^{(1)}(\boldsymbol{\beta}^*, T_i) - S_E^{(1)}(\boldsymbol{\beta}^*, T_i)] + [\hat{S}_V^{(1)}(\boldsymbol{\beta}^*, T_i) - S_V^{(1)}(\boldsymbol{\beta}^*, T_i)] \boldsymbol{\beta}_X^* \end{pmatrix}$$

where

$$\begin{aligned} & \hat{S}_Z^{(1)}(\boldsymbol{\beta}^*, T_i) - S_Z^{(1)}(\boldsymbol{\beta}^*, T_i) \\ &= \frac{1}{n} \sum_{j=1}^n Y_j(T_i) Z_j \exp(\boldsymbol{\beta}_Z^* Z_j) \{ \exp[\boldsymbol{\beta}_X^{*T} \hat{\mathbf{E}}_j(T_i) + \frac{1}{2} \boldsymbol{\beta}_X^{*T} \hat{\mathbf{V}}_j(T_i) \boldsymbol{\beta}_X^*] - \exp[\boldsymbol{\beta}_X^{*T} \mathbf{E}_j(T_i) + \frac{1}{2} \boldsymbol{\beta}_X^{*T} \mathbf{V}_j(T_i) \boldsymbol{\beta}_X^*] \} \\ &= \frac{1}{n} \sum_{j=1}^n Y_j(T_i) Z_j \exp[\boldsymbol{\beta}_Z^* Z_j + \boldsymbol{\beta}_X^{*T} \mathbf{E}_j(T_i) + \frac{1}{2} \boldsymbol{\beta}_X^{*T} \mathbf{V}_j(T_i) \boldsymbol{\beta}_X^*] \end{aligned}$$

$$\begin{aligned}
& \times \{\boldsymbol{\beta}_X^{*T} [\hat{\mathbf{E}}_j(T_i) - \mathbf{E}_j(T_i)] + \frac{1}{2} \boldsymbol{\beta}_X^{*T} [\hat{\mathbf{V}}_j(T_i) - \mathbf{V}_j(T_i)] \boldsymbol{\beta}_X^*\} + o_p(1) \\
& \hat{S}_E^{(1)}(\boldsymbol{\beta}^*, T_i) - S_E^{(1)}(\boldsymbol{\beta}^*, T_i) \\
& = n^{-1} \sum_{j=1}^n Y_j(T_i) \exp(\boldsymbol{\beta}_Z^* Z_j) \\
& \quad \times \{\hat{\mathbf{E}}_j(T_i) \exp[\boldsymbol{\beta}_X^{*T} \hat{\mathbf{E}}_j(T_i) + \frac{1}{2} \boldsymbol{\beta}_X^{*T} \hat{\mathbf{V}}_j(T_i) \boldsymbol{\beta}_X^*] - \mathbf{E}_j(T_i) \exp[\boldsymbol{\beta}_X^{*T} \mathbf{E}_j(T_i) + \frac{1}{2} \boldsymbol{\beta}_X^{*T} \mathbf{V}_j(T_i) \boldsymbol{\beta}_X^*]\} \\
& = n^{-1} \sum_{j=1}^n Y_j(T_i) \exp[\boldsymbol{\beta}_Z^* Z_j + \boldsymbol{\beta}_X^{*T} \mathbf{E}_j(T_i) + \frac{1}{2} \boldsymbol{\beta}_X^{*T} \mathbf{V}_j(T_i) \boldsymbol{\beta}_X^*] \\
& \quad \times \{[I + \mathbf{E}_j(T_i) \boldsymbol{\beta}_X^{*T}]^T [\hat{\mathbf{E}}_j(T_i) - \mathbf{E}_j(T_i)] + \frac{1}{2} \mathbf{E}_j(T_i) \boldsymbol{\beta}_X^{*T} [\hat{\mathbf{V}}_j(T_i) - \mathbf{V}_j(T_i)] \boldsymbol{\beta}_X^*\} + o_p(1) \\
& [\hat{S}_V^{(1)}(\boldsymbol{\beta}^*, T_i) - S_V^{(1)}(\boldsymbol{\beta}^*, T_i)] \boldsymbol{\beta}_X^* \\
& = n^{-1} \sum_{j=1}^n Y_j(T_i) \exp(\boldsymbol{\beta}_Z^* Z_j) \\
& \quad \times \{\hat{\mathbf{V}}_j(T_i) \exp[\boldsymbol{\beta}_X^{*T} \hat{\mathbf{E}}_j(T_i) + \frac{1}{2} \boldsymbol{\beta}_X^{*T} \hat{\mathbf{V}}_j(T_i) \boldsymbol{\beta}_X^*] - \mathbf{V}_j(T_i) \exp[\boldsymbol{\beta}_X^{*T} \mathbf{E}_j(T_i) + \frac{1}{2} \boldsymbol{\beta}_X^{*T} \mathbf{V}_j(T_i) \boldsymbol{\beta}_X^*]\} \boldsymbol{\beta}_X^* \\
& = n^{-1} \sum_{j=1}^n Y_j(T_i) \exp[\boldsymbol{\beta}_Z^* Z_j + \boldsymbol{\beta}_X^{*T} \mathbf{E}_j(T_i) + \frac{1}{2} \boldsymbol{\beta}_X^{*T} \mathbf{V}_j(T_i) \boldsymbol{\beta}_X^*] \\
& \quad \times \{\mathbf{V}_j(T_i) \boldsymbol{\beta}_X^* \boldsymbol{\beta}_X^{*T} [\hat{\mathbf{E}}_j(T_i) - \mathbf{E}_j(T_i)] + [\frac{1}{2} \mathbf{V}_j(T_i) \boldsymbol{\beta}_X^* \boldsymbol{\beta}_X^{*T} + I] [\hat{\mathbf{V}}_j(T_i) - \mathbf{V}_j(T_i)] \boldsymbol{\beta}_X^*\} + o_p(1)
\end{aligned}$$

Note that $\hat{\mathbf{E}}_j(T_i) - \mathbf{E}_j(T_i)$ is a 2×1 vector, and $\hat{\mathbf{V}}_j(T_i) - \mathbf{V}_j(T_i)$ is a symmetric 2×2 matrix.

We now denote

$$\text{vec}(\hat{\mathbf{E}}_j(T_i) - \mathbf{E}_j(T_i), \hat{\mathbf{V}}_j(T_i) - \mathbf{V}_j(T_i)) = \begin{pmatrix} \hat{E}_{0j}(T_i) - E_{0j}(T_i) \\ \hat{E}_{1j}(T_i) - E_{1j}(T_i) \\ \hat{V}_{0j}(T_i) - V_{0j}(T_i) \\ \hat{V}_{1j}(T_i) - V_{1j}(T_i) \\ \hat{V}_{2j}(T_i) - V_{2j}(T_i) \end{pmatrix}$$

where E_{0j}, E_{1j} are the two components of \mathbf{E}_j , and V_{0j}, V_{1j}, V_{2j} are the $(1, 1), (1, 2), (2, 2)$ components of \mathbf{V}_j . Since $\hat{\mathbf{E}}_j(T_i) - \mathbf{E}_j(T_i)$ and $\hat{\mathbf{V}}_j(T_i) - \mathbf{V}_j(T_i)$ are both functions of $\mathcal{A}(I_l)$ for $i = n_{l-1} + 1, \dots, n_l$, we can write

$$\text{vec}(\hat{\mathbf{E}}_j(T_i) - \mathbf{E}_j(T_i), \hat{\mathbf{V}}_j(T_i) - \mathbf{V}_j(T_i)) = \mathcal{T}_j(T_i) \text{vec}(\hat{\mathcal{A}}(I_l) - \mathcal{A}(I_l)) + o_p(1),$$

where $\mathcal{T}_j(T_i)$ is the $5 \times m$ partial derivative matrix of $\text{vec}(\hat{\mathbf{E}}_j(T_i) - \mathbf{E}_j(T_i), \hat{\mathbf{V}}_j(T_i) - \mathbf{V}_j(T_i))$ with respect to $\mathcal{A}(I_l)$, if we denote the dimension of $\mathcal{A}(I_l)$ as m . Now we can simplify (4.18)

to

$$\begin{aligned}
& n^{1/2}[U_{n,l}(\boldsymbol{\beta}^*, \hat{\mathcal{A}}) - U_{n,l}(\boldsymbol{\beta}^*, \mathcal{A})] \\
&= \frac{1}{\sqrt{n}} \sum_{i=n_1-n_l+1}^{n_1-n_l+1} \delta_i \left\{ P^T \mathcal{T}_i(T_i) - \frac{\frac{1}{n} \sum_{j=1}^n Y_j(T_i) Z_j \exp[\beta_Z^* Z_j + \boldsymbol{\beta}_X^{*T} \mathbf{E}_j(T_i) + \frac{1}{2} \boldsymbol{\beta}_X^{*T} \mathbf{V}_j(T_i) \boldsymbol{\beta}_X^*] Q_j(T_i)^T \mathcal{T}_j(T_i)}{S^{(0)}(\boldsymbol{\beta}^*, T_i)} \right. \\
&\quad \left. + \frac{S^{(1)}(\boldsymbol{\beta}^*, T_i) \frac{1}{n} \sum_{j=1}^n Y_j(T_i) \exp[\beta_Z^* Z_j + \boldsymbol{\beta}_X^{*T} \mathbf{E}_j(T_i) + \frac{1}{2} \boldsymbol{\beta}_X^{*T} \mathbf{V}_j(T_i) \boldsymbol{\beta}_X^*] R^T \mathcal{T}_j(T_i)}{[S^{(0)}(\boldsymbol{\beta}^*, T_i)]^2} \right\} \\
&\quad \times \text{vec}(\hat{\mathcal{A}}(I_l) - \mathcal{A}(I_l)) + o_p(1)
\end{aligned}$$

where $\boldsymbol{\beta}_X^* = (\beta_0^*, \beta_1^*)$,

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & \beta_0^* & 0 \\ 0 & \beta_1^* & \beta_0^* \\ 0 & 0 & \beta_1^* \end{pmatrix} \quad R = \begin{pmatrix} \beta_0^* \\ \beta_1^* \\ \frac{1}{2} \beta_0^{*2} \\ \beta_0^* \beta_1^* \\ \frac{1}{2} \beta_1^{*2} \end{pmatrix}$$

$$Q_j(t) = \begin{pmatrix} \beta_0^* & 1 + \beta_0^* E_{0j}(t) + \beta_0^* [\beta_0^* V_{0j}(t) + \beta_1^* V_{1j}(t)] & \beta_0^* E_{1j}(t) + \beta_0^* [\beta_0^* V_{1j}(t) + \beta_1^* V_{2j}(t)] \\ \beta_1^* & \beta_1^* E_{0j}(t) + \beta_1^* [\beta_0^* V_{0j}(t) + \beta_1^* V_{1j}(t)] & 1 + \beta_1^* E_{1j}(t) + \beta_1^* [\beta_0^* V_{1j}(t) + \beta_1^* V_{2j}(t)] \\ \frac{1}{2} \beta_0^{*2} & \frac{1}{2} \beta_0^{*2} [E_{0j}(t) + \beta_0^* V_{0j}(t) + \beta_1^* V_{1j}(t)] + \beta_0^* & \frac{1}{2} \beta_0^{*2} [E_{1j}(t) + \beta_0^* V_{1j}(t) + \beta_1^* V_{2j}(t)] \\ \beta_0^* \beta_1^* & \beta_0^* \beta_1^* [E_{0j}(t) + \beta_0^* V_{0j}(t) + \beta_1^* V_{1j}(t)] + \beta_1^* & \beta_0^* \beta_1^* [E_{1j}(t) + \beta_0^* V_{1j}(t) + \beta_1^* V_{2j}(t)] + \beta_0^* \\ \frac{1}{2} \beta_1^{*2} & \frac{1}{2} \beta_1^{*2} [E_{0j}(t) + \beta_0^* V_{0j}(t) + \beta_1^* V_{1j}(t)] & \frac{1}{2} \beta_1^{*2} [E_{1j}(t) + \beta_0^* V_{1j}(t) + \beta_1^* V_{2j}(t)] + \beta_1^* \end{pmatrix}$$

As $n \rightarrow \infty$,

$$\begin{aligned}
& \frac{1}{n} \sum_{j=1}^n Y_j(T_i) Z_j \exp[\beta_Z^* Z_j + \boldsymbol{\beta}_X^{*T} \mathbf{E}_j(T_i) + \frac{1}{2} \boldsymbol{\beta}_X^{*T} \mathbf{V}_j(T_i) \boldsymbol{\beta}_X^*] Q_j(T_i)^T \mathcal{T}_j(T_i) \rightarrow \mathcal{Q}(\boldsymbol{\beta}^*, T_i) \\
& \frac{1}{n} \sum_{j=1}^n Y_j(T_i) \exp[\beta_Z^* Z_j + \boldsymbol{\beta}_X^{*T} \mathbf{E}_j(T_i) + \frac{1}{2} \boldsymbol{\beta}_X^{*T} \mathbf{V}_j(T_i) \boldsymbol{\beta}_X^*] R^T \mathcal{T}_j(T_i) \rightarrow \mathcal{R}(\boldsymbol{\beta}^*, T_i)
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{Q}(\boldsymbol{\beta}^*, t) &= E\{Y_1(t) Z_1 \exp[\beta_Z^* Z_1 + \boldsymbol{\beta}_X^{*T} \mathbf{E}_1(t) + \frac{1}{2} \boldsymbol{\beta}_X^{*T} \mathbf{V}_1(t) \boldsymbol{\beta}_X^*] Q_1(t)^T \mathcal{T}_1(t)\} \\
\mathcal{R}(\boldsymbol{\beta}^*, t) &= E\{Y_1(t) \exp[\beta_Z^* Z_1 + \boldsymbol{\beta}_X^{*T} \mathbf{E}_1(t) + \frac{1}{2} \boldsymbol{\beta}_X^{*T} \mathbf{V}_1(t) \boldsymbol{\beta}_X^*] R^T \mathcal{T}_1(t)\}
\end{aligned}$$

So,

$$\begin{aligned}
& n^{1/2}[U_{n,l}(\boldsymbol{\beta}^*, \hat{\mathcal{A}}) - U_{n,l}(\boldsymbol{\beta}^*, \mathcal{A})] \\
&= n^{-1/2} \sum_{i=n_1-n_l+1}^{n_1-n_l+1} \delta_i \left\{ P^T \mathcal{T}_i(T_i) - \frac{\mathcal{Q}(\boldsymbol{\beta}^*, T_i)}{S^{(0)}(\boldsymbol{\beta}^*, T_i)} + \frac{S^{(1)}(\boldsymbol{\beta}^*, T_i) \mathcal{R}(\boldsymbol{\beta}^*, T_i)}{[S^{(0)}(\boldsymbol{\beta}^*, T_i)]^2} \right\} \text{vec}(\hat{\mathcal{A}}(I_l) - \mathcal{A}(I_l)) + o_p(1) \\
&= \frac{1}{n} \sum_{i=1}^n \int_{I_l}^{I_{l+1}} \left\{ P^T \mathcal{T}_i(u) - \frac{\mathcal{Q}(\boldsymbol{\beta}^*, u)}{S^{(0)}(\boldsymbol{\beta}^*, u)} + \frac{S^{(1)}(\boldsymbol{\beta}^*, u) \mathcal{R}(\boldsymbol{\beta}^*, u)}{[S^{(0)}(\boldsymbol{\beta}^*, u)]^2} \right\} dN_i(u) \sqrt{n} \text{vec}(\hat{\mathcal{A}}(I_l) - \mathcal{A}(I_l)) + o_p(1) \\
&= D_l(\boldsymbol{\beta}^*) \{ \sqrt{n} \text{vec}(\hat{\mathcal{A}}(I_l) - \mathcal{A}(I_l)) \} + o_p(1)
\end{aligned}$$

where

$$D_l(\boldsymbol{\beta}^*) = E \left\{ \int_{I_l}^{I_{l+1}} \left[P^T \mathcal{T}_1(u) - \frac{\mathcal{Q}(\boldsymbol{\beta}^*, u)}{S^{(0)}(\boldsymbol{\beta}^*, u)} + \frac{S^{(1)}(\boldsymbol{\beta}^*, u) \mathcal{R}(\boldsymbol{\beta}^*, u)}{[S^{(0)}(\boldsymbol{\beta}^*, u)]^2} \right] dN_1(u) \right\}.$$

Now we focus on discussing $\sqrt{n} \text{vec}(\hat{\mathcal{A}}(I_l) - \mathcal{A}(I_l))$. For this section, we focus on the use of an external dataset to estimate distribution parameters $\mathcal{A}(\cdot)$. Similar arguments apply to the situation of using an internal dataset. $\hat{\mathcal{A}}(I_l)$ are estimated through method of moments based on subjects still at risk at I_l . In general, we can write

$$\sqrt{n} \text{vec}(\hat{\mathcal{A}}(I_l) - \mathcal{A}(I_l)) = C(I_l) \{ \sqrt{n_E}(\hat{q} - q) \} + \frac{1}{\sqrt{n}} \sum_{i=1}^n F_l(\mathbf{W}_i, Z_i, \mathcal{A}(\sigma_e^2, I_l)),$$

where n_E is the sample size of the external dataset, q are the parameters borrowed from external data (i.e., $q = (\rho_0, \rho_1)$ with uncorrelated measurement error, and $q = (\rho, r_0, r_1)$ with correlated measurement error), and $\sqrt{n_E}(\hat{q} - q) \xrightarrow{d} N(0, V_q)$. This expression consists of two parts: one from the bias of q estimated from external data, one from the main study data at risk at I_l . These two parts contribute to the total variability independently. We will discuss the detailed expression of $C(I_l)$ and $F_l(\mathbf{W}_i, Z_i, \mathcal{A}(I_l))$ in Section 4.5.

