2019

Measurement Error And Missing Data Methods In Biomarker Research

Carrie Caswell
University of Pennsylvania

Measurement Error And Missing Data Methods In Biomarker Research

Abstract

Measurement error and missing data are two phenomena which prevent researchers from observing essential quantities in their studies. Measurement error occurs when data are subject to variability which masks an underlying value. Recognition of measurement error is essential to preventing bias in an analysis, and methods to handle it have been well-developed in recent years. However, in time-to-event analyses, competing risks is another important consideration which can invalidate study results if not properly accounted for. Current methods to accommodate competing risks do not account for measurement error, and, as a result, incur a large amount of bias when using covariates measured with error. We first propose a novel method which combines the intuition of the subdistribution model for competing risks with risk set regression calibration, which corrects for measurement error in Cox regression by recalibrating at each failure time. We show through simulations that the proposed estimator removes bias that occurs when measurement error is ignored. The second part of this dissertation addresses missing outcome data in longitudinal models. While this is a well-studied area of research, some current missing data methods are subject to misspecification, while others are not suited to handle a large amount of missing data. We propose a novel method to account for missing longitudinal outcome data in the situation where some patients have no recorded outcomes. We accomplish this through use of an auxiliary outcome available for all patients, and avoid the pitfall of misspecification by estimating its relationship with the data nonparametrically. We show that this method is more efficient than conventional methods and robust to misspecification. For both proposed methods, we show that the estimators are asymptotically normal, and provide consistent variance estimates. We also show that the estimator for the second method is consistent. We apply both proposed methods to neurodegenerative disease data. Finally, we introduce an R package to implement the first proposed method and make it widely available for regular use.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Epidemiology & Biostatistics

First Advisor

Sharon X. Xie

Keywords

competing risks, longitudinal analysis, measurement error, missing data, mixed-effects models, survival analysis

Subject Categories

Biostatistics

MEASUREMENT ERROR AND MISSING DATA METHODS IN BIOMARKER RESEARCH

Carrie Caswell

A DISSERTATION

in

Epidemiology and Biostatistics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2019

Supervisor of Dissertation

____

Sharon X. Xie, Professor of Biostatistics, University of Pennsylvania

Graduate Group Chairperson

____

Nandita Mitra, Professor of Biostatistics, University of Pennsylvania

Dissertation Committee

Warren B. Bilker, Professor of Biostatistics, University of Pennsylvania

Murray Grossman, Professor of Neurology, University of Pennsylvania

Lauren M. Massimo, Assistant Professor of Nursing, University of Pennsylvania

Justine Shults, Professor of Biostatistics, University of Pennsylvania

ACKNOWLEDGEMENT

I would like to thank my advisor, Dr. Sharon Xie, for years of unrelenting patience and pedagogy. Sharon, you have been understanding, flexible, professional, and caring, and have taught me to be a discerning and thorough statistician. I am forever grateful for your expertise. I would like to thank my committee members. Thank you to Drs. Murray Grossman and Lauren Massimo, not only for fulfilling committee duties with enthusiasm and grace, but for presenting me with uniquely invaluable opportunities to collaborate on clinical analyses and explore areas of research beyond the scope of this dissertation. I will always be grateful that you welcomed me into your lab and gave me the same opportunities that your own students have. Thank you to Drs. Warren Bilker and Justine Shults for offering statistical insights and advice which elevated the quality of this dissertation, and for sacrificing your time and energy to serve on my committee. I would also like to thank the professors, staff, and students in the Biostatistics Department for cultivating a welcoming and stimulating learning environment throughout my time in this program.

Above all, I'd like to thank my family and friends, particularly my mother, father, brother, sister, Brian, Taylor, and Todd, for the unconditional love and unwavering support that carried me through all these years of school. Thank you all for believing in me when I didn't believe in myself, for pushing me when the road ahead seemed impossible, and for celebrating little victories of this dissertation with me even when you had no idea what I was talking about.

ABSTRACT

MEASUREMENT ERROR AND MISSING DATA METHODS IN BIOMARKER RESEARCH

Carrie Caswell

Sharon X. Xie

Measurement error and missing data are two phenomena which prevent researchers from observing essential quantities in their studies. Measurement error occurs when data are subject to variability which masks an underlying value. Recognition of measurement error is essential to preventing bias in an analysis, and methods to handle it have been well-developed in recent years. However, in time-to-event analyses, competing risks is another important consideration which can invalidate study results if not properly accounted for. Current methods to accommodate competing risks do not account for measurement error, and, as a result, incur a large amount of bias when using covariates measured with error. We first propose a novel method which combines the intuition of the subdistribution model for competing risks with risk set regression calibration, which corrects for measurement error in Cox regression by recalibrating at each failure time. We show through simulations that the proposed estimator removes bias that occurs when measurement error is ignored. The second part of this dissertation addresses missing outcome data in longitudinal models. While this is a well-studied area of research, some current missing data methods are subject to misspecification, while others are not suited to handle a large amount of missing data. We propose a novel method to account for missing longitudinal outcome data in the situation where some patients have no recorded outcomes. We accomplish this through use of an auxiliary outcome available for all patients, and avoid the pitfall of misspecification by estimating its relationship with the data nonparametrically. We show that this method is more efficient than conventional methods and robust to misspecification. For both proposed methods, we show that the estimators are asymptotically normal, and provide consistent variance estimates. We also show that the estimator for the second method is consistent. We apply both proposed methods to neurodegenerative disease data. Finally, we introduce an R package to implement the first proposed method and make it widely available for regular use.

TABLE OF CONTENTS

ACKNOWLEDGEMENT iii

ABSTRACT iv

LIST OF TABLES viii

LIST OF ILLUSTRATIONS ix

CHAPTER 1 : INTRODUCTION 1

CHAPTER 2 : ADJUSTING FOR COVARIATE MEASUREMENT ERROR IN FAILURE TIME MODELS UNDER COMPETING RISKS 3

2.1 Introduction 3

2.2 Proposed Methods and Asymptotic Results 6

2.3 Simulation 9

2.4 Analysis of Association Between CSF Biomarkers and Risk of Conversion to AD . . 13

2.5 Discussion 15

CHAPTER 3 : EFFICIENT LINEAR LONGITUDINAL MODELS IN THE PRESENCE OF MISSING OUTCOMES 18

3.1 Introduction 18

3.2 Proposed Methods and Asymptotic Results 21

3.3 Simulations 31

3.4 Data Example 32

3.5 Discussion 36

CHAPTER 4 : CMPRRC - AN R PACKAGE FOR COVARIATE MEASUREMENT ERROR AND COMPETING RISKS IN FAILURE TIME MODELS 39

4.1 Introduction 39

4.2 Data 40

4.3 Implementation 40

4.4 Performance 42
4.5 Discussion 42

CHAPTER 5 : DISCUSSION 44

APPENDICES 47

BIBLIOGRAPHY 102

LIST OF TABLES

TABLE 2.1 : $\widehat{\beta}_{C-RRC}$ is the proposed estimator; $\widehat{\beta}_{SHR}$ is the logarithm of the subdistribution hazard ratio estimator. Bias = $\widehat{\beta} - \beta_0$. SD = empirical standard deviation of estimates across the 1000 simulation runs. $\widehat{\text{SE}}$ = estimated standard error. MSE = mean squared error. COV = coverage of 95% confidence intervals. In all simulations, β_2 (not shown) is equal to β_0. Z is a N(0,1) random variable. NSR = noise-to-signal ratio, defined as (σ^2/k)/Var(Z), where each subject has k = 4 replicates. HR = hazard ratio. 11

TABLE 2.2 : $\widehat{\beta}_{C-RRC}$ is the proposed estimator; $\widehat{\beta}_{SHR}$ is the logarithm of the subdistribution hazard ratio estimator. Bias = $\widehat{\beta} - \beta_0$. SD = standard deviation of estimates across the 1000 simulation runs. SE = standard error. MSE = mean squared error. COV = coverage of 95% confidence intervals. In all simulations, β_2 (not shown) is equal to β_0. Z is a uniform(0, $\sqrt{12}$) random variable. NSR = noise-to-signal ratio, defined as (σ^2/k)/Var(Z), where each subject has k = 4 replicates. HR = hazard ratio. 12

TABLE 2.3 : Continuous variables summarized with mean and standard deviation in parentheses; categorical variables summarized with counts and percentages in parentheses. 14

TABLE 2.4 : $\widehat{\beta}_{C-RRC}$ is the proposed estimator; $\widehat{\beta}_{SHR}$ is the logarithm of the subdistribution hazard ratio estimator. $\widehat{\text{SE}}$ = estimated standard error. CI = confidence interval. 15

TABLE 3.1 : Results for timepoint coefficient from 500 simulations. SD = standard deviation; $\widehat{SE}$ = estimated standard error; MSE = mean squared error; RE = relative efficiency. $\widehat{\beta}_{HS}$ = estimator using both hybrid and validation subsample; $\widehat{\beta}_{VS}$ = estimator using an internal validation sample; $\widehat{\beta}_{CC}$ = complete-case estimator. 33

TABLE 3.2 : Results for baseline covariate coefficient from 500 simulations. SD = standard deviation; $\widehat{SE}$ = estimated standard error; MSE = mean squared error; RE = relative efficiency. $\widehat{\beta}_{HS}$ = estimator using both hybrid and validation subsample; $\widehat{\beta}_{VS}$ = estimator using an internal validation sample; $\widehat{\beta}_{CC}$ = complete-case estimator. 34

TABLE 3.3 : Summary statistics of baseline measurements. Continuous variables summarized with mean and standard deviation in parentheses; categorical variables summarized with counts and percentages in parentheses. GM = gray matter. True GM volume is standardized by dividing by intracranial volume for each subject, then multiplied by 1000 for clarity of presentation. Composite GM volume is the auxiliary outcome predicted by clinical assessment scores. MoCA = Montreal Cognitive Assessment; QUIP-RS = Questionnaire for Impulsive-Compulsive Disorders in Parkinson's Disease Rating Scale; SDM = Symbol Digit Modalities Score; SFT = Semantic Fluency Total Score; STAI = State-Trait Anxiety Inventory; UPDRS-Cog = cognitive score of Unified Parkinson's Disease Rating Scale. 35

TABLE 3.4 : $\widehat{\beta}_{HS}, \widehat{\beta}_{VS}$ denote the proposed estimators; $\widehat{\beta}_{CC}$ denotes the complete-case estimator; $\widehat{SE}$ = estimated standard error; Aux. Out. = Auxiliary Outcome; CGMAV = Composite gray matter auxiliary variable; SDM = Symbol Digit Modalities Score. 36

TABLE 4.1 : Average run time in seconds of function with varying sample sizes. Each model used 3 covariates. The first column displays results from 1 covariate having measurement error (ME); the second displays results from 2 covariates having ME. Each cell is the result of 10 runs. 42

TABLE B.1 : Results for timepoint coefficient from 500 simulations with a useless auxiliary outcome and a perfect auxiliary outcome. SD = standard deviation; $\widehat{SE}$ = estimated standard error; MSE = mean squared error; RE = relative efficiency. $\widehat{\beta}_{HS}$ = estimator based on both hybrid and validation subsample; $\widehat{\beta}_{VS}$ = estimator based on internal validation sample; $\widehat{\beta}_{CC}$ = complete-case estimator. 76

TABLE B.2 : Results for baseline covariate coefficient from 500 simulations with a useless auxiliary outcome and a perfect auxiliary outcome. SD = standard deviation; $\widehat{SE}$ = estimated standard error; MSE = mean squared error; RE = relative efficiency. $\widehat{\beta}_{HS}$ = estimator based on both hybrid and validation subsample; $\widehat{\beta}_{VS}$ = estimator based on internal validation sample; $\widehat{\beta}_{CC}$ = complete-case estimator. 77

TABLE B.3 : Results for timepoint coefficient from 500 simulations comparing proposed estimators with multiple imputation and predictive mean matching estimators. SD = standard deviation; $\widehat{SE}$ = estimated standard error; MSE = mean squared error; RE = relative efficiency. $\widehat{\beta}_{HS}$ = estimator based on both hybrid and validation subsample; $\widehat{\beta}_{VS}$ = estimator based on internal validation sample; $\widehat{\beta}_{PMM3}$ = predictive mean matching estimator using 3 donors; $\widehat{\beta}_{PMM10}$ = predictive mean matching estimator using 10 donors; $\widehat{\beta}_{MI}$ = multiple imputation estimator; $\widehat{\beta}_{CC}$ = complete-case estimator. 79

TABLE B.4 : Results for baseline covariate coefficient from 500 simulations comparing proposed estimators with multiple imputation and predictive mean matching estimators. SD = standard deviation; $\widehat{SE}$ = estimated standard error; MSE = mean squared error; RE = relative efficiency. $\widehat{\beta}_{HS}$ = estimator based on both hybrid and validation subsample; $\widehat{\beta}_{VS}$ = estimator based on internal validation sample; $\widehat{\beta}_{PMM3}$ = predictive mean matching estimator using 3 donors; $\widehat{\beta}_{PMM10}$ = predictive mean matching estimator using 10 donors; $\widehat{\beta}_{MI}$ = multiple imputation estimator; $\widehat{\beta}_{CC}$ = complete-case estimator. 80

TABLE B.5 : Results for timepoint coefficient from 500 simulations evaluating impact of MAR outcomes in the hybrid set. SD = standard deviation; $\widehat{SE}$ = estimated standard error; MSE = mean squared error; RE = relative efficiency. $\widehat{\beta}_{HS}$ = estimator based on both hybrid and validation subsample; $\widehat{\beta}_{CC}$ = complete-case estimator. 81

TABLE B.6 : Results for baseline covariate coefficient from 500 simulations evaluating impact of MAR outcomes in hybrid set. SD = standard deviation; $\widehat{SE}$ = estimated standard error; MSE = mean squared error; RE = relative efficiency. $\widehat{\beta}_{HS}$ = estimator based on both hybrid and validation subsample; $\widehat{\beta}_{CC}$ = complete-case estimator. 82

CHAPTER 1

Introduction

Uncertainty is simultaneously a statistician's best friend and greatest nemesis. It is the foundation upon which all statistical principles are founded; to accept that all measured quantities have some inherent uncertainty is what separates statistics from mathematics. On the other hand, too much uncertainty can challenge the reliability and validity of statistical analyses. This issue of too much uncertainty can manifest in a variety of ways. We explore here two of the most insidious forms of uncertainty in data collection: measurement error and missing data.

Measurement error is a ubiquitous phenomenon which is often overlooked in practical situations. It occurs when a metric is subject to variability which masks the true value of a collected data point. For example, it is widely accepted that blood pressure should not be taken in some circumstances, such as when a person is afraid, nervous, or has recently taken certain medications. These situations can elevate a person's blood pressure so that a correct measurement of the person's true average blood pressure on a typical day will not be attainable, regardless of the accuracy of measuring instruments. Blood pressure is only measured in controlled situations because it is known to be prone to measurement error. Measurement error can occur for a variety of reasons, such as human error, variability of testing conditions, or biological fluctuations. Other biomarkers can be either incorrectly treated as not having measurement error, or impossible to collect without measurement error. Cerebrospinal fluid (CSF) is a very useful biomarker in the study of neurodegenerative diseases, and proper handling of CSF data in statistical analyses is crucial to developing deeper understanding of these diseases. This biomarker is often recognized as having measurement error, but its measurement error is difficult to control. One useful way of controlling for the variability that causes measurement error is to collect replicate data. Risk set regression calibration (RRC; Xie, Wang, and Prentice 2001) is an appropriate statistical method to account for measurement error through utility of replicate data in time-to-event analyses.

One of the biggest challenges in analyzing neurodegenerative disease data with a time-to-event model is competing risks. A competing risk is an event which can potentially affect many patients in a population and which precludes the event of interest. For example, when studying time to some disease-related event in a population with a high-mortality disease, death should be considered a competing risk, because patients who die will not experience the event of interest. In this sense, competing risks can be viewed as a form of missing data, because they prevent investigators from observing outcomes for all patients on a study. Current literature does not offer a statistical method to account for both competing risks and measurement error in survival analysis. In Chapter 2, we present a novel method to do this.

The other form of uncertainty in this dissertation, missing data (specifically missing outcomes), is addressed in Chapter 3. Here, we leave the realm of survival analysis in favor of a related setting, longitudinal studies. Missing outcomes is a very common problem in these studies, primarily because patient dropout is nearly inevitable. However, in this chapter, we explore an additional form of missingness in longitudinal studies. When the outcome is difficult or impossible to measure due to limitations of the study design or resources, then the amount of missing data on the study can be significantly greater than that of a typical patient dropout scenario. For example, the outcome of the study in Chapter 3 is measured by MRI; however, many study sites do not have adequate facilities to perform MRI scans. Patients randomized to these sites are missing outcome data at all timepoints, and have only covariate data to be entered into an analysis model. While statistical methods to account for missing data is an extensive and ever-evolving area of research, the current conventional methods to handle missing data are not suited to address datasets with such a large amount of missing data in a longitudinal setting. Therefore, we present a novel method to address this dearth of outcome data by using an auxiliary outcome which is correlated with the missing outcome.

Finally, Chapter 4 presents a software package which implements the method of Chapter 2, making it widely available for users of all levels of expertise. The Appendices give proofs for the asymptotic properties of all proposed estimators of Chapters 2 and 3, as well as additional simulations to supplement Chapter 3. Concluding remarks are given in Chapter 5.

CHAPTER 2

ADJUSTING FOR COVARIATE MEASUREMENT ERROR IN FAILURE TIME MODELS UNDER COMPETING RISKS

2.1. Introduction

Biomarker research can be fraught with challenges, particularly when biomarkers of a certain disease are known to be measured with error. Failure time analysis is a popular model choice for these situations, necessitating methods which can reliably correct for measurement error and other data features in a survival analysis setting. For example, in the study of neurodegenerative diseases such as Alzheimer's disease (AD), biomarkers are extracted from cerebrospinal fluid (CSF) and analyzed through assays. Measurement error can be incurred in these samples through a variety of sources, including, but not limited to, day-to-day fluctuations in the CSF, plate-to-plate variability in assay development, variable storage conditions, and operator error. These conditions render the true value of the biomarker unobservable, and inhibit the ability of statisticians to estimate true regression coefficients accurately in Cox models. As medical research continually advances our understanding of CSF biomarkers, reliable and accurate estimation methods are of the utmost importance for efficient and effective treatment of patients. Statistical advancements in this field can be applied to the study of any disease marked by measurement error-prone biomarkers.

A number of methods for correcting measurement error have been published in recent years. Ordinary regression calibration (Prentice, 1982) operates by attenuating the hazard toward its true value, using the conditional expectation of the true covariates given the observed covariates. Prentice (1982) also demonstrated that ignoring measurement error in the covariates can produce considerable bias. Many applications of this method have been widely cited, notably by Tsiatis, DeGruttola, and Wulfsohn (1995). Xie, Wang, and Prentice (2001) proposed an expansion of ordinary regression calibration, risk set regression calibration (RRC), which recalibrates at each failure time using only the subjects in the risk set at that time. RRC adjusts time-independent covariates, creating new true covariate estimates at each failure timepoint. It has been shown to perform well in simulations, including in a joint modeling framework (Ye, Lin, and Taylor, 2008).

Neurodegenerative diseases and other terminal illnesses are often studied among populations with a high mortality rate. As AD tends to affect the elderly population, many patients who have not yet developed clinical AD, but may do so at a future time, will pass away of unrelated causes prior to symptom onset. Therefore, it is appropriate to analyze these data with a competing risks framework. Lu (2017) has developed a Bayesian framework to address these concerns in a joint survival longitudinal model setting. However, Lu's method cannot be directly applied to the Cox model, which is the focus of our paper.

Moeschberger, Tordoff, and Kochar (2008) provide a thorough review of competing risks methods. There are two primary methods most often used in practice. Cause-specific hazard models use a hazard which only considers the event of interest in its calculation. Models for this method are interpreted in a similar way to a model which does not account for competing risks (Dignam, Zhang, and Kocherginsky, 2013). These models are useful when the question of interest involves multi-state or transition models, or when multiple event types carry equal interest. However, when dealing with an elderly population and disease diagnosis, often the diagnosis is the only event of interest, and death is a nuisance event. Furthermore, clinicians and researchers are often interested not only in the hazard associated with an event of interest among competing events, but also in the incidence of this event. Despite the intuitive construction of the cause-specific hazard, it has no direct association with cumulative incidence. Therefore, analyses performed within the cause-specific framework must be interpreted with caution and must not be extended to cumulative incidence. On the other hand, the regression coefficient for a covariate in a proportional subdistribution hazards model provides information about the direction and significance of the covariate effect on cumulative incidence, because this coefficient can be thought of as coming from a generalized linear model for the cumulative incidence function with a complementary log-log link (Austin and Fine, 2017). Therefore, while estimators must always be interpreted carefully under competing risks, it is beneficial to have some insight about the cumulative incidence in the original analysis. For these reasons, we chose to use the subdistribution hazard ratio estimator, proposed by Fine and Gray (1999), for our model.

Fine and Gray used a modified hazard based on a subdistribution of the event of interest. Let T be the failure time for any event and $\epsilon \in 1, 2, \ldots, H$ denote the type of event among H possible events, with $\epsilon = 1$ the event of interest. Define a vector $\boldsymbol{Z}$ as an s x 1 time-independent vector of covariates, where s is the number of covariates in the model. The subdistribution hazard for failure type 1 is defined as

$$\lambda_1(t;\boldsymbol{Z}) = lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t \leq T \leq t + \Delta t, \epsilon = 1 | T \geq T \cup \{T \leq t \cap \epsilon \neq 1\}, \boldsymbol{Z}). \tag{2.1}$$

This induces a subdistribution $F_1(t;\boldsymbol{Z}) = P(T \leq t, \epsilon = 1)$, which has an upper limit of $P(\epsilon = 1)$ as $t \to \infty$, in contrast with the traditional distribution function's upper limit of 1. In this sense, $\lambda_1(t;\boldsymbol{Z})$ is the hazard function for the improper random variable $T^* = I(\epsilon = 1) \times T + \{1 - I(\epsilon = 1)\} \times \infty$, where I is the indicator function. The risk set for this hazard function includes the usual subjects who have not failed or been censored at time t, in addition to all subjects that have failed of events $2, \ldots, H$ at any time on the interval $[0, \infty)$. This improper risk set is the key tenant of the subdistribution model, and it is what sets it apart from other competing risks methods. This also makes it an ideal method for combination with RRC: the estimates produced by RRC can become unstable if the risk set becomes too small. If fewer than 20 subjects are at risk, then estimates from previous timepoints are carried through (Xie, Wang, and Prentice, 2001). However, improper risk sets are unlikely to shrink very much toward the end of a study, because failures due to nuisance events remain at risk throughout the study period.

There are five new contributions to the literature from this paper. First, we propose an estimator for the regression coefficients in a Cox regression model (Cox, 1972) which accounts for covariate measurement error in the presence of competing risks. Our estimator uses the time-dependent RRC estimates for the true covariates at each timepoint, thereby assuming possible measurement error but not requiring it. Covariates measured without error can be included in the model as well. Furthermore, our method substitutes the traditional hazard function from the Cox model with $\lambda_1(t;Z)$ and its corresponding improper risk set. This risk set offers stability to the RRC estimates as well as an avenue for competing risks estimation in the presence of measurement error. Second, this paper makes a substantial contribution to the scientific field of biomarker research in AD. Previously, researchers have examined associations between CSF biomarkers and risk of developing AD among ADNI data, because this question is important for drug discovery and clinical trials (Jack et al., 2010). However, typically these studies have not utilized multiple replicates per subject, opting instead to reduce the set of replicates to a single statistic (such as median or mean) for each subject. Data reduction by this naive approach results in a significant loss of information as well as opportunity to analyze data more rigorously. We are among the first to use the complete set of replicates in this data, thereby contributing a novel data analysis which is consistent with current scientific knowledge and may help to elucidate further discoveries in this therapeutic area. Third, our proposed estimator reduces bias incurred by competing risks estimators when measurement error is ignored. The effects of ignoring measurement error in the subdistribution hazard ratio estimator are directly compared to the reduced bias of our proposed estimator through simulations. We examined a variety of practical conditions, including different magnitudes of hazard ratios and measurement error variance, as well as different sample sizes. Fourth, failure to recognize the threat of competing risks can inhibit accurate risk estimation in elderly populations. Often when biomarkers are known to be measured with error, competing risks are ignored in favor of reducing the bias that comes with ignoring measurement error. Our proposed method offers greater flexibility than traditional RRC estimation by accounting for multiple failure types. Our proposed estimator can handle random right censoring, and can be extended to incorporate left truncation and time-dependent covariates as well, making it a versatile option in many traditional survival analysis studies. Finally, we develop and provide rigorous proof for the asymptotic distribution theory on a continuous time interval at all possible timepoints, subject to regularity conditions outlined in the appendices. We also provide a consistent estimator for the asymptotic variance.

Section 2.2 of this paper outlines the details of the proposed new method. Sections 2.3 and 2.4 corroborate claims with simulation results and a worked data example. Section 2.5 discusses possible extensions of the proposed method.

2.2. Proposed Methods and Asymptotic Results

For each subject i, let T_i denote the failure time, and C_i the censoring time, $i = 1, \ldots, n$. Assuming there are H observable failure types, T_i is the common failure time notation for any failure of type $1, \ldots, H$, with $\epsilon_i \in (1, \ldots, H)$ to distinguish the respective failure type. Consider failure type 1 to be the event of interest. We observe X_i = min$\{T_i, C_i\}$, the minimum of the subject's failure time and censoring time. Let δ_i = $I(\epsilon_i = 1, T_i \leq C_i)$. We denote for each subject an s x 1 true covariate vector $\boldsymbol{Z}_i$, which is time-independent. Adopting the subdistribution function framework, let $N_i(t) = I(X_i \leq t, \epsilon_i = 1)$, the counting process for the event of interest subject i. Thus, the at-risk process is $Y_i(t) = 1 - N_i(t-)$. For any time t, the risk set for failure type 1 includes subjects whose failure or censoring time has not yet been observed right before time t as well as subjects who have already failed of any event type $\epsilon_i \neq 1$. Assume that we observe data on an interval of continuous time from 0 to some maximum timepoint M, that C_i is independent of T_i conditional on the covariates $\boldsymbol{Z}_i$, and that the censoring distribution does not depend on the covariates.

Under the classical measurement error model, we do not observe the true covariate $\boldsymbol{Z}_i$ for any subject i. Instead, for $j = 1, \ldots, k_i$ replicates of the covariate measurement for subject i, $k_i = 1, \ldots, \ell$, where ℓ is the maximum number of replicates that can be attained by a single subject on the study, we observe

$$\boldsymbol{W}_{ij} = \boldsymbol{Z}_i + \boldsymbol{\zeta}_{ij}.$$

In the above, $\boldsymbol{\zeta}_{ij}$ are independent and identically distributed random variables with mean 0 and variance $\boldsymbol{\sigma}^2$, and $\boldsymbol{\zeta}_{ij}$ are independent of $\boldsymbol{Z}_i$ values.

The risk set regression calibration estimator of Xie, Wang, and Prentice (2001) relies on within-subject and between-subject variance estimation at each timepoint. First, the within-subject variance is estimated by

$$\widehat{\boldsymbol{\Delta}} = \widetilde{n}^{-1} \sum_{i=1}^{n} \sum_{j=1}^{k_i} (\boldsymbol{W}_{ij} - \overline{\boldsymbol{W}}_i)(\boldsymbol{W}_{ij} - \overline{\boldsymbol{W}}_i)',$$

where $\overline{\boldsymbol{W}}_i$ is the average of the k_i replicates for subject i, and $\widetilde{n} = \sum_{i=1}^{n}(k_i - 1)$. Next, the between-subject covariance matrix at each timepoint is

$$\widehat{\boldsymbol{\Sigma}}(t) = \left\{ \sum_{i=1}^{n} Y_i(t) - 1 \right\}^{-1} \left[\sum_{i=1}^{n} Y_i(t) \{\overline{\boldsymbol{W}}_i - \widehat{\boldsymbol{\mu}}(t)\}\{\overline{\boldsymbol{W}}_i - \widehat{\boldsymbol{\mu}}(t)\}' \right] - \left\{ \sum_{i=1}^{n} Y_i(t) \right\}^{-1} \sum_{i=1}^{n} Y_i(t) \widehat{\boldsymbol{\Delta}}/k_i,$$

where

$$\widehat{\boldsymbol{\mu}}(t) = \sum_{i=1}^{n} Y_i(t) \overline{\boldsymbol{W}}_i \bigg/ \sum_{i=1}^{n} Y_i(t).$$

Then the measurement error-corrected estimate for covariate $\boldsymbol{Z}_i$ at timepoint t can be constructed in the following way:

$$\widehat{\boldsymbol{Z}}_i(t) = \widehat{\boldsymbol{\eta}}_{k_i}(t) + \widehat{\boldsymbol{\tau}}_{k_i}(t)\overline{\boldsymbol{W}}_i, \tag{2.2}$$

where

$$\widehat{\boldsymbol{\eta}}_{k_i}(t) = \widehat{\boldsymbol{\Delta}}k_i^{-1}\{\widehat{\boldsymbol{\Sigma}}(t) + \widehat{\boldsymbol{\Delta}}k_i^{-1}\}^{-1}\widehat{\boldsymbol{\mu}}(t)$$

and

$$\widehat{\boldsymbol{\tau}}_{k_i}(t) = \widehat{\boldsymbol{\Sigma}}(t)\{\widehat{\boldsymbol{\Sigma}}(t) + \widehat{\boldsymbol{\Delta}}k_i^{-1}\}^{-1}.$$

The estimate $\widehat{\boldsymbol{Z}}_i(t)$ can be thought of heuristically as the sum of two quantities: first, the observed overall covariate mean at time t attenuated by the proportion of the total variance attributable to within-subject variability; second, the observed covariate mean for subject i attenuated by the proportion of the total variance attributable to between-subject variability. Xie, Wang, and Prentice (2001) note that the above can be obtained through a least squares estimation procedure, or by temporarily assuming that $(\boldsymbol{Z}_i, \overline{\boldsymbol{W}}_i)$ have a joint normal distribution, in which case $\widehat{\boldsymbol{Z}}_i(t)$ is the conditional expectation of $\boldsymbol{Z}_i$ given $\overline{\boldsymbol{W}}_i$ at time t. Furthermore, this method can accommodate covariates measured without error along with those measured with error. For simplicity of notation, covariates measured without error can be included in the vector $\overline{\boldsymbol{W}}_i$, so that the off-diagonal elements of $\widehat{\boldsymbol{\Sigma}}(t)$ capture the covariance among all covariates. Since there is no within-subject variability for those covariates measured without error, the corresponding elements of $\widehat{\boldsymbol{\Delta}}$ are 0.