Hence, when the intervals are chosen as quantiles of event times, we write $U_l(\boldsymbol{\beta}^*, \hat{\mathcal{A}})$ as

$$\begin{aligned}
& n^{1/2} U_l(\boldsymbol{\beta}^*, \hat{\mathcal{A}}) \\
&= n^{1/2} U_l(\boldsymbol{\beta}^*, \mathcal{A}) + n^{1/2} [U_{n,l}(\boldsymbol{\beta}^*, \hat{\mathcal{A}}) - U_{n,l}(\boldsymbol{\beta}^*, \mathcal{A})] \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_{I_l}^{I_{l+1}} \left[\begin{pmatrix} Z_i \\ \hat{\mathbf{E}}_i(I_l) + \hat{\mathbf{V}}_i(I_l) \boldsymbol{\beta}_X^* \end{pmatrix} - \frac{\hat{S}^{(1)}(\boldsymbol{\beta}^*, u)}{\hat{S}^{(0)}(\boldsymbol{\beta}^*, u)} \right] dN_i(u) + \frac{1}{\sqrt{n}} \sum_{i=1}^n D_l(\boldsymbol{\beta}^*) F_l(\mathbf{W}_i, Z_i, \mathcal{A}(I_l)) \\
&+ D_l(\boldsymbol{\beta}^*) C(I_l) \{ \sqrt{n_E}(\hat{\rho} - \rho) \} + o_p(1)
\end{aligned}$$

and sum over intervals, we get

$$\begin{aligned}
& n^{1/2}U(\boldsymbol{\beta}^*, \hat{\mathcal{A}}) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^1 \left[\begin{pmatrix} Z_i \\ \mathbf{E}_i(u) + \mathbf{V}_i(u)\boldsymbol{\beta}_X^* \end{pmatrix} - \frac{S^{(1)}(\boldsymbol{\beta}^*, u)}{S^{(0)}(\boldsymbol{\beta}^*, u)} \right] dN_i(u) + \sum_{l=1}^L \left\{ \sum_{i=1}^n D_l(\boldsymbol{\beta}^*) F_l(\mathbf{W}_i, Z_i, \mathcal{A}(I_l)) \right\} \\
&+ \sum_{l=1}^L D_l(\boldsymbol{\beta}^*) C(I_l) \{\sqrt{n_E}(\hat{\rho} - \rho)\} + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \int_0^1 \left[\begin{pmatrix} Z_i \\ \mathbf{E}_i(u) + \mathbf{V}_i(u)\boldsymbol{\beta}_X^* \end{pmatrix} - \frac{S^{(1)}(\boldsymbol{\beta}^*, u)}{S^{(0)}(\boldsymbol{\beta}^*, u)} \right] dN_i(u) + \sum_{l=1}^L D_l(\boldsymbol{\beta}^*) F_l(\mathbf{W}_i, Z_i, \mathcal{A}(I_l)) \right\} \\
&+ \sum_{l=1}^L D_l(\boldsymbol{\beta}^*) C(I_l) \{\sqrt{n_E}(\hat{\rho} - \rho)\} + o_p(1) \\
&= \sum_{i=1}^n \mathcal{V}_i + \left[\sum_{l=1}^L D_l(\boldsymbol{\beta}^*) C(I_l) \right] \{\sqrt{n_E}(\hat{\rho} - \rho)\} + o_p(1)
\end{aligned}$$

where

$$\mathcal{V}_i = n^{-1/2} \left\{ \int_0^1 \left[\begin{pmatrix} Z_i \\ \mathbf{E}_i(u) + \mathbf{V}_i(u)\boldsymbol{\beta}_X^* \end{pmatrix} - \frac{s^{(1)}(\boldsymbol{\beta}^*, u)}{s^{(0)}(\boldsymbol{\beta}^*, u)} \right] dM_i^*(u) + \sum_{l=1}^L D_l(\boldsymbol{\beta}^*) F_l(\mathbf{W}_i, Z_i, \mathcal{A}(I_l)) \right\}.$$

Now we form a triangle array $\{\mathcal{X}_{n,j} : j \geq 1, n \geq 1\}$ where $\mathcal{X}_{n,j} = \mathcal{V}_j$. It is not hard to prove that $\{\mathcal{X}_{n,j} : j \geq 1, n \geq 1\}$ is a martingale difference array, and

$$\begin{aligned}
E(\mathcal{V}_i) &= 0 \\
\text{var}(\mathcal{V}_i) &= n^{-1} E \left\{ \int_0^1 \left[\begin{pmatrix} Z_1 \\ \mathbf{E}_1(u) + \mathbf{V}_1(u)\boldsymbol{\beta}_X^* \end{pmatrix} - \frac{s^{(1)}(\boldsymbol{\beta}^*, u)}{s^{(0)}(\boldsymbol{\beta}^*, u)} \right] dM_1^*(u) \right. \\
&\quad \left. + \sum_{l=1}^L F_l(\mathbf{W}_1, Z_1, \mathcal{A}(I_l)) D_l(\boldsymbol{\beta}^*) \right\}^{\otimes 2} \\
&= \frac{1}{n} B(\boldsymbol{\beta}^*)
\end{aligned}$$

With Condition 7 and the Lindeberg Central Limit Theorem, we have

$$\sum_{i=1}^n \mathcal{V}_i \xrightarrow{D} N(0, B(\boldsymbol{\beta}^*)).$$

In addition, we have

$$\left[\sum_{l=1}^L D_l(\boldsymbol{\beta}^*) C(I_l) \right] \{\sqrt{n_E}(\hat{q} - q)\} \xrightarrow{d} N(0, D(\boldsymbol{\beta}^*)),$$

where

$$D(\beta^*) = \left[\sum_{l=1}^L D_l(\beta^*) C(I_l) \right] V_q \left[\sum_{l=1}^L D_l(\beta^*) C(I_l) \right]^T,$$

and this part is independent of the previous part. Thus

$$n^{1/2} U(\beta^*, \hat{\mathcal{A}}) \xrightarrow{d} N(0, B(\beta^*) + D(\beta^*))$$

This completes the proof of Claim 2.

Combining Claim 1 and 2, from Equation (4.14), it implies that

$$n^{1/2}(\hat{\beta}^* - \beta^*) \xrightarrow{D} N(0, \Sigma^{-1}(\beta^*) [B(\beta^*) + D(\beta^*)] \Sigma^{-1}(\beta^*)).$$

■

In practice, we estimate $\Sigma(\beta^*)$ by $\mathcal{I}_n(\beta^*, \hat{\mathcal{A}})$, and $B(\beta^*)$, $D(\beta^*)$ by plugging in the empirical estimates.

4.5 Details with Uncorrelated and Correlated Measurement Errors

In this section, we will discuss the details of the previous proof in uncorrelated and correlated measurement error special cases separately. This includes providing forms of $C(I_l)$ and $F_l(W, Z, \mathcal{A}(I_l))$, and examining regularity conditions.

4.5.1 Uncorrelated Measurement Errors

As described in Chapter 2, in the uncorrelated measurement errors setting, the distribution parameters are

$$\begin{aligned} \mathcal{A}(t) = & (\sigma_e^2, \mu_0(0, t), \mu_1(0, t), \sigma_0^2(0, t), \rho_0(0, t), \\ & \mu_0(1, t), (\mu_1 + d)(1, t), \sigma_0^2(1, t), \sigma_1^2(1, t), \rho_1(1, t)) \end{aligned}$$

Here we allow all the other parameters to change over time, but the measurement error variance σ_e^2 is constant over time. σ_e^2 is estimated at the beginning of the study. Hence it will carry variability from both the samples and the $q = (\rho_0, \rho_1)$ estimates from the external data. To be more specific, $\sqrt{n}(\hat{\mathcal{A}}(I_l) - \mathcal{A}(I_l))$ should be written as $\sqrt{n}(\hat{\mathcal{A}}(\hat{\sigma}_e^2, I_l) - \mathcal{A}(\sigma_e^2, I_l))$, since there are two steps of estimation involved: the first is to estimate σ_e^2 , and the second

is to estimate other distribution parameters based on $\hat{\sigma}_e^2$. Here $\mathcal{A}(\sigma_e^2, I_l)$ and $\hat{\mathcal{A}}(\sigma_e^2, I_l)$ are functions of σ_e^2 (note that $\hat{\mathcal{A}}(\sigma_e^2, I_l)$ are slightly different at time $I_l = 0$ and $I_l \neq 0$), where

$$\mathcal{A}(\sigma_e^2, I_l) = \left(\begin{array}{c} \sigma_e^2 \\ \mu_0(0, I_l) \\ \mu_1(0, I_l) \\ (\sigma_0^2 + \sigma_e^2)(0, I_l) - \sigma_e^2 \\ \frac{(\rho_0 \sigma_0^2)(0, I_l)}{(\sigma_0^2 + \sigma_e^2)(0, I_l) - \sigma_e^2} \\ \mu_0(1, I_l) \\ (\mu_1 + d)(1, I_l) \\ (\sigma_0^2 + \sigma_e^2)(1, I_l) - \sigma_e^2 \\ (\sigma_1^2 + \sigma_e^2)(1, I_l) - \sigma_e^2 \\ \frac{(\rho_1 \sigma_0 \sigma_1)(1, I_l)}{\sqrt{(\sigma_0^2 + \sigma_e^2)(1, I_l) - \sigma_e^2} \sqrt{(\sigma_1^2 + \sigma_e^2)(1, I_l) - \sigma_e^2}} \end{array} \right),$$

$$\hat{\mathcal{A}}(\sigma_e^2, 0) = \left(\begin{array}{c} \sigma_e^2 \\ \frac{1}{n} \sum_{i=1}^n W_{0i} \\ \frac{1}{n_0} \sum_{i=1}^n W_{1i}(1-Z_i) \\ \frac{\frac{n}{n+n_0} \frac{1}{n-1} \sum_{i=1}^n [W_{0i} - \bar{W}_0]^2 + \frac{n_0}{n+n_0} \frac{1}{n_0-1} \sum_{i=1}^n [W_{1i} - \bar{W}_1(1-Z)]^2 (1-Z) - \sigma_e^2}{\frac{1}{n_0-1} \sum_{i=1}^n [W_{0i} - \bar{W}_0(1-Z)][W_{1i} - \bar{W}_1(1-Z)](1-Z_i)} \\ \frac{\frac{n}{n+n_0} \frac{1}{n-1} \sum_{i=1}^n [W_{0i} - \bar{W}_0]^2 + \frac{n_0}{n+n_0} \frac{1}{n_0-1} \sum_{i=1}^n [W_{1i} - \bar{W}_1(1-Z)]^2 (1-Z) - \sigma_e^2}{\frac{1}{n} \sum_{i=1}^n W_{0i}} \\ \frac{1}{n_1} \sum_{i=1}^n W_{1i} Z_i \\ \frac{\frac{n}{n+n_0} \frac{1}{n-1} \sum_{i=1}^n [W_{0i} - \bar{W}_0]^2 + \frac{n_0}{n+n_0} \frac{1}{n_0-1} \sum_{i=1}^n [W_{1i} - \bar{W}_1(1-Z)]^2 (1-Z) - \sigma_e^2}{\frac{1}{n_1-1} \sum_{i=1}^n (W_{1i} - \bar{W}_1 Z)^2 Z_i - \sigma_e^2} \\ \frac{\frac{1}{n_1-1} \sum_{i=1}^n [W_{0i} - \bar{W}_0 Z][W_{1i} - \bar{W}_1 Z] Z_i}{\sqrt{\frac{n}{n+n_0} \frac{1}{n-1} \sum_{i=1}^n [W_{0i} - \bar{W}_0]^2 + \frac{n_0}{n+n_0} \frac{1}{n_0-1} \sum_{i=1}^n [W_{1i} - \bar{W}_1(1-Z)]^2 (1-Z) - \sigma_e^2} \sqrt{\frac{1}{n_1-1} \sum_{i=1}^n (W_{1i} - \bar{W}_1 Z)^2 Z_i - \sigma_e^2}} \end{array} \right),$$

and

$$\hat{\mathcal{A}}(\sigma_e^2, I_l) = \left(\begin{array}{c} \sigma_e^2 \\ \frac{1}{n_{0l}} \sum_{i=1}^n W_{0i}(1-Z_i)Y_i(I_l) \\ \frac{1}{n_{1l}} \sum_{i=1}^n W_{1i}(1-Z_i)Y_i(I_l) \\ \frac{\frac{1}{2(n_{0l}-1)} \sum_{i=1}^n \{[W_{0i}-\overline{W_0(1-Z)Y(I_l)}]^2 + [W_{1i}-\overline{W_1(1-Z)Y(I_l)}]^2\} (1-Z_i)Y_i(I_l) - \sigma_e^2}{\frac{1}{n_{0l}-1} \sum_{i=1}^n [W_{0i}-\overline{W_0(1-Z)Y(I_l)}][W_{1i}-\overline{W_1(1-Z)Y(I_l)}](1-Z_i)Y_i(I_l)} \\ \frac{\frac{1}{2(n_{0l}-1)} \sum_{i=1}^n \{[W_{0i}-\overline{W_0(1-Z)Y(I_l)}]^2 + [W_{1i}-\overline{W_1(1-Z)Y(I_l)}]^2\} (1-Z_i)Y_i(I_l) - \sigma_e^2}{\frac{1}{n_{1l}} \sum_{i=1}^n W_{0i}Z_iY_i(I_l)} \\ \frac{1}{n_{1l}} \sum_{i=1}^n W_{1i}Z_iY_i(I_l) \\ \frac{\frac{1}{n_{1l}-1} \sum_{i=1}^n [W_{0i}-\overline{W_0ZY(I_l)}]^2 Z_iY_i(I_l) - \sigma_e^2}{\frac{1}{n_{1l}-1} \sum_{i=1}^n [W_{1i}-\overline{W_1ZY(I_l)}]^2 Z_iY_i(I_l) - \sigma_e^2} \\ \frac{\frac{1}{n_{1l}-1} \sum_{i=1}^n [W_{0i}-\overline{W_0ZY(I_l)}][W_{1i}-\overline{W_1ZY(I_l)}]Z_iY_i(I_l)}{\sqrt{\frac{1}{n_{1l}-1} \sum_{i=1}^n [W_{0i}-\overline{W_0ZY(I_l)}]^2 Z_iY_i(I_l) - \sigma_e^2} \sqrt{\frac{1}{n_{1l}-1} \sum_{i=1}^n [W_{1i}-\overline{W_1ZY(I_l)}]^2 Z_iY_i(I_l) - \sigma_e^2}} \end{array} \right)$$

for $I_l > 0$. Thus,

$$\sqrt{n}[\hat{\mathcal{A}}(\hat{\sigma}_e^2, I_l) - \mathcal{A}(\sigma_e^2, I_l)] = \sqrt{n}[\hat{\mathcal{A}}(\hat{\sigma}_e^2, I_l) - \hat{\mathcal{A}}(\sigma_e^2, I_l)] + \sqrt{n}[\hat{\mathcal{A}}(\sigma_e^2, I_l) - \mathcal{A}(\sigma_e^2, I_l)] \quad (4.19)$$

The first term on the right side of equation (4.19) is

$$\begin{aligned} & \sqrt{n}[\hat{\mathcal{A}}(\hat{\sigma}_e^2, I_l) - \hat{\mathcal{A}}(\sigma_e^2, I_l)] \\ &= \sqrt{n}\hat{\mathcal{A}}'(\sigma_e^2, I_l)[\hat{\sigma}_e^2(\hat{\rho}_0, \hat{\rho}_1) - \sigma_e^2] + o_p(1) \\ &= \sqrt{n}\hat{\mathcal{A}}'(\sigma_e^2, I_l)\{[\hat{\sigma}_e^2(\hat{\rho}_0, \hat{\rho}_1) - \hat{\sigma}_e^2(\rho_0, \rho_1)] + [\hat{\sigma}_e^2(\rho_0, \rho_1) - \sigma_e^2]\} + o_p(1) \\ &= \sqrt{n}\hat{\mathcal{A}}'(\sigma_e^2, I_l)\{\hat{\sigma}_e^2(\rho_0, \rho_1)'\begin{pmatrix} \hat{\rho}_0 - \rho_0 \\ \hat{\rho}_1 - \rho_1 \end{pmatrix} + \frac{1}{n} \sum_{i=1}^n G(\mathbf{W}_i, Z_i, \mathcal{A}((\rho_0, \rho_1), 0))\} + o_p(1) \\ &= \mathcal{A}'(\sigma_e^2, I_l)\sigma_e^2(\rho_0, \rho_1)'\{\sqrt{n}\begin{pmatrix} \hat{\rho}_0 - \rho_0 \\ \hat{\rho}_1 - \rho_1 \end{pmatrix}\} + \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathcal{A}'(\sigma_e^2, I_l)G(\mathbf{W}_i, Z_i, \mathcal{A}((\rho_0, \rho_1), 0))\} + o_p(1) \end{aligned}$$

where

$$\begin{aligned} & G(\mathbf{W}_i, Z_i, \mathcal{A}((\rho_0, \rho_1), 0)) \\ &= \frac{1}{2}[W_{0i} - \mu_0(0, 0)]^2 + \frac{1}{2}[W_{1i} - \mu_1(0, 0)]^2(1-Z_i) + \frac{1}{2}[W_{1i} - (\mu_1 + d)(1, 0)]^2 Z_i \\ & - \frac{1+p_0}{2p_0\rho_0}[W_{0i} - \mu_0(0, 0)][W_{1i} - \mu_1(0, 0)](1-Z_i) - \frac{\sigma_1}{2\rho_1\sigma_0}[W_{0i} - \mu_0(1, 0)][W_{0i} - (\mu_1 + d)(1, 0)]Z_i - \sigma_e^2, \end{aligned}$$

and $p_0 = n_0/n$, $p_1 = n_1/n$. The second term on the right side of equation (4.19) can be written as the sum of functions of each sample, that is

$$\sqrt{n}[\hat{\mathcal{A}}(\sigma_e^2, I_l) - \mathcal{A}(\sigma_e^2, I_l)] = \frac{1}{\sqrt{n}} \sum_{i=1}^n H_l(\mathbf{W}_i, Z_i, \mathcal{A}(\sigma_e^2, I_l)) + o_p(1).$$

Due to different $\hat{\mathcal{A}}(\sigma_e^2, I_l)$ definitions in the first interval and the remaining intervals, $H_1(\cdot)$ and $H_l(\cdot)$, $l \geq 2$ are defined separately. For the interval $(I_1, I_2]$,

$$H_1(\mathbf{W}_i, Z_i, \mathcal{A}(\sigma_e^2, 0)) = \begin{pmatrix} 0 \\ W_{0i} - \mu_0(0, 0) \\ \frac{1}{p_0} W_{1i}(1 - Z_i) - \mu_1(0, 0) \\ \frac{1}{1+p_0} \{ [W_{0i} - \mu_0(0, 0)]^2 + [W_{1i} - \mu_1(0, 0)]^2 (1 - Z_i) \} - \sigma_0^2(0, 0) - \sigma_e^2 \\ \frac{\frac{1}{p_0} (W_{0i} - \mu_0)(W_{1i} - \mu_1)(1 - Z_i)}{(\sigma_0^2 + \sigma_e^2)(0, 0) - \sigma_e^2} - \rho_0(0, 0) \\ W_{0i} - \mu_0(1, 0) \\ \frac{1}{1-p_0} W_{1i} Z_i - (\mu_1 + d)(1, 0) \\ \frac{1}{1+p_0} \{ [W_{0i} - \mu_0(1, 0)]^2 + [W_{1i} - \mu_1(0, 0)]^2 (1 - Z_i) \} - \sigma_0^2(0, 0) - \sigma_e^2 \\ \frac{1}{1-p_0} [W_{1i} - (\mu_1 + d)(1, 0)]^2 Z_i - \sigma_1^2(1, 0) - \sigma_e^2 \\ \frac{\frac{1}{1-p_0} [W_{0i} - \mu_0(1, 0)][W_{1i} - (\mu_1 + d)(1, 0)]}{\sqrt{\sigma_0^2(1, 0)\sigma_1^2(1, 0)}} - \rho_1(1, 0) \end{pmatrix}$$

where p_0 is the probability that a subject is in the control group. For other intervals, denote p_l as the probability that subjects survived to interval I_l , p_{0l} as the probability that subjects are in the control group and survived to interval I_l . Then for $l \geq 2$,

$$H_l(\mathbf{W}_i, Z_i, \mathcal{A}(\sigma_e^2, I_l)) = \begin{pmatrix} 0 \\ \frac{1}{p_{0l}} W_{0i}(1 - Z_i) Y_i(I_l) - \mu_0(0, I_l) \\ \frac{1}{p_{0l}} W_{1i}(1 - Z_i) Y_i(I_l) - \mu_1(0, I_l) \\ \frac{1}{2p_{0l}} \{ [W_{0i} - \mu_0(0, I_l)]^2 + [W_{1i} - \mu_1(0, I_l)]^2 \} (1 - Z_i) Y_i(I_l) - \sigma_0^2(0, I_l) - \sigma_e^2 \\ \frac{\frac{1}{p_{0l}} [W_{0i} - \mu_0(0, I_l)][W_{1i} - \mu_1(0, I_l)](1 - Z_i) Y_i(I_l)}{(\sigma_0^2 + \sigma_e^2)(0, I_l) - \sigma_e^2} - \rho_0(0, I_l) \\ \frac{1}{p_l - p_{0l}} W_{0i} Z_i Y_i(I_l) - \mu_0(1, I_l) \\ \frac{1}{p_l - p_{0l}} W_{1i} Z_i Y_i(I_l) - (\mu_1 + d)(1, I_l) \\ \frac{1}{p_l - p_{0l}} [W_{0i} - \mu_0(1, I_l)]^2 Z_i Y_i(I_l) - \sigma_0^2(1, I_l) - \sigma_e^2 \\ \frac{1}{p_l - p_{0l}} [W_{1i} - (\mu_1 + d)(1, I_l)]^2 Z_i Y_i(I_l) - \sigma_1^2(1, I_l) - \sigma_e^2 \\ \frac{\frac{1}{p_l - p_{0l}} [W_{0i} - \mu_0(1, I_l)][W_{1i} - (\mu_1 + d)(1, I_l)] Z_i Y_i(I_l)}{\sqrt{\sigma_0^2(1, I_l)\sigma_1^2(1, I_l)}} - \rho_1(1, I_l) \end{pmatrix}$$

To sum up, we have

$$\begin{aligned}
& \sqrt{n}[\hat{\mathcal{A}}(\hat{\sigma}_e^2, I_l) - \mathcal{A}(\sigma_e^2, I_l)] \\
&= \mathcal{A}'(\sigma_e^2, I_l) \sigma_e^2(\rho_0, \rho_1)' \left\{ \sqrt{n} \begin{pmatrix} \hat{\rho}_0 - \rho_0 \\ \hat{\rho}_1 - \rho_1 \end{pmatrix} \right\} \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n \{ \hat{\mathcal{A}}'(\sigma_e^2, I_l) G(\mathbf{W}_i, Z_i, \mathcal{A}(\sigma_e^2, 0)) + H(\mathbf{W}_i, Z_i, \mathcal{A}(\sigma_e^2, I_l)) \} + o_p(1) \\
&= \frac{1}{\sqrt{c}} \mathcal{A}'(\sigma_e^2, I_l) \sigma_e^2(\rho_0, \rho_1)' \left\{ \sqrt{n_E} \begin{pmatrix} \hat{\rho}_0 - \rho_0 \\ \hat{\rho}_1 - \rho_1 \end{pmatrix} \right\} \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n \{ \mathcal{A}'(\sigma_e^2, I_l) G(\mathbf{W}_i, Z_i, \mathcal{A}(\sigma_e^2, 0)) + H(\mathbf{W}_i, Z_i, \mathcal{A}(\sigma_e^2, I_l)) \} + o_p(1) \\
&= C(I_l) \left\{ \sqrt{n_E} \begin{pmatrix} \hat{\rho}_0 - \rho_0 \\ \hat{\rho}_1 - \rho_1 \end{pmatrix} \right\} + \frac{1}{\sqrt{n}} \sum_{i=1}^n F_l(\mathbf{W}_i, Z_i, \mathcal{A}(\sigma_e^2, I_l))
\end{aligned}$$

where

$$\begin{aligned}
& \sqrt{n_E} \begin{pmatrix} \hat{\rho}_0 - \rho_0 \\ \hat{\rho}_1 - \rho_1 \end{pmatrix} \rightarrow N(0, V_q), \\
& C(I_l) = \frac{1}{\sqrt{c}} \mathcal{A}'(\sigma_e^2, I_l) \sigma_e^2(\rho_0, \rho_1)', \\
& F_l(\mathbf{W}_i, Z_i, \mathcal{A}(\sigma_e^2, I_l)) = \mathcal{A}'(\sigma_e^2, I_l) G(\mathbf{W}_i, Z_i, \mathcal{A}(\sigma_e^2, 0)) + H(\mathbf{W}_i, Z_i, \mathcal{A}(\sigma_e^2, I_l)),
\end{aligned}$$

if the sample size of the external data set (n_E) increases proportionally as the sample size of the main study (n) increases, that is

$$\lim_{n \rightarrow \infty} n_E/n = c.$$

Now we examine the regularity conditions. Condition 2 is obvious according to the estimating procedure mentioned above. Condition 3 is also easy to see due to the definition of \mathbf{E}_i and \mathbf{V}_i . Condition 4 can be proven by central limit theorem. Condition 1, 5, 6 and 7 are common regularity conditions, and are true in most scenarios.

4.5.2 Correlated Measurement Errors

For correlated measurement error scenario, the distribution parameters are

$$\mathcal{A}(I_l) = (\sigma_{e0}^2, \sigma_{e1}^2, r_0, r_1, \mu_0(0, I_l), \mu_1(0, I_l), \sigma_0^2(0, I_l), \\ \mu_0(1, I_l), (\mu_1 + d)(1, I_l), \sigma_0^2(1, I_l), \sigma_1^2(1, I_l), \rho(1, I_l)).$$

Measurement error distribution is assumed to be constant over time (i.e., $\sigma_{e0}^2, \sigma_{e1}^2, r_0, r_1$ are constant), while other parameters may change with time. At the beginning of the study, we estimate $(\hat{\sigma}_{e0}^2, \hat{\sigma}_{e1}^2, \hat{r}_0, \hat{r}_1)$. As mentioned in Chapter 3, we first get estimates of $q = (\rho, r_0, r_1)$ from external data, and then use this information to estimate $(\sigma_{e0}^2, \sigma_{e1}^2)$. This is slightly more complicated than in the uncorrelated measurement error case, since measurement error distribution involves more parameters. To be more specific, $\mathcal{A}(I_l)$ should be considered as a function of $\theta = (\sigma_{e0}^2, \sigma_{e1}^2, r_0, r_1)$, thus $\sqrt{n}(\hat{\mathcal{A}}(I_l) - \mathcal{A}(I_l))$ can be written as

$$\sqrt{n}[\hat{\mathcal{A}}(\hat{\theta}, I_l) - \mathcal{A}(\theta, I_l)],$$

and

$$\sqrt{n}[\hat{\mathcal{A}}(\hat{\theta}, I_l) - \mathcal{A}(\theta, I_l)] = \sqrt{n}[\hat{\mathcal{A}}(\hat{\theta}, I_l) - \hat{\mathcal{A}}(\theta, I_l)] + \sqrt{n}[\hat{\mathcal{A}}(\theta, I_l) - \mathcal{A}(\theta, I_l)]. \quad (4.20)$$

This decomposition demonstrates that there are two sources of variabilities in this estimating procedure: one is from estimating $\theta = (\sigma_{e0}^2, \sigma_{e1}^2, r_0, r_1)$, another is from estimating other parameters based on θ . In this section, we focus on using the method of moments to estimate parameters as described in Chapter 3. Although the MLE may have more stable performance, it is not as straightforward to use for asymptotic variance estimate. As discussed before, the estimate of θ only contributes to part of the variability. Hence, although the method of moments will result in slightly larger variance than the MLE, it would not affect the total variance of the the Cox model parameters too much. Hence, with method

of moments, $\mathcal{A}(\cdot, I_l)$ and $\hat{\mathcal{A}}(\cdot, I_l)$ are (again $\hat{\mathcal{A}}(\cdot, I_l)$ are different for $I_l = 0$ and $I_l \neq 0$)

$$\mathcal{A}(\theta, I_l) = \begin{pmatrix} \sigma_{e0}^2 \\ \sigma_{e1}^2 \\ r_0 \\ r_1 \\ \mu_0(0, I_l) \\ \mu_1(0, I_l) \\ (\sigma_0^2 + \sigma_{e0}^2)(0, I_l) - \sigma_{e0}^2 \\ \mu_0(1, I_l) \\ (\mu_1 + d)(1, I_l) \\ (\sigma_0^2 + \sigma_{e0}^2)(1, I_l) - \sigma_{e0}^2 \\ (\sigma_1^2 + \sigma_{e1}^2)(1, I_l) - \sigma_{e1}^2 \\ \frac{(\rho\sigma_0\sigma_1 + r_1\sigma_{e0}\sigma_{e1})(1, I_l) - r_1\sigma_{e0}\sigma_{e1}}{\sqrt{(\sigma_0^2 + \sigma_{e0}^2)(1, I_l) - \sigma_{e0}^2} \sqrt{(\sigma_1^2 + \sigma_{e1}^2)(1, I_l) - \sigma_{e1}^2}} \end{pmatrix},$$