We propose a new estimating equation to simultaneously account for covariate measurement error and competing risks. Our proposed estimating equation is based on the subdistribution hazard ratio estimator proposed by Fine and Gray for competing risks, with the modification that each covariate $\boldsymbol{Z}_i$ is replaced with its RRC estimate, $\widehat{\boldsymbol{Z}}_i(t)$, which is the estimated expected value of $\boldsymbol{Z}_i$ at each timepoint. Let $r_i(t) = I\{C_i \geq (X_i \wedge t)\}$ and $w_i(t) = r_i(t)\widehat{G}(t)/\widehat{G}(X_i \wedge t)$, as defined in Fine and Gray (1999), where $G(t)$ is the survival function associated with the censoring process, and $\widehat{G}(t)$ is its Kaplan-Meier estimator. Consider the Cox model $\lambda_1(t;\boldsymbol{Z}) = \lambda_{10}(t)exp(\boldsymbol{\beta}^T\boldsymbol{Z})$, where $\lambda_{10}(t)$ is the baseline hazard for failure type 1, and $\boldsymbol{\beta}$ is a s x 1 vector of regression coefficients. The new estimator $\widehat{\boldsymbol{\beta}}_{C-RRC}$ is the solution to

$$\boldsymbol{U}_{C-RRC}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^M \left\{ \widehat{\boldsymbol{Z}}_i(t) - \frac{\sum_{j=1}^n w_j(t) Y_j(t) \widehat{\boldsymbol{Z}}_j(t) exp(\boldsymbol{\beta}^T \widehat{\boldsymbol{Z}}_j(t))}{\sum_{j=1}^n w_j(t) Y_j(t) exp(\boldsymbol{\beta}^T \widehat{\boldsymbol{Z}}_j(t))} \right\} w_i(t) dN_i(t). \tag{2.3}$$

A shorthand for this equation can be created by defining the following:

$$\widehat{\boldsymbol{S}}^{(p)}(\boldsymbol{\beta}, t) = n^{-1} \sum_{j=1}^{n} w_j(t) Y_j(t) \widehat{\boldsymbol{Z}}_j(t)^{\otimes p} exp(\boldsymbol{\beta}^T \widehat{\boldsymbol{Z}}_j(t)), \text{ for } p = 0, 1, 2.$$

A substitution is then made in the equation for p = 0,1, producing

$$\boldsymbol{U}_{C-RRC}(\beta) = \sum_{i=1}^{n} \int_0^M \left\{ \widehat{\boldsymbol{Z}}_i(t) - \frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}, t)}{\widehat{S}^{(0)}(\boldsymbol{\beta}, t)} \right\} w_i(t) dN_i(t).$$

Using established convergence in Xie, Wang, and Prentice (2001), it can be noted that $\widehat{\boldsymbol{\beta}}_{C-RRC} \xrightarrow{p} \boldsymbol{\beta}^*$ as $n \to \infty$, where $\boldsymbol{\beta}^* \neq \boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_0$ is the true value of $\boldsymbol{\beta}$. Therefore, our new estimator will retain some asymptotic bias in estimation. However, we will show through simulations that this bias is very small under a variety of practical situations. We also show that our new estimator significantly reduces the bias of the subdistribution hazard ratio estimator.

Additionally, we show in Appendix A.1 that $n^{-1/2}\boldsymbol{U}_{C-RRC}(\boldsymbol{\beta})$ is asymptotically equivalent to a sum of independent, identically distributed random variables. Therefore,
$n^{1/2}(\widehat{\boldsymbol{\beta}}_{C-RRC} - \boldsymbol{\beta}^*)$ is asymptotically normally distributed with mean 0 and covariance matrix $\boldsymbol{A}^{-1}(\boldsymbol{\beta}^*)\boldsymbol{B}(\boldsymbol{\beta}^*)\boldsymbol{A}^{-1}(\boldsymbol{\beta}^*)$ as $n \to \infty$. Definitions of matrices A and B along with their consistent estimates are provided in A.1. This asymptotic variance is used to obtain standard error estimates of the proposed estimator in the simulations and data example.

2.3. Simulation

A cohort of n = 200 subjects was generated along with failure times, censoring times, and failure types according to the algorithm described in Fine and Gray (1999). To detail the algorithm, a single covariate was generated for each subject, and H = 2 failure types were considered. Denote β_0 the true regression coefficient for type 1 failures, the event of interest; β_2 is the true regression coefficient for type 2 failures, the nuisance event. The subdistribution for type 1 failure was generated from an exponential mixture distribution as

$$P(T_i \leq t, \epsilon_i = 1|Z_i) = 1 - [1 - p\{1 - exp(-t)\}]^{exp(\beta_0^T Z_i)}.$$

where p is a parameter specifying the mixture weight. Following this, $P(\epsilon_i = 1|Z_i)$ can be obtained by taking the limit as $t \to \infty$ to obtain

$$P(\epsilon_i = 1|Z_i) = 1 - (1 - p)^{exp(\beta_0^T Z_i)}.$$

Then, t can be solved for using the relation $P(T_i \leq t, \epsilon_i = 1|Z_i) = P(T_i \leq t|\epsilon_i = 1, Z_i)P(\epsilon_i = 1|Z_i)$ and the fact that $P(T_i \leq t|\epsilon_i = 1, Z_i)$ is a uniform(0,1) random variable. Type 2 failure times were generated from an exponential distribution with rate parameter $exp(\beta_2^T Z_i)$. Failure type was ultimately decided using a Bernoulli distribution with probability parameter $P(\epsilon_i = 1|Z_i)$. All values for $\widehat{\beta}_{SHR}$, the logarithm of the subdistribution hazard ratio estimator, were obtained using the `cmprsk` package in R version `3.3.2`. Since this package does not accommodate replicate data, we used the median of the replicate covariate values for each subject in the calculation of β_{SHR}. Varying degrees of measurement error were applied using the noise-to-signal ratio (NSR), calculated as (σ^2/k)/Var(Z), where k is the number of replicates per subject. In all simulations, k = 4 and the distribution of Z was chosen such that Var(Z) = 1. Therefore, in these simulations, we vary the NSR by varying only σ^2. An NSR value of 0.05 was chosen to emulate the magnitude of measurement error variance in our data example. Values of 0.2, 1, and 4 were chosen for σ^2, to produce respective NSR values of 0.05, 0.25, and 1. All results were obtained using 1000 simulation replicates. The standard error for $\widehat{\beta}_{C-RRC}$ was calculated using the formula for the asymptotic variance of the estimator found in Appendix A.1. Tables 2.1 and 2.2 display the bias ($\widehat{\beta} - \beta_0$), empirical standard deviation of estimates across the 1000 simulation runs (SD), estimated standard error ($\widehat{\text{SE}}$) obtained from asymptotic variance, mean squared error (MSE), and coverage of 95% confidence intervals (COV).

The results of the first simulation are displayed in Table 2.1. Z is distributed as N(0,1). Censoring times were generated from a uniform(0.5,2) distribution and p set to 0.3, to produce approximately 30% censoring. Values of β_0 were chosen to give a hazard ratio of 2 and 3, respectively. In all simulations, β_0 and β_2 (not shown) are equal. It is apparent that the bias of $\widehat{\beta}_{SHR}$ exceeds that of $\widehat{\beta}_{C-RRC}$ in every parameter combination. In fact, the bias of $\widehat{\beta}_{SHR}$ increases to roughly 50% of

Table 2.1: $\widehat{\beta}_{C-RRC}$ is the proposed estimator; $\widehat{\beta}_{SHR}$ is the logarithm of the subdistribution hazard ratio estimator. Bias = $\widehat{\beta} - \beta_0$. SD = empirical standard deviation of estimates across the 1000 simulation runs. $\widehat{\text{SE}}$ = estimated standard error. MSE = mean squared error. COV = coverage of 95% confidence intervals. In all simulations, β_2 (not shown) is equal to β_0. Z is a N(0,1) random variable. NSR = noise-to-signal ratio, defined as (σ^2/k)/Var(Z), where each subject has k = 4 replicates. HR = hazard ratio.

		n=200					
		NSR = 0.05		NSR = 0.25		NSR = 1.00	
β_0	Parameter	$\widehat{\beta}_{C-RRC}$	$\widehat{\beta}_{SHR}$	$\widehat{\beta}_{C-RRC}$	$\widehat{\beta}_{SHR}$	$\widehat{\beta}_{C-RRC}$	$\widehat{\beta}_{SHR}$
0.6931 (HR = 2)	Bias	0.010	-0.027	0.012	-0.140	0.360	-0.355
	SD	0.161	0.152	0.183	0.139	0.276	0.109
	Mean($\widehat{\text{SE}}$)	0.156	0.147	0.176	0.134	0.263	0.105
	MSE	0.024	0.022	0.031	0.038	0.070	0.137
	95% COV	0.943	0.930	0.938	0.784	0.957	0.107
1.0986 (HR = 3)	Bias	0.007	-0.054	-0.005	-0.255	0.013	-0.603
	SD	0.170	0.158	0.194	0.140	0.324	0.104
	Mean($\widehat{\text{SE}}$)	0.159	0.149	0.184	0.133	0.308	0.101
	MSE	0.025	0.025	0.034	0.083	0.095	0.374
	95% COV	0.939	0.917	0.940	0.483	0.938	0.001

the value of β_0 when σ^2 increases. When the NSR is 0.05 (very small measurement error), the bias from $\widehat{\beta}_{SHR}$ is small. The standard error from $\widehat{\beta}_{C-RRC}$ is only marginally larger than that from $\widehat{\beta}_{SHR}$. The MSE are almost identical for both methods. When the NSR is 0.25 or higher (moderate to large measurement error), $\widehat{\beta}_{SHR}$ exhibits considerable bias and the corresponding confidence interval is very inaccurate. $\widehat{\beta}_{C-RRC}$ performs well with little bias and excellent confidence interval coverage. $\widehat{\beta}_{C-RRC}$ has a larger standard error than $\widehat{\beta}_{SHR}$ due to an additional component to be estimated in its score equation. However, MSE($\widehat{\beta}_{C-RRC}$) is smaller than MSE($\widehat{\beta}_{SHR}$), indicating that the increase in bias of $\widehat{\beta}_{SHR}$ is large enough to overcome its smaller standard error. This is evident in the fact that the coverage probability for $\widehat{\beta}_{SHR}$ plummets to 0 quite quickly, particularly at large values of β_0 and large σ^2. The coverage probability for $\widehat{\beta}_{C-RRC}$ holds near or above 90% for all simulations. This stark difference in coverage reveals that confidence intervals for $\widehat{\beta}_{SHR}$ should generally not be trusted in practice when the dataset contains a moderate to large amount of measurement error, despite the fact that the confidence intervals for $\widehat{\beta}_{C-RRC}$ under the same conditions will be slightly wider. Therefore, we can consider both the confidence intervals and the point estimates from $\widehat{\beta}_{C-RRC}$ to be more robust to unfavorable conditions than $\widehat{\beta}_{SHR}$.

Table 2.2: $\widehat{\beta}_{C-RRC}$ is the proposed estimator; $\widehat{\beta}_{SHR}$ is the logarithm of the subdistribution hazard ratio estimator. Bias = $\widehat{\beta} - \beta_0$. SD = standard deviation of estimates across the 1000 simulation runs. SE = standard error. MSE = mean squared error. COV = coverage of 95% confidence intervals. In all simulations, β_2 (not shown) is equal to β_0. Z is a uniform(0, $\sqrt{12}$) random variable. NSR = noise-to-signal ratio, defined as (σ^2/k)/Var(Z), where each subject has k = 4 replicates. HR = hazard ratio.

		n=200					
		NSR = 0.05		NSR = 0.25		NSR = 1.00	
β_0	Parameter	$\widehat{\beta}_{C-RRC}$	$\widehat{\beta}_{SHR}$	$\widehat{\beta}_{C-RRC}$	$\widehat{\beta}_{SHR}$	$\widehat{\beta}_{C-RRC}$	$\widehat{\beta}_{SHR}$
0.6931 (HR = 2)	Bias	0.003	-0.031	-0.014	-0.156	-0.009	-0.369
	SD	0.209	0.198	0.218	0.169	0.302	0.126
	Mean($\widehat{\text{SE}}$)	0.198	0.188	0.209	0.162	0.287	0.122
	MSE	0.039	0.036	0.044	0.051	0.083	0.151
	95% COV	0.945	0.942	0.935	0.814	0.944	0.172
1.0986 (HR = 3)	Bias	-0.014	-0.073	-0.049	-0.285	-0.039	-0.620
	SD	0.159	0.149	0.177	0.127	0.307	0.093
	Mean($\widehat{\text{SE}}$)	0.157	0.147	0.170	0.123	0.277	0.089
	MSE	0.025	0.027	0.031	0.097	0.078	0.393
	95% COV	0.952	0.907	0.927	0.350	0.905	0.001

In Table 2.2, the same simulation was performed with covariates generated from a uniform(0,$\sqrt{12}$) distribution, to assess the performance of the estimator under non-normally distributed covariates. The upper limit of the uniform distribution was chosen to guarantee that the covariate has a variance of 1. Censoring was maintained at approximately 25%, with censoring times generated from a uniform(0.2,1) distribution and p = 0.1. Once again, we see that the 95% coverage probability of $\widehat{\beta}_{SHR}$ falls well below 95%, particularly at increasing magnitudes of β_0 and σ^2. This implies that confidence intervals obtained from this method without accounting for measurement error will not give the desired coverage. In contrast, $\widehat{\beta}_{C-RRC}$ once again performs well, with a larger standard error but stable MSE. All simulations were performed with n = 200 subjects and repeated with n = 100 subjects to demonstrate the performance of the estimators at a smaller sample size (n=100 results are included in Appendix A.2). All results were similar between the two sample sizes.

2.4. Analysis of Association Between CSF Biomarkers and Risk of Conversion to AD

Motiving data comes from the Alzheimer's Disease Neuroimaging Initiative (ADNI), a multisite, ongoing longitudinal study which validates the use of biomarkers for AD diagnosis and clinical trials (Weiner et al., 2012). This example analyzes data collected from the start of the study in 2004 up to September 1, 2016. Our primary scientific question is how CSF biomarkers are related to the risk of conversion to AD. Each patient underwent at least one CSF collection; most patients returned annually for subsequent collections. Samples of CSF were divided into many subsamples, and all but one were frozen immediately after collection. When a patient returned for the next collection, one of the previous subsamples was processed and CSF biomarkers were measured. Therefore, each patient's number of replicates of the baseline measurement is equal to the number of visits. In addition to biomarker collection, patients also underwent a clinical exam at each visit. Conversion to AD was recorded based on clinical exams. The event of interest is first diagnosis of AD; death is considered the competing event. Time 0 is defined as study entry. A total of 1064 subjects entered the study without AD; of those, 211 converted to AD during the study, and 79 died without conversion to AD. This results in a ratio of death to AD of 0.37, which is comparable to the ratio of nuisance event to event of interest featured in Fine and Gray (1999), which was 0.42. The sample of patients is largely elderly, with mean age 73 (SD = 7.1). Patients without a record of death or conversion to AD were considered right-censored at the date of the most recent clinical cognitive exam. Demographics are summarized by gender in Table 2.3.

The methods of Section 2.2 were implemented to answer the above scientific question. In this situation, censoring does not include death, but includes random occurrences, such as moving geographic locations, that may lead to a loss to follow-up. Therefore, the study design does not lead us to believe that the censoring distribution depends on the covariates. Because recent literature cites evidence that abnormalities in amyloid-β (Aβ) are detectable earlier in the disease stage than abnormalities of other CSF biomarkers (Jack et al., 2010), Aβ was chosen as the CSF biomarker covariate for this model. The original unit of measurement for Aβ is pg/mL, but the variable was rescaled by dividing each value by 100 to allow easier interpretation. Tables 2.3 and 2.4 reflect the rescaled values. The remaining covariates, which are not assumed to have measurement error, are gender, age (years) at study entry, education (years), and presence of ApoE4 gene. The

Table 2.3: Continuous variables summarized with mean and standard deviation in parentheses; categorical variables summarized with counts and percentages in parentheses.

Predictor	Male	Female
N	587	477
CSF Aβ (100 pg/mL)	1.76 (0.56)	1.80 (0.55)
Age (years)	74.0 (7.0)	72.2 (7.1)
Education (years)	16.6 (2.8)	15.5 (2.7)
ApoE4		
0	325 (55.4%)	271 (56.8%)
1	204 (34.8%)	169 (35.4%)
2	58 (9.9%)	37 (7.8%)
Conversions to AD (Yes/No)	126	85
Deaths (Yes/No)	53	26
Ratio of Death Counts to AD Counts	0.42	0.31

corresponding elements of $\widehat{\boldsymbol{\Delta}}$ are 0 for these covariates measured without error; thus, it can be seen in equation 2.2 that $\widehat{\boldsymbol{Z}}_i(t) = \overline{\boldsymbol{W}}_i, \forall i, t$, for all covariates except Aβ. To calculate the NSR for this dataset, we obtained $\widehat{\boldsymbol{\Sigma}}(t)$ evaluated at t = 0, the baseline measurement. We consider $\widehat{\boldsymbol{\Sigma}}(t)$ evaluated at t = 0 to be an appropriate estimate for Var($\boldsymbol{Z}$), because all subjects are in the risk set at baseline. We calculated $\widehat{\text{Var}}(\boldsymbol{Z})$ = 0.2725 and $\widehat{\boldsymbol{\Delta}}$=0.0198. The NSR is then $\widehat{\boldsymbol{\Delta}}/\widehat{\text{Var}}(\boldsymbol{Z})$ = 0.07, as noted in Section 2.3. This low NSR is consistent with NSRs found across normal, MCI, and AD patients in other published studies using ADNI data (White, Shaw, and Xie, 2016). Results can be found in Table 2.4. Ties in event times were accounted for using Breslow's approximation (Kalbfleisch and Prentice, 1980). For the subdistribution model, each subject's single Aβ value was taken to be the median of his/her baseline replicate values. To assess the assumption of additive measurement error, we employed the procedure of (White and Xie, 2013), by plotting intra-subject standard deviations against means (not shown). We found no discernible patterns, indicating that the additive error assumption is not violated.

We have seen in Section 2.3 that $\widehat{\beta}_{SHR}$ has small bias when the measurement error in the dataset is small, a condition which holds in this example. In this analysis, both $\widehat{\beta}_{C-RRC}$ and $\widehat{\beta}_{SHR}$ are comparable in both magnitude and standard error; however, we do see attenuation toward 0 of $\widehat{\beta}_{SHR}$. Furthermore, the confidence intervals for the Aβ coefficient do not include 0 by either method. Whether an investigator uses the proposed method or the SHR method, he or she would conclude that lower levels of Aβ are significantly associated with risk of conversion to AD, a conclusion which

Table 2.4: $\widehat{\beta}_{C-RRC}$ is the proposed estimator; $\widehat{\beta}_{SHR}$ is the logarithm of the subdistribution hazard ratio estimator. $\widehat{\text{SE}}$ = estimated standard error. CI = confidence interval.

	$\widehat{\beta}_{C-RRC}$			$\widehat{\beta}_{SHR}$		
Predictor	Estimate	$\widehat{\text{SE}}$	95% CI	Estimate	$\widehat{\text{SE}}$	95% CI
CSF Aβ	-1.525	0.190	(-1.897, -1.152)	-1.426	0.175	(-1.768, -1.083)
Age	-0.013	0.011	(-0.035, 0.009)	-0.012	0.011	(-0.034, 0.010)
Gender	0.081	0.148	(-0.209, 0.372)	0.088	0.148	(-0.202, 0.378)
Education	-0.003	0.026	(-0.054, 0.048)	-0.004	0.026	(-0.055, 0.047)
ApoE4	0.169	0.116	(-0.057, 0.396)	0.210	0.112	(-0.009, 0.430)

is consistent with previous studies (Shaw et al., 2011). The similarity between these two estimates is confirmed in the simulation results of Section 2.3, where it was shown that bias for $\widehat{\beta}_{SHR}$ is low and the coverage probability is above 90% for both methods with a low NSR of 0.05. As we can see from Section 2.3, the proposed method works well, removing bias and leading to valid conclusions, regardless of the magnitude of measurement error variance. While the subdistribution model has similar conclusions in this data example due to the small amount of measurement error, it does not make full use of the available replicate dataset and runs the risk of generating incorrect conclusions if the measurement error were larger. Thus, we recommend using the proposed method whenever covariate replicate data is available in practice.

2.5. Discussion

The proposed method implemented in this paper removes bias incurred by the subdistribution hazard ratio estimator caused by covariates measured with error. It accomplishes this by recalibrating at each time where an event of interest occurred, using only the subjects in the risk set at that time. It is evident in Section 2.3 that this bias of $\widehat{\beta}_{SHR}$ is significant, especially when the amount of measurement error is moderate or large. Furthermore, the coverage probability of the subdistribution model is very sensitive to both measurement error and the magnitude of the true regression coefficients. Although the subdistribution model has been groundbreaking in addressing competing risks, this bias can be removed almost completely by simultaneously accounting for covariate measurement error. Our proposed method has robust coverage probability that is only marginally impacted by extreme conditions, and remains high under reasonable circumstances. Additionally, the MSE for our proposed method is better than or on par with that of $\widehat{\beta}_{SHR}$. This indicates that the increased variance that results from estimating an extra component of the score equation does not eclipse the decrease in bias. We recommend use of our proposed method whenever replicate data is available, as our method will provide valid conclusions in the presence of measurement error whose variance ranges from small to large in magnitude.

Although not presented in this paper, our proposed method can be extended to incorporate left-truncated data, which often occurs in studies where patients have delayed entry into the study (i.e., entering after time 0). In this situation, the risk set can be redefined to include only subjects who entered after the truncation time, in addition to the risk set constraints described in Section 2.2. Although we present the proposed method in the context of time independent covariates, it is straightforward to model time-dependent covariates by recalibrating in a similar fashion. There are some limitations to our method which require further exploration. First, our method requires replicate data. Not all subjects need to have more than one replicate, but it is necessary that a subset of them do. Replicate data is available from ADNI, but other studies may have more difficulty accessing or collecting replicate data. However, our method can be extended to include external validation datasets to circumvent this problem. Second, our method assumes a classical measurement error model, where errors are additive and independent. Our method can be extended to encompass correlated measurement errors, for situations where the classical measurement error model is not appropriate. However, even when we use the classical measurement error model, we do not require that the measurement error or covariates be normally distributed. This is because the RRC estimates can be derived using a least squares method, rather than assuming normality, as mentioned in Section 2.2 (Xie, Wang, and Prentice, 2001).

The methods in this paper carry important clinical implications. Failure to employ a robust estimation method will often result in misleading or incorrect conclusions. In a clinical setting, this can lead to squandering valuable resources by treating too many patients whose risk for conversion to AD was overestimated, or neglecting to treat patients whose true risk for AD was underestimated. In either case, the burden that AD imposes on patients, families, caregivers, and society will potentially increase. Additionally, as the intricate relationship between biomarkers and AD becomes better understood, it is paramount that statistical methods can remain on par with ongoing medical advances. By accounting for covariate measurement error in the presence of competing risks, we can utilize a reliable and accurate estimation procedure which addresses these challenges. Our method can be applied to any disease study with measurement error-prone biomarkers and a population subject to multiple failure types.

CHAPTER 3

Efficient Linear Longitudinal Models in the Presence of Missing Outcomes

3.1. Introduction

Linear mixed-effects models have long been an invaluable tool in analyzing longitudinal data. Ever since the theory and estimation for these models was made widely available by Laird and Ware (1982), they have provided a crucial pathway for characterizing individual effects on population means, and remain a popular choice of analysis for clinical trials. However, the persistent challenge of missing data in longitudinal studies does not discriminate by model type. This ongoing problem demands continued improvements in research, in order to ensure that investigators can evaluate the effects of the predictors in their studies in the most efficient and reliable way possible.

Despite the many strides made in the approach to handling missing outcomes, complete case analysis (CCA) remains the default approach in many software packages. CCA estimates are unbiased when the outcomes are missing completely at random (MCAR); however, the omission of observations from the analysis results in a loss of efficiency and power that can make the difference between success and failure in a clinical trial. Such a critical flaw necessitates appropriate methods to incorporate the incomplete cases into the analysis.

A very popular choice is multiple imputation (MI), first introduced by Rubin (1987). MI has proved to be an efficient way to analyze data (Ibrahim et al., 2005), and many variations of the approach have been developed over the years, including predictive mean matching (Schenker and Taylor, 1996) and multiple imputation with chained equations (Erler et al., 2016). In the absence of a definitive "gold standard" for missing data analyses, MI and its variations remain the conventional choice for handling missing data in a variety of situations. While these methods improve efficiency compared to CCA (Ibrahim et al., 2005; Qi and Sun, 2014), they require parametric assumptions in addition to the parameters of the regression model. For example, MI requires specifying a relationship between the covariates and missing outcome in order to impute missing values; predictive mean matching (PMM) uses an assumed model in order to compute the predicted means, which are used to select an imputed value from the set of complete cases. Misspecification of the assumed model in MI methods can lead to biased results (Schenker and Taylor, 1996; Tomita, Fujisawa, and Henmi, 2018).

Other analysis choices include weighted estimating equations, Bayesian methods, and inverse probability weighting (IPW). These analyses are all either parametric or semiparametric, and thus subject to the same vulnerabilities (Ibrahim et al., 2005; Qi and Sun, 2014). IPW can be used nonparametrically, by estimating the weights from the data; however, IPW methods are subject to inefficiency and bias in a variety of situations (Clayton et al., 1998; Carpenter, Kenward, and Vansteelandt, 2006). An augmented IPW (Robins, Rotnitzky, and Zhao, 1994; Scharfstein, Rotnitzky, and Robins, 1999) has an attractive doubly robust property that the other methods lack. These methods have been extensively evaluated in literature (Seaman and Vansteelandt, 2018; Kang and Schafer, 2007). We offer an alternative robust approach that avoids placing parametric restrictions when accounting for missing data by using an auxiliary outcome. In other words, our new approach does not require additional parametric assumptions besides the standard assumptions from the mixed-effects model.

Use of an auxiliary outcome offers a way to incorporate all observations into an analysis, complete or incomplete, by taking advantage of the information contained in the relationship between the auxiliary outcome and the true outcome. Pepe (1992) developed a method which takes advantage of an auxiliary outcome by nonparametrically estimating its relationship to the true outcome, thereby avoiding the pitfall of misspecification altogether. Pepe, Reilly, and Fleming (1994)'s mean score method similarly takes advantage of auxiliary data without placing restrictions on its relationship to the missing outcome. Both methods were developed for cross-sectional studies.

This article is motivated by the Parkinson's Progression Markers Initiative (PPMI), the largest ongoing, prospective, biomarker-rich longitudinal study in early Parkinson's disease (PD). The data from this study present a promising avenue for advances in PD research. The study collects a variety of useful endpoints, one of which is magnetic resonance imaging (MRI). Our scientific aim is to estimate the longitudinal change in gray matter volume over time measured by MRI. However, some participating study sites lack the facilities to perform research-quality MRI scans. As a result, many study subjects (about 60%) are missing an MRI scan at all visits. Removing these subjects from the analysis would greatly reduce the sample size of the study and the efficiency of any regression estimates. Additionally, while MI is a possible choice for analyzing this data, the significant proportion of subjects without MRI scans calls into question the efficiency of MI or PMM estimates, as datasets with a small number of complete cases can lend to instability in MI or PMM estimates (Schenker and Taylor, 1996).