$$\hat{\mathcal{A}}(\theta, 0) = \begin{pmatrix} \sigma_{e0}^2 \\ \sigma_{e1}^2 \\ r_0 \\ r_1 \\ \frac{1}{n} \sum_{i=1}^n W_{i0} \\ \frac{1}{n_0} \sum_{i=1}^n W_{i1}(1 - Z_i) \\ \frac{1}{n+n_0} \sum_{i=1}^n \{ [W_{0i} - \bar{W}_0]^2 + [W_{1i} - \bar{W}_1(1 - Z)]^2 (1 - Z_i) \} - \sigma_{e0}^2 \\ \frac{1}{n} \sum_{i=1}^n W_{i0} \\ \frac{1}{n_1} \sum_{i=1}^n W_{i1} Z_i \\ \frac{1}{n+n_0} \sum_{i=1}^n \{ [W_{0i} - \bar{W}_0]^2 + [W_{1i} - \bar{W}_1(1 - Z)]^2 (1 - Z_i) \} - \sigma_{e0}^2 \\ \frac{1}{n_1 - 1} \sum_{i=1}^n (W_{i1} - \bar{W}_1 Z)^2 Z_i - \sigma_{e1}^2 \\ \frac{1}{n_1 - 1} \sum_{i=1}^n [W_{i0} - \bar{W}_0 Z] [W_{i1} - \bar{W}_1 Z] Z_i - r_1 \sigma_{e0} \sigma_{e1} \\ \frac{1}{\sqrt{\frac{1}{n+n_0} \sum_{i=1}^n \{ [W_{0i} - \bar{W}_0]^2 + [W_{1i} - \bar{W}_1(1 - Z)]^2 (1 - Z_i) \} - \sigma_{e0}^2} \sqrt{\frac{1}{n_1 - 1} \sum_{i=1}^n (W_{i1} - \bar{W}_1 Z)^2 Z_i - \sigma_{e1}^2}} \end{pmatrix}$$

and

$$\hat{A}(\theta, I_l) = \begin{pmatrix} \sigma_{e0}^2 \\ \sigma_{e1}^2 \\ r_0 \\ r_1 \\ \frac{1}{n_{0l}} \sum_{i=1}^n W_{i0}(1-Z_i)Y_i(I_l) \\ \frac{1}{n_{0l}} \sum_{i=1}^n W_{i1}(1-Z_i)Y_i(I_l) \\ \frac{1}{2(n_{0l}-1)} \sum_{i=1}^n \{ [W_{i0} - \overline{W_0(1-Z)Y(I_l)}]^2 + [W_{i1} - \overline{W_1(1-Z)Y(I_l)}]^2 \} (1-Z_i)Y_i(I_l) - \sigma_{e0}^2 \\ \frac{1}{n_{1l}} \sum_{i=1}^n W_{i0}Z_iY_i(I_l) \\ \frac{1}{n_{1l}} \sum_{i=1}^n W_{i1}Z_iY_i(I_l) \\ \frac{1}{n_{1l}-1} \sum_{i=1}^n [W_{i0} - \overline{W_0ZY(I_l)}]^2 Z_iY_i(I_l) - \sigma_{e0}^2 \\ \frac{1}{n_{1l}-1} \sum_{i=1}^n [W_{i1} - \overline{W_1ZY(I_l)}]^2 Z_iY_i(I_l) - \sigma_{e1}^2 \\ \frac{1}{n_{1l}-1} \sum_{i=1}^n [W_{i0} - \overline{W_0ZY(I_l)}][W_{i1} - \overline{W_1ZY(I_l)}]Z_iY_i(I_l) - r_1\sigma_{e0}\sigma_{e1} \\ \frac{1}{\sqrt{\frac{1}{n_{1l}-1} \sum_{i=1}^n [W_{i0} - \overline{W_0ZY(I_l)}]^2 Z_iY_i(I_l)}} \sqrt{\frac{1}{n_{1l}-1} \sum_{i=1}^n [W_{i1} - \overline{W_1ZY(I_l)}]^2 Z_iY_i(I_l)}} \end{pmatrix}.$$

The first term on the right side of (4.20) is

$$\begin{aligned} & \sqrt{n}[\hat{\mathcal{A}}(\hat{\theta}, I_l) - \hat{\mathcal{A}}(\theta, I_l)] \\ &= \sqrt{n}\hat{\mathcal{A}}'_\theta(\theta, I_l)[\hat{\theta}(\hat{\rho}, \hat{r}_0, \hat{r}_1) - \theta] + o_p(1) \\ &= \sqrt{n}\hat{\mathcal{A}}'_\theta(\theta, I_l)\{[\hat{\theta}(\hat{\rho}, \hat{r}_0, \hat{r}_1) - \hat{\theta}(\rho, r_0, r_1)] + [\hat{\theta}(\rho, r_0, r_1) - \theta]\} + o_p(1) \\ &= \sqrt{n}\hat{\mathcal{A}}'_\theta(\theta, I_l)\{\hat{\theta}'_{(\rho, r_0, r_1)} \begin{pmatrix} \hat{\rho} - \rho \\ \hat{r}_0 - r_0 \\ \hat{r}_1 - r_1 \end{pmatrix} + \frac{1}{n} \sum_{i=1}^n G(\mathbf{W}_i, Z_i, \mathcal{A}((\rho, r_0, r_1), 0))\} + o_p(1) \\ &= \mathcal{A}'_\theta(\theta, I_l)\theta'_{(\rho, r_0, r_1)}\{\sqrt{n} \begin{pmatrix} \hat{\rho} - \rho \\ \hat{r}_0 - r_0 \\ \hat{r}_1 - r_1 \end{pmatrix}\} + \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathcal{A}'_\theta(\theta, I_l)G(\mathbf{W}_i, Z_i, \mathcal{A}((\rho, r_0, r_1), 0))\} + o_p(1) \end{aligned}$$

where $G(\mathbf{W}_i, Z_i, \mathcal{A}((\rho, r_0, r_1), 0)) = (G_{\sigma_{e0}^2}(\cdot), G_{\sigma_{e1}^2}(\cdot), 0, 0)^T$, and

$$G_{\sigma_{e0}^2}(\cdot) = \frac{A_i - C_i}{1 - r_0} - \sigma_{e0}^2,$$

$$\begin{aligned} G_{\sigma_{e1}^2}(\cdot) &= \frac{\rho\sigma_0\sigma_{e1}}{(\rho^2\sigma_0^2 + r_1^2\sigma_{e0}^2)(\rho\sigma_0\sigma_{e1} - r_1\sigma_{e0}\sigma_1)} \{ -(\rho^2\sigma_0^2 + r_1^2\sigma_{e0}^2)(B_i - \sigma_1^2 - \sigma_{e1}^2) \\ &\quad + 2(\rho\sigma_0\sigma_1 + r_1\sigma_{e0}\sigma_{e1})(D_i - \rho\sigma_0\sigma_1 - r_1\sigma_{e0}\sigma_{e1}) \} \end{aligned}$$

and

$$\begin{aligned}
A_i &= \frac{1}{1+p_0} [(W_{i0} - \mu_0)^2 + (W_{i1} - \mu_1)^2] (1 - Z_i), & B_i &= \frac{1}{p_1} (W_{i1} - \mu_0 - d)^2 Z_i, \\
C_i &= \frac{1}{p_0} (W_{i0} - \mu_0)(W_{i1} - \mu_1)(1 - Z_i), & D_i &= \frac{1}{p_1} (W_{i0} - \mu_0)(W_{i1} - \mu_1 - d) Z_i.
\end{aligned}$$

The second term on the right side of (4.20) can also be written as the sum of functions of each sample, that is

$$\sqrt{n}(\hat{\mathcal{A}}(\theta, I_l) - \mathcal{A}(\theta, I_l)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n H_l(\mathbf{W}_i, Z_i, \mathcal{A}(\theta, I_l)) + o_p(1)$$

where $H_l(\cdot)$ is defined as

$$H_1(\mathbf{W}_i, Z_i, \mathcal{A}(\theta, 0)) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ W_{0i} - \mu_0(0,0) \\ \frac{1}{p_0} W_{1i}(1-Z_i) - \mu_1(0,0) \\ \frac{1}{1+p_0} \{ [W_{i0} - \bar{W}_0]^2 + [W_{i1} - \bar{W}_1(1-Z)]^2 (1-Z_i) \} - \sigma_0^2(0,0) - \sigma_{e0}^2 \\ W_{0i} - \mu_0(1,0) \\ \frac{1}{1-p_0} W_{1i} Z_i - (\mu_1 + d)(1,0) \\ \frac{1}{1+p_0} \{ [W_{i0} - \bar{W}_0]^2 + [W_{i1} - \bar{W}_1(1-Z)]^2 (1-Z_i) \} - \sigma_0^2(1,0) - \sigma_{e0}^2 \\ \frac{1}{1-p_0} (W_{i1} - \bar{W}_1 Z)^2 Z_i - \sigma_1^2(1,0) - \sigma_{e1}^2 \\ \frac{1}{1-p_0} (W_{i0} - \bar{W}_0 Z)(W_{i1} - \bar{W}_1 Z) Z_i - r_1 \sigma_{e0} \sigma_{e1} \\ \frac{1}{\sqrt{\sigma_0^2(1,0) \sigma_1^2(1,0)}} \end{pmatrix}$$

and for $l > 1$,

$$H_l(\mathbf{W}_i, Z_i, \mathcal{A}(\theta, 0)) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \frac{1}{p_{0l}} W_{0i}(1-Z_i)Y_i(I_l) - \mu_0(0, I_l) \\ \frac{1}{p_0} W_{1i}(1-Z_i)Y_i(I_l) - \mu_1(0, I_l) \\ \frac{1}{p_{0l}} \{ [W_{i0} - \overline{W_0(1-Z)Y(I_l)}]^2 + [W_{i1} - \overline{W_1(1-Z)Y(I_l)}]^2 (1-Z_i)Y(I_l) \} - \sigma_0^2(0, 0, I_l) - \sigma_{e0}^2 \\ \frac{1}{p_l - p_{0l}} W_{0i} Z_i Y_i(I_l) - \mu_0(1, I_l) \\ \frac{1}{p_l - p_{0l}} W_{1i} Z_i Y_i(I_l) - (\mu_1 + d)(1, I_l) \\ \frac{1}{p_l - p_{0l}} \{ (W_{i0} - \overline{W_0 Z Y(I_l)})^2 Z_i Y_i(I_l) \} - \sigma_0^2(1, I_l) - \sigma_{e0}^2 \\ \frac{1}{p_l - p_{0l}} [W_{i1} - \overline{W_1 Z Y(I_l)}]^2 Z_i Y_i(I_l) - \sigma_1^2(1, I_l) - \sigma_{e1}^2 \\ \frac{1}{p_l - p_{0l}} [W_{i0} - \overline{W_0 Z Y(I_l)}][W_{i1} - \overline{W_1 Z Y(I_l)}] Z_i Y_i(I_l) - r_1 \sigma_{e0} \sigma_{e1} \\ \sqrt{\sigma_0^2(1, I_l) \sigma_1^2(1, I_l)} \end{pmatrix}.$$

So now we can write

$$\begin{aligned} & \sqrt{n}[\hat{\mathcal{A}}(\hat{\theta}, I_l) - \mathcal{A}(\theta, I_l)] \\ &= C(I_l) \left\{ \sqrt{n_E} \begin{pmatrix} \hat{\rho} - \rho \\ \hat{r}_0 - r_0 \\ \hat{r}_1 - r_1 \end{pmatrix} \right\} + \frac{1}{\sqrt{n}} \sum_{i=1}^n F_l(\mathbf{W}_i, Z_i, \mathcal{A}(\theta, I_l)) \end{aligned}$$

where

$$\sqrt{n_E} \begin{pmatrix} \hat{\rho} - \rho \\ \hat{r}_0 - r_0 \\ \hat{r}_1 - r_1 \end{pmatrix} \rightarrow N(0, V_q),$$

$$C(I_l) = \frac{1}{\sqrt{c}} \mathcal{A}'_{\theta}(\theta, I_l) \theta'_{(\rho, r_0, r_1)},$$

$$F_l(\mathbf{W}_i, Z_i, \mathcal{A}(\theta, I_l)) = \mathcal{A}'_{\theta}(\theta, I_l) G(\mathbf{W}_i, Z_i, \mathcal{A}(\theta, 0)) + H(\mathbf{W}_i, Z_i, \mathcal{A}(\theta, I_l)),$$

if the sample size of the external data set (n_E) is asymptotically proportionally to the main study sample size (n), that is

$$\lim_{n \rightarrow \infty} n_E/n = c.$$

For regularity conditions, they can be justified exactly the same as in the uncorrelated measurement error situation.

4.6 *Simulation Studies*

In this section, we conduct simulation studies to evaluate the performance of our proposed variance estimator with both types of measurement error.

For the uncorrelated measurement error case, we look at the situation where $\sigma_1^2 = 0.8$, $\sigma_e^2 = 0.2$, and suppose (ρ_0, ρ_1) are known. This corresponds to the top rows of Table 2.3 and 2.4. Simulation standard errors (SE) from previous simulations and the mean standard error ($\text{mean}(\hat{SE})$) from our proposed variance estimator are summarized in Table 4.1. Our proposed variance estimates are quite close to the simulation standard errors for all the scenarios. For the MVC, our variance estimates are slightly smaller than simulation standard errors. But for the FUC with larger L , the under-estiamte trend disappears.

For the correlated measurement error case, we present results for when $\sigma_1^2 = 0.8$, $\sigma_{e_0}^2 = 0.5$, $\sigma_{e_1}^2 = 0.75$, and suppose (ρ, r_0, r_1) are known. This also corresponds to the top rows of Table 3.1 and 3.2. Simulation standard errors (SE) and mean standard error estimates ($\text{mean}(\hat{SE})$) are summarized in Table 4.2. Our proposed variance estimator performs well in the variances of all the three parameters. The difference between the mean standard error and the simulation standard error can be due to the difference of using the MLE or using the method of moments at the first step in the calibration. But the influence is generally small.

To sum up, our proposed variance estimator are quite close to simulation standard errors, and can be used in practice.

bias($\times 10^3$)		$\beta_Z = \log(0.8)$			$\beta_Z = 0$			$\beta_Z = \log(1.2)$			$\beta_Z = \log(1.5)$		
method		$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$
SE	MVC	0.216	0.229	0.120	0.211	0.226	0.118	0.208	0.224	0.117	0.204	0.220	0.116
	FUC4	0.216	0.228	0.121	0.211	0.225	0.120	0.207	0.222	0.119	0.202	0.217	0.117
	FUC8	0.216	0.228	0.121	0.211	0.225	0.120	0.208	0.222	0.119	0.202	0.217	0.118
mean(\hat{SE})	MVC	0.211	0.224	0.108	0.207	0.221	0.109	0.205	0.220	0.111	0.202	0.218	0.113
	FUC4	0.212	0.224	0.116	0.209	0.222	0.116	0.207	0.221	0.117	0.204	0.219	0.119
	FUC8	0.213	0.225	0.117	0.210	0.223	0.118	0.207	0.221	0.118	0.205	0.219	0.120

Table 4.1: Summary of simulation standard errors (SE) and mean standard error estimates (mean(\hat{SE})) for parameter estimates, with the uncorrelated measurement errors.

bias($\times 10^3$)		$\beta_Z = \log(0.8)$			$\beta_Z = 0$			$\beta_Z = \log(1.2)$			$\beta_Z = \log(1.5)$		
method		$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$
SE	MVC	0.294	0.316	0.137	0.289	0.314	0.135	0.285	0.312	0.135	0.277	0.307	0.133
	FUC4	0.300	0.324	0.146	0.296	0.322	0.144	0.292	0.320	0.144	0.284	0.315	0.141
	FUC8	0.301	0.325	0.148	0.297	0.324	0.146	0.293	0.322	0.146	0.285	0.316	0.143
mean(\hat{SE})	MVC	0.307	0.310	0.130	0.307	0.287	0.131	0.306	0.319	0.133	0.284	0.311	0.136
	FUC4	0.285	0.305	0.139	0.275	0.296	0.131	0.268	0.290	0.132	0.262	0.285	0.134
	FUC8	0.286	0.304	0.135	0.277	0.296	0.135	0.270	0.290	0.136	0.264	0.285	0.137

Table 4.2: Summary of simulation standard errors (SE) and mean standard error estimates (mean(\hat{SE})), with the correlated measurement errors.

Chapter 5

**MULTIPLE BIOMARKERS AND COHORT SAMPLING
GENERALIZATIONS**

In the previous chapters, two correction methods (MVC and FUC) based on the hazard function induced by the observed biomarker values and the treatment assignment were introduced. These methods were illustrated separately for two measurement error specifications: the uncorrelated and the correlated measurement errors, using simulated data sets. Both methods demonstrated success in reducing biases of parameter estimates. However, we restricted our discussion to a single biomarker in a cohort study setting. This chapter considers a generalization of both MVC and FUC to allow multiple biomarkers. A second important generalization is also considered, namely, that of replacing a cohort study by a case-cohort study design.

5.1 Multiple Biomarkers

When there are K ($K > 1$) biomarkers measured for each subject, a question of scientific interest is whether the joint effect of these biomarkers' changes from baseline to post-randomization mediates the relationship between the treatment assignment and the outcome. Now the underlying Cox proportional hazard model becomes

$$\lambda(t; \mathbf{X}_0, \mathbf{X}_1, Z) = \lambda_0(t) \exp(\boldsymbol{\beta}_0^T \mathbf{X}_0 + \boldsymbol{\beta}_1^T \mathbf{X}_1 + \beta_Z Z),$$

where $\mathbf{X}_0 = (X_0^{(1)}, X_0^{(2)}, \dots, X_0^{(K)})^T$ and $\mathbf{X}_1 = (X_1^{(1)}, X_1^{(2)}, \dots, X_1^{(K)})^T$ are both $K \times 1$ vectors, and $(X_0^{(k)}, X_1^{(k)})$ denotes the baseline and the post-randomization values of the k^{th} biomarker. Instead of $(\mathbf{X}_0, \mathbf{X}_1, Z)$, we observe $(\mathbf{W}_0, \mathbf{W}_1, Z)$, where

$$\mathbf{W}_0 = \mathbf{X}_0 + \mathbf{U}_0, \quad \mathbf{W}_1 = \mathbf{X}_1 + \mathbf{U}_1,$$

and $(\mathbf{U}_0, \mathbf{U}_1)$ are independent of $(\mathbf{X}_0, \mathbf{X}_1)$ when given Z . The induced hazard is defined similarly as before:

$$\lambda(t; \mathbf{W}_0, \mathbf{W}_1, Z) = \lambda_0(t) E[\exp(\beta_0^T \mathbf{X}_0 + \beta_1^T \mathbf{X}_1 + \beta_Z Z) | \tilde{T} \geq t, \mathbf{W}_0, \mathbf{W}_1, Z],$$

and the corresponding partial likelihood is

$$PL(\beta) = \prod_{i=1}^k \frac{E[\exp(\beta_0^T \mathbf{X}_{i0} + \beta_1^T \mathbf{X}_{i1} + \beta_Z Z_i) | \tilde{T} \geq t, \mathbf{W}_{i0}, \mathbf{W}_{i1}, Z_i]}{\sum_{j \in R(t_i)} E[\exp(\beta_0^T \mathbf{X}_{j0} + \beta_1^T \mathbf{X}_{j1} + \beta_Z Z_j) | \tilde{T} \geq t, \mathbf{W}_{j0}, \mathbf{W}_{j1}, Z_j]}.$$

Let $\mathbf{X} = (\mathbf{X}_0^T, \mathbf{X}_1^T)^T$, $\mathbf{W} = (\mathbf{W}_0^T, \mathbf{W}_1^T)^T$, and $\beta_X = (\beta_0^T, \beta_1^T)^T$. Similar as with a single biomarker, the induced hazard can be approximated by

$$PL(\beta) \approx \prod_{i=1}^k \frac{\exp[\beta_Z Z_i + \beta_X^T E(\mathbf{X}_i | \mathbf{W}_i, Z_i) + \frac{1}{2} \beta_X^T V(\mathbf{X}_i | \mathbf{W}_i, Z_i) \beta_X]}{\sum_{j \in R(t_i)} \exp[\beta_Z Z_j + \beta_X^T E(\mathbf{X}_j | \mathbf{W}_j, Z_j) + \frac{1}{2} \beta_X^T V(\mathbf{X}_j | \mathbf{W}_j, Z_j) \beta_X]}$$

with MVC, and by

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

with FUC of L intervals: $[I_1, I_2), [I_2, I_3), \dots, [I_L, I_{L+1})$. The challenge here is to efficiently estimate $E(\mathbf{X} | \tilde{T} \geq t, \mathbf{W}, Z)$ and $V(\mathbf{X} | \tilde{T} \geq t, \mathbf{W}, Z)$ (MVC can be considered as $t = 0$).

The best way to estimate the conditional means and variances for the K biomarkers is through estimating their joint distribution. However, there are several difficulties with joint modeling. First, as known from previous discussions, we need to be able to estimate some distribution parameters to perform the calibration. For example, with the uncorrelated measurement error, correlation structure of the K biomarkers at both baseline and post-randomization is required for calibration, which means we need to know $2K$ correlation coefficients. No matter using an internal or external data set, a complicated biomarker process modeling is needed for estimating these parameters. With a relatively small internal or external data set, fitting such a model may lead to unstable performance, which can seriously influence the calibration performance. Second, even if we are able to provide reasonable estimates of these parameters, we may experience difficulties in estimating other parameters in the calibration procedure. Due to the specific structures we impose on the

variances, we would need to maximize the partial likelihood to estimate the other distribution parameters. This procedure can be time consuming and unstable near the boundaries of the parameter space. For these reasons, in this chapter we only consider the simplest modeling procedure where each biomarker is separately calibrated.

When calibrating the K biomarkers separately, we temporarily consider the biomarkers as independent of each other given treatment Z . Then we can fit separate biomarker process models for each biomarker, and get similar information as described in Chapter 2 and 3. The calibration procedure is also done separately for each biomarker. We plug the following $\hat{E}(\mathbf{X}|\tilde{T} \geq t, \mathbf{W}, Z)$ and $\hat{V}(\mathbf{X}|\tilde{T} \geq t, \mathbf{W}, Z)$ into the partial likelihood to estimate (β_X, β_Z) :

$$\hat{E}(\mathbf{X}|\tilde{T} \geq t, \mathbf{W}, Z) = (\hat{E}(\mathbf{X}^{(1)}|\tilde{T} \geq t, \mathbf{W}, Z)^T, \hat{E}(\mathbf{X}^{(2)}|\tilde{T} \geq t, \mathbf{W}, Z)^T, \dots, \hat{E}(\mathbf{X}^{(K)}|\tilde{T} \geq t, \mathbf{W}, Z)^T)^T$$

and

$$V(\mathbf{X}|\tilde{T} \geq t, \mathbf{W}, Z) \approx \begin{pmatrix} \hat{V}(\mathbf{X}^{(1)}|\tilde{T} \geq t, \mathbf{W}, Z) & 0 & \dots & 0 \\ 0 & \hat{V}(\mathbf{X}^{(2)}|\tilde{T} \geq t, \mathbf{W}, Z) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \hat{V}(\mathbf{X}^{(K)}|\tilde{T} \geq t, \mathbf{W}, Z)^T \end{pmatrix}.$$

A simulation study was conducted to assess the performance of this proposed generalization of MVC and FUC to multiple biomarkers. We assume two potential mediating biomarkers A and B.

With the uncorrelated measurement error, the biomarkers' correlations with each other are 0.8 without treatment and 0.7 with treatment. Both biomarkers follow standard normal distribution without treatment, and treatment changes the mean to 0.3 for A and to 0.5 for B, and the corresponding variances become 1.2 and 1.5. For each biomarker, their baseline and post-randomization correlations are 0.95 without treatment and 0.9 with treatment. Measurement error variances are 0.3 for A and 0.5 for B. We let $\beta_0 = \beta_1 = 1$ for the underlying Cox model, and data are administratively censored with rate 95%. Calibration uses the true ρ_0 and ρ_1 values. Results are summarized in Table 5.1. Although we ignore the high correlation between the two biomarkers in the calibration, both MVC and FUC still show the ability to reduce biases, especially for β_Z . For (β_0, β_1) , bias reductions are not as dramatic as with a single biomarker. The goodness of parameter estimations

depends on the accuracy of both biomarkers' values. Although each individual biomarker is calibrated to have similar accuracy as in the single biomarker case, they may not be accurate simultaneously. Hence there are still considerable remaining biases in (β_0, β_1) estimates. Standard errors of all the parameters are larger than in a single biomarker case. Although they are not directly comparable, the magnitude of the standard error elevations with multiple biomarkers suggests that, a larger sample size may be required for estimation accuracy.

With the correlated measurement error, the biomarkers' correlation with each other is assumed to be 0.9 and 0.8. These values are chosen to be higher than those in the uncorrelated measurement error case, since most correlations between the two biomarkers are from the subject-specific mean part not from the temporal variation. Hence when temporal variation is considered as part of the measurement error, the correlation of the true biomarker values increases. The means and variances of the two biomarkers are assumed similarly as before. Measurement errors are assumed to be correlated within each biomarker. The correlations are (0.7, 0.5) for biomarker A and (0.6, 0.4) for biomarker B. Results are summarized in Table 5.2. Similar to the uncorrelated measurement error situation, MVC and FUC perform the best in correcting the bias of β_Z . Biases of other parameter are also reduced, but with considerable remaining biases. Standard errors are also higher than those in the single biomarker case, due to the increased number of parameters.

To sum up, MVC and FUC can be applied to calibrate multiple biomarkers. The easiest way is to separately calibrate each biomarker. MVC and FUC perform well in recovering the treatment effect. For other parameters, our proposed methods are still associated with some remaining biases. More sophisticated methods, such as joint calibration, are needed if these parameters are of scientific interest.

5.2 Case-Cohort Study Design

Prentice (1986) proposed the case-cohort design as a way to reduce data collection burden for large cohort studies with rare outcomes. With the case-cohort design, 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 recorded for this subcohort members and the cases. The

ρ_{AB}	method	$\beta_Z = 0$					$\beta_Z = \log(1.2)$				
		$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
bias($\times 10^3$) (0.8, 0.7)	True	1.9	7.0	2.0	2.0	-3.6	3.2	6.5	1.0	2.0	-2.2
	Naive	-221.8	-404.7	-191.1	-325.2	102.4	-222.6	-405.6	-191.6	-325.1	64.1
	MVC	-56.0	-199.9	-117.4	-232.2	-57.8	-53.5	-196.2	-121.4	-236.9	-90.5
	FUC4	-46.7	-178.2	-110.8	-231.8	-20.3	-44.4	-174.6	-114.5	-236.3	-46.4
	FUC8	-46.2	-176.5	-108.4	-229.9	-16.0	-43.8	-172.5	-112.3	-234.7	-40.6
SE (0.8, 0.7)	True	0.122	0.126	0.115	0.098	0.133	0.120	0.126	0.114	0.097	0.132
	Naive	0.071	0.059	0.068	0.055	0.117	0.071	0.059	0.067	0.055	0.116
	MVC	0.263	0.286	0.247	0.248	0.267	0.260	0.285	0.243	0.246	0.262
	FUC4	0.262	0.281	0.244	0.243	0.267	0.260	0.281	0.241	0.241	0.263
	FUC8	0.263	0.281	0.245	0.242	0.269	0.261	0.281	0.242	0.241	0.265

Table 5.1: Summary of estimated biases ($\times 10^3$) and SEs in parameter estimate with 95% censoring at the end of study, with the uncorrelated measurement error and two biomarkers.

ρ_{AB}	method	$\beta_Z = 0$					$\beta_Z = \log(1.2)$				
		$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
bias($\times 10^3$) (0.9, 0.8)	True	-3.7	10.3	10.0	-4.7	-1.0	-2.1	7.6	9.0	-3.1	-1.8
	Naive	-374.2	-287.0	-522.9	-572.9	238.0	-373.8	-285.7	-524.7	-575.6	183.0
	MVC	-278.4	-2.5	-153.7	-379.3	-45.5	-278.7	-1.1	-155.3	-382.5	-99.4
	FUC4	-264.7	16.8	-129.8	-361.4	14.7	-264.8	17.6	-131.2	-363.8	-28.1
	FUC8	-262.5	19.5	-125.7	-357.8	23.5	-262.7	20.5	-126.9	-360.2	-17.3
SE (0.9, 0.8)	True	0.156	0.176	0.133	0.125	0.131	0.154	0.174	0.131	0.123	0.132
	Naive	0.076	0.071	0.068	0.058	0.117	0.076	0.070	0.067	0.057	0.118
	MVC	0.336	0.343	0.328	0.306	0.291	0.333	0.338	0.325	0.301	0.287
	FUC4	0.331	0.337	0.325	0.301	0.285	0.327	0.332	0.321	0.295	0.282
	FUC8	0.332	0.339	0.326	0.302	0.287	0.328	0.333	0.322	0.296	0.284

Table 5.2: Summary of estimated biases ($\times 10^3$) and SEs in parameter estimate with 95% censoring at the end of study, with the correlated measurement error and two biomarkers.

subcohort may contain subjects who later become cases. If the Cox proportional hazards model holds for the entire cohort,

$$\lambda(t; \mathbf{X}, Z) = \lambda_0(t) \exp(\boldsymbol{\beta}_X^T \mathbf{X} + \beta_Z Z),$$

and the covariates \mathbf{X}, Z are measured correctly on cases and the subcohort members, the partial likelihood can be written similarly as that with the full cohort:

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

where t_i denotes the i^{th} uncensored event time. But now the definition of $R(t)$ is different. As discussed in Prentice (1986), $R(t)$ includes the failing subject and subcohort members who are still at risk at t . Due to this difference, this likelihood is termed as pseudolikelihood. Self and Prentice (1988) defined a slightly different version of the pseudolikelihood which does not include the case in the denominator if the case is not a member of the subcohort. They developed the asymptotic distribution theory for the corresponding estimator. These two estimators generally are quite close if the subcohort size n_{sc} is large. Barlow (1994) viewed the case-cohort design as a weighted version of the random sample. At failing time t_i , case i has weight 1, members in subcohort surviving at t_i have weight equal to the sampling rate n_{sc}/n , and other subjects have weight 0. The corresponding pseudolikelihood is

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

This approach also provides consistent and asymptotically normal parameter estimates.

When \mathbf{X} is measured with error, the previously proposed mean-variance regression calibration (MVC) and follow-up time calibration (FUC) can be easily adopted into this case-cohort design. Here we use the pseudolikelihood in Barlow (1994) as it has a form that can be generalized to the other two versions of pseudolikelihood by different weighting choices. The induced pseudolikelihood becomes

$$PL^*(\beta|\mathbf{W}, Z) = \prod_{i=1}^k \frac{w_i(t_i) E[\exp(\boldsymbol{\beta}_X^T \mathbf{X}_i + \beta_Z Z_i) | \tilde{T} \geq t_i, \mathbf{W}_i, Z_i]}{\sum_{j \in R(t_i)} w_j(t_i) E[\exp(\boldsymbol{\beta}_X^T \mathbf{X}_j + \beta_Z Z_j) | \tilde{T} \geq t_i, \mathbf{W}_j, Z_j]}.$$

This pseudolikelihood can be further approximated by

$$\prod_{i=1}^k \frac{w_i(t_i) \exp[\beta_Z Z_i + \boldsymbol{\beta}_X^T E(\mathbf{X}_i | \mathbf{W}_i, Z_i) + \frac{1}{2} \boldsymbol{\beta}_X^T V(\mathbf{X}_i | \mathbf{W}_i, Z_i) \boldsymbol{\beta}_X]}{\sum_{j \in R(t_i)} w_j(t_i) \exp[\beta_Z Z_j + \boldsymbol{\beta}_X^T E(\mathbf{X}_j | \mathbf{W}_j, Z_j) + \frac{1}{2} \boldsymbol{\beta}_X^T V(\mathbf{X}_j | \mathbf{W}_j, Z_j) \boldsymbol{\beta}_X]}$$

with MVC, and

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

with FUC of L intervals: $[I_1, I_2), [I_2, I_3), \dots, [I_L, I_{L+1})$.

When computing $E(\mathbf{X} | \mathbf{W}, Z)$, $V(\mathbf{X} | \mathbf{W}, Z)$, and $E(\mathbf{X} | \tilde{T} \geq t, \mathbf{W}, Z)$, $V(\mathbf{X} | \tilde{T} \geq t, \mathbf{W}, Z)$, we only use the subcohort members who are still at risk at the time of calibration. When the subcohort size is large, this calibration procedure is expected to show the ability to reduce biases with both uncorrelated and correlated measurement errors. But the performance can be unstable when the subcohort size is not large enough.