We propose a novel method for longitudinal data with a missing continuous outcome which enables use of all observed values in the dataset in a linear mixed-effects model, by making use of an auxiliary outcome. There are three new contributions to the literature in this article which we summarize here. First, we avoid imposing extraneous parametric assumptions, specifications, or restrictions on the analysis. Instead, we estimate the relationship between the auxiliary outcome and the true outcome nonparametrically, and develop a regression estimator by extending the estimated likelihood method (Pepe, 1992) to longitudinal settings. This nonparametric approach produces estimates that are robust to misspecification and thus desirable over previously described parametric and semi-parametric methods. Second, our proposed estimator is shown to be more efficient than CCA, yet still consistent. The efficiency of the proposed estimator is dependent on the correlation between auxiliary and true outcomes, with a stronger correlation leading to greater efficiency gains. However, we will show that this property does not lead to estimates that are less efficient than CCA estimates, even if the auxiliary and true outcomes are completely uncorrelated. Therefore, use of our proposed estimator carries the potential benefit of efficiency gains, and does not carry risk of efficiency loss compared to CCA. Like existing methods, our proposed estimator can potentially lose efficiency when faced with a large amount of missing data. However, unlike methods which do not use an auxiliary outcome, much of this efficiency can be regained by choosing an auxiliary outcome that is strongly correlated with the true outcome. Third, the estimated maximum likelihood procedure outlined by Pepe (1992) has remained an attractive option for analyzing missing data since its inception, but it was developed for cross-sectional data and addresses only two types of subjects: those with an observed outcome and those without. Our proposed method addresses a third type of subject previously unrecognized in an estimated maximum likelihood method, subjects with an outcome observed at some timepoints, but not all. We develop a novel approach to incorporate these subjects into the analysis and thus use all available data in our internal validation procedure. Our proposed method is robust both to non-monotone missingness, and to staggered timepoints, where subjects may not necessarily attend visits at the same times. We develop and provide rigorous proof for the asymptotic distribution theory, subject to regularity conditions outlined in the appendices. We also provide a consistent estimator for the asymptotic variance.

Section 3.2 of this article outlines the details of the proposed new method. Sections 3.3 and 3.4 corroborate claims with simulation results and a worked data example from the PPMI study. Section 3.5 discusses special properties and possible extensions of the proposed method. Additional simulation results and asymptotic theory are available in Appendix B.2.

3.2. Proposed Methods and Asymptotic Results

Let $\boldsymbol{Y}_\ell$ denote the T x 1 vector of true outcomes for subject $\ell = 1, \ldots, N$, where T is the total number of timepoints on the study and N the total number of subjects. We assume the standard framework for a linear mixed-effects model (Laird and Ware, 1982), namely that

$$\boldsymbol{Y}_\ell = \boldsymbol{X}_\ell \boldsymbol{\beta} + \boldsymbol{b}_\ell \boldsymbol{Z}_\ell + \boldsymbol{\epsilon}_\ell, \tag{3.1}$$

where for subject ℓ, $\boldsymbol{X}_\ell$ is the design matrix of fixed effects of size T x p, $\boldsymbol{\beta}$ is the p x 1 vector of regression coefficients, $\boldsymbol{Z}_\ell$ is the T x q matrix of covariates for the q x 1 vector $\boldsymbol{b}_\ell$ of random effects, and $\boldsymbol{\epsilon}_\ell$ is the T x 1 vector of random errors. We make two typical assumptions for this type of model. First, we assume that $\boldsymbol{b}_\ell$ and $\boldsymbol{\epsilon}_\ell$ are independent from each other. Second, we assume that both $\boldsymbol{\epsilon}_\ell$ and $\boldsymbol{b}_\ell$ have multivariate normal distributions, with mean 0 and variance-covariance matrices $\sigma^2 \boldsymbol{I}$ and $\boldsymbol{D}$, respectively. Here, $\boldsymbol{D}(\gamma)$ is a matrix of variance parameters γ for the distribution of $\boldsymbol{b}_\ell$. Therefore, the full parameter set for the model is $\boldsymbol{\Theta} = \{\boldsymbol{\beta}, \sigma^2, \boldsymbol{D}(\gamma)\}$. None of the variance parameters in $\boldsymbol{\Theta}$ are assumed to be known; we will focus on $\boldsymbol{\beta}$ as the parameters of interest, but all parameters in $\boldsymbol{\Theta}$ will be estimated simultaneously in the following procedure.

We partition the total set of subjects into three subsets. Denote V the validation set, containing subjects $i = 1, ..., n_V$ who have a true outcome observed at all T timepoints. The second partition is $\underline{V}$, the hybrid set, containing subjects $k = 1, ..., n_{\underline{V}}$ having a true outcome observed at least once but not at all timepoints. Denote $\underline{\boldsymbol{Y}}_k$ the vector of observed outcomes for subject k, of length $t_k < T$. Finally, $\overline{V}$ is the nonvalidation set, containing subjects $j = 1, ..., n_{\overline{V}}$ with true outcome missing at all T timepoints. $n_V, n_{\underline{V}}$, and $n_{\overline{V}}$ are the total number of subjects in each respective set. Throughout this article, we will strictly adhere to the indices i for subjects in V, k for subjects in $\underline{V}$, and j for subjects in $\overline{V}$, with ℓ reserved for indexing quantities that may apply to more than one subset.

There are two mechanisms of missing data in the PPMI dataset. First, subjects in $\underline{V}$ may miss one or more MRI scans mainly due to scheduling conflicts. Second, the fact that subjects in $\overline{V}$ have no MRI scans is due to study design and not related to patient characteristics. A key assumption of the following proposed methods is that V is a representative subsample of the total population; this is accomplished if membership in $\overline{V}$ is determined by an MCAR mechanism. For these subjects, MRI data are missing because the participating sites do not have the capability to perform high research quality MRI scans. Thus, we dont have any evidence against the assumption that the validation set is a random sample of the whole cohort. To further verify this, Schuff et al. (2015) found that the MRI cohort and non-MRI cohorts were similar in their demographic and disease stage characteristics, suggesting that the random sample assumption is appropriate for the validation set in the PPMI study.

Denote $\boldsymbol{S}$ the vector of auxiliary outcomes. We require that all subjects have S recorded at every timepoint; therefore, $\boldsymbol{S}_\ell$ is of length T for each subject ℓ. We also assume no missing covariates, and so $\boldsymbol{X}_\ell$ has size T x p for p regression coefficients in the model. Both of these assumptions are reasonable for the PPMI study, as our auxiliary outcome is constructed from neuropsychological assessment scores, which are readily available without MRI equipment.

The entire observed data are $(\boldsymbol{Y}_i, \boldsymbol{S}_i, \boldsymbol{X}_i)$ for subjects in the validation set, $(\underline{\boldsymbol{Y}}_k, \boldsymbol{S}_k, \boldsymbol{X}_k)$ for subjects in the hybrid set, and $(\boldsymbol{S}_j, \boldsymbol{X}_j)$ for subjects in the nonvalidation set. The estimated likelihood is then the product of estimated probabilities of all observed data:

$$\mathscr{L}(\boldsymbol{\beta}) = \prod_{i \in n_V} \widehat{P}_{\boldsymbol{\beta}}(\boldsymbol{Y}_i, \boldsymbol{S}_i | \boldsymbol{X}_i) \prod_{k \in n_{\underline{V}}} \widehat{P}_{\boldsymbol{\beta}}(\underline{\boldsymbol{Y}}_k, \boldsymbol{S}_k | \boldsymbol{X}_k) \prod_{j \in n_{\overline{V}}} \widehat{P}_{\boldsymbol{\beta}}(\boldsymbol{S}_j | \boldsymbol{X}_j).$$

Using the subscript $\boldsymbol{\beta}$ to indicate which quantities depend on $\boldsymbol{\beta}$, we can express the likelihood as

$$\mathscr{L}(\boldsymbol{\beta}) = \prod_{i \in n_V} P_{\boldsymbol{\beta}}(\boldsymbol{Y}_i | \boldsymbol{X}_i) \widehat{P}(\boldsymbol{S}_i | \boldsymbol{Y}_i, \boldsymbol{X}_i) \prod_{k \in n_{\underline{V}}} \widehat{P}_{\boldsymbol{\beta}}(\underline{\boldsymbol{Y}}_k, \boldsymbol{S}_k | \boldsymbol{X}_k) \prod_{j \in n_{\overline{V}}} \widehat{P}_{\boldsymbol{\beta}}(\boldsymbol{S}_j | \boldsymbol{X}_j). \tag{3.2}$$

Since there is no missing data in V, $\widehat{P}(\boldsymbol{S}_i | \boldsymbol{Y}_i, \boldsymbol{X}_i)$ in the first term of the likelihood is a constant with respect to $\boldsymbol{\beta}$ and can be ignored. We can express the estimated terms in the likelihood by integrating over the missing outcomes. For the nonvalidation set,

$$\widehat{P}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j) = \int \widehat{P}_{\boldsymbol{\beta}}(\boldsymbol{S}_j, \boldsymbol{y}|\boldsymbol{X}_j)d\boldsymbol{y} = \int P_{\beta}(\boldsymbol{y}|\boldsymbol{X}_j)\widehat{P}(\boldsymbol{S}_j|\boldsymbol{y}, \boldsymbol{X}_j)d\boldsymbol{y}.$$

As noted in Section 3.1, we aim to develop a nonparametric approach to estimate the relationship between $\boldsymbol{Y}$ and $\boldsymbol{S}$, in order to avoid assuming or specifying the form of $\widehat{P}(\boldsymbol{S}_j|\boldsymbol{y}, \boldsymbol{X}_j)$. Using similar arguments as in Pepe (1992), we compute this estimated probability as a ratio of empirical probabilities,

$$\widehat{P}(\boldsymbol{S}_j|\boldsymbol{y}, \boldsymbol{X}_j) = \frac{\widehat{P}(\boldsymbol{S}_j, \boldsymbol{y}, \boldsymbol{X}_j)}{\widehat{P}(\boldsymbol{y}, \boldsymbol{X}_j)}.$$

3.2.1. Estimation Using an Internal Validation Subsample

Because the validation set is a representative subset of the total population, we can estimate the probability of observing the values of $\boldsymbol{S}_j, \boldsymbol{y}$, and $\boldsymbol{X}_j$ in the population by calculating the probability of observing these values in the validation set. The probabilities of interest are

$$\widehat{P}(\boldsymbol{y}, \boldsymbol{X}_j) = \frac{1}{n_V}\sum_{i \in n_V} |\boldsymbol{H}_1|^{-1/2}\phi(\boldsymbol{L}_1(\boldsymbol{y} - \boldsymbol{Y}_i, \boldsymbol{X}_j - \boldsymbol{X}_i))$$

and

$$\widehat{P}(\boldsymbol{S}_j, \boldsymbol{y}, \boldsymbol{X}_j) = \frac{1}{n_V}\sum_{i \in n_V} |\boldsymbol{H}_2|^{-1/2}\phi(\boldsymbol{L}_2(\boldsymbol{S}_j - \boldsymbol{S}_i, \boldsymbol{y} - \boldsymbol{Y}_i, \boldsymbol{X}_j - \boldsymbol{X}_i))$$

where $(\boldsymbol{S}_j - \boldsymbol{S}_i, \boldsymbol{y} - \boldsymbol{Y}_i, \boldsymbol{X}_j - \boldsymbol{X}_i)$ represents the vectors $\boldsymbol{S}_i, \boldsymbol{Y}_i$, and $\boldsymbol{X}_i$ concatenated together, and subtracted from the vectors $\boldsymbol{S}_j, \boldsymbol{y}, \boldsymbol{X}_j$ concatenated together. For a model having s covariates, these vectors are each of length $2T + s$. Likewise, the vectors in the construction of $\widehat{P}(\boldsymbol{y}, \boldsymbol{X}_j)$ are of length $T + s$. $\boldsymbol{H}_1, \boldsymbol{H}_2$ are bandwidth matrices for the kernel function ϕ, and $\boldsymbol{L}_1, \boldsymbol{L}_2$ are the Cholesky decompositions of $\boldsymbol{H}_1^{-1}, \boldsymbol{H}_2^{-1}$, respectively, such that $\boldsymbol{L}_1^T\boldsymbol{L}_1 = \boldsymbol{H}_1^{-1}$ and $\boldsymbol{L}_2^T\boldsymbol{L}_2 = \boldsymbol{H}_2^{-1}$. All together, the above expressions quantify how well each subject i is "matched" with subject j using a multivariate kernel function. In all simulations and computations in this manuscript, we use the multivariate Gaussian kernel, but our proposed method requires only that the chosen kernel function is symmetric in all dimensions. If the model uses any discrete covariates $\boldsymbol{X}_D$, these discrete covariates can be removed from ϕ and $\boldsymbol{H}$ and replaced with an indicator function; the probabilities then become

$$\widehat{P}(\boldsymbol{y}, \boldsymbol{X}_j) = \frac{1}{n_V} \sum_{i \in n_V} I(\boldsymbol{X}_{Dj} = \boldsymbol{X}_{Di}) |\boldsymbol{H}_1|^{-1/2} \phi(\boldsymbol{L}_1(\boldsymbol{y} - \boldsymbol{Y}_i, \boldsymbol{X}_j - \boldsymbol{X}_i))$$

and

$$\widehat{P}(\boldsymbol{S}_j, \boldsymbol{y}, \boldsymbol{X}_j) = \frac{1}{n_V} \sum_{i \in n_V} I(\boldsymbol{X}_{Dj} = \boldsymbol{X}_{Di}) |\boldsymbol{H}_2|^{-1/2} \phi(\boldsymbol{L}_2(\boldsymbol{S}_j - \boldsymbol{S}_i, \boldsymbol{y} - \boldsymbol{Y}_i, \boldsymbol{X}_j - \boldsymbol{X}_i))$$

where $\boldsymbol{X}_D$ are the discrete covariates, and all remaining covariates $\boldsymbol{X}$ are continuous.

The contribution to the likelihood for the hybrid set can be constructed in an analogous manner, integrating over the missing values of $\boldsymbol{Y}_k$. Denote this vector of missing values $\boldsymbol{y}_{m_k}$. For the probabilities defined thus far for subjects in the nonvalidation set, their counterparts for the hybrid set are

$$\widehat{P}(\underline{\boldsymbol{Y}}_k, \boldsymbol{S}_k | \boldsymbol{X}_k) = \int \widehat{P}_{\boldsymbol{\beta}}(\underline{\boldsymbol{Y}}_k, \boldsymbol{S}_k, \boldsymbol{y}_{m_k} | \boldsymbol{X}_k) d\boldsymbol{y}_{m_k} = \int P_{\boldsymbol{\beta}}(\underline{\boldsymbol{Y}}_k, \boldsymbol{y}_{m_k}, \boldsymbol{X}_k) \widehat{P}(\boldsymbol{S}_k | \underline{\boldsymbol{Y}}_k, \boldsymbol{y}_{m_k}, \boldsymbol{X}_k) d\boldsymbol{y}_{m_k},$$

$$\widehat{P}(\boldsymbol{S}_k | \underline{\boldsymbol{Y}}_k, \boldsymbol{y}_{m_k}, \boldsymbol{X}_k) = \frac{\widehat{P}(\boldsymbol{S}_k, \underline{\boldsymbol{Y}}_k, \boldsymbol{y}_{m_k}, \boldsymbol{X}_k)}{\widehat{P}(\underline{\boldsymbol{Y}}_k, \boldsymbol{y}_{m_k}, \boldsymbol{X}_k)},$$

$$\widehat{P}(\underline{\boldsymbol{Y}}_k, \boldsymbol{y}_{m_k}, \boldsymbol{X}_k) = \frac{1}{n_V} \sum_{i \in n_V} \frac{1}{|\boldsymbol{H}_1|^{1/2}} \phi(\boldsymbol{L}_1(\underline{\boldsymbol{Y}}_k - \boldsymbol{Y}_{T_k i}, \boldsymbol{y}_{m_k} - \boldsymbol{Y}_{m_k i}, \boldsymbol{X}_k - \boldsymbol{X}_i)),$$

and

$$\widehat{P}(\boldsymbol{S}_k, \underline{\boldsymbol{Y}}_k, \boldsymbol{y}_{m_k}, \boldsymbol{X}_k) = \frac{1}{n_V} \sum_{i \in n_V} \frac{1}{|\boldsymbol{H}_2|^{1/2}} \phi(\boldsymbol{L}_2(\boldsymbol{S}_k - \boldsymbol{S}_i, \underline{\boldsymbol{Y}}_k - \boldsymbol{Y}_{T_k i}, \boldsymbol{y}_{m_k} - \boldsymbol{Y}_{m_k i}, \boldsymbol{X}_k - \boldsymbol{X}_i)),$$

where $\boldsymbol{Y}_{T_k i}$ denotes the vector of $\boldsymbol{Y}_i$ values recorded at the same timepoints as $\underline{\boldsymbol{Y}}_k$, and $Y_{m_k i}$ denotes the counterpart Y_i values recorded at the timepoints missed by subject k. For ease of notation, we use $\boldsymbol{H}_2$ and $\boldsymbol{H}_1$ to represent any possible bandwidth matrices for the numerator and denominator, respectively, of either $\widehat{P}(\boldsymbol{S}_k | \underline{\boldsymbol{Y}}_k, \boldsymbol{y}_{m_k}, \boldsymbol{X}_k)$ or $\widehat{P}(\boldsymbol{S}_j | \boldsymbol{y}, \boldsymbol{X}_j)$, since the bandwidth can be chosen arbitrarily.

Using $\widehat{P}(\underline{\boldsymbol{Y}}_k, \boldsymbol{y}_{m_k}, \boldsymbol{X}_k)$ and $\widehat{P}(\boldsymbol{S}_k, \underline{\boldsymbol{Y}}_k, \boldsymbol{y}_{m_k}, \boldsymbol{X}_k)$ to compute $\widehat{P}_{\boldsymbol{\beta}}(\underline{\boldsymbol{Y}}_k, \boldsymbol{S}_k | \boldsymbol{X}_k)$, as well as $\widehat{P}(\boldsymbol{y}, \boldsymbol{X}_j)$ and $\widehat{P}(\boldsymbol{S}_j, \boldsymbol{y}, \boldsymbol{X}_j)$ to compute $\widehat{P}_{\boldsymbol{\beta}}(\boldsymbol{S}_j | \boldsymbol{X}_j)$, the estimated likelihood in (3.2) is fully determined. Let the subscript VS denote estimation using a validation subsample, and set $\mathscr{L}_{VS}(\boldsymbol{\beta})$ equal to (3.2). Then the parameter estimates $\widehat{\boldsymbol{\Theta}}_{VS}$ are the solution to the corresponding estimated score equations, obtained by taking the first derivative of the log of $\mathscr{L}_{VS}(\boldsymbol{\beta})$. Writing $\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{Y}|\boldsymbol{X}) = \frac{\partial}{\partial \boldsymbol{\beta}} P_{\boldsymbol{\beta}}(\boldsymbol{Y}|\boldsymbol{X})$, $\widehat{\boldsymbol{D}}_{\boldsymbol{\beta}}(\underline{\boldsymbol{Y}}_k, \boldsymbol{S}_k|\boldsymbol{X}_k) = \int \frac{\partial}{\partial \boldsymbol{\beta}} P_{\boldsymbol{\beta}}(\boldsymbol{y}_{m_k}, \underline{\boldsymbol{Y}}_k|\boldsymbol{X}_k)\widehat{P}(\boldsymbol{S}_k|\boldsymbol{y}_{m_k}, \underline{\boldsymbol{Y}}_k, \boldsymbol{X}_k)d\boldsymbol{y}_{m_k}$, and $\widehat{\boldsymbol{D}}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j) = \int \frac{\partial}{\partial \boldsymbol{\beta}} P_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X}_j)\widehat{P}(\boldsymbol{S}_j|\boldsymbol{X}_j)d\boldsymbol{y}$, these equations are

$$\boldsymbol{U}_{VS}(\boldsymbol{\beta}) = \sum_{i \in n_V} \frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}{P_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)} \sum_{k \in n_{\underline{V}}} \frac{\widehat{\boldsymbol{D}}_{\boldsymbol{\beta}}(\underline{\boldsymbol{Y}}_k, \boldsymbol{S}_k|\boldsymbol{X}_k)}{\widehat{P}_{\boldsymbol{\beta}}(\underline{\boldsymbol{Y}}_k, \boldsymbol{S}_k|\boldsymbol{X}_k)} \sum_{j \in n_{\overline{V}}} \frac{\widehat{\boldsymbol{D}}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)}{\widehat{P}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)}.$$

$\widehat{\boldsymbol{\beta}}_{VS}$ is the vector of estimators for the regression coefficients.

3.2.2. Estimation Using Both Hybrid and Validation Subsample

The proposed method as presented in Section 3.2.1 provides an appealing alternative to CCA when the validation set is not small (see Section 3.3 for simulation results). However, it is not always practical to assume that a substantial proportion of subjects in a study will have no missing data. A large number of subjects missing a small number of visits, which results in a relatively small n_V and larger $n_{\underline{V}}$, is a common scenario encountered in practice. In such a situation, the internal validation subsample of Section 3.2.1 would neglect to use most of the observed data. Therefore, we propose an improvement on this method which takes advantage of all available data to further improve efficiency of the regression estimates.

For timepoints $t = 1, ..., T$, we compute probabilities by timepoint using the expressions

$$\widehat{P}(S'_{jt}, y_t, X'_{jt}) = \frac{1}{n_{V_t}} \sum_{\ell \in V_t} |\boldsymbol{H}_2|^{-1/2} \phi(\boldsymbol{L}_2(S'_{jt} - S'_{\ell t}, y_t - Y'_{\ell t}, X'_{jt} - X'_{\ell t}))$$

and

$$\widehat{P}(y_t, X'_{jt}) = \frac{1}{n_{V_t}} \sum_{\ell \in V_t} |\boldsymbol{H}_1|^{-1/2} \phi(\boldsymbol{L}_1(y_t - Y'_{\ell t}, X'_{jt} - X'_{\ell t})),$$

where V_t is the set of all subjects from n_V and $n_{\underline{V}}$ having a true outcome observed at timepoint t, and n_{V_t} is the size of V_t. For any subject, $\boldsymbol{S}', \boldsymbol{Y}'$, and $\boldsymbol{X}'$ are the original variables transformed to induce independence across timepoints, while preserving the original correlations among $\boldsymbol{S}, \boldsymbol{Y}$, and $\boldsymbol{X}$ within the same timepoint. This can be accomplished in the following steps. We first impose complete independence of $\boldsymbol{S}', \boldsymbol{Y}'$, and $\boldsymbol{X}'$ by computing the global variance-covariance matrix, $\boldsymbol{G}$, for these variables, and multiplying by the Cholesky decomposition of its inverse. Second, we impose the desired covariance structure by specifying a block matrix, $\boldsymbol{B}$, where the blocks consist of the variance-covariance structure within each timepoint, with zeros filled in between timepoints. To summarize, for $\boldsymbol{L}_G^T\boldsymbol{L}_G = \boldsymbol{G}^{-1}$ and $\boldsymbol{L}_B^T\boldsymbol{L}_B = \boldsymbol{B}^{-1}$, we define an overall matrix $\boldsymbol{\Sigma}^{'} = \boldsymbol{L}_B^{-1}\boldsymbol{L}_G$, and

$$(\boldsymbol{S}^{'}, \boldsymbol{Y}^{'}, \boldsymbol{X}^{'}) = \boldsymbol{\Sigma}^{'}(\boldsymbol{S}, \boldsymbol{Y}, \boldsymbol{X}).$$

Because we have some missing observations in $\boldsymbol{Y}$, computation of a variance-covariance matrix using all subjects is not possible. Therefore, taking advantage of the property that V is a representative subsample of the population, we use only data from V to compute $\boldsymbol{\Sigma}^{'}$. The estimates for the variances and covariances contained in $\boldsymbol{\Sigma}^{'}$ are consistent for the true variances and covariances in the population. After imposing independence across timepoints, we can now multiply the estimated probabilities for each timepoint together to obtain a valid estimate of the overall probability of observing the values of subject j's auxiliary outcome and covariates. This estimate is

$$\widehat{P'}(\boldsymbol{S}_j|\boldsymbol{y}, \boldsymbol{X}_j) = \frac{\widehat{P'}(\boldsymbol{S}_j, \boldsymbol{y}, \boldsymbol{X}_j)}{\widehat{P'}(\boldsymbol{y}, \boldsymbol{X}_j)},$$

where the components are computed as

$$\widehat{P'}(\boldsymbol{y}, \boldsymbol{X}_j) = \prod_{t=1}^{T} \widehat{P}(y_t, X_{jt}^{'})$$

and

$$\widehat{P'}(\boldsymbol{S}_j, \boldsymbol{y}, \boldsymbol{X}_j) = \prod_{t=1}^{T} \widehat{P}(S_{jt}^{'}, y_t, X_{jt}^{'}).$$

since y represents the variable over which we are integrating, there is no need to write $y^{'}$, as the domain of integration is infinite. Finally, we can define subject j's contribution to a new likelihood using this new notation. This contribution is

$$\widehat{P'}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j) = \int P_{\boldsymbol{\beta}}(\boldsymbol{y}, \boldsymbol{X}_j)\widehat{P'}(\boldsymbol{S}_j, \boldsymbol{y}, \boldsymbol{X}_j)d\boldsymbol{y}.$$

Just as in Section 3.2.1, the construction of $\widehat{P'}_{\boldsymbol{\beta}}(\underline{\boldsymbol{Y}}_k, \boldsymbol{S}_k|\boldsymbol{X}_k)$ for the hybrid set is analogous to the construction of $\widehat{P'}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)$ described here for the nonvalidation set. An important distinction in the calculation of $\widehat{P'}_{\boldsymbol{\beta}}(\underline{\boldsymbol{Y}}_k, \boldsymbol{S}_k|\boldsymbol{X}_k)$ is that subject k should be left out of the kernel density estimation.

If this is the case, then n_{V_t} in the denominator is replaced with $n_{V_t} - 1$.

For subject k,

$$\widehat{P}(Y_{kt}^{'}, \boldsymbol{X}_{kt}^{\boldsymbol{'}}) = \frac{1}{n_{V_t}} \sum_{\substack{\ell \in V_t \\ \ell \neq k}} |\boldsymbol{H}_1|^{-1/2} \phi(\boldsymbol{L}_1(Y_{kt}^{'} - Y_{\ell t}^{'}, \boldsymbol{X}_k^{\boldsymbol{'}} - \boldsymbol{X}_\ell^{\boldsymbol{'}})),$$

and

$$\widehat{P}(S_{kt}^{'}, Y_{kt}^{'}, \boldsymbol{X}_{kt}^{\boldsymbol{'}}) = \frac{1}{n_t} \sum_{\substack{\ell \in V_t \\ \ell \neq k}} |\boldsymbol{H}_2|^{-1/2} \phi(\boldsymbol{L}_2(S_{kt}^{'} - S_{\ell t}^{'}, Y_{kt}^{'} - Y_{\ell t}^{'}, \boldsymbol{X}_k^{\boldsymbol{'}} - \boldsymbol{X}_\ell^{\boldsymbol{'}})),$$

replacing $Y_{kt}^{'}$ with y_t if subject k is missing Y_{kt}. Furthermore,

$$\widehat{P^{\boldsymbol{'}}}(\boldsymbol{S}_k | \underline{\boldsymbol{Y}}_k, y_{m_k}, \boldsymbol{X}_k) = \frac{\prod_{t=1}^{T} \widehat{P}(S_{kt}^{'}, Y_{kt}^{'}, \boldsymbol{X}_{kt}^{\boldsymbol{'}})}{\prod_{t=1}^{T} \widehat{P}(Y_{kt}^{'}, \boldsymbol{X}_{kt}^{\boldsymbol{'}})},$$

making the same replacement at each t if Y_{kt} is missing. Finally,

$$\widehat{P^{\boldsymbol{'}}}_{\boldsymbol{\beta}}(\boldsymbol{S}_k, \underline{\boldsymbol{Y}}_k | \boldsymbol{X}_k) = \int P_{\boldsymbol{\beta}}(\underline{\boldsymbol{Y}}_k, \boldsymbol{y}_{m_k} | \boldsymbol{X}_k) \widehat{P^{\boldsymbol{'}}}(\boldsymbol{S}_k | \underline{\boldsymbol{Y}}_k, \boldsymbol{y}_{m_k}, \boldsymbol{X}_k) d\boldsymbol{y}_{m_k}.$$

Having completed the new construction of the estimated portions of the likelihood, we can define a new likelihood with the following notation.

$$\mathscr{L}_{HS}(\boldsymbol{\beta}) = \prod_{i \in n_V} P_{\boldsymbol{\beta}}(\mathbf{Y}_i | \boldsymbol{X}_i) \prod_{k \in n_{\underline{V}}} \widehat{P^{\boldsymbol{'}}}_{\boldsymbol{\beta}}(\underline{\boldsymbol{Y}}_k, \boldsymbol{S}_k | \boldsymbol{X}_k) \prod_{j \in n_{\overline{V}}} \widehat{P^{\boldsymbol{'}}}_{\boldsymbol{\beta}}(\boldsymbol{S}_j | \boldsymbol{X}_j) \tag{3.3}$$

The parameter estimates $\widehat{\boldsymbol{\Theta}}_{HS}$, where HS denotes estimation using both a hybrid and validation subsample, are the solution to the corresponding estimated score equations, which are

$$\boldsymbol{U}_{HS}(\boldsymbol{\beta}) = \sum_{i \in n_V} \frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\mathbf{Y}_i | \boldsymbol{X}_i)}{P_{\boldsymbol{\beta}}(\mathbf{Y}_i | \boldsymbol{X}_i)} \sum_{k \in n_{\underline{V}}} \frac{\widehat{\boldsymbol{D}^{\boldsymbol{'}}}_{\boldsymbol{\beta}}(\underline{\boldsymbol{Y}}_k, \boldsymbol{S}_k | \boldsymbol{X}_k)}{\widehat{P^{\boldsymbol{'}}}_{\boldsymbol{\beta}}(\underline{\boldsymbol{Y}}_k, \boldsymbol{S}_k | \boldsymbol{X}_k)} \sum_{j \in n_{\overline{V}}} \frac{\widehat{\boldsymbol{D}^{\boldsymbol{'}}}_{\boldsymbol{\beta}}(\boldsymbol{S}_j | \boldsymbol{X}_j)}{\widehat{P^{\boldsymbol{'}}}_{\boldsymbol{\beta}}(\boldsymbol{S}_j | \boldsymbol{X}_j)}.$$

As in Section 3.2.1, $\widehat{\boldsymbol{\beta}}_{HS}$ is the vector of estimators for the regression coefficients.