A simulation study was performed to demonstrate the performance of MVC and FUC with the case-cohort design. In this simulation, the full cohort has 10,000 subjects and 500 expected events, and we randomly choose 3,000 and 1,000 subjects to form the subcohort. The full cohort study can be viewed as a special case of case-cohort design where $n_{sc} = 10,000$. Parameter choices are similar as before: $\mu_0 = \mu_1 = 0$, $d = 0.5$, $\sigma_0^2 = 1$, $\sigma_1^2 = 1.2$, $\lambda_0(t) = 1$, $\beta_0 = \beta_1 = 1$, $\beta_Z = \log(1.5)$. With the uncorrelated measurement error, we assume $\sigma_{e_0}^2 = \sigma_{e_1}^2 = 0.5$, $\rho_0 = 0.95$, $\rho_1 = 0.9$, while with the correlated measurement error, $\sigma_{e_0}^2 = 0.5$, $\sigma_{e_1}^2 = 1$, $\rho = 0.9$, $r_0 = 0.7$, $r_1 = 0.5$. For the correlated measurement error case, when the subcohort is of sample size 1,000, several simulations show extreme distribution parameter estimates. This is because that, with the correlated measurement error, the MLE is used at the first step of calibration. Thus large sample sizes are needed to provide reasonable estimates. Hence we only report the results with a subcohort of sample size 3,000 for the correlated measurement case. Simulation results for the two measurement error specifications are summarized in Table 5.3 and 5.4.

For both the uncorrelated and correlated measurement errors, MVC and FUC successfully reduce biases in all the three parameters. The trend of further reduction in biases with more calibration still holds. However, when the subcohort size is small (such as $n_{sc} = 1,000$), there can be over-correction in all the three parameters. Standard errors increase rapidly as the subcohort size decreases. Considering that the observed extreme values in the calibration with the correlated measurement error with 1,000 subcohort members, if calibration is

		full cohort			case-cohort, $n_{sc} = 3000$			case-cohort, $n_{sc} = 1000$		
meth.		$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$
bias($\times 10^3$)	True	-1.4	2.7	-1.8	-0.9	9.1	0.1	-0.1	19.1	-3.4
	Naive	-290.2	-251.5	75.8	-291.3	-246.3	76.8	-278.2	-239.1	75.7
	MVC	-60.8	-118.1	-17.8	-67.2	-103.2	-14.8	-30.3	-101.5	14.5
	FUC4	-51.8	-108.6	2.0	-35.2	-24.8	5.3	13.9	5.0	25.5
	FUC8	-50.6	-106.5	5.4	-29.4	2.6	8.1	23.6	51.7	27.7
SE	True	0.109	0.103	0.113	0.145	0.137	0.150	0.206	0.196	0.212
	Naive	0.049	0.048	0.101	0.061	0.060	0.125	0.087	0.087	0.180
	MVC	0.285	0.269	0.193	0.359	0.343	0.233	0.518	0.501	0.319
	FUC4	0.277	0.260	0.192	0.389	0.366	0.270	0.566	0.537	0.398
	FUC8	0.279	0.261	0.194	0.407	0.383	0.283	0.613	0.582	0.428

Table 5.3: Summary of estimated biases($\times 10^3$) and SEs in parameter estimate with 95% censoring at the end of study, with the uncorrelated measurement errors. Studies are designed with full cohort, case-cohort sampling with subcohort sample size 3000 and 1000.

known to be needed beforehand, one may plan a case-cohort study with a relatively larger subcohort compared to if the true \mathbf{X} can be measured. Otherwise, calibration can lead to estimates with large variability. When the subcohort size is relatively large, all properties of MVC and FUC for the cohort studies are preserved with a case-cohort study design.

		full cohort			case-cohort, $n_{sc} = 3000$		
	meth.	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_Z$
bias($\times 10^3$)	True	-2.1	3.1	-3.0	-2.8	11.0	-1.1
	Naive	-187.9	-615.2	201.5	-190.1	-609.4	202.2
	MVC	-93.6	-159.0	-47.5	-77.5	-162.9	-32.2
	FUC4	-76.2	-136.2	-43.0	-64.0	-66.3	-10.6
	FUC8	-72.7	-133.1	-42.4	-53.2	-49.4	-3.6
SE	True	0.120	0.112	0.118	0.164	0.154	0.157
	Naive	0.061	0.053	0.105	0.074	0.065	0.126
	MVC	0.300	0.297	0.205	0.510	0.504	0.321
	FUC4	0.308	0.306	0.209	0.483	0.487	0.315
	FUC8	0.310	0.307	0.210	0.492	0.497	0.322

Table 5.4: Summary of estimated biases($\times 10^3$) and SEs in parameter estimate with 95% censoring at the end of study, with the correlated measurement errors. Studies are designed with full cohort, case-cohort sampling with subcohort sample size 3000.

Chapter 6

APPLICATIONS

In this chapter, we apply our proposed methods to the Women's Health Initiative (WHI) hormone therapy trials. The aim of this data analysis is to understand the mediation effects of blood biomarkers in explaining the relationship between hormone therapy and stroke. The specific biomarkers of interest is insulin-like growth factor-binding protein 4 (IGFBP4).

6.1 Data Description

The Women's Health Initiative (WHI) is a national epidemiologic research program focused on the prevention of chronic disease in postmenopausal women. A total of 161,808 postmenopausal women aged 50 to 79 were enrolled into this study from 1993 to 1998. Among these participants, 68,132 were enrolled in the randomized clinical trial and the remaining 93,676 women were enrolled in the observational study. There are three main components of this randomized clinical trial, including hormone therapy trials, dietary modification trial and Calcium/Vitamin D trial. Our focus is on the hormone therapy trials.

As mentioned in Chapter 1, a total of 10,739 post-hysterectomy women were randomized to 0.625 mg/d conjugated equine estrogens (E-alone) or placebo, and 16,608 with an intact uterus were randomized to 0.625 mg/d conjugated equine estrogens plus 2.5 mg/d medroxyprogesterone acetate (E+P) or placebo. Both trials were stopped early due to elevated adverse events risks, including stroke, chronic heart disease and venous thromboembolism. The estimated hazard ratios for stroke were 1.31(1.02, 1.68) for E+P compared to placebo, and 1.37(1.09, 1.73) for E-alone compared to placebo.

Several nested case-control studies within the hormone therapy trials cohorts were conducted to identify biomarkers that potentially mediate the observed hormone therapy effect on stroke and other clinical outcomes. A Cardiovascular Disease Biomarker Study (CVD) focused on the inflammation markers, blood lipids and lipoprotein markers, coagu-

lation/thrombosis markers, and genetic variants (Kooperberg et al. (2007), Rossouw et al. (2008)). This study identified several blood biomarkers, such as low-density lipoprotein (LDL) and high-density lipoprotein (HDL) change significantly following the treatment intervention. These biomarkers are promising mediators that potentially explain mechanism of the intervention effect.

An additional proteomic study was performed to identify other promising mediators (Prentice et al. (2010)). This study focused on protein expressions. Concentrations of these biomarkers based on blood drawn at baseline and 1-year after randomization were compared on a subcohort of hormone therapy trials participants. 169 of the 378 proteins investigated changed significantly following either E-alone or E+P treatment. 37 and 47 proteins were identified to be significantly related to CHD and stroke, respectively. The overlap biomarkers are novel candidates for mediators of the relationship between hormone therapy and corresponding clinical outcomes.

Among the mediator candidates for stroke, insulin-like growth factor binding protein 4 (IGFBP4) was confirmed for its association with both hormone therapy and stroke. Insulin-like growth factors (IGFs) are polypeptides with high sequence similarity to insulin, and they stimulate cellular growth, proliferation and differentiation (Zhou et al. (2003)). Insulin-like growth factor binding protein (IGFBP) regulates IGF actions. IGFBP4 is the smallest of the 6 IGFBPs, and it inhibits IGF actions. There are several biological pathways linking IGF and stroke (Bondanelli et al. (2006), Denti et al. (2004), Johnsen et al. (2005)). Lower IGF level is related to abnormal lipoprotein metabolism, and low NO production. Low tissue IGF level is found in atherosclerotic plaque, and atherosclerosis is a major cause of stroke. Low IGF level also corresponds to low insulin sensitivity, which may cause diabetes. Although there is no literature for the specific mechanism linking IGF and stroke, the above mentioned relationship between IGFBP and stroke and the inhibiting effect of IGFBP4 on IGF provide scientific evidence for the mediation effect of IGFBP4.

However, traditional statistical methods comparing the two models with and without the IGFBP4 change from baseline to post-randomization do not support the mediation effect. As discussed before, IGFBP4 was measured based on blood drawn. Thus both technical measurement error and temporal variation could be eliminating the mediation effect of

IGFBP4. In the next section, we apply our proposed methods in an attempt to recover the true mediation effect.

6.2 Results

In this section, we analyze the IGFBP4 data with previously proposed methods. We hope to recover the mediation effects of IGFBP4 in the relationship between hormone therapy and stroke, by taking into account the measurement errors.

In the original nested case-control study (Koopberg et al. (2007)), 205 stroke cases and 878 matching controls were included in the proteomic study. For this re-analysis, we included some later cases, and the total number of cases is now 258. A total of 258 controls were matched in a 1 : 1 fashion on age, former history of stroke and randomization year to the cases. For the total of 516 subjects, their IGFBP4 values at both baseline and 1-year post-randomization were measured based on their blood samples. The Cox proportional hazards model is fitted to assess the hormone therapy effects adjusting for IGFBP4.

Note that this study is a nested case-control study. However, our proposed method is for a cohort study or case-cohort study. Due to the nature of the nested case-control study, it is not easy to find a random sample of the population. A common way to deal with a nested case-control study is to consider it as a stratified case-cohort study, with each pair of case and control as a separate stratum. In our data, cases and controls are 1:1 matched. If a pair of cases and controls is randomized to the same treatment, this pair will drop out of the study. Considering our sample is of small sample size, we decide to treat the matching variables as confounders instead of using the exact stratification. This approach will make use of all the samples. However, due to the fact that the true study design is a nested case-control rather than a case-cohort study, the results we provide here can be biased. Also, due to the control selection, our estimates of hazard ratios are different from those reported by Koopberg et al. (2007). The aim of this re-analysis is to show that our proposed methods are useful in recovering the mediation effects, but we are aware that the above described issues can make our results biased.

When treating the data as from a case-cohort study, the total number of cases is 258 and the subcohort is of sample size 258 as well. Due to the small sample size, we decide to only

use the mean-variance regression calibration (MVC). Even with FUC of 4 calibrations, on average there will be only around 30 subjects in each arm with the last calibration, which will not provide stable calibration results. Hence all the results below are based on the mean-variance regression calibration. All models are fitted with log-transformed IGFBP4 values. Without the post-randomization IGFBP4 value in the model, the estimated HRs are 1.84 for E-alone versus placebo, and 1.59 for E+P versus placebo. Without adjustment for measurement errors, the HRs with post-randomization IGFBP4 values in the models are 1.80 and 1.53, respectively. Either E-alone or E+P effect does not seem to be mediated by IGFBP4 changes following the intervention.

When considering the uncorrelated measurement error for each study (E-alone vs. placebo, and E+P vs. placebo), there are two parameters we need to specify (ρ_0, ρ_1) , which are the correlations of the biomarker values at baseline and post-randomization for the two treatment arms. Currently we do not have any additional information about this novel biomarker, thus we need to do a sensitivity study. With no prior information on the biomarker process of IGFBP4, it is difficult to find a range of (ρ_0, ρ_1) that is meaningful in practice. An alternative is to specify $\delta_0^2 = \frac{\sigma_{e0}^2}{\sigma_0^2}$ and $\delta_1^2 = \frac{\sigma_{e1}^2}{\sigma_1^2}$, which are the variance ratios of the measurement errors. There is a 1-1 correspondence between (δ_0^2, δ_1^2) and (ρ_0, ρ_1) that

$$\begin{aligned} \text{cor}(W_0, W_1 | Z = 0) &= \frac{\rho_0 \sigma_0^2}{\sigma_0^2 + \sigma_{e0}^2} = \frac{\rho_0}{1 + \delta_0^2} \\ \text{cor}(W_0, W_1 | Z = 1) &= \frac{\rho_1 \sigma_0 \sigma_1}{\sqrt{(\sigma_0^2 + \sigma_{e0}^2)(\sigma_1^2 + \sigma_{e1}^2)}} = \frac{\rho_1}{\sqrt{(1 + \delta_0^2)(1 + \delta_1^2)}}. \end{aligned} \quad (6.1)$$

δ_0^2 and δ_1^2 have straightforward interpretation in practice. For example, $\delta_0^2 = 0.5$ means $0.5/(1 + 0.5) = 1/3$ of the total variance observed in the baseline biomarker values is due to measurement error. Equations (6.1) also suggest some restrictions on δ_0^2 and δ_1^2 for (ρ_0, ρ_1) to be in the range of $(0, 1)$. For E-alone data, the observed correlations between $\log(\text{IGFBP4})$ at baseline and 1-year are 0.715 and 0.709, thus $\delta_0^2 \in (0, 0.398)$ and $\delta_1^2 \in (0, 0.410)$. For E+P data, the two correlations are 0.738 and 0.588, thus δ_0^2 and δ_1^2 should be in $(0, 0.354)$ and $(0, 0.701)$, respectively. We choose several combinations of (δ_0^2, δ_1^2) , and results are summarized in Table 6.2. With $\delta_0^2 = \delta_1^2 = 0$, after calibration, the estimated hazard ratios can be slightly different from those without adjustment. This is because that calibration

δ_0^2	δ_1^2	E-alone			E+P		
		ρ_0	ρ_1	HR	ρ_0	ρ_1	HR
0	0	0.715	0.709	1.829	0.738	0.588	1.530
0.1	0.1	0.787	0.780	1.811	0.812	0.647	1.516
0.2	0.2	0.859	0.851	1.776	0.886	0.706	1.493
0.3	0.3	0.930	0.922	1.675	0.960	0.764	1.443
0.3	0.5	—	—	—	0.960	0.821	1.424
0.3	0.7	—	—	—	0.960	0.874	1.417

Table 6.1: Summary of hazard ratios, adjusting for the uncorrelated ME.

assumes some specific mean and variance structures, which do not necessarily agree with the data. Hence even when assuming $\delta_0^2 = \delta_1^2 = 0$, calibrated $E(X|W)$ can be slightly different from W and may result in slightly different HRs. When we assume the measurement error variances to increase, hazard ratios of E-alone change from 1.829 to 1.675. Comparing to 1.84 without adjusting for IGFBP4 at year 1, we may conclude a mediation effect with the $HR = 1.675$. With a broader range of δ_1^2 in E+P data, the HRs after calibration ranges from 1.530 to 1.417. We may also conclude a mediation effect.

When considering the correlated measurement error, three parameters are needed: (ρ, r_0, r_1) . Again, without any additional information, it is easier to work on $(\rho, \delta_0^2, \delta_1^2)$ instead. The correspondence between (ρ, r_0, r_1) and $(\rho, \delta_0^2, \delta_1^2)$ are:

$$\begin{aligned} \text{cor}(W_0, W_1|Z = 0) &= \frac{r_0\sigma_{e0}^2}{\sigma_0^2 + \sigma_{e0}^2} = \frac{1 + r_0\delta_0^2}{1 + r_0} \\ \text{cor}(W_0, W_1|Z = 1) &= \frac{r_1\sigma_{e0}\sigma_{e1}}{\sqrt{(\sigma_0^2 + \sigma_{e0}^2)(\sigma_1^2 + \sigma_{e1}^2)}} = \frac{\rho + r_1\delta_0\delta_1}{\sqrt{(1 + \delta_0^2)(1 + \delta_1^2)}}. \end{aligned}$$

Results are summarized in Table 6.2. HRs adjusting for the correlated measurement errors are complicated functions of the three parameters, and no obvious trend is observed. HR ranges from 1.175 to 2.008 for E-alone study and from 1.248 to 1.690 for E+P study. The changes due to the correlated measurement error are much larger than those due to the uncorrelated measurement error. We are very likely to conclude mediation effects with small HRs. There are also situations where estimated HRs are larger than before. In these situations, IGFBP4 can even be on the opposite side of a mediator. Further investigation is needed to understand this mechanism.

ρ	δ_0^2	δ_1^2	E-alone			E+P		
			r_0	r_1	HR	r_0	r_1	HR
0.95	1	0.7	0.431	0.427	1.175	0.476	0.160	1.311
		1	0.431	0.468	2.008	0.476	0.226	1.399
0.9	1	0.7	0.431	0.487	1.369	0.476	0.220	1.248
		1	0.431	0.518	1.580	0.476	0.275	1.349
0.85	1	0.7	0.431	0.547	1.933	0.476	0.289	1.690
		1	0.431	0.568	1.677	0.476	0.326	1.690

Table 6.2: Summary of hazard ratios, adjusting for the correlated ME.

6.3 Summary

In this chapter, we present a re-analysis of the IGFBP4 data from the WHI hormone therapy trials. When taking into account measurement errors, there are potentials for different conclusions about mediation effects. Due to the lack of information about the IGFBP4 biomarker, a sensitivity analysis is performed. With a more common biomarker, we recommend getting some information about the biomarker processes to better provide distribution parameters for the calibration procedure. This will increase the accuracy of calibration. Also, due to the small sample size of our data set, we are unable to perform follow-up time calibration. For a larger data set, follow-up time calibration can result in smaller biases in the HR estimates. Another issue of this data analysis is that we treated this nested case-control study as a case-cohort study. Further extension of MVC and FUC to nested case-control study will be discussed in Chapter 7.

Chapter 7

DISCUSSION AND FUTURE WORK

7.1 Dissertation Contributions

Mediation analysis is important in medical and epidemiological studies to understand the mechanisms between the intervention and the outcome. Measurement error in the potential mediator may obscure the ability of this mediator to explain this mechanism. Thus, correction methods are needed to properly evaluate mediation effects.

In this dissertation, we focused on the failure time data with the underlying Cox proportional hazards model. Biomarkers change from baseline to 1-year post-randomization is a potential mediator for the relationship between the intervention and the outcome, and biomarker values can be mis-measured. Due to the fact that the measurement error distribution may depend on intervention, the unadjusted hazard ratio estimates are usually biased and need to be corrected. The motivation of this topic is that the mediation effects of the IGFBP4 biomarker in the WHI hormone therapy trials are not supported by traditional statistical methods, as discussed in Chapter 1 and Chapter 6.

We decomposed the biomarker process into four parts: population mean, subject-specific mean, temporal variation and technical measurement error. Based on this decomposition, we specified two types of measurement error: the uncorrelated and the correlated measurement error. The uncorrelated measurement error includes only technical measurement error, while the correlated measurement error includes both temporal variation and technical measurement error. The choice of these two specifications depends on the scientific question. If the long-term average biomarker level is more relevant to disease risk, the correlated measurement error should be used to account for temporal variation. If the biomarker values at specific clinic visits are of interest, the uncorrelated measurement error should be chosen.

When the underlying model is the Cox proportional hazards model, the hazard function

induced by the observed biomarker values and conditional on subject still at risk at t is used for the development of correction methods. The partial likelihood of the induced hazard involves the baseline hazard function due to the conditioning event $\{\tilde{T} \geq t\}$. For the induced partial likelihood to maintain the feature of not involving the baseline hazard function, we make the rare disease assumption. Under the rare disease assumption, we can approximate the induced partial likelihood as a function of the conditional means and variances of the true biomarker value \mathbf{X} given the observed \mathbf{W} and intervention Z , with similar forms as the regular partial likelihood. We call this method as a mean-variance regression calibration (MVC). MVC is easy to use, and is expected to correct for biases. But it is restricted by the rare disease assumption.

We further developed a follow-up time calibration (FUC) to relax the sensitivity on the rare disease assumption. This approach is similar as MVC, but instead of ignoring the condition $\{\tilde{T} \geq t\}$ completely, we approximate this condition by $\{\tilde{T} \geq I_l\}$, where I_l are pre-specified time interval cutoff points. Thus the induced partial likelihood becomes a product of partial likelihood within each interval. For each interval, the conditional means and variances of subjects who survive to the beginning of the interval are used to approximate the partial likelihood. This approach largely reduces the rare disease assumption. By choosing an appropriate number of intervals, we can balance the accuracy and the computational burden.

Both MVC and FUC require some prior information on the distribution parameters. We discussed two approaches of estimating the distribution parameters, including using an internal or an external data set with longitudinal measurements. Biomarker process modeling is usually needed for estimating these parameters. In the situation of no additional information, a sensitivity study covering a range of distribution parameters is used.

Simulation studies were conducted to evaluate the performances of our proposed methods. Both MVC and FUC show the ability in reducing biases with both types of measurement errors. FUC generally has more stable performance than MVC, and is less sensitivity to the censoring probabilities. However, with large number of calibrations, FUC may over-correct the biases. Both methods are robust to distribution parameter specification, and robust to mild violation of normality assumption.

Distribution theory of our proposed methods was derived through multivariate counting process and martingale theories. We proved that the estimates of MVC and FUC converge to a target β^* . This target β^* can be different from the true β , but the difference is generally small as illustrated by simulation studies. A sandwich variance estimator is derived. The middle part of the variance estimator takes into account the variability contributed by estimating the distribution parameters from the internal or external data set. Simulation studies were conducted and the mean standard deviations are quite similar to the simulation standard errors.

We generalized MVC and FUC to allow multiple biomarkers. The best solution for multiple biomarkers is to jointly calibrate the multiple biomarkers. However, this involves knowing the correlation matrix of the K biomarkers at both baseline and post-randomization. Either internal or external data set with longitudinal data are usually of small sample sizes. Modeling the multiple biomarkers simultaneously may not provide stable estimates of the correlation structure. Hence we propose to use separate calibration for these biomarkers, and then compute the induced hazard similarly as described before. This method may not capture all the information in these biomarkers, but is more practical. Simulation results showed that biases for intervention effects are quite small. There are remaining biases in the parameters in front of these biomarkers. Hence if accuracy in all the parameters is needed, some more complicated methods, such as joint calibration, are needed. If only the treatment effect is of interest, separate calibration performs well.

We also generalize MVC and FUC to the case-cohort design. With case-cohort design, the induced pseudo-likelihood can be approximated similar as the induced partial likelihood. For the computation of the conditional means and variances, we only use the subcohort subjects as the reference population. Since subcohort consists of random samples of the full cohort, calibration based on subcohort is equivalent to that based on the full cohort.

Finally, we illustrated our proposed methods by applying them to the IGFBP4 data with sensitivity analysis. Our proposed methods discover the potential mediation effects of this biomarker.

7.2 Future Work

The correction methods based on the induced hazard are developed and evaluated in this dissertation. They perform well in most scenarios. However, there are several additional research areas that need to be further explored for this topic.

7.2.1 Interaction Models

In this dissertation, we assumed that the underlying Cox proportional hazards model is a main effects model. When there are interactions between the biomarker values \mathbf{X} and the treatment Z , the mean-variance regression calibration applies easily. The induced hazard also has the form of an interaction model. That is, the underlying model is

$$\lambda(t; X_0, X_1, Z) = \lambda_1(t) \exp(\beta_0 X_0 + \beta_1 X_1 + \beta_2 Z + \beta_3 X_0 Z + \beta_4 X_1 Z),$$

and the induced hazard is

$$\lambda(t; W_0, W_1, Z) = \lambda_2(t) \exp(b_0 W_0 + b_1 W_1 + b_2 Z + b_3 W_0 Z + b_4 W_1 Z).$$

There is a 1 – 1 correspondence between $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$ and $(b_0, b_1, b_2, b_3, b_4)$. With this approach, we can simply fit the naive model and then easily find the corresponding true parameters. Simulation studies have been performed in this setting. Similar performance is found as with the main effects model. Due to the 1-1 correspondence between the two sets of parameters, the Delta method can be applied for the development of the asymptotic distribution. Tests for the interaction terms can also be developed easily. This approach is also sensitive to the rare disease assumption. To relax this assumption, FUC can be applied similarly, and it is expected to have good performance.

7.2.2 Relaxation of the Proportional Hazard Assumption

One fundamental assumption of our proposed methods is the proportional hazards assumption. In practice, this assumption may not hold. One way to reduce this assumption is to allow hazard to change between pre-specified time intervals. Here, we first divide time

into M intervals: $[S_1, S_2), \dots, [S_M, \infty)$, then include interactions between treatment and indicators of these intervals into the underlying Cox model:

$$\lambda(t) = \lambda_0(t) \exp[\beta_X^T \mathbf{X} + \sum_{m=1}^M \beta_Z^m Z I\{t \in [S_m, S_{m+1})\}].$$

With this model, hazard ratio of treatment may differ between intervals. The corresponding induced partial likelihood becomes

$$PL(\beta) = \prod_{i=1}^K \frac{E[\exp(\beta_X^T \mathbf{X}_i + \sum_{m=1}^M \beta_Z^m Z_i I\{t_i \in [S_m, S_{m+1})\}) | \tilde{T}_i \geq t_i, \mathbf{W}_i, Z_i]}{\sum_{j \in R(t_i)} E[\exp(\beta_X^T \mathbf{X}_j + \sum_{m=1}^M \beta_Z^m Z_j I\{t_j \in [S_m, S_{m+1})\}) | \tilde{T}_j \geq t_i, \mathbf{W}_j, Z_j]},$$

and it can be approximated by FUC (or MVC with the special case that $L = 1$)

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

This approximated induced partial likelihood is of a rather complicated form. However, in the special case of S_l coincide with I_l ($M = L$, $S_l = I_l$), the induced partial likelihood will be largely reduced, and then similar calibration procedure applies. This provides an easy approach to deal with non-proportional hazard situation.

Another approach to avoid the proportional hazards assumption is to assume the underlying model is an accelerated failure time model rather than a Cox proportional hazards model. Accelerated failure time model is another common choice for modeling failure time. Instead of modeling the hazard function, accelerated failure time model focuses on the log of failure time. This model has not been well studied in mediation analysis, and measurement error in the mediator may also bias the parameter estimates in a complicated form. Correction methods are also needed.

7.2.3 Generalization to Other Study Designs

In Chapter 6, we treated the nested case-control study as a stratified case-cohort study. We pointed out that this might result in biases. With a nested case-control study, the

denominator of the partial likelihood only consists of the cases and the matching controls. This can be approximated similarly as in the full cohort study and case-cohort study. The difficulty is in calibration. There is no obvious random sample of the full cohort. Only using the controls as the reference population to calibrate will result in the calibrated biomarker values of the cases biased towards controls, thus will over-estimate the treatment effects. A more reasonable way is to weight the cases and controls in the calibration. For example, cases have weight 1 and controls have weight n/n_c , where n is the sample size of the full cohort, and n_c is the number of controls in the study. This is expected to have better performance than treating a nested case-control study as a stratified case-cohort study.

In this dissertation, we did not provide a distribution theory of MVC and FUC for case-cohort designs. Similar arguments in Self and Prentice (1988) can be used to prove the distribution theory.

7.2.4 Measurement Error in the Independent Variable

In a mediation analysis, measurement error may affect both the mediator and the independent variable. In this dissertation, we focused on measurement error in the mediator. When the independent variable is not a categorical variable such as the treatment assignment, it is also subject to measurement error. For example, in the WHI dietary modification trial, we are interested in understanding how energy consumption affects coronary heart disease and stroke risks, while accounting for the mediation effect of BMI change. In this example, the potential mediator BMI is usually accurately recorded at each clinic visit. The independent variable energy intake is often based on food frequency questionnaire (FFQ). It is well known that people tend to under-report in FFQ, and the under-reporting can be more serious in the intervention group. This differential measurement error may distort the relationships between these variables. This area needs similar correction methods that can account for the measurement errors.

7.2.5 Other Approaches

Our proposed methods are based on the induced hazard and assume normality on the biomarker distributions. Although we showed through simulation studies that our methods are robust to mild to moderate violation of normality, methods independent of normality assumption are also of interest. Multiple imputation usually have less distribution assumptions, and will work well when normality is violated. The performance of the multiple imputation with application to mediation analysis needs to be evaluated.

BIBLIOGRAPHY

- Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Annals of Statistics*, 10(4):1100–1120.
- Anderson, G. L., Limacher, M., Assaf, A. R., Bassford, T., Beresford, S. A., Black, H., ..., and Women's Health Initiative Steering Committee (2004). Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the women's health initiative randomized controlled trial. *Journal of the American Medical Association*, 291(14):1701–1712.
- Barlow, W. E. (1994). Robust variance estimation for the case-cohort design. *Biometrics*, 50(4):1064–1072.
- Baron, R. M. and Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6):1173–1182.
- Bondanelli, M., Ambrosio, M. R., Onofri, A., Bergonzoni, A., Lavezzi, S., Zatelli, M. C., Valle, D., Basaglia, N., and degli Uberti, E. C. (2006). Predictive value of circulating insulin-like growth factor I levels in ischemic stroke outcome. *Journal of Clinical Endocrinology & Metabolism*, 91(10):3928–3934.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B*, 34(2):187–220.
- Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62(2):269–276.
- Denti, L., Annoni, V., Cattadori, E., Salvagnini, M. A., Visioli, S., Merli, M. F., Corradi, F., Ceresini, G., Valenti, G., Hoffman, A. R., and Ceda, G. P. (2004). Insulin-like growth factor 1 as a predictor of ischemic stroke outcome in the elderly. *American Journal of Medicine*, 117(5):312–317.

- Diggle, P. J., Heagerty, P., Liang, K. Y., and Zegar, S. L. (2002). *Analysis of Longitudinal Data*. Oxford University Press.
- Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. New Jersey: Wiley.
- Hoyle, R. H. and Kenny, D. A. (1999). Sample size, reliability, and tests of statistical mediation. *Statistical strategies for small sample research*, pages 195–222.
- Hu, P., Tsiatis, A. A., and Davidian, M. (1998). Estimating the parameters in the Cox model when covariate variables are measured with error. *Biometrics*, 54(4):1407–1419.
- Huang, Y. and Wang, C. Y. (2000). Cox regression with accurate covariates unascertainable: a nonparametric correction approach. *Journal of the American Statistical Association*, 95(452):1209–1219.
- Huang, Y. and Wang, C. Y. (2006). Error-in-covariates effect on estimating functions: Additivity in limit and nonparametric correction. *Statistica Sinica*, 16:861–881.
- Hughes, M. D. (1993). Regression dilution in the proportional hazards model. *Biometrics*, 49(4):1056–1066.
- Johnsen, S. P., Hundborg, H. H., Srensen, H. T., Orskov, H., Tjnneland, A., Overvad, K., and Jrgensen, J. O. L. (2005). Insulin-like growth factor (IGF) I, -II, and IGF binding protein-3 and risk of ischemic stroke. *Journal of Clinical Endocrinology & Metabolism*, 90(11):5937–5941.
- Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. New Jersey: Wiley, second edition.
- Kooperberg, C., Cushman, M., Hsia, J., Robinson, J. G., Aragaki, A. K., Lynch, J. K., ..., and Rodriguez, B. (2007). Can biomarkers identify women at increased stroke risk? the women’s health initiative hormone trials. *PLoS Clin Trials*, 2(6):e28.
- Kundua, S., Majumdarb, S., and Mukherjee, K. (2000). Central limit theorems revisited. *Statistics & Probability Letters*, 47(3):265–275.

- Lange, T. and Hansen, J. V. (2011). Direct and indirect effects in a survival context. *Epidemiology*, 22(4):575–581.
- MacKinnon, D. P. (2008). *Introduction to Statistical Mediation Analysis*. Taylor & Francis.
- Preacher, K. J. and Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, 36(4):717–731.
- Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, 69(2):331–342.
- Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*, 73(1):pp. 1–11.
- Prentice, R. L., Paczesny, S., Aragaki, A., Amon, L. M., Chen, L., Pitteri, S. J., ..., and Hanash, S. M. (2010). Novel proteins associated with risk for coronary heart disease or stroke among postmenopausal women identified by in-depth plasma proteome profiling. *Genome Medicine*, 2(7):48–48.
- Rao, R. R. (1963). The law of large numbers for $d[0, 1]$ -valued random variables. *Theory of Probability and Its Application*, 8:70–74.
- Rossouw, J. E., Anderson, G. L., Prentice, R. L., LaCroix, A. Z., Kooperberg, C., Stefanick, M. L., ..., and Writing Group for the Women’s Health Initiative Investigators (2002). 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*, 288(3):321–333.
- Rossouw, J. E., Cushman, M., Greenland, P., Lloyd-Jones, D. M., Bray, P., Kooperberg, C., ..., and Hsia, J. (2008). Inflammatory, lipid, thrombotic, and genetic markers of coronary heart disease risk in the Women’s Health Initiative trials of hormone therapy. *Archives of Internal Medicine*, 168(20):2245–2253.