It is important to note that the construction of the estimated portions of $\mathscr{L}_{VS}(\boldsymbol{\beta})$ in Section 3.2.1 relies on the assumption that missing data at the subject level is MCAR; that is, membership in the nonvalidaton set is based on study randomization and not on the values of the covariates or outcome. As we explained in Section 3.1, this assumption is reasonable for the PPMI data.

The MCAR assumption for exclusion from the nonvalidation set ensures that the validation set is a representative subsample of the total population, and furthermore, ensures that the estimated probabilities in Section 3.2.1 are valid. However, the estimated probabilities described in this section also use the hybrid set for kernel density estimation. In the hybrid set, missing values may possibly be MCAR, but we recognize that assuming an MCAR mechanism in the hybrid set is very restrictive and not likely to apply in practical settings. However, an MAR (missing at random) assumption is reasonable for patient dropout in many situations, and the traditional linear mixed-effects model also carries this assumption. The simulations in Section 3.3 were generated using an MCAR mechanism for all missing data; however, we evaluated the robustness of $\widehat{\boldsymbol{\beta}}_{HS}$ to MAR data in Tables B.5 and B.6 of Appendix B.1. These tables demonstrate that MAR data in the hybrid set has little to no impact on the bias or efficiency of $\widehat{\boldsymbol{\beta}}_{HS}$. Therefore, the MCAR assumption is necessary only for exclusion from the nonvalidation set; missing visits in the hybrid set can be MAR.

3.2.3. Asymptotic Properties of the Proposed Estimators

Both $\sqrt{N}(\widehat{\boldsymbol{\beta}}_{VS} - \boldsymbol{\beta}_{VS})$ and $\sqrt{N}(\widehat{\boldsymbol{\beta}}_{HS} - \boldsymbol{\beta}_{HS})$ are asymptotically normally distributed as $N \to \infty$ with mean $\mathbf{0}$. We show in Appendix B.2 that, using arguments similar to those in Pepe (1992), the variance of the asymptotic distribution for $\widehat{\boldsymbol{\beta}}_{VS}$ is

$$\mathscr{I}^{-1}(\boldsymbol{\beta}) + \mathscr{I}^{-1}(\boldsymbol{\beta})\frac{(\rho_{\underline{V}} + \rho_{\overline{V}})^2}{\rho_V}\mathscr{K}_{VS}(\boldsymbol{\beta})\mathscr{I}^{-1}(\boldsymbol{\beta})$$

where the information matrix, $\mathscr{I}(\boldsymbol{\beta})$, is

$$\rho_V E\left[\frac{-\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\log P_{\boldsymbol{\beta}}(\boldsymbol{Y}|\boldsymbol{X})\right] + \rho_{\underline{V}} E\left[\frac{-\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\log P_{\boldsymbol{\beta}}(\boldsymbol{Y},\boldsymbol{S}|\boldsymbol{X})\right] + \rho_{\overline{V}} E\left[\frac{-\partial}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\log P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})\right],$$

and ρ_V, $\rho_{\underline{V}}$, and $\rho_{\overline{V}}$ are the proportions of V, $\underline{V}$, and $\overline{V}$ in the population, respectively. $\mathscr{K}_{VS}(\boldsymbol{\beta})$, the expectation of the variance of the estimated quantities given the observed data, is

$$E\left(Var\left[\left(\frac{D_{\boldsymbol{\beta}}(\boldsymbol{Y},\boldsymbol{S}|\boldsymbol{X})}{P_{\boldsymbol{\beta}}(\boldsymbol{Y},\boldsymbol{S}|\boldsymbol{X})} + \frac{D_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}{P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}\right)\middle|\boldsymbol{Y},\boldsymbol{X}\right]\right).$$

Consistent estimates of $\mathscr{I}(\boldsymbol{\beta})$ and $\mathscr{K}_{VS}(\boldsymbol{\beta})$ are given by

$$\widehat{\mathscr{I}}(\boldsymbol{\beta}) = \frac{1}{N}\frac{\partial^2 \log\widehat{\mathscr{L}}_{VS}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\Bigg|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}_{VS}}$$

and

$$\widehat{\mathscr{K}}_{VS}(\boldsymbol{\beta}) = \frac{1}{n_V} \sum_{i \in V} \widehat{\boldsymbol{Q}}_i \widehat{\boldsymbol{Q}}_i^T,$$

where

$$\widehat{\boldsymbol{Q}}_i = \frac{1}{n_{\underline{V}} + n_{\overline{V}}} \bigg\{ \sum_{k \in \underline{V}} \widehat{\boldsymbol{Q}}_{ik} + \sum_{j \in \overline{V}} \widehat{\boldsymbol{Q}}_{ij} \bigg\}.$$

$\boldsymbol{Q}_{ik}$ quantifies how the estimated portion of the score equation for the hybrid set differs from the true portion of the score equation for the hybrid set. It is estimated as

$$\widehat{\boldsymbol{Q}}_{ik} = \int \Bigg[\frac{|\boldsymbol{H}_2|^{-1/2} \phi(\boldsymbol{L}_2(\boldsymbol{S}_k - \boldsymbol{S}_i, \underline{\boldsymbol{Y}}_k - \boldsymbol{Y}_{T_k i}, \boldsymbol{y}_{m_k} - \boldsymbol{Y}_{m_k i}, \boldsymbol{X}_k - \boldsymbol{X}_i))}{\widehat{P}(\boldsymbol{y}_{m_k}, \underline{\boldsymbol{Y}}_k, \boldsymbol{X}_k)} - \frac{|\boldsymbol{H}_1|^{-1/2} \phi(\boldsymbol{L}_1(\underline{\boldsymbol{Y}}_k - \boldsymbol{Y}_{T_k i}, \boldsymbol{y}_{m_k} - \boldsymbol{Y}_{m_k i}, \boldsymbol{X}_k - \boldsymbol{X}_i)) \widehat{P}(\boldsymbol{S}, \boldsymbol{y}_{m_k}, \underline{\boldsymbol{Y}}_k, \boldsymbol{X}_k)}{\widehat{P}^2(\boldsymbol{y}_{m_k}, \underline{\boldsymbol{Y}}_k, \boldsymbol{X}_k)} \Bigg] \times \Bigg[\frac{\widehat{D}_{\boldsymbol{\beta}}(\boldsymbol{y}_{m_k}, \underline{\boldsymbol{Y}}_k | \boldsymbol{X}_k)}{\widehat{P}_{\boldsymbol{\beta}}(\underline{\boldsymbol{Y}}_k, \boldsymbol{S}_k | \boldsymbol{X}_k)} - \frac{\widehat{D}_{\boldsymbol{\beta}}(\underline{\boldsymbol{Y}}_k, \boldsymbol{S}_k | \boldsymbol{X}_k) \widehat{P}_{\boldsymbol{\beta}}(\boldsymbol{y}_{m_k}, \underline{\boldsymbol{Y}}_k | \boldsymbol{X}_k)}{\widehat{P}^2_{\boldsymbol{\beta}}(\underline{\boldsymbol{Y}}_k, \boldsymbol{S}_k | \boldsymbol{X}_k)} \Bigg] d\boldsymbol{y}_{m_k}.$$

Likewise, $\boldsymbol{Q}_{ij}$ captures how the estimated portion of the score equation for $\overline{V}$ differs from the true portion of the score equation for $\overline{V}$. Its consistent estimate is

$$\widehat{\boldsymbol{Q}}_{ij} = \int \Bigg[\frac{|\boldsymbol{H}_2|^{-1/2} \phi(\boldsymbol{L}_2(\boldsymbol{S}_j - \boldsymbol{S}_i, \boldsymbol{y} - \boldsymbol{Y}_i, \boldsymbol{X}_j - \boldsymbol{X}_j))}{\widehat{P}(\boldsymbol{y}, \boldsymbol{X}_j)} - \frac{|\boldsymbol{H}_1|^{-1/2} \phi(\boldsymbol{L}_1(\boldsymbol{y} - \boldsymbol{Y}_i, \boldsymbol{X}_j - \boldsymbol{X}_i)) \widehat{P}(\boldsymbol{S}, \boldsymbol{y}, \boldsymbol{X}_j)}{\widehat{P}^2(\boldsymbol{y}, \boldsymbol{X}_j)} \Bigg] \times \Bigg[\frac{\widehat{D}_{\boldsymbol{\beta}}(\boldsymbol{y}, | \boldsymbol{X}_j)}{\widehat{P}_{\boldsymbol{\beta}}(\boldsymbol{S}_j | \boldsymbol{X}_j)} - \frac{\widehat{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_j | \boldsymbol{X}_j) \widehat{P}_{\boldsymbol{\beta}}(\boldsymbol{y} | \boldsymbol{X}_j)}{\widehat{P}^2_{\boldsymbol{\beta}}(\boldsymbol{S}_j | \boldsymbol{X}_j)} \Bigg] d\boldsymbol{y}$$

Complete formulas and derivations for $\mathscr{K}_{VS}(\boldsymbol{\beta})$, $\boldsymbol{Q}_{ik}$, and $\boldsymbol{Q}_{ij}$ are given in Appendix B.2.

For the estimator $\widehat{\boldsymbol{\beta}}_{HS}$, the asymptotic variance is derived analogously. It is

$$\mathscr{I}^{-1}(\boldsymbol{\beta}) + \mathscr{I}^{-1}(\boldsymbol{\beta}) \frac{(\rho_{\underline{V}} + \rho_{\overline{V}})^2}{\rho_1} \mathscr{K}_{HS}(\boldsymbol{\beta}) \mathscr{I}^{-1}(\boldsymbol{\beta})$$

where $\rho_{V_1} = n_{V_t}/N$ for t=1, and the formulas for $\mathscr{I}(\boldsymbol{\beta})$ and $\mathscr{K}_{HS}(\boldsymbol{\beta})$ are similar to those for $\widehat{\boldsymbol{\beta}}_{VS}$. $\mathscr{I}(\boldsymbol{\beta})$ is asymptotically equivalent for both estimators, so it can be consistently estimated using either version of the estimated likelihood. In this case,

$$\widehat{\mathscr{I}}(\boldsymbol{\beta}) = \frac{1}{N}\frac{\partial^2 \log \widehat{\mathscr{L}}_{HS}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\Bigg|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}_{HS}}$$

and $\mathscr{K}_{HS}(\boldsymbol{\beta})$ is estimated in the same manner as $\mathscr{K}_{VS}(\boldsymbol{\beta})$.

$$\widehat{\mathscr{K}_{HS}}(\boldsymbol{\beta}) = \frac{1}{n_{V_t}-1}\sum_{\ell\in T_t}\widehat{\boldsymbol{Q'}}_\ell \widehat{\boldsymbol{Q'}}_\ell^T\Bigg|_{t=1},$$

where $n_{V_t}-1$ accounts for the exclusion of each subject k from its own kernel estimation, and

$$\widehat{\boldsymbol{Q'}}_\ell = \frac{1}{n_{\overline{V}}+n_{\underline{V}}}\left\{\sum_{\substack{k\in\underline{V}\\ k\neq\ell}}\widehat{\boldsymbol{Q'}}_{\ell k} + \sum_{j\in\overline{V}}\widehat{\boldsymbol{Q'}}_{\ell j}\right\}.$$

As before, $\widehat{\boldsymbol{Q'}}_{\ell j}$ and $\widehat{\boldsymbol{Q'}}_{\ell k}$ capture the estimation in the score equations.

$$\begin{aligned}\widehat{\boldsymbol{Q'}}_{\ell j} = \int\Bigg[&\frac{|\boldsymbol{H}_2|^{-1/2}\phi(\boldsymbol{L}_2(S'_{j1}-S'_{\ell 1}, y_1 - Y'_{\ell 1}, \boldsymbol{X'}_j - \boldsymbol{X'}_\ell))\prod_{t=2}^T \widehat{P}(S'_{jt}, y_t, \boldsymbol{X'}_{jt})}{\widehat{P'}(\boldsymbol{y},\boldsymbol{X}_j)} - \\ &\frac{|\boldsymbol{H}_1|^{-1/2}\phi(\boldsymbol{L}_1(y_1 - Y'_{\ell 1}, \boldsymbol{X'}_{j1}-\boldsymbol{X'}_{\ell 1}))\widehat{P'}(\boldsymbol{S}_j,\boldsymbol{y},\boldsymbol{X}_j)\prod_{t=2}^T\widehat{P}(y_t,\boldsymbol{X'}_{jt})}{\widehat{P'}^2(\boldsymbol{y},\boldsymbol{X}_j)}\Bigg]\times \\ &\left[\frac{\widehat{\boldsymbol{D'}}_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X}_j)}{\widehat{P'}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)} - \frac{\widehat{\boldsymbol{D'}}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)\widehat{P'}_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X}_j)}{\widehat{P'}^2_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)}\right]d\boldsymbol{y},\end{aligned}$$

and

$$\begin{aligned}\widehat{\boldsymbol{Q'}}_{\ell k} = \int\Bigg[&\frac{|\boldsymbol{H}_2|^{-1/2}\phi(\boldsymbol{L}_2(S'_{k1}-S'_{\ell 1}, Y'_{k1} - Y'_{\ell 1}, \boldsymbol{X'}_k - \boldsymbol{X'}_\ell))\prod_{t=2}^T \widehat{P}(S'_{kt}, Y'_{kt}, \boldsymbol{X'}_{kt})}{\widehat{P'}(\boldsymbol{y}_{m_k},\underline{\boldsymbol{Y}}_k,\boldsymbol{X}_k)} - \\ &\frac{|\boldsymbol{H}_1|^{-1/2}\phi(\boldsymbol{L}_1(Y'_{k1} - Y'_{\ell 1}, \boldsymbol{X'}_{k1}-\boldsymbol{X'}_{\ell 1}))\widehat{P'}(\boldsymbol{S}_k,\boldsymbol{y}_{m_k},\underline{\boldsymbol{Y}}_k,\boldsymbol{X}_k)\prod_{t=2}^T\widehat{P}(Y'_{kt},\boldsymbol{X'}_{kt})}{\widehat{P'}^2(\boldsymbol{y}_{m_k},\underline{\boldsymbol{Y}}_{kt},\boldsymbol{X}_k)}\Bigg]\times \\ &\left[\frac{\widehat{\boldsymbol{D'}}_{\boldsymbol{\beta}}(\boldsymbol{y}_{m_k},\underline{\boldsymbol{Y}}_k|\boldsymbol{X}_k)}{\widehat{P'}_{\boldsymbol{\beta}}(\underline{\boldsymbol{Y}}_k,\boldsymbol{S}_k|\boldsymbol{X}_k)} - \frac{\widehat{\boldsymbol{D'}}_{\boldsymbol{\beta}}(\underline{\boldsymbol{Y}}_k,\boldsymbol{S}_k|\boldsymbol{X}_k)\widehat{P'}_{\boldsymbol{\beta}}(\boldsymbol{y}_{m_k},\underline{\boldsymbol{Y}}_k|\boldsymbol{X}_k)}{\widehat{P'}^2_{\boldsymbol{\beta}}(\underline{\boldsymbol{Y}}_k,\boldsymbol{S}_k|\boldsymbol{X}_k)}\right]d\boldsymbol{y}_{m_k}.\end{aligned} \tag{3.4}$$

As before, we interchange $y_{m_k t}$ and Y'_{kt} as appropriate at each t in equation (3.4), dependent on whether Y_{kt} is observed.

3.3. Simulations

We conducted a variety of simulations in order to assess the performance of our proposed estimator. We specified a model with uncorrelated random intercepts and slopes, with γ_1 and γ_2 denoting the variances of the random intercepts and slopes, respectively. The true outcome $\boldsymbol{Y}_\ell$, $\ell = 1, ..., N$, was simulated according to the model given by (3.1), with σ^2=0.9, γ_1=1, and γ_2=0.5. We included one continuous baseline covariate, $X_2 \sim N(3, 0.16)$, and a discrete timepoint covariate, X_1 = 1,..,T. Since the amount of missing data can vary in two ways, with the dropout rate in $\underline{V}$ and the number of subjects in $\overline{V}$, we present a model with two timepoints for clarity of presentation, so that each subject in $\underline{V}$ is missing exactly one timepoint.

Define ρ_V, $\rho_{\underline{V}}$, and $\rho_{\overline{V}}$ as the proportions n_V/N, $n_{\underline{V}}/N$, and $n_{\overline{V}}/N$, respectively. We set N=350 and varied ρ_V between 0.3 and 0.8. Missing data are generated under MCAR. We then determined $\rho_{\underline{V}}$ and $\rho_{\overline{V}}$ by setting $\rho_{\underline{V}} = \rho_{\overline{V}} = \frac{(1-\rho_V)}{2}$. As explained in Section 3.1, we expect more efficiency gains with a stronger correlation between $\boldsymbol{Y}$ and $\boldsymbol{S}$. To illustrate this, we examined the model using a strong correlation, 0.8, and a moderate one, 0.4. In Table 3.1, we display both proposed estimators, $\widehat{\beta}_{VS}$ and $\widehat{\beta}_{HS}$. These are compared to the estimator from CCA, $\widehat{\beta}_{CC}$, as well as the Oracle estimator, $\widehat{\beta}_O$, which uses the full dataset as it would have been available if no data were missing. The relative efficiency (RE) in Tables 3.1 and 3.2 was calculated by dividing the respective estimated standard error by the Oracle estimated standard error. RE = 1.0 indicates optimal efficiency, while larger values of RE indicate efficiency loss. Both $\widehat{\beta}_{CC}$ and $\widehat{\beta}_O$ were computed using the `lme4` package in R.

Results from 500 simulations for $\widehat{\beta}_{HS}$, $\widehat{\beta}_{VS}$, and $\widehat{\beta}_{CC}$ are displayed in Tables 3.1 and 3.2. Table 3.1 illustrates the efficiency gains on estimates of β_1, the regression coefficient for time in the model, at a strong and a weaker correlation. The bias of all three estimators is very small, as expected. When the amount of missing data is small (ρ_V = 0.8), efficiency is close to optimal for $\widehat{\beta}_{HS}$ and $\widehat{\beta}_{VS}$. As the amount of missing data increases, $\widehat{\beta}_{CC}$ loses efficiency quite quickly. The two proposed estimators perform well even in this extreme situation, with $\widehat{\beta}_{HS}$ outperforming $\widehat{\beta}_{VS}$ in all situations. For the moderate correlation of 0.4, the efficiency is slightly less than optimal with a small amount of missing data. However, as the amount of missing data increases, both $\widehat{\beta}_{HS}$ and $\widehat{\beta}_{VS}$ tend to lose efficiency slightly more slowly than $\widehat{\beta}_{CC}$. Furthermore, $\widehat{\beta}_{HS}$ does not outperform $\widehat{\beta}_{VS}$ here, but it also does not do worse.

Table 3.2 demonstrates the results from the same models for β_2, the coefficient for the continuous baseline covariate. Of note on this table is that the percent bias is higher than that of β_1 for all estimators; however, $\widehat{\beta}_{HS}$ and $\widehat{\beta}_{VS}$ are still on par with $\widehat{\beta}_{CC}$ in terms of bias. The performance of $\widehat{\beta}_{HS}$ compared to $\widehat{\beta}_{VS}$ is about the same with β_2 as with β_1, at each respective correlation; that is, $\widehat{\beta}_{HS}$ gains more efficiency than $\widehat{\beta}_{VS}$ at a strong correlation, but the two are about the same at a weaker correlation.

Additional simulations are included in Appendix B.1. We demonstrate efficiency of the proposed estimators under no correlation and perfect correlation between the true and auxiliary outcomes in Tables B.1 and B.2. We evaluate the performance of PMM and MI estimators under misspecification of the imputation model and compare this to the robustness of our proposed estimators in Tables B.3 and B.4. Finally, we evaluate the performance of the proposed estimator $\widehat{\beta}_{HS}$ under MAR in Tables B.5 and B.6.

3.4. Data Example

Our method is motivated by the data characteristics of PPMI (data retrieved January 2, 2018). The true outcome, which is missing for some patients, is brain volume as measured by MRI. We illustrate our method using gray matter (GM) volume from the parietal and frontal lobes, each divided by total intracranial volume (ICV). Patients underwent MRI scans at baseline and at one year and two years post-baseline (T=3). Neuropsychological assessments are an ideal choice for auxiliary outcome, as they do not require specialized equipment to administer, and therefore will be available for most or all patients. We chose to take advantage of the plethora of information in the PPMI data by combining a number of these tests into one composite auxiliary variable. We accomplished this with a linear regression of complete cases at each timepoint, using observed MRI volume as the outcome and neuropsychological test scores as predictors. We then used the regression coefficients to predict MRI volume for each subject at each timepoint. This new composite auxiliary variable has a correlation of 0.52 with the true MRI volume in the parietal lobe, and 0.46 with the true MRI volume in the frontal lobe. The assessments used in this regression are the Questionnaire for Impulsive-Compulsive Disorders in Parkinson's Disease-Rating Scale (QUIP-RS; Weintraub et al., 2012), Symbol Digit Modalities Score (SDM; Barrett et al., 2019), Semantic Fluency Total Score

Table 3.1: Results for timepoint coefficient from 500 simulations. SD = standard deviation; $\widehat{SE}$ = estimated standard error; MSE = mean squared error; RE = relative efficiency. $\widehat{\beta}_{HS}$ = estimator using both hybrid and validation subsample; $\widehat{\beta}_{VS}$ = estimator using an internal validation sample; $\widehat{\beta}_{CC}$ = complete-case estimator.

Cor($\boldsymbol{Y}, \boldsymbol{S}$)	$(\rho_V, \rho_{\underline{V}}, \rho_{\overline{V}})$	Estimator	Bias (% Bias)	SD	$\widehat{SE}$	MSE	RE
0.8	(0.8, 0.1, 0.1)	$\widehat{\beta}_{HS}$	-0.001 (0.117)	0.085	0.086	0.007	1.069
		$\widehat{\beta}_{VS}$	-0.001 (0.111)	0.084	0.007	0.086	1.067
		$\widehat{\beta}_{CC}$	-0.001 (0.138)	0.087	0.090	0.008	1.115
	(0.7, 0.15, 0.15)	$\widehat{\beta}_{HS}$	-0.001 (0.134)	0.088	0.090	0.008	1.109
		$\widehat{\beta}_{VS}$	-0.001 (0.139)	0.088	0.090	0.008	1.110
		$\widehat{\beta}_{CC}$	-0.002 (0.164)	0.095	0.096	0.009	1.189
	(0.6, 0.2, 0.2)	$\widehat{\beta}_{HS}$	-0.003 (0.328)	0.093	0.094	0.009	1.164
		$\widehat{\beta}_{VS}$	-0.003 (0.328)	0.093	0.094	0.009	1.164
		$\widehat{\beta}_{CC}$	-0.004 (0.395)	0.102	0.103	0.010	1.281
	(0.5, 0.25, 0.25)	$\widehat{\beta}_{HS}$	-0.004 (0.417)	0.098	0.097	0.010	1.205
		$\widehat{\beta}_{VS}$	-0.004 (0.358)	0.097	0.100	0.009	1.238
		$\widehat{\beta}_{CC}$	-0.006 (0.590)	0.112	0.113	0.013	1.398
	(0.4, 0.3, 0.3)	$\widehat{\beta}_{HS}$	-0.006 (0.604)	0.108	0.102	0.012	1.267
		$\widehat{\beta}_{VS}$	-0.005 (0.535)	0.109	0.109	0.012	1.347
		$\widehat{\beta}_{CC}$	-0.009 (0.874)	0.125	0.126	0.016	1.562
	(0.3, 0.35, 0.35)	$\widehat{\beta}_{HS}$	-0.007 (0.670)	0.124	0.108	0.015	1.341
		$\widehat{\beta}_{VS}$	-0.006 (0.642)	0.127	0.123	0.016	1.521
		$\widehat{\beta}_{CC}$	-0.006 (0.633)	0.149	0.145	0.022	1.794
0.4	(0.8,0.1,0.1)	$\widehat{\beta}_{HS}$	-0.001 (0.124)	0.087	0.090	0.008	1.112
		$\widehat{\beta}_{VS}$	-0.001 (0.120)	0.087	0.090	0.008	1.111
		$\widehat{\beta}_{CC}$	-0.001 (0.138)	0.087	0.090	0.008	1.115
	(0.7,0.15,0.15)	$\widehat{\beta}_{HS}$	-0.001 (0.148)	0.094	0.096	0.009	1.182
		$\widehat{\beta}_{VS}$	-0.001 (0.141)	0.094	0.095	0.009	1.180
		$\widehat{\beta}_{CC}$	-0.002 (0.164)	0.095	0.096	0.009	1.189
	(0.6,0.2,0.2)	$\widehat{\beta}_{HS}$	-0.004 (0.398)	0.101	0.102	0.010	1.267
		$\widehat{\beta}_{VS}$	-0.004 (0.389)	0.101	0.102	0.010	1.265
		$\widehat{\beta}_{CC}$	-0.004 (0.395)	0.102	0.103	0.010	1.281
	(0.5,0.25,0.25)	$\widehat{\beta}_{HS}$	-0.006 (0.569)	0.110	0.111	0.012	1.373
		$\widehat{\beta}_{VS}$	-0.006 (0.552)	0.110	0.111	0.012	1.372
		$\widehat{\beta}_{CC}$	-0.006 (0.590)	0.112	0.113	0.013	1.398
	(0.4,0.3,0.3)	$\widehat{\beta}_{HS}$	-0.009 (0.880)	0.122	0.122	0.015	1.516
		$\widehat{\beta}_{VS}$	-0.009 (0.865)	0.122	0.123	0.015	1.521
		$\widehat{\beta}_{CC}$	-0.009 (0.874)	0.125	0.126	0.016	1.562
	(0.3,0.35,0.35)	$\widehat{\beta}_{HS}$	-0.007 (0.710)	0.143	0.138	0.020	1.708
		$\widehat{\beta}_{VS}$	-0.008 (0.790)	0.143	0.140	0.021	1.736
		$\widehat{\beta}_{CC}$	-0.006 (0.633)	0.149	0.145	0.022	1.794

Table 3.2: Results for baseline covariate coefficient from 500 simulations. SD = standard deviation; $\widehat{SE}$ = estimated standard error; MSE = mean squared error; RE = relative efficiency. $\widehat{\beta}_{HS}$ = estimator using both hybrid and validation subsample; $\widehat{\beta}_{VS}$ = estimator using an internal validation sample; $\widehat{\beta}_{CC}$ = complete-case estimator.

Cor($\boldsymbol{Y}, \boldsymbol{S}$)	$(\rho_V, \rho_{\underline{V}}, \rho_{\overline{V}})$	Estimator	Bias (% Bias)	SD	$\widehat{\text{SE}}$	MSE	RE
0.8	(0.8, 0.1, 0.1)	$\widehat{\beta}_{HS}$	-0.026 (3.688)	0.209	0.210	0.044	1.031
		$\widehat{\beta}_{VS}$	-0.022 (3.200)	0.209	0.210	0.044	1.033
		$\widehat{\beta}_{CC}$	-0.023 (3.267)	0.221	0.215	0.049	1.055
	(0.7, 0.15, 0.15)	$\widehat{\beta}_{HS}$	-0.031 (4.469)	0.210	0.212	0.045	1.041
		$\widehat{\beta}_{VS}$	-0.026 (3.779)	0.210	0.212	0.045	1.044
		$\widehat{\beta}_{CC}$	-0.027 (3.841)	0.228	0.221	0.053	1.089
	(0.6, 0.2, 0.2)	$\widehat{\beta}_{HS}$	-0.033 (4.762)	0.213	0.214	0.046	1.053
		$\widehat{\beta}_{VS}$	-0.026 (3.770)	0.214	0.215	0.046	1.057
		$\widehat{\beta}_{CC}$	-0.032 (4.515)	0.233	0.229	0.055	1.123
	(0.5, 0.25, 0.25)	$\widehat{\beta}_{HS}$	-0.035 (5.041)	0.218	0.216	0.049	1.062
		$\widehat{\beta}_{VS}$	-0.026 (3.764)	0.221	0.218	0.050	1.070
		$\widehat{\beta}_{CC}$	-0.027 (3.843)	0.245	0.237	0.061	1.163
	(0.4, 0.3, 0.3)	$\widehat{\beta}_{HS}$	-0.039 (5.621)	0.216	0.218	0.048	1.073
		$\widehat{\beta}_{VS}$	-0.029 (4.102)	0.218	0.221	0.048	1.087
		$\widehat{\beta}_{CC}$	-0.033 (4.754)	0.247	0.245	0.062	1.204
	(0.3, 0.35, 0.35)	$\widehat{\beta}_{HS}$	-0.037 (5.226)	0.226	0.221	0.052	1.085
		$\widehat{\beta}_{VS}$	-0.023 (3.219)	0.230	0.226	0.053	1.109
		$\widehat{\beta}_{CC}$	-0.030 (4.231)	0.261	0.255	0.069	1.254
0.4	(0.8, 0.1, 0.1)	$\widehat{\beta}_{HS}$	-0.023 (3.265)	0.219	0.217	0.048	1.065
		$\widehat{\beta}_{VS}$	-0.023 (3.230)	0.219	0.216	0.048	1.064
		$\widehat{\beta}_{CC}$	-0.023 (3.267)	0.221	0.215	0.049	1.055
	(0.7,0.15,0.15)	$\widehat{\beta}_{HS}$	-0.027 (3.838)	0.225	0.223	0.051	1.095
		$\widehat{\beta}_{VS}$	-0.026 (3.785)	0.225	0.222	0.051	1.093
		$\widehat{\beta}_{CC}$	-0.027 (3.841)	0.228	0.221	0.053	1.089
	(0.6,0.2,0.2)	$\widehat{\beta}_{HS}$	-0.031 (4.458)	0.230	0.229	0.054	1.127
		$\widehat{\beta}_{VS}$	-0.031 (4.367)	0.229	0.229	0.053	1.124
		$\widehat{\beta}_{CC}$	-0.032 (4.515)	0.233	0.229	0.055	1.123
	(0.5,0.25,0.25)	$\widehat{\beta}_{HS}$	-0.028 (3.934)	0.241	0.236	0.059	1.161
		$\widehat{\beta}_{VS}$	-0.027 (3.815)	0.239	0.235	0.058	1.157
		$\widehat{\beta}_{CC}$	-0.027 (3.843)	0.245	0.237	0.060	1.163
	(0.4,0.3,0.3)	$\widehat{\beta}_{HS}$	-0.033 (4.766)	0.242	0.244	0.060	1.198
		$\widehat{\beta}_{VS}$	-0.032 (4.541)	0.240	0.243	0.059	1.193
		$\widehat{\beta}_{CC}$	-0.033 (4.754)	0.247	0.245	0.062	1.204
	(0.3,0.35,0.35)	$\widehat{\beta}_{HS}$	-0.029 (4.181)	0.254	0.252	0.065	1.239
		$\widehat{\beta}_{VS}$	-0.027 (3.818)	0.253	0.251	0.065	1.232
		$\widehat{\beta}_{CC}$	-0.030 (4.231)	0.261	0.255	0.069	1.254

Table 3.3: Summary statistics of baseline measurements. Continuous variables summarized with mean and standard deviation in parentheses; categorical variables summarized with counts and percentages in parentheses. GM = gray matter. True GM volume is standardized by dividing by intracranial volume for each subject, then multiplied by 1000 for clarity of presentation. Composite GM volume is the auxiliary outcome predicted by clinical assessment scores. MoCA = Montreal Cognitive Assessment; QUIP-RS = Questionnaire for Impulsive-Compulsive Disorders in Parkinson's Disease Rating Scale; SDM = Symbol Digit Modalities Score; SFT = Semantic Fluency Total Score; STAI = State-Trait Anxiety Inventory; UPDRS-Cog = cognitive score of Unified Parkinson's Disease Rating Scale.