- Self, S. G. and Prentice, R. L. (1988). Asymptotic distribution theory and efficiency results for case-cohort studies. *The Annals of Statistics*, 16(1):pp. 64–81.
- Vale, C. and Maurelli, V. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, 48(3):465–471.
- Wang, C. Y., Hsu, L., Feng, Z. D., and Prentice, R. L. (1997). Regression calibration in failure time regression. *Biometrics*, 53(1):131–145.
- Xie, S. X. (1997). Covariate measurement error methods in failure time regression. *Thesis (Ph. D.)—University of Washington*.
- Xie, S. X., Wang, C. Y., and Prentice, R. L. (2001). A risk set calibration method for failure time regression by using a covariate reliability sample. *Journal of the Royal Statistical Society: Series B*, 63(4):855–870.
- Zhou, H. and Pepe, M. S. (1995). Auxiliary covariate data in failure time regression. *Biometrika*, 82(1):139–149.
- Zhou, H. and Wang, C. Y. (2000). Failure time regression with continuous covariates measured with error. *Journal of the Royal Statistical Society: Series B*, 62(4):657–665.
- Zhou, R., Diehl, D., Hoeflich, A., Lahm, H., and Wolf, E. (2003). IGF-binding protein-4: biochemical characteristics and functional consequences. *Journal of Endocrinology*, 178(2):177–193.

Appendix A

BIASES ASSOCIATED WITH NAIVE APPROACH

As discussed in Section (2.2), when the underlying model is

$$\lambda(t; \mathbf{X}, Z) = \lambda_1(t) \exp(\beta_0 X_0 + \beta_1 X_1 + \beta_Z Z),$$

the estimates of (b_0, b_1, b_Z) from the naive approach

$$\lambda(t; \mathbf{W}, Z) = \lambda_2(t) \exp(b_0 W_0 + b_1 W_1 + b_Z Z)$$

satisfy the following score equations:

$$\begin{cases} E_{\mathbf{W}, Z|T \leq C_{end}} W_0 - E_{t|t \leq C_{end}} \left[\frac{E_{\mathbf{W}, Z|T \geq t} W_0 \exp(b_0 W_0 + b_1 W_1 + b_Z Z)}{E_{\mathbf{W}, Z|T \geq t} \exp(b_0 W_0 + b_1 W_1 + b_Z Z)} \right] = 0 \\ E_{\mathbf{W}, Z|T \leq C_{end}} W_1 - E_{t|t \leq C_{end}} \left[\frac{E_{\mathbf{W}, Z|T \geq t} W_1 \exp(b_0 W_0 + b_1 W_1 + b_Z Z)}{E_{\mathbf{W}, Z|T \geq t} \exp(b_0 W_0 + b_1 W_1 + b_Z Z)} \right] = 0 \\ E_{\mathbf{W}, Z|T \leq C_{end}} Z - E_{t|t \leq C_{end}} \left[\frac{E_{\mathbf{W}, Z|T \geq t} Z \exp(b_0 W_0 + b_1 W_1 + b_Z Z)}{E_{\mathbf{W}, Z|T \geq t} \exp(b_0 W_0 + b_1 W_1 + b_Z Z)} \right] = 0 \end{cases} \quad (\text{A.1})$$

Consider first the expectations within the brackets of the second terms. To evaluate these expectations, the conditional density functions of $(\mathbf{W}, Z|T \geq t)$ and $(t|T \leq C_{end})$ are needed, given by

$$\begin{aligned} f(w_0, w_1, z|T \geq t) &= \iint_{\mathbb{R}^2} f(w_0, w_1, x_0, x_1, z|T \geq t) dx_0 dx_1 \\ &= \iint_{\mathbb{R}^2} \frac{S(t|w_0, w_1, x_0, x_1, z) f(w_0, w_1|x_0, x_1, z) f(x_0, x_1|z) f(z)}{S(t)} dx_0 dx_1 \\ &= \frac{\iint_{\mathbb{R}^2} S(t|x_0, x_1, z) f(w_0, w_1|x_0, x_1, z) f(x_0, x_1|z) f(z) dx_0 dx_1}{S(t)}, \end{aligned}$$

where the last “=” is due to the assumption that survival is determined by the underlying (\mathbf{X}, Z) values, and knowing \mathbf{W} does not add any additional information, and

$$f(t|T \leq C_{end}) = \begin{cases} 0, & \text{if } t > C_{end} \\ \frac{f(t)}{1-S(C_{end})} = \frac{\iint_{\mathbb{R}^3} f(t|x_0, x_1, z) f(x_0, x_1, z) f(z) dx_0 dx_1 dz}{1 - \iint_{\mathbb{R}^3} S(C_{end}|x_0, x_1, z) f(x_0, x_1, z) f(z) dx_0 dx_1 dz}, & \text{if } t \leq C_{end} \end{cases} \quad (\text{A.2})$$

Under the uncorrelated measurement error distribution (2.1), with some careful simplification and making the change of variable $\nu = \Lambda_0(t)$, so we have $d\nu = \lambda_0(t)dt$ and

$$\begin{aligned}
& E_{t|t \leq C_{end}} \left[\frac{E_{\mathbf{W}, Z|T \geq t} W_0 \exp(b_0 W_0 + b_1 W_1 + b_Z Z)}{E_{\mathbf{W}, Z|T \geq t} \exp(b_0 W_0 + b_1 W_1 + b_Z Z)} \right] \\
&= \int_0^{\Lambda_0(C_{end})} \frac{(1-p)e^{\frac{1}{2}b_1^2 \delta_0^2 \sigma_0^2} \times A + pe^{b_Z + \frac{1}{2}b_1^2 \delta_1^2 \sigma_1^2} \times B}{(1-p)e^{\frac{1}{2}b_1^2 \delta_0^2 \sigma_0^2} \times E + pe^{b_Z + \frac{1}{2}b_1^2 \delta_1^2 \sigma_1^2} \times F} \times \frac{(1-p) \times G + p \times H}{1 - (1-p) \times I - p \times J} d\nu \\
& E_{t|t \leq C_{end}} \left[\frac{E_{\mathbf{W}, Z|T \geq t} W_1 \exp(b_0 W_0 + b_1 W_1 + b_Z Z)}{E_{\mathbf{W}, Z|T \geq t} \exp(b_0 W_0 + b_1 W_1 + b_Z Z)} \right] \\
&= \int_0^{\Lambda_0(C_{end})} \frac{(1-p)e^{\frac{1}{2}b_1^2 \delta_0^2 \sigma_0^2} \times C + pe^{b_Z + \frac{1}{2}b_1^2 \delta_1^2 \sigma_1^2} \times D}{(1-p)e^{\frac{1}{2}b_1^2 \delta_0^2 \sigma_0^2} \times E + pe^{b_Z + \frac{1}{2}b_1^2 \delta_1^2 \sigma_1^2} \times F} \times \frac{(1-p) \times G + p \times H}{1 - (1-p) \times I - p \times J} d\nu \\
& E_{t|t \leq C_{end}} \left[\frac{E_{\mathbf{W}, Z|T \geq t} Z \exp(b_0 W_0 + b_1 W_1 + b_Z Z)}{E_{\mathbf{W}, Z|T \geq t} \exp(b_0 W_0 + b_1 W_1 + b_Z Z)} \right] \\
&= \int_0^{\Lambda_0(C_{end})} \frac{pe^{b_Z + \frac{1}{2}b_1^2 \delta_1^2 \sigma_1^2} \times F}{(1-p)e^{\frac{1}{2}b_1^2 \delta_0^2 \sigma_0^2} \times E + pe^{b_Z + \frac{1}{2}b_1^2 \delta_1^2 \sigma_1^2} \times F} \times \frac{(1-p) \times G + p \times H}{1 - (1-p) \times I - p \times J} d\nu.
\end{aligned}$$

where

$$\begin{aligned}
A &= \iint_{\mathbb{R}^2} (b_0 \delta_0^2 \sigma_0^2 + x_0) \exp[b_0 x_0 + b_1 x_1 - \nu \exp(\beta_0 x_0 + \beta_1 x_1)] f(x_0, x_1 | z = 0) dx_0 dx_1 \\
B &= \iint_{\mathbb{R}^2} (b_0 \delta_0^2 \sigma_0^2 + x_0) \exp[b_0 x_0 + b_1 x_1 - \nu \exp(\beta_0 x_0 + \beta_1 x_1 + \beta_Z)] f(x_0, x_1 | z = 1) dx_0 dx_1 \\
C &= \iint_{\mathbb{R}^2} (b_1 \delta_1^2 \sigma_1^2 + x_1) \exp[b_0 x_0 + b_1 x_1 - \nu \exp(\beta_0 x_0 + \beta_1 x_1)] f(x_0, x_1 | z = 0) dx_0 dx_1 \\
D &= \iint_{\mathbb{R}^2} (b_1 \delta_1^2 \sigma_1^2 + x_1) \exp[b_0 x_0 + b_1 x_1 - \nu \exp(\beta_0 x_0 + \beta_1 x_1 + \beta_Z)] f(x_0, x_1 | z = 1) dx_0 dx_1 \\
E &= \iint_{\mathbb{R}^2} \exp[b_0 x_0 + b_1 x_1 - \nu \exp(\beta_0 x_0 + \beta_1 x_1)] f(x_0, x_1 | z = 0) dx_0 dx_1 \\
F &= \iint_{\mathbb{R}^2} \exp[b_0 x_0 + b_1 x_1 - \nu \exp(\beta_0 x_0 + \beta_1 x_1 + \beta_Z)] f(x_0, x_1 | z = 1) dx_0 dx_1 \\
G &= \iint_{\mathbb{R}^2} \exp[\beta_0 x_0 + \beta_1 x_1 - \nu \exp(\beta_0 x_0 + \beta_1 x_1)] f(x_0, x_1 | z = 0) dx_0 dx_1 \\
H &= \iint_{\mathbb{R}^2} \exp[\beta_0 x_0 + \beta_1 x_1 - \nu \exp(\beta_0 x_0 + \beta_1 x_1 + \beta_Z)] f(x_0, x_1 | z = 1) dx_0 dx_1 \\
I &= \iint_{\mathbb{R}^2} \exp[-\nu \exp(\beta_0 x_0 + \beta_1 x_1)] f(x_0, x_1 | z = 0) dx_0 dx_1 \\
J &= \iint_{\mathbb{R}^2} \exp[-\nu \exp(\beta_0 x_0 + \beta_1 x_1 + \beta_Z)] f(x_0, x_1 | z = 1) dx_0 dx_1,
\end{aligned}$$

and $f(x_0, x_1|z = 0)$ and $f(x_0, x_1|z = 1)$ are the joint normal density functions,

$$f(x_0, x_1|z = 0) = \frac{1}{2\pi\sigma_0^2\sqrt{1-\rho_0^2}} \exp\left\{-\frac{1}{2\sqrt{1-\rho_0^2}}\left[\frac{(x_0-\mu_0)^2}{\sigma_0^2} - \frac{2\rho_0(x_0-\mu_0)(x_1-\mu_1)}{\sigma_0^2} + \frac{(x_1-\mu_1)^2}{\sigma_0^2}\right]\right\}$$

$$f(x_0, x_1|z = 1) = \frac{1}{2\pi\sigma_0\sigma_1\sqrt{1-\rho_1^2}} \exp\left\{-\frac{1}{\sqrt{1-\rho_1^2}}\left[\frac{(x_0-\mu_0)^2}{\sigma_0^2} - \frac{2\rho_0(x_0-\mu_0)(x_1-\mu_1)}{\sigma_0\sigma_1} + \frac{(x_1-\mu_1)^2}{\sigma_1^2}\right]\right\}$$

To compute the first terms, the conditional distribution function for $\mathbf{W}, Z|T \leq C_{end}$ is needed

$$\begin{aligned} f(w_0, w_1, z|T \leq C_{end}) &= \iint_{\mathbb{R}^2} f(w_0, w_0, x_0, x_1, z|T \leq C_{end}) dx_0 dx_1 \\ &= \iint_{\mathbb{R}^2} \frac{[1 - S(C_{end}|w_0, w_1, x_0, x_1, z)]f(w_0, w_1|x_0, x_1, z)f(x_0, x_1|z)f(z)}{1 - S(C_{end})} dx_0 dx_1 \\ &= \frac{\iint_{\mathbb{R}^2} [1 - S(C_{end}|x_0, x_1, z)]f(w_0, w_1|x_0, x_1, z)f(x_0, x_1|z)f(z) dx_0 dx_1}{1 - S(C_{end})} \\ &= \frac{\iint_{\mathbb{R}^2} [1 - S(C_{end}|x_0, x_1, z)]f(w_0, w_1|x_0, x_1, z)f(x_0, x_1|z)f(z) dx_0 dx_1}{1 - \iiint_{\mathbb{R}^3} S(C_{end}|x_0, x_1, z)f(x_0, x_1, z)f(z) dx_0 dx_1 dz}. \end{aligned}$$

Thus after simplification, the corresponding expectations are

$$\begin{aligned} E_{\mathbf{W}, Z|T \leq C_{end}} W_0 &= \frac{\mu_0 - (1-p) \times K - p \times L}{1 - (1-p) \times I - p \times J} \\ E_{\mathbf{W}, Z|T \leq C_{end}} W_1 &= \frac{\mu_1(1-p) + (\mu_1 + d)p - (1-p) \times M - p \times N}{1 - (1-p) \times I - p \times J} \\ E_{\mathbf{W}, Z|T \leq C_{end}} Z &= \frac{p - p \times J}{1 - (1-p) \times I - p \times J}, \end{aligned} \quad (\text{A.3})$$

where

$$\begin{aligned} K &= \iint_{\mathbb{R}^2} x_0 \exp\{-\Lambda_0(C_{end})e^{\beta_0 x_0 + \beta_1 x_1}\} f(x_0, x_1|z = 0) dx_0 dx_1 \\ L &= \iint_{\mathbb{R}^2} x_0 \exp\{-\Lambda_0(C_{end})e^{\beta_0 x_0 + \beta_1 x_1 + \beta_2 z}\} f(x_0, x_1|z = 1) dx_0 dx_1 \\ M &= \iint_{\mathbb{R}^2} x_1 \exp\{-\Lambda_0(C_{end})e^{\beta_0 x_0 + \beta_1 x_1}\} f(x_0, x_1|z = 0) dx_0 dx_1 \\ N &= \iint_{\mathbb{R}^2} x_1 \exp\{-\Lambda_0(C_{end})e^{\beta_0 x_0 + \beta_1 x_1 + \beta_2 z}\} f(x_0, x_1|z = 1) dx_0 dx_1 \end{aligned} \quad (\text{A.4})$$

Substituting (A.2) and (A.3) into (A.1) gives the estimating equation for (b_0, b_1, b_Z) . Usually analytical solution can be computed iteratively. It is computationally intensive due to the integrations involved in the estimating equations. An easier way is to simulate data to find naive estimates $(\hat{b}_0, \hat{b}_1, \hat{b}_2)$. When sample size gets bigger, $(\hat{b}_0, \hat{b}_1, \hat{b}_2)$ converge to (b_0, b_1, b_Z) . This can be confirmed by plugging $(\hat{b}_0, \hat{b}_1, \hat{b}_2)$ into the estimating equations.

VITA

Shanshan Zhao was born in Fuzhou, Fujian, China and grew up in Changzhou, Jiangsu, China. She received a Bachelor of Science in Probability and Statistics from Peking University, China in 2005, and a Master of Science in Biostatistics from the University of Iowa, IA in 2007. She received a Doctor of Philosophy in Biostatistics from the University of Washington in the year 2012.