Variable	V	$\underline{V}$	$\overline{V}$	Total
N	119	20	221	360
Gender				
Male	77 (64.7%)	15 (75.0%)	146 (66.1%)	238 (66.1%)
Female	42 (35.3%)	5 (25.0%)	75 (33.9%)	122 (33.9%)
GM Volume - Frontal				
True	110.9 (11.4)	111.6 (9.3)	N/A	111.0 (11.1)
Composite	110.9 (3.5)	112.5 (4.4)	110.2 (4.0)	110.5 (3.9)
GM Volume - Parietal				
True	77.3 (8.2)	78.1 (6.9)	N/A	77.4 (8.0)
Composite	77.3 (3.2)	78.5 (3.7)	76.7 (3.7)	77.0 (3.5)
Clinical Assessments				
MoCA	27.4 (2.1)	28.5 (1.7)	26.8 (2.5)	27.0 (2.3)
QUIP-RS	0.27 (0.56)	0.25 (0.55)	0.29 (0.65)	0.28 (0.62)
SDM	41.4 (9.2)	44.1 (11.0)	40.7 (9.3)	41.1 (9.4)
SFT	49.6 (10.5)	45.0 (10.5)	49.0 (12.6)	49.0 (11.8)
STAI	65.9 (17.8)	66.9 (21.8)	65.3 (18.7)	65.6 (18.6)
UPDRS-Cog	0.32 (0.55)	0.25 (0.64)	0.29 (0.51)	0.29 (0.53)

(SFT; Dadgar, Khatoonabadi, and Bakhtiyari, 2013), State-Trait Anxiety Inventory (STAI; Yang et al., 2019), the cognitive score of the Unified Parkinson's Disease Rating Scale (UPDRS; Ibarretxe-Bilbao et al., 2009; Nagano-Saito et al., 2005), and the Montreal Cognitive Assessment (MoCA; Nazem et al., 2009). We used separate models for the parietal and frontal lobes. Following this, we controlled for gender by including it as a covariate in the analysis model. Our scientific aim is to quantify the longitudinal decline in brain volume in PD patients.

After excluding subjects missing the auxiliary outcome or covariates, we had a cohort of N = 360 subjects, with $n_{\overline{V}}$ = 221 patients missing all three MRI measurements ($\rho_{\overline{V}}$ = 0.61). 15 patients were missing one MRI scan, and 5 missing two, leaving $n_{\underline{V}}$ = 20 and $\rho_{\underline{V}}$ = 0.06. The remaining n_V = 119 patients had no missing data ($\rho_V = 0.33$). Demographic characteristics are summarized in Table 3.3.

Table 3.4: $\widehat{\beta}_{HS}, \widehat{\beta}_{VS}$ denote the proposed estimators; $\widehat{\beta}_{CC}$ denotes the complete-case estimator; $\widehat{SE}$ = estimated standard error; Aux. Out. = Auxiliary Outcome; CGMAV = Composite gray matter auxiliary variable; SDM = Symbol Digit Modalities Score.

Lobe	Aux. Out.	Cor($\boldsymbol{S}, \boldsymbol{Y}$)	Predictor	$\widehat{\beta}_{HS}(\widehat{SE})$	$\widehat{\beta}_{VS}(\widehat{SE})$	$\widehat{\beta}_{CC}(\widehat{SE})$
Parietal	SDM	0.41	Visit	-0.754 (0.123)	-0.772 (0.121)	-0.765 (0.156)
			Gender	6.510 (1.143)	6.568 (1.129)	6.754 (1.302)
	CGMAV	0.52	Visit	-0.775 (0.121)	-0.784 (0.121)	-0.765 (0.156)
			Gender	6.616 (1.077)	6.655 (1.081)	6.754 (1.302)
Frontal	SDM	0.36	Visit	-0.611 (0.145)	-0.626 (0.141)	-0.700 (0.188)
			Gender	6.865 (1.607)	7.128 (1.584)	7.127 (1.807)
	CGMAV	0.47	Visit	-0.639 (0.146)	-0.644 (0.147)	-0.700 (0.188)
			Gender	7.042 (1.542)	7.352 (1.532)	7.127 (1.807)

Results from the analysis model are summarize in Table 3.4. In addition to the composite auxiliary variable, we also calculated the proposed estimators using the raw Symbol Digit Modalities Score as the auxiliary variable, since it is the single clinical assessment having the strongest correlation with MRI volume. The regression coefficient for visit represents annual change in GM volume measured by MRI. Of note is that both proposed estimators are more efficient than $\widehat{\beta}_{CC}$, while the values of $\widehat{\beta}_{HS}$ and $\widehat{\beta}_{VS}$ are similar to $\widehat{\beta}_{CC}$, which indicates that all three have similar bias. These results are consistent with the simulations of Section 3.3; for the models where the correlation between true GM volume and auxiliary variable is slightly better than the moderate correlation of 0.4 used in Section 3.3, we see better efficiency gains than estimators using 0.4 correlation in the simulations. Furthermore, we expect that $\widehat{\beta}_{HS}$ and $\widehat{\beta}_{VS}$ will perform similarly to each other when the size of $\underline{V}$ is small, as is the case in this example.

3.5. Discussion

In this article, we presented a novel methodology to capture the relationship between a true outcome and predictors in a longitudinal setting which is consistent and more efficient than the conventional method used by most software packages. We have shown through simulations that by using an auxiliary outcome correlated with the missing outcome, we are able to employ every subject in the dataset, even those with no true outcome recorded. Additionally, by estimating the relationship between auxiliary and true outcomes nonparametrically, we avoid over-specifying or imposing additional restrictions on the model.

We presented two novel estimators in this article; they differ by the manner of nonparametric estimation. Section 3.3 demonstrates that both estimators are unbiased, and that, with respect to efficiency gained, $\widehat{\beta}_{HS}$ is at worst roughly equivalent to and at best superior to $\widehat{\beta}_{VS}$ in all simulated situations. This result is intuitive and expected, because $\widehat{\beta}_{HS}$ utilizes more observed data in its construction.

The methodology described in this article has important implications for clinical trials. While resources for clinical trials are notoriously costly, much of this cost can be mitigated with a conservative study design. Typically, investigators anticipate dropout prior to enrollment, and account for this by enrolling more subjects than the number necessary to achieve desired power. By employing a method such as ours, which enables use of all enrolled subjects regardless of missing status, the number of subjects required for enrollment will be smaller. This will lead to valuable savings in time, money, and other resources, which can translate to better patient care on the trial. Additionally, some outcomes are difficult and expensive to collect on all patients. For example, lumbar punctures to collect cerebrospinal fluid are painful and often avoided by patients if possible. Neuropsychological assessments are an inexpensive, painless procedure that can be performed on the majority of PD patients. By using our proposed method and an auxiliary outcome, investigators can not only save costs, but also reach more patients.

Our estimator carries the attractive property described in Pepe (1992) that it is not less efficient than CCA even when the chosen auxiliary outcome is complete uncorrelated with the true outcome. In the situation where the auxiliary and true outcomes are perfectly correlated, the proposed estimator is theoretically fully efficient as $N \to \infty$. However, as shown in the simulation results in Appendix B.1, in finite samples the estimator falls just short of full efficiency. This is easily explained by the fact that the asymptotic variances of the respective distributions of $\widehat{\beta}_{HS}$ and $\widehat{\beta}_{VS}$ are $\mathscr{O}(\boldsymbol{H})$, shown in Appendix B.2.

There are some limitations of this method. Because the validation set needs to be representative of the population in order to avoid bias, the method is best applied to studies where the MCAR assumption holds. It is reasonable to assume that the missingness mechanism is independent of covariates in the PPMI study, because inclusion in $\overline{V}$ is determined by participating study sites. However, this may not be the case on all clinical trials. Furthermore, we make the assumption that all subjects will have no missing covariates or missing auxiliary outcome. This implies that, if the chosen auxiliary outcome needs to be collected in person and cannot be inferred, patient dropout can negatively affect the utility of the proposed estimators.

As mentioned previously, the asymptotic variance is $\mathcal{O}(\boldsymbol{H})$. This result has two implications for the analysis. First, the convergence rate of the estimators is affected by the dimensionality of the kernel density estimation (Stone, 1980). This dimensionality increases not only with an increasing number of timepoints, but also with inclusion of additional continuous covariates. In practice, this may necessitate larger sample sizes for studies with more timepoints or continuous covariates, particularly if the amount of missing data is large. Secondly, the results of the analysis are sensitive to choice of $\boldsymbol{H}$. Pepe (1992) originally claimed that with cross-sectional data, choice of bandwidth merely affects the rescaling of data, implying that different bandwidths should not result in different estimates. In our simulations, we found that choice of bandwidth does not affect the bias, but does impact the efficiency (results not shown). This contrast between cross-sectional and longitudinal methods suggests that sensitivity to bandwidth increases with an increasing number of timepoints. In practice, we've found that using the optimal bandwidth suggested by Simonoff (1998) achieves ideal efficiency for the proposed estimators in the situation where the true and auxiliary outcomes are perfectly correlated (see Appendix B.1). However, at lower correlations, a higher bandwidth is required to keep the efficiency of the proposed estimator at or above the efficiency of CCA. We've accounted for this anomaly by using an adapted bandwidth, which increases as a function of the correlation.

While our proposed estimator offers an attractive means to improve efficiency over CCA by using all available data, we recognize that implementing the method carries practical considerations. Estimation of the likelihood requires computation of multidimensional integrals with no analytical form. A wide variety of methods are available to handle such integrations, any of which will require more computing time with an increasing number of timepoints. For our analysis, we chose to approximate the integrand using natural cubic splines (Schoenberg, 1946; Birkhoff and Garabedian, 1960), which can easily be extended to multiple dimensions, then compute the analytical integral of the approximated integrand.

CHAPTER 4

`cmprrc` - An R Package for Covariate Measurement Error and Competing Risks in Failure Time Models

4.1. Introduction

Competing risks have been a popular topic of discussion in time-to-event analyses for a number of years (Fine and Gray 1999; Moeschberger, Tordoff, and Kochar 2008; Dignam, Zhang, and Kocherginsky 2013). Because it is inherently impossible to measure all survival times in a competing risks model, there has been some debate about which approach to a competing risks problem is the appropriate one (Austin and Fine, 2017). The cause-specific hazard approach considers subjects to be at risk for the event of interest if they have not been censored or experienced competing events; this approach treats subjects who have experienced competing events the same as censored subjects. The subdistribution hazard approach (Fine and Gray, 1999) considers all subjects who have not experienced the event of interest to be at risk for it. This leads to a counterintuitive risk set, it includes subjects who experienced a competing event (e.g., subjects who died), and are not at risk for the event of interest in reality. However, the subdistribution hazard approach also allows for distinction between censored subjects and subjects who experienced a competing event. This approach also has the attractive quality that as long as the number of subjects experiencing competing events is bounded away from 0 as $N \to \infty$, the size of the risk set will not approach 0 even at the end of the study. This is a useful property for risk set regression calibration (RRC; Xie, Wang, and Prentice 2001), as its estimates may become unstable with small risk set sizes.

There are currently a number of R packages in rotation to analyze data in a competing risks framework. Most notably, since cause-specific hazards are calculated identically as the hazard from a standard Cox model (Cox, 1972), these models can be analyzed using the `survival` package (Therneau, 2017). Subdistribution hazard models can be analyzed using the `cmprsk` package (Gray, 2014), developed by the authors of the methodology it implements. However, as noted in Chapter 2, these packages do not address measurement error, nor do they have the means to accommodate replicate data. In this chapter, we outline the `cmprrc` package, developed to implement the methods of Chapter 2 by addressing competing risks and measurement error in one model. Furthermore, although the methods of Chapter 2 focus on the subdistribution hazard approach, our package offers the option to analyze replicate data using RRC combined with cause-specific hazards, should users feel that a subdistribution hazard approach is not appropriate for the dataset at hand.

4.2. Data

Here we describe the type of data suitable for analysis by this software package. The package can handle either discrete or continuous failure times T; however, all failure times must be observed. Observations without a failure time will be removed from the dataset automatically. As with any competing risks software package, the user will also need to input the failure type, δ, and the event type of interest. The model can account for any number of covariates. Although the primary aim of the package is to account for covariate measurement error, covariates measured without error can be entered into the model without consequence. The user does not need to specify whether each covariate is assumed to be measured with error or not; all covariates are treated as if they are measured with error. If any are truly measured without error, treating them as if they contain measurement error will not affect the regression estimates.

Furthermore, this package handles replicate data. Each covariate that is assumed to be measured with error need to have replicates; covariates measured without error do not, similar to the way baseline covariates need only one recording while time-varying covariates need to be recorded at each timepoint. Subjects do not need to have the same number of replicates as each other or at each timepoint. Additionally, the model can handle some subjects having only one replicate. However, since measurement error is accounted for through replicate data, the number of subjects having only one replicate should be kept to a minimum, if possible. All data should be entered into the function in long format; that is, if subject i has covariates observed at M timepoints and k_i replicates at each timepoint, then the subject will have $M \times k_i$ observations in the dataset.

4.3. Implementation

The primary use of the function `cmprrc` is to maximize the log-partial likelihood

$$\mathscr{L}_{C-RRC}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^M \left\{ \widehat{\boldsymbol{Z}}_i(t)\boldsymbol{\beta} - log\Big(\sum_{j=1}^{n} w_j(t) Y_j(t) exp\Big\{ \widehat{\boldsymbol{Z}}_j(t)\boldsymbol{\beta} \Big\} \Big) \right\} w_i(t) Y_i(t) dN_i(t)$$

where the quantities $w_j(t)$, $Y_j(t)$ for subjects $j = 1, ..., n$ are the weights and at-risk process, respectively, as defined in Chapter 2. $\widehat{\boldsymbol{Z}}_j(t)$ are the RRC estimates for covariate $\boldsymbol{Z}$ for subject j at time t. As noted in Chapter 2, the package converts baseline, time-independent covariates into a time-dependent RRC covariate estimate. The current version of the package does not accommodate time-varying covariates in the model; however, future iterations should introduce this update. The parameter estimates $\widehat{\boldsymbol{\beta}}_{C-RRC}$ are those which maximize this estimated likelihood. The package uses the existing R package `optimx` to maximize the estimated likelihood, with the Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimization method specified.

The following arguments are required for the function `cmprrc`.

Argument	Description
`data`	The dataset containing covariates, failure times, and failure types in long format. Can be a data frame or matrix.
`ftimecol`	An integer indicating which column in the dataset contains the failure times for each subject.
`ftypecol`	An integer indicating which column in the dataset contains the failure types for each subject.
`type1`	An integer indicating which failure type is the event of interest.
`covcols`	A vector of integers denoting which columns in the dataset contain the covariates for each subject.
`subid`	An integer denoting which column contains the subject identifiers.
`cmpmethod`	Method to account for competing risks. Options are "SH" or "FG" for subdistribution hazards or "CS" for cause-specific hazards. Defaults to "SH".
`siglevel`	Significance level. Defaults to 0.05.

The function returns the following values after running.

Value	Description
`coefmat`	Matrix containing one of each of the following columns: parameter estimate, standard error, z-statistic, p-value. One row for each coefficient.
`iterations`	The number of iterations used by optimization.
`convcode`	Convergence code returned from optimization.
`ll`	Log-likelihood value returned from optimization.

4.4. Performance

The runtime of the function `cmprrc` was assessed using the R package `rbenchmark` version 1.0.0 (Kusnierczyk, 2012). We generated data with three covariates, two normally distributed and one binomially distributed. We specified two situations; in the first, only one of the continuous covariates is measured with error; in the second, both continuous covariates are. The measurement error variance σ^2 was set at 0.5. We tested these situations with a variety of sample sizes (200, 1000, and 2000). Results are displayed in Table 4.1. This table displays the average run time from 10 replications for each situation, run on `R 3.5.0`. The run time increases with increasing number of subjects; however, accounting for additional covariates with measurement error does not have a noticeable impact on run time. This implies that datasets with a large number of covariates with replicates will not adversely impact the run time of the analysis.

Table 4.1: Average run time in seconds of function with varying sample sizes. Each model used 3 covariates. The first column displays results from 1 covariate having measurement error (ME); the second displays results from 2 covariates having ME. Each cell is the result of 10 runs.

	Number of Covariates with ME	
N	1	2
200	1.7	1.9
1000	45.8	41.9
2000	246.1	253.2

4.5. Discussion

We introduced a new software package for R to implement the regression methods described in Chapter 2. The package offers a novel method to account for competing risks and covariate measurement error through replicate data. As current competing risks software packages do not handle replicate data or covariate measurement error, this is a new contribution to the repertoire available software tools for statisticians and researchers. Dissemination of this R package will make the method described in Chapter 2 widely available and easy to implement regardless of skill level or expertise.

We recognize that although the subdistribution hazard method to account for competing risks was chosen carefully and thoughtfully, it is not the preferred competing risks method for all users. Therefore, we include an option in `cmprrc` to account for competing risks by the cause-specific hazard method instead. This method handles multiple failure types by treating all competing events as censored observations. However, we do not outline this option in detail in previous sections, because the model is mathematically equivalent to the RRC method of Xie, Wang, and Prentice (2001).

The package leaves room for improvement in some areas. First, we only account for baseline covariates in the model. Future versions of the package will include time-varying covariates. Secondly, measurement error may also be present in the outcome, where some subjects may have uncertain failure times. This is a common scenario when the event of interest is not able to be precisely measured, such as conversion to Alzheimer's disease in Chapter 2. This is a limitation of the statistical method, but practically speaking, users of the package may be affected.

CHAPTER 5

Discussion

Measurement error and missing data present similar but unique challenges to statistical analyses by rendering any number of essential quantities unobservable. Although both phenomena have received some attention in literature for a variety of situations, we identified data properties and circumstances which remain unaddressed by current methodology. In this dissertation, we offered contributions to these respective areas of research by developing novel methods that can be applied to these previously unaccounted-for circumstances.

The first is measurement error in the presence of competing risks for time-to-event analyses. Both measurement error and competing risks require recognition by investigators and choice of appropriate model prior to data analysis; ignoring either one may lead to unexpected or invalid results. Analyzing data from patients with a neurodegenerative disease such as Alzheimer's requires proper framework for both competing risks and measurement error. When a patient exhibits symptoms of early AD, it is often of interest for the patient and the family to know when that patient can expect to transition from mild cognitive impairment (MCI) to AD. However, the population of AD patients is largely elderly, and patients of both normal and MCI status who pass away from comorbidities prior to converting to AD will not have an observable conversion time. Literature has offered a number of useful approaches to account for these risks (Fine and Gray 1999; Moeschberger, Tordoff, and Kochar 2008). Furthermore, the biomarker CSF is prone to measurement error. We also have adequate methods to account for measurement error in the biomarker (Prentice 1982; Xie, Wang, and Prentice 2001). However, the lack of a method to account for both will force a researcher to choose between ignoring measurement error or ignoring competing risks, both of which can lead to biased results.

The proposed method of Chapter 2 addresses this gap in the literature. We showed that our proposed estimator almost completely removed the bias incurred by ignoring measurement error in a competing risks framework, and maintained proper coverage probabilities. This new estimator was applied to ADNI data to demonstrate its utility in neurodegenerative disease research. We also assert that this method will prove useful for the research of any disease with high mortality and error-prone biomarker measurements.

Our method is not without limitations. Clinical diagnosis of AD is known to be imprecise, and sometimes even inaccurate (Beach et al., 2012). This can be viewed as a form of measurement error on the outcome, which is the time to conversion to AD. Since RRC was developed for covariate measurement error, our method will require a more advanced strategy for addressing measurement error in both outcome and covariates in order to address the uncertainty in the time to conversion to AD. Furthermore, our method is only applicable to replicate data, which can be difficult to obtain if resources are limited. Finally, the regression coefficients of the subdistribution hazard method can sometimes be difficult to interpret (Austin and Fine, 2017), and may not be preferred by all investigators. We have addressed this limitation by providing an option in the R package `cmprrc` to use the cause-specific hazards approach instead. When combined with RRC, this approach is mathematically equivalent to using RRC with no competing risks and thus is not acknowledged in Chapter 2.

The second gap in literature addressed in this dissertation is missing outcome data in longitudinal settings. Unlike the situations of Chapter 2, missing longitudinal outcome data has not been overlooked in literature, as many missing data methods are applicable to longitudinal studies (Spratt et al., 2010). However, we recognize shortcomings with many of these methods, such as the need to specify a fully parametric or semi-parametric imputation model. Furthermore, these methods assume a relatively small amount of missing data, which is usually a result only of patient dropout. Our data are missing at all timepoints for many patients, which implies that use of an auxiliary outcome available for all patients would be an ideal approach for modeling this data. By estimating the relationship between auxiliary and true outcomes nonparametrically, we avoid placing any restrictions on the imputation model and develop a method which is more robust than standard multiple imputation. This result is demonstrated in Appendix B.1. Additionally, the simulations of Chapter 3 and Appendix B.1 indicate that our proposed method is more efficient than the conventional CCA in all situations, and approaches optimal efficiency when the correlation between true and auxiliary outcomes is 1. Our proposed method has useful implications for neurodegenerative disease research, where true outcomes are often costly and difficult to obtain, such as CSF samples and MRI scans, but auxiliary outcomes are widely available for all subjects, such as neuropsychological exams.

The limitations of this method are also addressed in Chapter 3. One limitation is that the MCAR assumption is necessary for exclusion from the nonvalidation set. Furthermore, the method suffers from the curse of dimensionality in two ways. First, the estimated likelihood is afflicted by costly calculations, which can be demanding on computing resources. This pitfall occurs regardless of the choice of numerical integration method. With a small number of timepoints, the computing time is reasonable, but quickly becomes impractical as the number of timepoints increases. Planned improvements to this method involve approximating the integrals in the estimated likelihood by replacing each multivariate integral with a product of univariate integrals. This approximation currently requires more extensive testing before it can be implemented in an R package. The second effect of high dimensionality on the method lies in the kernel estimation. It has been shown that the asymptotic rate of convergence slows as the dimension of the kernel increases (Stone, 1980). This result is corroborated in Appendices B.1 and B.2, where we show both theoretically and through simulations that the proposed method comes close to, but cannot reach, optimal efficiency in finite samples due to the bandwidth matrix $\boldsymbol{H}$. This shortcoming can be mitigated with large sample sizes, but investigators should be cautious about the length of a study if resources on enrollment are limited.

APPENDIX A

APPENDICES FOR CHAPTER 2

A.1. Consistency and Definitions

Using Lemma 1 (Xie, Wang, and Prentice, 2001), consistency and asymptotic normality is achieved with the estimated quantities used to derive the RRC estimator. $\widehat{\boldsymbol{\mu}}(t)$, $\widehat{\boldsymbol{\Delta}}$, $\widehat{\boldsymbol{\Sigma}}(t)$, $\widehat{\boldsymbol{\mu}}(t)$, $\widehat{\boldsymbol{\tau}}_j(t)$, and $\widehat{\boldsymbol{\eta}}_j(t)$ each converge uniformly to their respective probability limits $\boldsymbol{\mu}(t)$, $\boldsymbol{\Delta}$, $\boldsymbol{\Sigma}(t)$, $\boldsymbol{\tau}_j(t)$, $\boldsymbol{\eta}_j(t)$, for $j = 1, 2, .., \ell$.

Furthermore, the following proofs will utilize the fact that $\boldsymbol{s}^{(p)}(\boldsymbol{\beta}, t)$, continuous functions of $\boldsymbol{\beta} \in \boldsymbol{\mathscr{B}}$, are bounded on $\boldsymbol{\mathscr{B}} \times [0, M]$ for $p = 0, 1, 2$, again following Lemma 2 in the same reference. $s^{(0)}(\beta, t)$ is bounded away from zero, allowing us to consider $\boldsymbol{s}^{(1)}(\beta, t)/s^{(0)}(\beta, t)$ and $\boldsymbol{s}^{(2)}(\beta, t)/s^{(0)}(\beta, t)$ bounded functions on $\mathscr{B} \times [0, M]$.

Finally, we will take advantage of the fact that $\sqrt{n}(\widehat{\boldsymbol{\eta}}_k(t) - \boldsymbol{\eta}_k(t))$ and $\sqrt{n}(\widehat{\boldsymbol{\tau}}_k(t) - \boldsymbol{\tau}_k(t))$ converge in distribution to normal random variables, $\boldsymbol{C}_{i(k)}^{k_i}(t)$ and $\boldsymbol{D}_{i(k)}^{k_i}(t)$, respectively, with mean $\mathbf{0}$ and finite variance. Define the notation:

$$
\begin{aligned}
\sqrt{n}(\widehat{\boldsymbol{\eta}}_k(t) - \boldsymbol{\eta}_k(t)) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{C}_{i(k)}^{k_i}(t) + o_p(1) \\
\sqrt{n}(\widehat{\boldsymbol{\tau}}_k(t) - \boldsymbol{\tau}_k(t)) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{D}_{i(k)}^{k_i}(t) + o_p(1)
\end{aligned}
$$

The above lemmas will be used to establish convergence in probability of $\widehat{\boldsymbol{S}}^{(p)}(\boldsymbol{\beta}, t)$ for $p = 0, 1, 2$. Denote $Q_k = \{i : k_i = k,$ for $i = 1, ..., n\}$ and $q_k = \sum_{i=1}^{n} I(k_i = k)$. Further define a_k as the probability limit of a subject having k replicates, estimated as q_k/n. Finally, define $\boldsymbol{Z}_k(t)$ as $\widehat{\boldsymbol{Z}}_i(t)$ with its estimated components replaced with their probability limits $\boldsymbol{\eta}_k(t)$ and $\boldsymbol{\tau}_k(t)$. According to Fine and Gray (1999),

$$
\frac{1}{n} \sum_{i=1}^{n} Y_i(t) w_i(t) \boldsymbol{Z}_i(t)^{\otimes p} exp\{\boldsymbol{\beta}' \boldsymbol{Z}_i(t)\} \xrightarrow{p} G(t) \boldsymbol{s}^{(p)}(\boldsymbol{\beta}, t) \text{ for } p = 0, 1, 2
$$

Therefore, using the continuous mapping theorem and Lemma 1,

$$
\begin{aligned}
\widehat{\boldsymbol{S}}^{(1)}(\beta,t) &= \frac{1}{n}\sum_{i=1}^{n} Y_i(t)w_i(t)\widehat{\boldsymbol{Z}}_i(t)exp\big\{\boldsymbol{\beta}'\widehat{\boldsymbol{Z}}_i(t)\big\} \\
&= \frac{1}{n}\sum_{k=1}^{\ell}\sum_{i\in Q_k} Y_i(t)w_i(t)\big[\widehat{\boldsymbol{\eta}}_k(t) + \widehat{\boldsymbol{\tau}}_k(t)\overline{\boldsymbol{W}}_i\big]exp\big\{\boldsymbol{\beta}'\widehat{\boldsymbol{Z}}_i(t)\big\} \\
&= \frac{1}{n}\sum_{k=1}^{\ell}\sum_{i\in Q_k} Y_i(t)w_i(t)\big[\boldsymbol{\eta}_k(t) + (\widehat{\boldsymbol{\eta}}_k(t) - \boldsymbol{\eta}_k(t)) + \widehat{\boldsymbol{\tau}}_k(t)\overline{\boldsymbol{W}}_i\big]exp\big\{\boldsymbol{\beta}'\widehat{\boldsymbol{Z}}_i(t)\big\} \\
&= \frac{1}{n}\sum_{k=1}^{\ell}\sum_{i\in Q_k} Y_i(t)w_i(t)\big[\boldsymbol{\eta}_k(t) + \widehat{\boldsymbol{\tau}}_k(t)\overline{\boldsymbol{W}}_i\big]exp\big\{\boldsymbol{\beta}'\widehat{\boldsymbol{Z}}_i(t)\big\} + o_p(1) \\
&= \frac{1}{n}\sum_{k=1}^{\ell}\sum_{i\in Q_k} Y_i(t)w_i(t)\big[\boldsymbol{\eta}_k(t) + \big(\boldsymbol{\tau}_k(t) + (\widehat{\boldsymbol{\tau}}_k(t) - \boldsymbol{\tau}_k(t))\big)\overline{\boldsymbol{W}}_i\big]exp\big\{\boldsymbol{\beta}'\widehat{\boldsymbol{Z}}_i(t)\big\} + o_p(1) \\
&= \frac{1}{n}\sum_{k=1}^{\ell}\sum_{i\in Q_k} Y_i(t)w_i(t)\big[\boldsymbol{\eta}_k(t) + \boldsymbol{\tau}_k(t)\overline{\boldsymbol{W}}_i\big]exp\big\{\boldsymbol{\beta}'\widehat{\boldsymbol{Z}}_i(t)\big\} + o_p(1) \\
&= \frac{1}{n}\sum_{k=1}^{\ell}\sum_{i\in Q_k} Y_i(t)w_i(t)\boldsymbol{Z}_k(t)exp\big\{\boldsymbol{\beta}'\big[\boldsymbol{\eta}_k(t) + (\widehat{\boldsymbol{\eta}}_k(t) - \boldsymbol{\eta}_k(t)) + \widehat{\boldsymbol{\tau}}_k(t)\overline{\boldsymbol{W}}_i\big]\big\} + o_p(1) \\
&= \frac{1}{n}\sum_{k=1}^{\ell}\sum_{i\in Q_k} Y_i(t)w_i(t)Z_k(t)exp\big\{\boldsymbol{\beta}'\big[\boldsymbol{\eta}_k(t) + \widehat{\boldsymbol{\tau}}_k(t)\overline{\boldsymbol{W}}_i\big]\big\} + o_p(1) \\
&= \frac{1}{n}\sum_{k=1}^{\ell}\sum_{i\in Q_k} Y_i(t)w_i(t)\boldsymbol{Z}_k(t)exp\big\{\boldsymbol{\beta}'\big[\boldsymbol{\eta}_k(t) + \big(\boldsymbol{\tau}_k(t) + (\widehat{\boldsymbol{\tau}}_k(t) - \boldsymbol{\tau}_k(t)\big)\overline{\boldsymbol{W}}_i\big]\big\} + o_p(1) \\
&= \frac{1}{n}\sum_{k=1}^{\ell}\sum_{i\in Q_k} Y_i(t)w_i(t)\boldsymbol{Z}_k(t)exp\big\{\boldsymbol{\beta}'\big[\boldsymbol{\eta}_k(t) + \boldsymbol{\tau}_k(t)\overline{\boldsymbol{W}}_i\big]\big\} + o_p(1) \\
&= \frac{1}{n}\sum_{k=1}^{\ell}\sum_{i\in Q_k} Y_i(t)w_i(t)\boldsymbol{Z}_k(t)exp\big\{\boldsymbol{\beta}'\boldsymbol{Z}_k(t)\big\} + o_p(1) \\
&\xrightarrow{p} G(t)\boldsymbol{s}^{(1)}(\boldsymbol{\beta},t)
\end{aligned}
$$

Similar results can be established for $\widehat{\boldsymbol{S}}^{(2)}(\boldsymbol{\beta},t) \xrightarrow{p} G(t)\boldsymbol{s}^{(2)}(\boldsymbol{\beta},t)$ and $\widehat{S}^{(0)}(\boldsymbol{\beta},t) \xrightarrow{p} G(t)s^{(0)}(\boldsymbol{\beta},t)$.

A.2. Martingale Form of Censoring Weights

Assume that time is continuous, and let $\tilde{w}_i(t) = r_i(t)G(t)/G(X_i \wedge t)$, which is the true value of the censoring weight for subject i. According to Gill (1980), the Kaplan-Meier estimator $\widehat{G}(t)$ can be written as a martingale in the following way:

$$
\frac{\widehat{G}(t)}{G(t)} - 1 = -\int_0^t \frac{\widehat{G}(u-)}{G(u)Y_n(u)}dM_n^c(u) \qquad \text{(A.1)}
$$

where:

$G(t)$ is the survival function associated with the censoring process

$\widehat{G}(t)$ is the corresponding Kaplan-Meier estimator

$M_n^c(u) = \sum_{j=1}^{n} M_j^c(u)$ is the martingale for the censoring process

$$M_j^c(u) = I(X_j \le u, C_j < T_j) - \int_0^u I(X_j \ge s) d\Lambda^c(s)$$

$\Lambda^c(t)$ is the hazard function for the censoring process

$$Y_n(u) = \sum_{j=1}^{n} I(X_j \ge u)$$

Manipulating equation A.1,

$$\begin{aligned}\frac{\widehat{G}(X_i \wedge t)}{G(X_i \wedge t)} &= -\int_0^{X_i \wedge t} \frac{\widehat{G}(u-)}{G(u)Y_n(u)} dM_n^c(u) \\ &= -I(X_i < t)\int_0^{X_i} \frac{\widehat{G}(u-)}{G(u)Y_n(u)} dM_n^c(u) - I(X_i \ge t)\int_0^{t} \frac{\widehat{G}(u-)}{G(u)Y_n(u)} dM_n^c(u)\end{aligned}$$

$$\begin{aligned}\frac{\widehat{G}(t)}{G(t)} - \frac{\widehat{G}(X_i \wedge t)}{G(X_i \wedge t)} &= -I(X_i < t)\left[\int_0^{t} \frac{\widehat{G}(u-)}{G(u)Y_n(u)} dM_n^c(u) - \int_0^{X_i} \frac{\widehat{G}(u-)}{G(u)Y_n(u)} dM_n^c(u)\right] \\ &\quad + -I(X_i \ge t)\left[\int_0^{t} \frac{\widehat{G}(u-)}{G(u)Y_n(u)} dM_n^c(u) - \int_0^{t} \frac{\widehat{G}(u-)}{G(u)Y_n(u)} dM_n^c(u)\right] \\ &= -I(X_i < t)\int_{X_i}^{t} \frac{\widehat{G}(u-)}{G(u)Y_n(u)} dM_n^c(u)\end{aligned}$$

Multiply by G(t):

$$\widehat{G}(t) - \frac{G(t)\widehat{G}(X_i \wedge t)}{G(X_i \wedge t)} = -G(t)IX_i < t)\int_{X_i}^{t} \frac{\widehat{G}(u-)}{G(u)Y_n(u)} dM_n^c(u)$$

Divide by $\widehat{G}(X_i \wedge t)$:

$$\frac{\widehat{G}(t)}{\widehat{G}(X_i \wedge t)} - \frac{G(t)}{G(X_i \wedge t)} = \frac{-G(t)I(X_i < t)}{\widehat{G}(X_i \wedge t)}\int_{X_i}^{t} \frac{\widehat{G}(u-)}{G(u)Y_n(u)} dM_n^c(u)$$

Therefore, the exact form of $w_i(t) - \tilde{w}_i(t)$ is:

$$r_i(t)\left[\frac{\widehat{G}(t)}{\widehat{G}(X_i \wedge t)} - \frac{G(t)}{G(X_i \wedge t)}\right] = \frac{-r_i(t)G(t)I(X_i < t)}{\widehat{G}(X_i \wedge t)} \int_{X_i}^{t} \frac{\widehat{G}(u-)}{G(u)Y_n(u)} dM_n^c(u) \tag{A.2}$$

(A.2) is the form that will be used in the proofs to follow, because in this form it relies on no asymptotic assumptions. If we assume $n \to \infty$, then we can use the uniform consistency of $\widehat{G}(\cdot) \to G(\cdot)$ established by Fleming and Harrington (1991):

$$\begin{aligned}(w_i(t) - \tilde{w}_i(t)) &= \frac{-G(t)I(X_i < t)r_i(t)}{G(X_i \wedge t)} \int_{X_i}^{t} \frac{G(u-)}{G(u)Y_n(u)} dM_n^c(u) + o_p(1) \\ &= -\tilde{w}_i(t)I(X_i < t) \int_{X_i}^{t} \frac{G(u-)}{G(u)Y_n(u)} dM_n^c(u) + o_p(1)\end{aligned}$$

With continuous time, $G(u) = G(u-)$:

$$\begin{aligned}(w_i(t) - \tilde{w}_i(t)) &= -\tilde{w}_i(t)I(X_i < t) \int_{X_i}^{t} \frac{1}{Y_n(u)} dM_n^c(u) + o_p(1) \\ &= -\tilde{w}_i(t)I(X_i < t) \int_{X_i}^{t} \frac{1}{\pi(u)} \frac{1}{n} \sum_{j=1}^{n} dM_j^c(u) + o_p(1)\end{aligned}$$

where $\pi(u) = lim_{n\to\infty} \frac{1}{n} Y_n(u)$ as defined in Fine and Gray.

A.3. Asymptotic Normality of $\sqrt{n}(\overline{R}_n(t) - \tilde{R}(t))$

Here we will show that $\sqrt{n}(\overline{R}_n(t) - \tilde{R}(t))$ is a normal variate with mean 0 and finite variance. Define:

$$R_i(t) = w_i(t)N_i(t)$$

$$\overline{R}_n(t) = \frac{1}{n}\sum_{i=1}^{n} R_i(t), \text{ the sample mean of } R_i(t)$$

$$\tilde{R}(t) = E[\overline{R}_n(t)], \text{ the asymptotic mean of } \overline{R}_n(t)$$

First, establish that

$$\sqrt{n}(\overline{R}_n(t) - \tilde{R}(t)) = \frac{1}{\sqrt{n}} \sum_{i=1} \left\{ \overline{R}_n(t) - \tilde{R}(t) \right\}$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \left(\tilde{w}_i(t) N_i(t) - \tilde{R}(t) \right) + \left([w_i(t) - \tilde{w}_i(t)] N_i(t) - \tilde{R}(t) \right) \right\} \quad \text{(A.3)}$$

For the first term in (A.3), the central limit theorem applies because the terms are independent and identically distributed, and because $E(\overline{R}_n(t)) = \tilde{R}(t)$. Therefore, the first term is normally distributed with mean zero.

We can apply the martingale form of the Kaplan-Meier estimator (A.2) to the second term in (A.3):

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{w_i(t) - \tilde{w}_i(t)\} N_i(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{-N_i(t) r_i(t) G(t)}{\widehat{G}(X_i)} \int_0^\infty \frac{\widehat{G}(u-) I(t \geq u > X_i)}{G(u) \frac{1}{n} Y_n(u)} \frac{1}{n} \sum_{j=1} dM_j^c(u) \right]$$

Assuming that time is continuous, and using uniform consistency of $\widehat{G}(\cdot) \to G(\cdot)$, the above expression becomes:

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \int_0^M \frac{E\big[N(t)\tilde{w}(t) I(t \geq u > X)\big]}{\pi(u)} dM_j^c(u) + o_p(1)$$
$$= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \int_0^M H(u) dM_j^c(u)$$

We use the standard assumption that $\pi(u) > 0$ to guarantee boundedness of $H(u)$, along with the additional assumption that $\frac{1}{\pi(u)}$ is predictable with respect to the filtration $\mathscr{F}_t$. Therefore, by the martingale central limit theorem (Fleming and Harrington, 1991), the predictable variation process of the above integral converges to a finite value, and $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big([w_i(t) - \tilde{w}_i(t)] N_i(t) \big)$ is an asymptotically normal random variable with finite variance. The same conclusion applies to $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big([w_i(t) - \tilde{w}_i(t)] N_i(t) - \tilde{R}(t) \big)$, and $\sqrt{n}(\overline{R}_n(t) - \tilde{R}(t))$ converges in distribution to a normal variate with mean 0 and finite variance.

A.4. Asymptotic Normality of $\widehat{\boldsymbol{\beta}}_{C-RRC}$

In order to replace $\widehat{\boldsymbol{S}}^{(\cdot)}$ with $\boldsymbol{s}^{(\cdot)}$, we follow the procedure for the robust Cox proportional hazards model estimator laid out in Theorem II of Lin and Wei (1989). Define $\tilde{R}(t)$ and $\overline{R}_n(t)$ as in Appendix A.2. $n^{-1/2} \boldsymbol{U}_{C-RRC}(\boldsymbol{\beta}^*)$ can be rewritten as:

$$
\begin{aligned}
n^{-1/2}\boldsymbol{U}_{C-RRC}(\boldsymbol{\beta}^*) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^M \Big(\widehat{\boldsymbol{Z}}_i(t) - \frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}^*,t)}{\widehat{S}^{(0)}(\boldsymbol{\beta}^*),t)}\Big) w_i(t)dN_i(t) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^M \widehat{\boldsymbol{Z}}_i(t)dR_i(t) - \sqrt{n}\int_0^M \frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}^*,t)}{\widehat{S}^{(0)}(\boldsymbol{\beta}^*,t)} d\overline{R}_n(t) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^M \widehat{\boldsymbol{Z}}_i(t)dR_i(t) - \sqrt{n}\int_0^M \frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}^*,t)}{\widehat{S}^{(0)}(\boldsymbol{\beta}^*,t)} d\overline{R}_n(t) \\
&\pm \sqrt{n}\int_0^M \frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}^*,t)}{\widehat{S}^{(0)}(\boldsymbol{\beta}^*,t)} d\tilde{R}(t) \pm \sqrt{n}\int_0^M \frac{\boldsymbol{\mathscr{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{\mathscr{s}^{(0)}(\boldsymbol{\beta}^*,t)} d\tilde{R}(t) \\
&\pm \sqrt{n}\int_0^M \frac{\boldsymbol{\mathscr{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{\mathscr{s}^{(0)}(\boldsymbol{\beta}^*,t)} d\overline{R}_n(t)
\end{aligned}
$$

$$
\begin{aligned}
n^{-1/2}\boldsymbol{U}_{C-RRC}(\beta^*) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \widehat{\boldsymbol{Z}}_i(t)dR_i(t) - \sqrt{n}\int_0^M \frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}^*,t)}{\widehat{S}^{(0)}(\boldsymbol{\beta}^*,t)} d\tilde{R}(t) \\
&- \sqrt{n}\int_0^M \left[\frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}^*,t)}{\widehat{S}^{(0)}(\boldsymbol{\beta}^*,t)} - \frac{\boldsymbol{\mathscr{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{\mathscr{s}^{(0)}(\boldsymbol{\beta}^*,t)}\right] d\{\overline{R}_n(t) - \tilde{R}(t)\} \\
&- \sqrt{n}\int_0^M \frac{\boldsymbol{\mathscr{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{\mathscr{s}^{(0)}(\boldsymbol{\beta}^*,t)} d[\overline{R}_n(t) - \tilde{R}(t)]
\end{aligned}
$$

The final term is $o_p(1)$ because $\sqrt{n}\{\overline{R}_n(t) - \tilde{R}(t)\}$ is a mean-zero normally distributed random variable (see Appendix A.2) and because $\frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}^*,t)}{\widehat{S}^{(0)}(\boldsymbol{\beta}^*,t)} - \frac{\boldsymbol{\mathscr{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{\mathscr{s}^{(0)}(\boldsymbol{\beta}^*,t)} \xrightarrow{p} 1$.

Further manipulating the second term,

$$
\begin{aligned}
\sqrt{n}\int_0^M \frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}^*,t)}{\widehat{S}^{(0)}(\boldsymbol{\beta}^*,t)} d\tilde{R}(t) &= \sqrt{n}\int_0^M \left[\frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}^*,t)}{\widehat{S}^{(0)}(\boldsymbol{\beta}^*,t)} \pm \frac{\boldsymbol{\mathscr{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{\mathscr{s}^{(0)}(\boldsymbol{\beta}^*,t)}\right] d\tilde{R}(t) \\
&= \sqrt{n}\int_0^M \left[\frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}^*,t)}{\widehat{S}^{(0)}(\boldsymbol{\beta}^*,t)} - \frac{\boldsymbol{\mathscr{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{\mathscr{s}^{(0)}(\boldsymbol{\beta}^*,t)}\right]\left[\frac{\widehat{S}^{(0)}(\boldsymbol{\beta}^*,t)}{G(t)\mathscr{s}^{(0)}(\boldsymbol{\beta}^*,t)}\right] d\tilde{R}(t) \\
&+ \sqrt{n}\int_0^M \frac{\boldsymbol{\mathscr{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{\mathscr{s}^{(0)}(\boldsymbol{\beta}^*,t)}\left[\frac{\mathscr{s}^{(0)}(\boldsymbol{\beta}^*,t)}{\mathscr{s}^{(0)}(\boldsymbol{\beta}^*,t)}\right] d\tilde{R}(t) + o_p(1)
\end{aligned}
$$

Note: $\frac{\widehat{S}^{(0)}(\boldsymbol{\beta}^*,t)}{G(t)s^{(0)}(\boldsymbol{\beta}^*,t)} \xrightarrow{p} 1$, and by Slutsky's Theorem (Casella and Berger, 2002),

$$\sqrt{n}\left(\frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}^*,t)}{\widehat{S}^{(0)}(\boldsymbol{\beta}^*,t)} - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s^{(0)}(\boldsymbol{\beta}^*,t)}\right)\left(\frac{\widehat{S}^{(0)}(\boldsymbol{\beta}^*,t)}{G(t)s^{(0)}(\boldsymbol{\beta}^*,t)}\right) \xrightarrow{d} \sqrt{n}\left(\frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}^*,t)}{\widehat{S}^{(0)}(\boldsymbol{\beta}^*,t)} - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s^{(0)}(\boldsymbol{\beta}^*,t)}\right)$$

Therefore, multiplication by $\frac{\widehat{S}^{(0)}(\boldsymbol{\beta}^*,t)}{G(t)s^{(0)}(\boldsymbol{\beta}^*,t)}$ does not change the asymptotic properties of the term.

$$\begin{aligned}
\sqrt{n}\int_0^M \frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}^*,t)}{\widehat{S}^{(0)}(\boldsymbol{\beta}^*,t)}d\tilde{R}(t) = &\int_0^M \left[\frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}^*,t)}{G(t)s^{(0)}(\boldsymbol{\beta}^*,t)} - \frac{\widehat{S}^{(0)}(\boldsymbol{\beta}^*,t)\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{G(t)s^{(0)}(\boldsymbol{\beta}^*,t)^2}\right.\\
&\left. + \frac{G(t)\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)s^{(0)}(\boldsymbol{\beta}^*,t)}{G(t)s^{(0)}(\boldsymbol{\beta}^*,t)^2}\right]d\tilde{R}(t) + o_p(1)\\
=&\sqrt{n}\int_0^M \frac{1}{G(t)s^{(0)}(\boldsymbol{\beta}^*,t)}\left[\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}^*,t)\right.\\
&\left. - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s^{(0)}(\boldsymbol{\beta}^*,t)}\left\{\widehat{S}^{(0)}(\boldsymbol{\beta}^*,t) - G(t)s^{(0)}(\boldsymbol{\beta}^*,t)\right\}\right]d\tilde{R}(t) + o_p(1) \qquad \text{(A.4)}
\end{aligned}$$

Replacing the original term in the score equation with A.4, we have

$$\begin{aligned}
\frac{1}{\sqrt{n}}\boldsymbol{U}_{C-RRC}(\boldsymbol{\beta}^*,t) =&\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^M \widehat{\boldsymbol{Z}}_i(t)dR_i(t) - \sqrt{n}\int_0^M \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s^{(0)}(\boldsymbol{\beta}^*,t)}d\overline{R}_n(t)\\
&+ \sqrt{n}\int_0^M \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s^{(0)}(\boldsymbol{\beta}^*,t)}d\tilde{R}(t)\\
&- \sqrt{n}\int_0^M \left[\frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}^*,t)}{G(t)s^{(0)}(\boldsymbol{\beta}^*,t)} - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s(0)(\boldsymbol{\beta}^*,t)}\left\{\frac{\widehat{S}^{(0)}(\boldsymbol{\beta}^*,t)}{G(t)s^{(0)}(\boldsymbol{\beta}^*,t)}\right\}\right]d\tilde{R}(t)\\
&- \sqrt{n}\int_0^M \frac{G(t)s^{(0)}(\boldsymbol{\beta}^*,t)}{G(t)s^{(0)}(\boldsymbol{\beta}^*,t)}\left[\frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s(0)(\boldsymbol{\beta}^*,t)}\right]d\tilde{R}(t) + o_p(1)\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^M \widehat{\boldsymbol{Z}}_i(t)dR_i(t) - \sqrt{n}\int_0^M \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s^{(0)}(\boldsymbol{\beta}^*,t)}d\overline{R}_n(t)\\
&- \sqrt{n}\int_0^M \left[\frac{\widehat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}^*,t)}{G(t)s^{(0)}(\boldsymbol{\beta}^*,t)} - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s(0)(\boldsymbol{\beta}^*,t)}\left\{\frac{\widehat{S}^{(0)}(\boldsymbol{\beta}^*,t)}{G(t)s^{(0)}(\boldsymbol{\beta}^*,t)}\right\}\right]d\tilde{R}(t)\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^M \left[\widehat{\boldsymbol{Z}}_i(t) - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s(0)(\boldsymbol{\beta}^*,t)}\right]dR_i(t)\\
&- \frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^M \frac{1}{G(t)s^{(0)}(\boldsymbol{\beta}^*,t)}Y_i(t)w_i(t)exp(\boldsymbol{\beta}^{*'}\widehat{\boldsymbol{Z}}_i(t))\left[\widehat{\boldsymbol{Z}}_i(t)\right.\\
&\left. - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s(0)(\boldsymbol{\beta}^*,t)}\right]d\tilde{R}(t) + o_p(1) \qquad \text{(A.5)}
\end{aligned}$$

The goal is to express (A.5) as a sum of independent, identically distributed random variables. This can be accomplished by taking steps to replace $\widehat{\boldsymbol{Z}}_i(t)$ and $w_i(t)$ with their known, fixed counterparts $\boldsymbol{Z}_i(t)$ and $\tilde{w}_i(t)$. Recall that $\widehat{\boldsymbol{Z}}_i(t) = \widehat{\boldsymbol{\eta}}_{k_i}(t) + \widehat{\boldsymbol{\tau}}_{k_i}(t)\overline{\boldsymbol{W}}_i$. A Taylor expansion of (A.5) around $\widehat{\boldsymbol{\eta}}_{k_i}(t)$ and $\widehat{\boldsymbol{\tau}}_{k_i}(t)$ is now performed. We first break (A.5) into subjects who have the same number of replicates. Following the procedure in Xie, Wang, and Prentice (2001), we expand the following expression:

$$\begin{aligned}\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\boldsymbol{\phi}_i(\boldsymbol{\beta}^*) &= \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M\left[\widehat{\boldsymbol{Z}}_i(t) - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s(0)(\boldsymbol{\beta}^*,t)}\right]dR_i(t)\\ &- \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M\frac{Y_i(t)w_i(t)exp(\boldsymbol{\beta}^{*'}\widehat{\boldsymbol{Z}}_i(t))}{G(t)s^{(0)}(\boldsymbol{\beta}^*,t)}\left[\widehat{\boldsymbol{Z}}_i(t) - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s(0)(\boldsymbol{\beta}^*,t)}\right]d\tilde{R}(t)\\ &= \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\boldsymbol{\phi}_{1i}(\boldsymbol{\beta}^*) - \boldsymbol{\phi}_{2i}(\boldsymbol{\beta}^*)\end{aligned}$$

Let $\boldsymbol{\eta}_k(t)$ and $\boldsymbol{\tau}_k(t)$ (dropping subscript i) denote the value of $\boldsymbol{\eta}(t)$ and $\boldsymbol{\tau}(t)$ common to all subjects in Q_k. Furthermore, let $\boldsymbol{Z}_{i(k)}(t)$ be the known covariate value for subject i with k replicates.

$$\begin{aligned}\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\boldsymbol{\phi}_{1i}(\boldsymbol{\beta}^*) =& \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M\left[\widehat{\boldsymbol{\eta}}_k(t) + \widehat{\boldsymbol{\tau}}_k(t)\overline{\boldsymbol{W}}_i - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s(0)(\boldsymbol{\beta}^*,t)}\right]dR_i(t)\\ =& \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\left[\widehat{\boldsymbol{\eta}}_k(t) - \boldsymbol{\eta}_k(t) + (\widehat{\boldsymbol{\tau}}_k(t) + \boldsymbol{\tau}_k(t))\overline{\boldsymbol{W}}_i + \boldsymbol{Z}_{i(k)}(t) - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s(0)(\boldsymbol{\beta}^*,t)}\right]dR_i(t)\\ =& \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M(\widehat{\boldsymbol{\eta}}_k(t) - \boldsymbol{\eta}_k(t))dR_i(t)\\ &+ \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M(\widehat{\boldsymbol{\tau}}_k(t) - \boldsymbol{\tau}_k(t))\overline{\boldsymbol{W}}_i dR_i(t)\\ &+ \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M\left[\boldsymbol{Z}_{i(k)}(t) - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s(0)(\boldsymbol{\beta}^*,t)}\right]dR_i(t)\\ =& \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\boldsymbol{S}_{1i}(\boldsymbol{\beta}^*) + \boldsymbol{S}_{2i}(\beta^*) + \boldsymbol{S}_{3i}(\boldsymbol{\beta}^*)\end{aligned}$$

$$\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\boldsymbol{\phi}_{2i}(\boldsymbol{\beta}^*) = \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M\frac{Y_i(t)w_i(t)exp(\boldsymbol{\beta}^{*'}\widehat{\boldsymbol{Z}}_{i(k)}(t))}{G(t)s^{(0)}(\boldsymbol{\beta}^*,t)}\left[\widehat{\boldsymbol{Z}}_{i(k)}(t) - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s^{(0)}(\boldsymbol{\beta}^*,t)}\right]d\tilde{R}(t)$$

Now we perform the Taylor series expansion around $\boldsymbol{\eta}_k(t)$ and $\boldsymbol{\tau}_k(t)$. We ignore $w_i(t)$ for now, since it is not the case that $w_i(t) \xrightarrow{p} \tilde{w}_i(t)$. For some $|\boldsymbol{\beta} - \boldsymbol{\beta}^*| < |\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}|$ and $|\boldsymbol{\tau} - \boldsymbol{\tau}^*| < |\boldsymbol{\tau} - \widehat{\boldsymbol{\tau}}|$,

$$f(\boldsymbol{\eta}_k(t), \boldsymbol{\tau}_k(t)) = exp(\boldsymbol{\beta}^{*'} \widehat{\boldsymbol{Z}}_{i(k)}(t))\Big[\widehat{\boldsymbol{Z}}_{i(k)}(t) - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*, t)}{s(0)(\boldsymbol{\beta}^*, t)}\Big]$$

$$= exp(\boldsymbol{\beta}^{*'}[\boldsymbol{\eta}_k(t) + \boldsymbol{\tau}_k(t)\overline{\boldsymbol{W}}_i])\Big[\widehat{\beta}_k(t) + \widehat{\boldsymbol{\tau}}_k(t) - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*, t)}{s(0)(\boldsymbol{\beta}^*, t)}\Big]$$

$$f(\widehat{\boldsymbol{\eta}}_k(t), \widehat{\boldsymbol{\tau}}_k(t)) = f(\boldsymbol{\eta}_k(t), \boldsymbol{\tau}_k(t)) + \frac{\partial f}{\partial \boldsymbol{\eta}}(\widehat{\boldsymbol{\eta}}_k(t) - \boldsymbol{\eta}_k(t))\Bigg|_{\boldsymbol{\eta}=\boldsymbol{\eta}^*, \boldsymbol{\tau}=\boldsymbol{\tau}^*} + \frac{\partial f}{\partial \boldsymbol{\tau}}(\widehat{\boldsymbol{\tau}}_k(t) - \boldsymbol{\tau}_k(t))\Bigg|_{\boldsymbol{\eta}=\boldsymbol{\eta}^*, \boldsymbol{\tau}=\boldsymbol{\tau}^*}$$

Assuming $\frac{\partial f}{\partial \boldsymbol{\eta}}$ and $\frac{\partial f}{\partial \boldsymbol{\tau}}$ are continuous, then by the continuous mapping theorem,

$$f(\widehat{\boldsymbol{\eta}}_k(t), \widehat{\boldsymbol{\tau}}_k(t)) = f(\boldsymbol{\eta}_k(t), \boldsymbol{\tau}_k(t)) + \frac{\partial f}{\partial \boldsymbol{\eta}}\Bigg|_{\boldsymbol{\eta}=\boldsymbol{\eta}, \boldsymbol{\tau}=\boldsymbol{\tau}}(\widehat{\boldsymbol{\eta}}_k(t) - \boldsymbol{\eta}_k(t)) + \frac{\partial f}{\partial \boldsymbol{\tau}}\Bigg|_{\boldsymbol{\eta}=\boldsymbol{\eta}, \boldsymbol{\tau}=\boldsymbol{\tau}}(\widehat{\boldsymbol{\tau}}_k(t) - \boldsymbol{\tau}_k(t)) + o_p(1)$$

$$\frac{\partial f}{\partial \boldsymbol{\eta}} = \boldsymbol{\beta}^* exp\big\{\boldsymbol{\beta}^{*'}(\boldsymbol{\eta}_k(t) + \boldsymbol{\tau}_k(t))\big\}\Big[\boldsymbol{\eta}_k(t) + \boldsymbol{\tau}_k(t)\overline{\boldsymbol{W}}_i - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*, t)}{s^{(0)}(\boldsymbol{\beta}^*, t)}\Big]$$
$$+ exp\big\{\boldsymbol{\beta}^{*'}(\boldsymbol{\eta}_k(t) + \boldsymbol{\tau}_k(t)\overline{\boldsymbol{W}}_i)\big\}$$

$$\frac{\partial f}{\partial \boldsymbol{\tau}} = \beta^{*'}\overline{\boldsymbol{W}}_i exp\big\{\boldsymbol{\beta}^{*'}(\boldsymbol{\eta}_k(t) + \boldsymbol{\tau}_k(t)\overline{\boldsymbol{W}}_i)\big\}\Big[\boldsymbol{\eta}_k(t) + \boldsymbol{\tau}_k(t)\overline{\boldsymbol{W}}_i - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*, t)}{s^{(0)}(\boldsymbol{\beta}^*, t)}\Big]$$
$$+ \overline{\boldsymbol{W}}_i exp\big\{\boldsymbol{\beta}^{*'}(\boldsymbol{\eta}_k(t) + \boldsymbol{\tau}_k(t)\overline{\boldsymbol{W}}_i)\big\}$$

$$\begin{aligned}
\implies f(\widehat{\boldsymbol{\eta}}_k(t), \widehat{\boldsymbol{\tau}}_k(t)) = &\left[\boldsymbol{Z}_{i(k)}(t) - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*, t)}{s^{(0)}(\boldsymbol{\beta}^*, t)}\right] exp\big\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\big\} \\
&+ (\widehat{\boldsymbol{\eta}}_k(t) - \boldsymbol{\eta}_k(t))\boldsymbol{\beta}^* exp\big\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\big\}\left[\boldsymbol{Z}_{i(k)}(t) - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*, t)}{s^{(0)}(\boldsymbol{\beta}^*, t)}\right] \\
&+ (\widehat{\boldsymbol{\eta}}_k(t) - \boldsymbol{\eta}_k(t)) exp\big\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\big\} \\
&+ (\widehat{\boldsymbol{\tau}}_k(t) - \boldsymbol{\tau}_k(t))\boldsymbol{\beta}^{*'}\overline{\boldsymbol{W}}_i exp\big\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\big\}\left[\boldsymbol{Z}_{i(k)}(t) - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*, t)}{s^{(0)}(\boldsymbol{\beta}^*, t)}\right] \\
&+ (\widehat{\boldsymbol{\tau}}_k(t) - \boldsymbol{\tau}_k(t))\overline{\boldsymbol{W}}_i exp\big\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\big\} + o_p(1)
\end{aligned}$$

Now we insert $f(\widehat{\boldsymbol{\eta}}_k(t), \widehat{\boldsymbol{\tau}}_k(t))$ back into $\boldsymbol{\phi}_{2i}(\boldsymbol{\beta}^*)$:

$$
\begin{aligned}
\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\boldsymbol{\phi}_{2i}(\boldsymbol{\beta}^*) &= \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M \frac{Y_i(t)w_i(t)exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\}}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\left[\boldsymbol{Z}_{i(k)}(t)-\frac{\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\right]d\tilde{R}(t)\\
&+\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M \frac{Y_i(t)w_i(t)exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\}}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\boldsymbol{\beta}^{*'}(\widehat{\boldsymbol{\eta}}_k(t)-\boldsymbol{\eta}_k(t))\left[\boldsymbol{Z}_{i(k)}(t)-\frac{\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\right]d\tilde{R}(t)\\
&+\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M \frac{Y_i(t)w_i(t)exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\}}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}(\widehat{\boldsymbol{\eta}}_k(t)-\boldsymbol{\eta}_k(t))d\tilde{R}(t)\\
&+\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M \frac{Y_i(t)w_i(t)exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\}}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\boldsymbol{\beta}^{*'}\overline{\boldsymbol{W}}_i\left[\boldsymbol{Z}_{i(k)}(t)-\frac{\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\right](\widehat{\boldsymbol{\tau}}_k(t)-\boldsymbol{\tau}_k(t))d\tilde{R}(t)\\
&+\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M \frac{Y_i(t)w_i(t)exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\}}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\overline{\boldsymbol{W}}_i(\widehat{\boldsymbol{\tau}}_k(t)-\boldsymbol{\tau}_k(t))d\tilde{R}(t)+o_p(1)\\
&=\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\boldsymbol{T}_{1i}(\boldsymbol{\beta}^*)+\boldsymbol{T}_{2i}(\beta^*)+\boldsymbol{T}_{3i}(\boldsymbol{\beta}^*)+\boldsymbol{T}_{4i}(\beta^*)+\boldsymbol{T}_{5i}(\beta^*)
\end{aligned}
$$

$$
\begin{aligned}
-\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}T_{3i}(\boldsymbol{\beta}^*)+\boldsymbol{S}_{1i}(\boldsymbol{\beta}^*) &= -\int_0^M \sqrt{n}(\widehat{\boldsymbol{\eta}}_k(t)-\boldsymbol{\eta}_k(t))\frac{1}{q_k}\sum_{i\in Q_k}\frac{Y_i(t)w_i(t)exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\}}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}d\tilde{R}(t)\\
&+\int_0^M\sqrt{n}(\widehat{\boldsymbol{\tau}}_k(t)-\boldsymbol{\tau}_k(t))d\tilde{R}(t)+o_p(1)\\
&=-\int_0^M\sqrt{n}(\widehat{\boldsymbol{\eta}}_k(t)-\boldsymbol{\eta}_k(t))\frac{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}d\tilde{R}(t)+\int_0^M\sqrt{n}(\widehat{\boldsymbol{\tau}}_k(t)-\boldsymbol{\tau}_k(t))d\tilde{R}(t)+o_p(1)\\
&=0+o_p(1)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\sqrt{n}}{q_k}\quad\sum_{i\in Q_k}\boldsymbol{T}_{2i}(\boldsymbol{\beta}^*) &= \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M\frac{Y_i(t)w_i(t)exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\}}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\boldsymbol{\beta}^{*'}(\widehat{\boldsymbol{\eta}}_k(t)\\
&-\boldsymbol{\eta}_k(t))\left[\boldsymbol{Z}_{i(k)}(t)-\frac{\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\right]d\tilde{R}(t)\\
=&\boldsymbol{\beta}^*\int_0^M\sqrt{n}(\widehat{\boldsymbol{\eta}}_k(t)-\boldsymbol{\eta}_k(t))\frac{1}{q_k}\sum_{i\in Q_k}\left[\frac{Y_i(t)w_i(t)\boldsymbol{Z}_{i(k)}(t)exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\}}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\right.\\
&\left.-\frac{Y_i(t)w_i(t)exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\}\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)^2}\right]d\tilde{R}(t)\\
=&\boldsymbol{\beta}^*\int_0^M\sqrt{n}(\widehat{\boldsymbol{\eta}}_k(t)-\boldsymbol{\eta}_k(t))\left[\frac{G(t)\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}-\frac{\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\left\{\frac{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\right\}\right]d\tilde{R}(t)+o_p(1)\\
=&\beta^*\int_0^M\frac{1}{\sqrt{n}}\sum_{i=1}^n\boldsymbol{C}_{i(k)}^{k_i}(t)\left[\frac{\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}-\frac{\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\right]d\tilde{R}(t)+o_p(1)\\
=&0+o_p(1)
\end{aligned}
$$

We now add and subtract $\tilde{w}_i(t)$ in the remaining components and simplify:

$$\begin{aligned}
\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\boldsymbol{S}_{3i}(\boldsymbol{\beta}^*) &= \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M (w_i(t)\pm\tilde{w}_i(t))\left[\boldsymbol{Z}_{i(k)}(t)-\frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s^{(0)}(\boldsymbol{\beta}^*,t)}\right]dN_i(t)\\
&= \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M \tilde{w}_i(t)\left[\boldsymbol{Z}_{i(k)}(t)-\frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s^{(0)}(\boldsymbol{\beta}^*,t)}\right]dN_i(t)\\
&+ \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M (w_i(t)-\tilde{w}_i(t))\left[\boldsymbol{Z}_{i(k)}(t)-\frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s^{(0)}(\boldsymbol{\beta}^*,t)}\right]dN_i(t)\\
&= \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\tilde{\boldsymbol{\phi}}_{1i}(\boldsymbol{\beta}^*)+\boldsymbol{\psi}_{1i}(\boldsymbol{\beta}^*)
\end{aligned}$$

Note that $\tilde{\boldsymbol{\phi}}_{1i}(\boldsymbol{\beta}^*)$ is similar in form to $\boldsymbol{\phi}_{1i}(\boldsymbol{\beta}^*)$, but $\tilde{\boldsymbol{\phi}}_{1i}(\boldsymbol{\beta}^*)$ are independent, identically distributed random variables, with each estimated component replaced with its known value. We now replace $(w_i(t)-\tilde{w}_i(t))$ with the expression derived in Appendix A.2.

$$\begin{aligned}
\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\boldsymbol{\psi}_{1i}(\boldsymbol{\beta}^*) &= \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M (w_i(t)-\tilde{w}_i(t))\left[\boldsymbol{Z}_{i(k)}(t)-\frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s^{(0)}(\boldsymbol{\beta}^*,t)}\right]dN_i(t)\\
= \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M &\frac{-r_i(t)G(t)I(X_i<t)}{\widehat{G}(X_i)}\int_{X_i}^t \frac{\widehat{G}_(u-)}{G(u)Y_n(u)}dM_n^c(u)\left[\boldsymbol{Z}_{i(k)}(t)-\frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s^{(0)}(\boldsymbol{\beta}^*,t)}\right]dN_i(t)
\end{aligned}$$

Invoking consistency of $\widehat{G}(\cdot)$ and continuous time as described in Appendix A.2,

$$
\begin{aligned}
\frac{\sqrt{n}}{q_k} \quad & \sum_{i \in Q_k} \boldsymbol{\psi}_{1i}(\boldsymbol{\beta}^*) = \frac{\sqrt{n}}{q_k} \sum_{i \in Q_k} \int_0^M \frac{-r_i(t)G(t)I(X_i < t)}{\widehat{G}(X_i)} \int_{X_i}^t \frac{\widehat{G}_(u-)}{G(u)Y_n(u)} dM_n^c(u) \Bigg[\boldsymbol{Z}_{i(k)}(t) \\
& - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*, t)}{s^{(0)}(\boldsymbol{\beta}^*, t)} \Bigg] dN_i(t) + o_p(1) \\
=& \frac{\sqrt{n}}{q_k} \sum_{i \in Q_k} \int_0^M \frac{-r_i(t)G(t)I(X_i < t)}{G(X_i)} \int_{X_i}^t \frac{1}{Y_n(u)} dM_n^c(u) \Bigg[\boldsymbol{Z}_{i(k)}(t) - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*, t)}{s^{(0)}(\boldsymbol{\beta}^*, t)} \Bigg] dN_i(t) + o_p(1) \\
=& \frac{\sqrt{n}}{q_k} \sum_{i \in Q_k} \int_0^M -\tilde{w}_i(t) \int_0^M \frac{I(X_i < u \leq t)}{\frac{1}{n} Y_n(u)} \frac{1}{n} \sum_{j=1}^n dM_j^c(u) \Bigg[\boldsymbol{Z}_{i(k)}(t) - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*, t)}{s^{(0)}(\boldsymbol{\beta}^*, t)} \Bigg] dN_i(t) + o_p(1) \\
=& \frac{\sqrt{n}}{q_k} \sum_{i \in Q_k} \frac{1}{n} \sum_{j=1}^n \int_0^M \int_0^M -\tilde{w}_i(t) \Bigg[\boldsymbol{Z}_{i(k)}(t) - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*, t)}{s^{(0)}(\boldsymbol{\beta}^*, t)} \Bigg] \frac{I(X_i < u \leq t)}{\pi(u)} dM_j^c(u) dN_i(t) + o_p(1) \\
=& \frac{1}{\sqrt{n}} \sum_{j=1}^n \frac{1}{q_k} \sum_{i \in Q_k} \int_0^M \int_0^M -\tilde{w}_i(t) \Bigg[\boldsymbol{Z}_{i(k)}(t) - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*, t)}{s^{(0)}(\boldsymbol{\beta}^*, t)} \Bigg] \frac{I(t \geq u > X_i)}{\pi(u)} dN_i(t) dM_j^c(u) + o_p(1) \\
=& \frac{1}{\sqrt{n}} \sum_{j=1}^n \int_0^M \int_0^M E \Bigg[\frac{-\tilde{w}(t) I(t \geq u > X)}{\pi(u)} \big[\boldsymbol{Z}_k(t) - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*, t)}{s^{(0)}(\boldsymbol{\beta}^*, t)} \big] dN(t) \Bigg] dM_j^c(u) + o_p(1)
\end{aligned}
$$

The above assumes that regularity conditions hold in the reversal of limits of integration for $dM_j^c(u)$ and $dN_i(t)$.

The same process is repeated for the remaining components of $\frac{\sqrt{n}}{q_k} \sum_{i \in Q_k} \boldsymbol{\phi}_{1i}(\boldsymbol{\beta}^*) - \boldsymbol{\phi}_{2i}(\boldsymbol{\beta}^*)$:

$$
\begin{aligned}
\frac{\sqrt{n}}{q_k} \quad & \sum_{i \in Q_k} \boldsymbol{T}_{1i}(\boldsymbol{\beta}^*) = \frac{\sqrt{n}}{q_k} \sum_{i \in Q_k} \int_0^M \frac{Y_i(t) exp\{\boldsymbol{\beta}^{*'} \boldsymbol{Z}_{i(k)}(t)\}}{G(t) s^{(0)}(\boldsymbol{\beta}^*, t)} \Bigg[w_i(t) - \tilde{w}_i(t) + \tilde{w}_i(t) \Bigg] \\
& \times \Bigg[\boldsymbol{Z}_{i(k)}(t) - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*, t)}{s^{(0)}(\boldsymbol{\beta}^*, t)} \Bigg] d\tilde{R}(t) \\
=& \frac{\sqrt{n}}{q_k} \sum_{i \in Q_k} \int_0^M \frac{Y_i(t) \tilde{w}_i(t) exp\{\boldsymbol{\beta}^{*'} \boldsymbol{Z}_{i(k)}(t)\}}{G(t) s^{(0)}(\boldsymbol{\beta}^*, t)} \Bigg[\boldsymbol{Z}_{i(k)}(t) - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*, t)}{s^{(0)}(\boldsymbol{\beta}^*, t)} \Bigg] d\tilde{R}(t) \\
& + \frac{\sqrt{n}}{q_k} \sum_{i \in Q_k} \int_0^M \frac{Y_i(t) exp\{\boldsymbol{\beta}^* \boldsymbol{Z}_{i(k)}(t)\}}{G(t) s^{(0)}(\boldsymbol{\beta}^*, t)} \Bigg[\boldsymbol{Z}_{i(k)}(t) - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*, t)}{s^{(0)}(\boldsymbol{\beta}^*, t)} \Bigg] \Bigg[w_i(t) - \tilde{w}_i(t) \Bigg] d\tilde{R}(t) \\
& = \frac{\sqrt{n}}{q_k} \sum_{i \in Q_k} \tilde{\boldsymbol{\phi}}_{2i}(\boldsymbol{\beta}^*) + \boldsymbol{\psi}_{2i}(\boldsymbol{\beta}^*)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\sqrt{n}}{q_k} \quad \sum_{i \in Q_k} \boldsymbol{\psi}_{2i}(\boldsymbol{\beta}^*) = {} & \frac{\sqrt{n}}{q_k} \sum_{i \in Q_k} \int_0^M \frac{Y_i(t) exp\{\boldsymbol{\beta}^{*'} \boldsymbol{Z}_{i(k)}(t)\}}{G(t) \mathcal{s}^{(0)}(\boldsymbol{\beta}^*, t)} \left[\boldsymbol{Z}_{i(k)}(t) - \frac{\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*, t)}{\mathcal{s}^{(0)}(\boldsymbol{\beta}^*, t)} \right] \\
& \times \left\{ w_i(t) - \tilde{w}_i(t) \right\} d\tilde{R}(t) \\
= & \frac{\sqrt{n}}{q_k} \sum_{i \in Q_k} \int_0^M \frac{Y_i(t) exp\{\boldsymbol{\beta}^{*'} \boldsymbol{Z}_{i(k)}(t)\}}{G(t) \mathcal{s}^{(0)}(\boldsymbol{\beta}^*, t)} \Bigg[\boldsymbol{Z}_{i(k)}(t) \\
& - \frac{\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*, t)}{\mathcal{s}^{(0)}(\boldsymbol{\beta}^*, t)} \Bigg] \left\{ \frac{-r_i(t) G(t) I(X_i < t)}{\widehat{G}(X_i)} \int_{X_i}^t \frac{\widehat{G}(u-)}{G(u) Y_n(u)} dM_n^c(u) \right\} d\tilde{R}(t) \\
= & \frac{\sqrt{n}}{q_k} \sum_{i \in Q_k} \int_0^M \frac{Y_i(t) exp\{\boldsymbol{\beta}^{*'} \boldsymbol{Z}_{i(k)}(t)\}}{G(t) \mathcal{s}^{(0)}(\boldsymbol{\beta}^*, t)} \Bigg[\boldsymbol{Z}_{i(k)}(t) \\
& - \frac{\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*, t)}{\mathcal{s}^{(0)}(\boldsymbol{\beta}^*, t)} \Bigg] \left\{ - \tilde{w}_i(t) I(X_i < t) \int_{X_i}^t \frac{1}{\frac{1}{n} Y_n(u)} \frac{1}{n} dM_n^c(u) \right\} d\tilde{R}(t) + o_p(1) \\
= & \frac{\sqrt{n}}{q_k} \sum_{i \in Q_k} \int_0^M \frac{Y_i(t) exp\{\boldsymbol{\beta}^{*'} \boldsymbol{Z}_{i(k)}(t)\Big\}}{G(t) \mathcal{s}^{(0)}(\boldsymbol{\beta}^*, t)} \Bigg[\boldsymbol{Z}_{i(k)}(t) \\
& - \frac{\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*, t)}{\mathcal{s}^{(0)}(\boldsymbol{\beta}^*, t)} \Bigg] \left\{ - \tilde{w}_i(t) \int_0^M \frac{I(X_i < u \leq t)}{\pi(u)} \frac{1}{n} \sum_{j=1}^n dM_j^c(u) \right\} d\tilde{R}(t) + o_p(1) \\
= & \frac{\sqrt{n}}{q_k} \sum_{i \in Q_k} \frac{1}{n} \sum_{j=1}^n \int_0^M \int_0^M \frac{Y_i(t) exp\{\beta^{*'} \boldsymbol{Z}_{i(k)}(t)\}}{G(t) \mathcal{s}^{(0)}(\boldsymbol{\beta}^*, t)} \Bigg[\boldsymbol{Z}_{i(k)}(t) \\
& - \frac{\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*, t)}{\mathcal{s}^{(0)}(\boldsymbol{\beta}^*, t)} \Bigg] \left\{ - \tilde{w}_i(t) \frac{I(X_i < u \leq t)}{\pi(u)} \right\} dM_j^c(u) d\tilde{R}(t) + o_p(1) \\
= & \frac{1}{\sqrt{n}} \sum_{j=1}^n \frac{1}{q_k} \sum_{i \in Q_k} \int_0^M \int_0^M \frac{Y_i(t) exp\{\boldsymbol{\beta}^{*'} \boldsymbol{Z}_{i(k)}(t)\}}{G(t) \mathcal{s}^{(0)}(\boldsymbol{\beta}^*, t)} \Bigg[\boldsymbol{Z}_{i(k)}(t) \\
& - \frac{\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*, t)}{\mathcal{s}^{(0)}(\boldsymbol{\beta}^*, t)} \Bigg] \left\{ \frac{-\tilde{w}(t) I(t \geq u > X_i)}{\pi(u)} \right\} d\tilde{R}(t) dM_j^c(u) + o_p(1) \\
= & \frac{1}{\sqrt{n}} \sum_{j=1}^n \int_0^M \int_0^M E \Bigg[\frac{-Y(t) exp\{\boldsymbol{\beta}^{*'} \boldsymbol{Z}_k(t)\} \tilde{w}(t) I(t \geq u > X)}{G(t) \mathcal{s}^{(0)}(\boldsymbol{\beta}^*, t) \pi(u)} \Big[\boldsymbol{Z}_k(t) \\
& - \frac{\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*, t)}{\mathcal{s}^{(0)}(\boldsymbol{\beta}^*, t)} \Big] \Bigg] d\tilde{R}(t) dM_j^c(u) + o_p(1)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\boldsymbol{S}_{2i}(\boldsymbol{\beta}^*) &= \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M(\widehat{\boldsymbol{\tau}}_k(t)-\boldsymbol{\tau}_k(t))\overline{\boldsymbol{W}}_i(w_i(t)\pm\tilde{w}_i(t))dN_i(t)\\
&= \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M(\widehat{\boldsymbol{\tau}}_k(t)-\boldsymbol{\tau}_k(t))\overline{\boldsymbol{W}}_i\tilde{w}_i(t)dN_i(t)\\
&+ \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M(\widehat{\boldsymbol{\tau}}_k(t)-\boldsymbol{\tau}_k(t))\overline{\boldsymbol{W}}_i(w_i(t)-\tilde{w}_i(t))dN_i(t)\\
&= \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\boldsymbol{\psi}_{3i}(\boldsymbol{\beta}^*)+\boldsymbol{\psi}_{4i}(\boldsymbol{\beta}^*)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\boldsymbol{\psi}_{3i}(\boldsymbol{\beta}^*) &= \frac{\sqrt{n}}{q_k}\int_0^M(\widehat{\boldsymbol{\tau}}_k(t)-\boldsymbol{\tau}_k(t))\overline{\boldsymbol{W}}_i\tilde{w}_i(t)dN_i(t)\\
&= \int_0^M\sqrt{n}(\widehat{\boldsymbol{\tau}}_k(t)-\boldsymbol{\tau}_k(t))\frac{1}{q_k}\sum_{i\in Q_k}\Big\{\overline{\boldsymbol{W}}_i\tilde{w}_i(t)dN_i(t)\Big\}\\
&= \int_0^M\frac{1}{\sqrt{n}}\sum_{i=1}^n\boldsymbol{D}_{i(k)}^{k_i}(t)E\Bigg[\overline{\boldsymbol{W}}(k)\tilde{w}(t)dN(t)\Bigg]+o_p(1)
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\boldsymbol{\psi}_{4i}(\boldsymbol{\beta}^*) = \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M(\widehat{\boldsymbol{\tau}}_k(t)-\boldsymbol{\tau}_k(t))\overline{\boldsymbol{W}}_i\{w_i(t)-\tilde{w}_i(t)\}dN_i(t)\\
&=\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M(\widehat{\boldsymbol{\tau}}_k(t)-\boldsymbol{\tau}_k(t))\overline{\boldsymbol{W}}_i\left\{\frac{-r_i(t)G(t)I(X_i<t)}{\widehat{G}(X_i)}\int_{X_i}^t\frac{\widehat{G}(u-)}{G(u)Y_n(u)}dM_n^c(u)\right\}dN_i(t)\\
&=\int_0^M\sqrt{n}(\widehat{\boldsymbol{\tau}}_k(t)-\boldsymbol{\tau}_k(t))\frac{1}{q_k}\sum_{i\in Q_k}\overline{\boldsymbol{W}}_i\left\{-\tilde{w}_i(t)\int_0^M\frac{I(X_i<u\le t)}{\frac{1}{n}\sum_{k=1}^n I(X_k\ge u)}\frac{1}{n}\sum_{j=1}^n dM_j^c(u)\right\}dN_i(t)\\
&\quad+o_p(1)\\
&=\int_0^M\int_0^M\frac{1}{\sqrt{n}}\sum_{i=1}^n\boldsymbol{D}_{i(k)}^{k_i}(t)E\left[\frac{\overline{\boldsymbol{W}}(k)\tilde{w}(t)I(t\ge u>X)dN(t)}{\pi(u)}\right]E\big[dM^c(u)\big]+o_p(1)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\boldsymbol{T}_{4i}(\boldsymbol{\beta}^*) &= \frac{\sqrt{n}}{q_k}\sum_{0}^{M}\frac{Y_i(t)[w_i(t)\pm\tilde{w}_i(t)]exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\}}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\\
&\times\left[\boldsymbol{Z}_{i(k)}(t)-\frac{\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\right]\boldsymbol{\beta}^{*'}\overline{\boldsymbol{W}}_i(\widehat{\boldsymbol{\tau}}_k(t)-\boldsymbol{\tau}_k(t))d\tilde{R}(t)\\
=&\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M\frac{Y_i(t)\tilde{w}_i(t)exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\}}{G(t)\mathcal{s}^{(0)}}\left[\boldsymbol{Z}_{i(k)}(t)-\frac{\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\right]\boldsymbol{\beta}^{*'}\overline{\boldsymbol{W}}_i(\widehat{\boldsymbol{\tau}}_k(t)-\boldsymbol{\tau}_k(t))d\tilde{R}(t)\\
=&\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M\frac{Y_i(t)\{w_i(t)-\tilde{w}_i(t)\}exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\}}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\left[\boldsymbol{Z}_{i(k)}(t)-\frac{\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\right]\\
&\times\boldsymbol{\beta}^{*'}\overline{\boldsymbol{W}}_i(\widehat{\boldsymbol{\tau}}_k(t)-\boldsymbol{\tau}_k(t))d\tilde{R}(t)\\
=&\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\boldsymbol{\psi}_{5i}(\boldsymbol{\beta}^*)+\boldsymbol{\psi}_{6i}(\boldsymbol{\beta}^*)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\sqrt{n}}{q_k}\quad\sum_{i\in Q_k}\boldsymbol{\psi}_{5i}(\boldsymbol{\beta}^*) &= \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M\frac{Y_i(t)\tilde{w}_i(t)exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\}}{G(t)\mathcal{s}^{(0)}}\left[\boldsymbol{Z}_{i(k)}(t)-\frac{\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\right]\\
&\times\boldsymbol{\beta}^{*'}\overline{\boldsymbol{W}}_i(\widehat{\boldsymbol{\tau}}_k(t)-\boldsymbol{\tau}_k(t))d\tilde{R}(t)\\
&=\int_0^M\sqrt{n}(\widehat{\boldsymbol{\tau}}_k(t)-\boldsymbol{\tau}_k(t))\frac{1}{q_k}\sum_{i\in Q_k}\left\{\frac{Y_i(t)\tilde{w}_i(t)exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\}}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\right.\\
&\times\left.\left[\boldsymbol{Z}_{i(k)}(t)-\frac{\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\right]\boldsymbol{\beta}^{*'}\overline{\boldsymbol{W}}_i\right\}d\tilde{R}(t)\\
&=\int_0^M\frac{1}{\sqrt{n}}\sum_{i=1}^n\boldsymbol{D}_{i(k)}^{k_i}(t)E\left[\frac{Y(t)\tilde{w}(t)exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_k(t)\}}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\left\{\boldsymbol{Z}_k(t)-\frac{\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\right\}\beta^{*'}\overline{\boldsymbol{W}}(k)\right]d\tilde{R}(t)\\
&+o_p(1)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\boldsymbol{\psi}_{6i}(\boldsymbol{\beta}^*) = &\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M(\widehat{\boldsymbol{\tau}}_k(t)-\boldsymbol{\tau}_k(t))\boldsymbol{\beta}^{*'}\overline{\boldsymbol{W}}_i\frac{Y_i(t)exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\}}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\Bigg[\boldsymbol{Z}_{i(k)}(t)\\
&-\frac{\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\Bigg]\Big(w_i(t)-\tilde{w}_i(t)\Big)d\tilde{R}(t)\\
=&\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M(\widehat{\boldsymbol{\tau}}_k(t)-\boldsymbol{\tau}_k(t))\boldsymbol{\beta}^{*'}\overline{\boldsymbol{W}}_i\frac{Y_i(t)exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\}}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\Bigg[\boldsymbol{Z}_{i(k)}(t)\\
&-\frac{\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\Bigg]\Big(-\tilde{w}_i(t)\int_0^M\frac{I(X_i<u\le t)}{\pi(u)}\frac{1}{n}\sum_{j=1}^n dM_j^c(u)\Big)d\tilde{R}(t)\\
=&\int_0^M\int_0^M\sqrt{n}(\widehat{\boldsymbol{\tau}}_k(t)-\boldsymbol{\tau}_k(t))\frac{1}{q_k}\sum_{i\in Q_k}\Bigg\{\boldsymbol{\beta}^{*'}\overline{\boldsymbol{W}}_i\frac{Y_i(t)exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\}}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\Bigg[\boldsymbol{Z}_{i(k)}(t)\\
&-\frac{\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\Bigg]-\tilde{w}_i(t)I(X_i<u\le t)\Bigg\}\frac{1}{n}\sum_{j=1}^n\Big\{\frac{dM_j^c(u)}{\pi(u)}\Big\}\\
=&\int_0^M\int_0^M\frac{1}{\sqrt{n}}\sum_{i=1}^n\boldsymbol{D}_{i(k)}^{k_i}(t)E\Bigg[\frac{-\beta^{*'}\overline{\boldsymbol{W}}(k)Y(t)exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_k(t)\}\tilde{w}(t)I(t\ge u>X)}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\Big\{\boldsymbol{Z}_k(t)\\
&-\frac{\boldsymbol{\mathcal{s}}^{(1)}(\boldsymbol{\beta}^*,t)}{\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\Big\}\Bigg]d\tilde{R}(t)E\Bigg[\frac{dM^c(u)}{\pi(u)}\Bigg]+o_p(1)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\boldsymbol{T}_{5i}(\boldsymbol{\beta}^*) =& \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M\frac{Y_i(t)[w_i(t)\pm\tilde{w}_i(t)]exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\}}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\overline{\boldsymbol{W}}_i(\widehat{\boldsymbol{\tau}}_k(t)-\boldsymbol{\tau}_k(t))d\tilde{R}(t)\\
=&\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M\frac{Y_i(t)\tilde{w}_i(t)exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\}}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\overline{\boldsymbol{W}}_i(\widehat{\boldsymbol{\tau}}_k(t)-\boldsymbol{\tau}_k(t))d\tilde{R}(t)\\
&+\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M\frac{Y_i(t)[w_i(t)-\tilde{w}_i(t)]exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\}}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\overline{\boldsymbol{W}}_i(\widehat{\boldsymbol{\tau}}_k(t)-\boldsymbol{\tau}_k(t))d\tilde{R}(t)\\
=&\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\boldsymbol{\psi}_{7i}(\boldsymbol{\beta}^*)+\boldsymbol{\psi}_{8i}(\boldsymbol{\beta}^*)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\boldsymbol{\psi}_{7i}(\boldsymbol{\beta}^*) =& \frac{\sqrt{n}}{q_k}\sum_{i\in Q_k}\int_0^M\frac{Y_i(t)\tilde{w}_i(t)exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\}}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\overline{\boldsymbol{W}}_i(\widehat{\boldsymbol{\tau}}_k(t)-\boldsymbol{\tau}_k(t))d\tilde{R}(t)\\
=&\int_0^M\sqrt{n}(\widehat{\boldsymbol{\tau}}_k(t)-\boldsymbol{\tau}_k(t))\frac{1}{q_k}\sum_{i\in Q_k}\Bigg\{\frac{Y_i(t)exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_{i(k)}(t)\}\overline{\boldsymbol{W}}_i\tilde{w}_i(t)}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\Bigg\}d\tilde{R}(t)\\
=&\int_0^M\frac{1}{\sqrt{n}}\sum_{i=1}^n\boldsymbol{D}_{i(k)}^{k_i}(t)E\Bigg[\frac{Y(t)exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_k(t)\}\overline{\boldsymbol{W}}(k)\tilde{w}(t)}{G(t)\mathcal{s}^{(0)}(\boldsymbol{\beta}^*,t)}\Bigg]d\tilde{R}(t)+o_p(1)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\sqrt{n}}{q_k} \quad \sum_{i \in Q_k} \boldsymbol{\psi}_{8i}(\boldsymbol{\beta}^*) &= \frac{\sqrt{n}}{q_k} \sum_{i \in Q_k} \int_0^M \frac{Y_i(t)[w_i(t) - \tilde{w}_i(t)]exp\{\boldsymbol{\beta}^{*'} Z_{i(k)}(t)\}}{G(t)s^{(0)}(\boldsymbol{\beta}^*, t)} \overline{\boldsymbol{W}}_i(\widehat{\boldsymbol{\tau}}_k(t) - \boldsymbol{\tau}_k(t))d\tilde{R}(t) \\
=& \frac{\sqrt{n}}{q_k} \sum_{i \in Q_k} \int_0^M (\widehat{\boldsymbol{\tau}}_k(t) - \boldsymbol{\tau}_k(t)) \frac{Y_i(t)exp\{\beta^{*'} \boldsymbol{Z}_{i(k)}(t)\}}{G(t)s^{(0)}(\boldsymbol{\beta}^*, t)} \overline{\boldsymbol{W}}_i \\
& \times \left[\frac{-r_i(t)G(t)}{\widehat{G}(X_i)} \int_0^M \frac{\widehat{G}(u-)I(X_i < u \leq t)}{G(u)\frac{1}{n}Y_n(u)} \frac{1}{n} \sum_{j=1}^n dM_j^c(u) \right] d\tilde{R}(t) \\
=& \frac{\sqrt{n}}{q_k} \sum_{i \in Q_k} \int_0^M (\widehat{\boldsymbol{\tau}}_k(t) - \boldsymbol{\tau}_k(t)) \frac{Y_i(t)exp\{\boldsymbol{\beta}^{*'} \boldsymbol{Z}_{i(k)}(t)\}}{G(t)s^{(0)}(\beta^*, t)} \overline{W}_i \\
& \times \left[- \tilde{w}_i(t) \int_0^M \frac{I(X_i < u \leq t)}{\pi(u)} \frac{1}{n} \sum_{j=1}^n dM_j^c(u) \right] d\tilde{R}(t) \\
=& \int_0^M \int_0^M \sqrt{n}(\widehat{\boldsymbol{\tau}}_k(t) - \boldsymbol{\tau}_k(t)) \frac{1}{q_k} \sum_{i \in Q_k} \left\{ \frac{-Y_i(t)exp\{\boldsymbol{\beta}^{*'} \boldsymbol{Z}_{i(k)}(t)\} \overline{\boldsymbol{W}}_i \tilde{w}_i(t) I(X_i < u \leq t)}{G(t)s^{(0)}(\boldsymbol{\beta}^*, t)\pi(u)} \right\} \\
& \times \frac{1}{n} \sum_{j=1}^n dM_j^c(u) \\
=& \int_0^M \int_0^M \frac{1}{\sqrt{n}} \boldsymbol{D}_{i(k)}^{k_i}(t) E\left[\frac{-Y(t)exp\{\boldsymbol{\beta}^{*'} \boldsymbol{Z}_k(t)\} \overline{\boldsymbol{W}}(k) \tilde{w}(t) I(X < u \leq t)}{G(t)s^{(0)}(\boldsymbol{\beta}^*, t)\pi(u)} \right] \\
& \times E[dM^c(u)]d\tilde{R}(t) + o_p(1)
\end{aligned}
$$

Now add up all components again and simplify. Denote the function $\boldsymbol{\psi}_{i(k)}^{k_i}(\boldsymbol{\beta}^*) = \sum_{s=1}^{8} \boldsymbol{\psi}_{si}(\boldsymbol{\beta}^*)$, which is delineated with a (k) to indicate dependence on k replicates, and superscript k_i to indicate that subject i has k_i replicates.

$$
\frac{\sqrt{n}}{q_k} \sum_{i \in Q_k} \phi(\beta^*) = \frac{\sqrt{n}}{q_k} \sum_{i \in Q_k} \tilde{\boldsymbol{\phi}}_i(\boldsymbol{\beta}^*) + \frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{\psi}_{i(k)}^{k_i}(\boldsymbol{\beta}^*) + o_p(1)
$$

where

$$
\begin{aligned}
&\boldsymbol{\psi}_{i(k)}^{k_i}(\boldsymbol{\beta}^*) = \sum_{s=1}^{8} \boldsymbol{\psi}_{si}(\boldsymbol{\beta}^*) \\
=&\int_0^M \Bigg\{ \int_0^M E\Bigg[\frac{-\tilde{w}(t)I(t \geq u > X)}{\pi(u)}\Big\{\boldsymbol{Z}_k(t) - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s^{(0)}(\boldsymbol{\beta}^*,t)}\Big\}dN(t)\Bigg] \\
&+ \int_0^M E\Bigg[\frac{-Y(t)exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_k(t)\}\tilde{w}(t)I(t \geq u > X)}{G(t)s^{(0)}(\boldsymbol{\beta}^*,t)\pi(u)}\Big\{\boldsymbol{Z}_k(t) - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s^{(0)}(\boldsymbol{\beta}^*,t)}\Big\}\Bigg]d\tilde{R}(t)\Bigg\}dM_i^c(u) \\
&+ \int_0^M \boldsymbol{D}_{i(k)}^{k_i}(t)E\Bigg[\overline{\boldsymbol{W}}(k)\tilde{w}(t)dN(t)\Bigg] \\
&+ \int_0^M \boldsymbol{D}_{i(k)}^{k_i}(t)E\Bigg[\frac{Y(t)\tilde{w}(t)exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_k(t)\}}{G(t)s^{(0)}(\boldsymbol{\beta}^*,t)}\Bigg\{\Big\{\boldsymbol{Z}_k(t) - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s^{(0)}(\boldsymbol{\beta}^*,t)}\Big\}\boldsymbol{\beta}^{*'}\overline{\boldsymbol{W}}(k) + \overline{\boldsymbol{W}}(k)\Bigg\}\Bigg]d\tilde{R}(t) \\
&+ \int_0^M \Bigg\{ \int_0^M \boldsymbol{D}_{i(k)}^{k_i}(t)E\Bigg[\frac{\overline{\boldsymbol{W}}(k)\tilde{w}(t)I(t \geq u > X)}{\pi(u)}dN(t)\Bigg] \\
&+ \int_0^M D_{i(k)}^{k_i}(t)E\Bigg[\frac{-Y(t)exp\{\boldsymbol{\beta}^{*'}\boldsymbol{Z}_k(t)\}\tilde{w}(t)I(t \geq u > X)}{G(t)s^{(0)}(\boldsymbol{\beta}^*,t)\pi(u)} \\
&\times \Bigg\{\Big\{\boldsymbol{Z}_k(t) - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s^{(0)}(\boldsymbol{\beta}^*,t)}\Big\}\boldsymbol{\beta}^{*'}\overline{\boldsymbol{W}}(k) + \overline{\boldsymbol{W}}(k)\Bigg\}\Bigg]d\tilde{R}(t)\Bigg\}E\big[dM^c(u)\big]
\end{aligned}
$$

Using $\boldsymbol{\phi}_i(\boldsymbol{\beta}^*)$, we can derive $\tilde{\boldsymbol{\phi}}_i(\boldsymbol{\beta}^*)$ by substituting all unknown quantities with their probability limits.

Therefore, following the example in Xie, Wang, and Prentice (2001),

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\boldsymbol{U}_{C-RRC}(\boldsymbol{\beta}^*) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Bigg\{\sum_{k=1}^{\ell} a_k \boldsymbol{\psi}_{i(k)}^{k_i}(\boldsymbol{\beta}^*) + \tilde{\boldsymbol{\phi}}_i(\boldsymbol{\beta}^*)\Bigg\} + o_p(1) \\
&= \sum_{m=1}^{\ell}\frac{\sqrt{q_m}}{\sqrt{n}}\frac{\sum_{i\in Q_m}\Bigg\{\sum_{k=1}^{\ell} a_k \boldsymbol{\psi}_{i(k)}^{m}(\boldsymbol{\beta}^*) + \tilde{\boldsymbol{\phi}}_i(\boldsymbol{\beta}^*)\Bigg\}}{\sqrt{q_m}} + o_p(1)
\end{aligned}
$$

Although $\tilde{\boldsymbol{\phi}}_i(\boldsymbol{\beta}^*)$ depends on k through $\boldsymbol{\eta}_k(t)$ and $\boldsymbol{\tau}_k(t)$, we do not denote it with the k subscript, because in the estimation procedure we replace $\boldsymbol{Z}_k(t)$ with $\widehat{\boldsymbol{Z}}_i(t)$ and do not require individual functions that depend on k.

Note that $\boldsymbol{U}_{C-RRC}(\boldsymbol{\beta}^*)$ is now a sum of independent, identically distributed random variables with mean 0 and finite variance. We define a matrix $\boldsymbol{B}(\boldsymbol{\beta}^*)$ as

$$\boldsymbol{B}(\boldsymbol{\beta}^*) = \sum_{m=1}^{\ell} a_m E_{i\in Q_m}\left\{\sum_{k=1}^{\ell} a_k \boldsymbol{\psi}_{i(k)}^{k_i}(\boldsymbol{\beta}^*) + \tilde{\boldsymbol{\phi}}_i(\boldsymbol{\beta}^*)\right\}^{\otimes 2}$$

and, according to the multivariate central limit theorem,

$$\frac{1}{\sqrt{n}}\boldsymbol{U}_{C-RRC}(\boldsymbol{\beta}^*) \xrightarrow{D} N(0, \boldsymbol{B}(\boldsymbol{\beta}^*))$$

A second-order Taylor series expansion of $\boldsymbol{U}_{C-RRC}(\widehat{\boldsymbol{\beta}}_{C-RRC})$ around $\boldsymbol{\beta}^*$ yields the following relation:

$$\begin{aligned}
\sqrt{n}(\widehat{\boldsymbol{\beta}}_{C-RRC} - \boldsymbol{\beta}^*) &= \boldsymbol{A}(\boldsymbol{\beta}^*)^{-1}\frac{1}{\sqrt{n}}\boldsymbol{U}_{C-RRC}(\boldsymbol{\beta}^*)\boldsymbol{A}^T(\boldsymbol{\beta}^*)^{-1} \\
\boldsymbol{A}(\boldsymbol{\beta}^*) &= -\frac{1}{n}\frac{\partial \boldsymbol{U}(\boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}^*} \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^M \left[\frac{\boldsymbol{\jmath}^{(2)}(\boldsymbol{\beta}^*, t)}{\jmath^{(0)}(\boldsymbol{\beta}^*, t)} - \left\{\frac{\boldsymbol{\jmath}^{(1)}(\boldsymbol{\beta}^*, t)}{\jmath^{(0)}(\boldsymbol{\beta}^*, t)}\right\}^{\otimes 2}\right] w_i(t)dN_i(t) + o_p(1)
\end{aligned}$$

By the central limit theorem and Slutsky's theorem, $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{C-RRC} - \boldsymbol{\beta}^*)$ converges in distribution to a vector of normal random variables with covariance matrix $\boldsymbol{A}(\boldsymbol{\beta}^*)^{-1}\boldsymbol{B}(\boldsymbol{\beta}^*)\boldsymbol{A}^T(\boldsymbol{\beta}^*)^{-1}$. $\boldsymbol{A}(\boldsymbol{\beta}^*)$ and $\boldsymbol{B}(\boldsymbol{\beta}^*)$ can be estimated by replacing $\boldsymbol{\beta}^*$ with $\widehat{\boldsymbol{\beta}}_{C-RRC}$ and each unknown value with its consistent estimate:

$$\widehat{\boldsymbol{A}}(\widehat{\boldsymbol{\beta}}) = \frac{1}{n}\sum_{i=1}^{n}\int_0^M \left[\frac{\widehat{\boldsymbol{S}}^{(2)}(\widehat{\boldsymbol{\beta}}, t)}{\widehat{S}^{(0)}(\widehat{\boldsymbol{\beta}}, t)} - \left\{\frac{\widehat{\boldsymbol{S}}^{(1)}(\widehat{\boldsymbol{\beta}}, t)}{\widehat{S}^{(0)}(\widehat{\boldsymbol{\beta}}, t)}\right\}^{\otimes 2}\right]$$

$$\widehat{\boldsymbol{B}}(\widehat{\boldsymbol{\beta}}) = \sum_{m=1}^{\ell}\frac{\widehat{a}_m}{q_m}\sum_{i\in Q_m}\left\{\sum_{k=1}^{\ell}\widehat{a}_k\widehat{\boldsymbol{\psi}}_{i(k)}^{m}(\widehat{\boldsymbol{\beta}}) + \widehat{\boldsymbol{\phi}}_i(\widehat{\boldsymbol{\beta}})\right\}^{\otimes 2}$$

Finally, we replace the unknown quantities of $\widehat{\psi}_{i(k)}^{m}(t)$ with their empirical estimators:

$$
\begin{aligned}
\widehat{\boldsymbol{\psi}}^{m}_{i(k)}(\widehat{\boldsymbol{\beta}}_{C-RRC}) = & \int_0^M \Bigg\{ \int_0^M \frac{1}{q_k} \sum_{j\in Q_k} \Big[\frac{-w_j(t) I(t \geq u > X_j)}{\hat{\pi}(u)} \Big\{ \widehat{\boldsymbol{Z}}_j(t) - \frac{\widehat{\boldsymbol{S}}^{(1)}(\widehat{\boldsymbol{\beta}},t)}{\widehat{S}^{(0)}(\widehat{\boldsymbol{\beta}},t)} \Big\} dN_j(t) \Big] \\
& + \int_0^M \frac{1}{q_k} \sum_{j\in Q_k} \Big[\frac{-Y_j(t) w_j(t) exp\{\widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{Z}}_j(t)\} I(t \geq u > X_j)}{\widehat{G}(t)\widehat{S}^{(0)}(\widehat{\boldsymbol{\beta}},t)\widehat{\pi}(u)} \Big\{ \widehat{\boldsymbol{Z}}_j(t) - \frac{\widehat{\boldsymbol{S}}^{(1)}(\widehat{\boldsymbol{\beta}},t)}{\widehat{S}^{(0)}(\widehat{\boldsymbol{\beta}},t)} \Big\} \Big] \\
& \times \frac{\sum_j^n w_j(t) dN_j(t)}{n} \Bigg\} d\hat{M}_i^c(u) \\
& + \int_0^M \widehat{\boldsymbol{D}}^{k_i}_{i(k)}(t) \frac{\sum_{j\in Q_k} \overline{\boldsymbol{W}}_j w_j(t) dN_j(t)}{q_k} \\
& + \int_0^M \widehat{\boldsymbol{D}}^{k_i}_{i(k)}(t) \frac{1}{q_k} \sum_{j\in Q_k} \Bigg[\frac{Y_j(t) w_j(t) exp\{\widehat{\boldsymbol{\beta}}' \widehat{\boldsymbol{Z}}_j(t)\}}{\widehat{G}(t)\widehat{S}^{(0)}(\widehat{\boldsymbol{\beta}},t)} \Big\{ \Big[\widehat{\boldsymbol{Z}}_j(t) - \frac{\widehat{\boldsymbol{S}}^{(1)}(\widehat{\boldsymbol{\beta}},t)}{\widehat{S}^{(0)}(\widehat{\boldsymbol{\beta}},t)} \Big] \\
& \times \widehat{\boldsymbol{\beta}}' \overline{\boldsymbol{W}}_j + \overline{\boldsymbol{W}}_j \Big\} \Bigg] \frac{\sum_{j=1}^n w_j(t) dN_j(t)}{n} \\
& + \int_0^M \Bigg\{ \widehat{\boldsymbol{D}}^{k_i}_{i(k)}(t) \frac{\sum_{j\in Q_k} \overline{\boldsymbol{W}}_j w_j(t) I(t \geq u > X_j) dN_j(t)}{q_k \widehat{\pi}(u)} \\
& + \int_0^M \widehat{\boldsymbol{D}}^{k_i}_{i(k)}(t) \frac{1}{q_k} \sum_{j=1}^n \Bigg[\frac{-Y_j(t) exp\{\widehat{\boldsymbol{\beta}}' \widehat{\boldsymbol{Z}}_j(t)\} w_j I(t \geq u > X_j)}{\widehat{G}(t)\widehat{S}^{(0)}(\widehat{\boldsymbol{\beta}},t)\widehat{\pi}(u)} \Big\{ \Big[\widehat{\boldsymbol{Z}}_j(t) \\
& - \frac{\widehat{\boldsymbol{S}}^{(1)}(\widehat{\boldsymbol{\beta}},t)}{\widehat{S}^{(0)}(\widehat{\boldsymbol{\beta}},t)} \Big] \widehat{\boldsymbol{\beta}}' \overline{\boldsymbol{W}}_j + \overline{\boldsymbol{W}}_j \Big\} \Bigg] \frac{\sum_{j=1}^n w_i(t) dN_i(t)}{n} \Bigg\} \frac{\sum_{j=1}^n d\widehat{M}_j^c(u)}{n}
\end{aligned}
$$

$$
\begin{aligned}
\widehat{\boldsymbol{\phi}}_i(\widehat{\boldsymbol{\beta}}) = & \int_0^M \Big[\widehat{\boldsymbol{Z}}_i(t) - \frac{\widehat{\boldsymbol{S}}^{(1)}(\widehat{\boldsymbol{\beta}},t)}{\widehat{S}^{(0)}(\widehat{\boldsymbol{\beta}},t)} \Big] dR_i(t) \\
& - \int_0^M \frac{Y_i(t) w_i(t) exp\{\widehat{\boldsymbol{\beta}}' \widehat{\boldsymbol{Z}}_i(t)\}}{\widehat{G}(t)\widehat{S}^{(0)}(\widehat{\boldsymbol{\beta}},t)} \Big[\widehat{\boldsymbol{Z}}_i(t) - \frac{\widehat{\boldsymbol{S}}^{(1)}(\widehat{\boldsymbol{\beta}},t)}{\widehat{S}^{(0)}(\widehat{\boldsymbol{\beta}},t)} \Big] \frac{\sum_{j=1}^n w_j(t) dN_j(t)}{n}
\end{aligned}
$$

where $\widehat{M}_j^c(u) = I(X_j \leq u, C_j < T_j) - \int_0^u I(X_j \geq t) d\widehat{\Lambda}^c(t)$, $\widehat{\pi}(u) = n^{-1} Y_n(u)$, and $\widehat{\boldsymbol{D}}^{k_i}_{i(k)}(t)$ is defined as in Xie, Wang, and Prentice:

$$
\begin{aligned}
\widehat{\boldsymbol{D}}^{k_i}_{i(k)}(t) =& \widehat{\boldsymbol{\Gamma}}_1(t)\Bigg(\widehat{\boldsymbol{\Delta}}j^{-1}\widehat{\boldsymbol{\Sigma}}^{-1}(t)\Bigg[Y_i(t)\{\overline{\boldsymbol{W}}_i - \widehat{\boldsymbol{\mu}}(t)\}^{\otimes 2} - \widehat{\boldsymbol{\Delta}}_i\widehat{E}(Y_1(t))\sum_{m=1}^{\ell} a_m/m - \widehat{\boldsymbol{\Delta}}Y_i(t)/k_i \\
&- Y_i(t)\widehat{\boldsymbol{\Sigma}}(t)\Bigg] - \widehat{\boldsymbol{\Delta}}_i j^{-1}\widehat{E}(Y_1(t))\Bigg)\widehat{\boldsymbol{\Gamma}}_2(t) \\
\widehat{\boldsymbol{\Delta}}_i =& \left\{\sum_{m=1}^{k_i}(\overline{\boldsymbol{W}}_{im} - \overline{bW}_i)^{\otimes 2} - (k_i - 1)\widehat{\boldsymbol{\Delta}}\right\}\Bigg/\sum_{m=1}^{\ell} a_m(m-1) \\
\widehat{\boldsymbol{\Gamma}}_1(t) =& \widehat{\boldsymbol{\Sigma}}(t)\widehat{\boldsymbol{\Sigma}}(t) + \widehat{\boldsymbol{\Delta}}j^{-1^{-1}}/\widehat{E}(Y_1(t)) \\
\widehat{\boldsymbol{\Gamma}}_2(t) =& \widehat{\boldsymbol{\Sigma}}(t) + \widehat{\boldsymbol{\Delta}}j^{-1^{-1}} \\
E(Y_1(t)) =& \frac{\sum_{k=1}^{n} Y_k(t)}{n}
\end{aligned}
$$

The asymptotic variance described in simulations is estimated by the above expressions.

A.5. Additional Derivations

The following section details the derivations for quantities described in Xie, Wang, and Prentice (2001) and used in the computation of $\widehat{\boldsymbol{\beta}}_{C-RRC}$.

$$
\begin{aligned}
\sqrt{n}(\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta}) =& \sqrt{n}\Bigg(\frac{1}{\sum_{i=1}^{n}(k_i - 1)}\sum_{i=1}^{n}\sum_{j=1}^{k_i}(\boldsymbol{W}_{ij} - \overline{\boldsymbol{W}}_i)(\boldsymbol{W}_{ij} - \overline{\boldsymbol{W}}_i)^T - \boldsymbol{\Delta}\Bigg) \\
=& \sqrt{n}\Bigg(\frac{1}{\frac{1}{n}\sum_{i=1}^{n}(k_i - 1)}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}(\boldsymbol{W}_{ij} - \overline{\boldsymbol{W}}_i)(\boldsymbol{W}_{ij} - \overline{\boldsymbol{W}}_i)^T - \boldsymbol{\Delta}\Bigg) \\
=& \sqrt{n}\Bigg(\frac{1}{\frac{1}{n}\sum_{i=1}^{n}(k_i - 1)}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}(\boldsymbol{W}_{ij} - \overline{\boldsymbol{W}}_i)(\boldsymbol{W}_{ij} - \overline{\boldsymbol{W}}_i)^T - \boldsymbol{\Delta} \\
&\pm \frac{1}{\sum_{m=1}^{\ell} a_m(m-1)}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}(\boldsymbol{W}_{ij} - \overline{\boldsymbol{W}}_i)(\boldsymbol{W}_{ij} - \overline{\boldsymbol{W}}_i)^T\Bigg) \\
=& \sqrt{n}(\boldsymbol{\nu}_1 - \boldsymbol{\nu}_2 + \boldsymbol{\nu}_3 - \boldsymbol{\nu}_4)
\end{aligned}
$$

where

$$\boldsymbol{\nu}_1 = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}(k_i-1)}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)^T,$$

$$\boldsymbol{\nu}_2 = \frac{1}{\sum_{m=1}^{\ell}a_m(m-1)}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)^T,$$

$\boldsymbol{\nu}_3 = \boldsymbol{\nu}_2,$ and $\boldsymbol{\nu}_4 = \boldsymbol{\Delta}$

$$\begin{aligned}
\sqrt{n}(\boldsymbol{\nu}_3-\boldsymbol{\nu}_4) =& \sqrt{n}\Bigg(\frac{1}{\sum_{m=1}^{\ell}a_m(m-1)}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)^T-\boldsymbol{\Delta}\Bigg)\\
=&\sqrt{n}\Bigg(\frac{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)^T-\boldsymbol{\Delta}\sum_{m=1}^{\ell}a_m(m-1)}{\sum_{m=1}^{\ell}a_m(m-1)}\Bigg)\\
=&\sqrt{n}\Bigg(\frac{\frac{1}{n}\sum_{m=1}^{\ell}\sum_{i\in Q_m}\sum_{j=1}^{m}(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)^T-\boldsymbol{\Delta}\sum_{m=1}^{\ell}a_m(m-1)}{\sum_{m=1}^{\ell}a_m(m-1)}\Bigg)\\
=&\sqrt{n}\Bigg(\frac{\sum_{m=1}^{\ell}\left[\frac{\sqrt{q_m}}{n}\frac{1}{\sqrt{q_m}}\sum_{i\in Q_m}\sum_{j=1}^{m}(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)^T-\boldsymbol{\Delta}a_m(m-1)\right]}{\sum_{m=1}^{\ell}a_m(m-1)}\Bigg)\\
=&\Bigg(\frac{\sum_{m=1}^{\ell}\left[\frac{\sqrt{q_m}}{\sqrt{n}}\frac{1}{\sqrt{q_m}}\sum_{i\in Q_m}\sum_{j=1}^{m}(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)^T-\sqrt{n}\boldsymbol{\Delta}a_m(m-1)\right]}{\sum_{m=1}^{\ell}a_m(m-1)}\Bigg)\\
=&\Bigg(\frac{\sum_{m=1}^{\ell}\left[\frac{\sqrt{q_m}}{\sqrt{n}}\frac{1}{\sqrt{q_m}}\sum_{i\in Q_m}\sum_{j=1}^{m}(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)^T-\sqrt{n}\boldsymbol{\Delta}a_m(m-1)\right]}{\sum_{m=1}^{\ell}a_m(m-1)}\\
&\pm\sum_{m=1}^{\ell}\frac{1}{\sqrt{n}}\sum_{i\in Q_m}\boldsymbol{\Delta}(m-1)\Bigg)\\
=&\boldsymbol{\gamma}_1+\boldsymbol{\gamma}_2
\end{aligned}$$

where

$$\boldsymbol{\gamma}_1 = \left(\frac{\sum_{m=1}^{\ell} \left[\frac{\sqrt{q_m}}{n} \frac{1}{\sqrt{q_m}} \sum_{i \in Q_m} \sum_{j=1}^{m} (\boldsymbol{W}_{ij} - \overline{\boldsymbol{W}}_i)(\boldsymbol{W}_{ij} - \overline{\boldsymbol{W}}_i)^T - \frac{1}{\sqrt{n}} \sum_{i \in Q_m} \boldsymbol{\Delta}(m-1) \right]}{\sum_{m=1}^{\ell} a_m(m-1)} \right),$$

and

$$\begin{aligned}
\boldsymbol{\gamma}_2 &= \left(\frac{\sum_{m=1}^{\ell} \left[\frac{1}{\sqrt{n}} \sum_{i \in Q_m} \boldsymbol{\Delta}(m-1) - \sqrt{n}\boldsymbol{\Delta} a_m(m-1) \right]}{\sum_{m=1}^{\ell} a_m(m-1)} \right) \\
&= \left(\frac{\sum_{m=1}^{\ell} \left[\frac{q_m}{\sqrt{n}} \boldsymbol{\Delta}(m-1) - \sqrt{n}\boldsymbol{\Delta} a_m(m-1) \right]}{\sum_{m=1}^{\ell} a_m(m-1)} \right) \\
&= \left(\frac{\sum_{m=1}^{\ell} \left[\sqrt{n} \frac{q_m}{n} \boldsymbol{\Delta}(m-1) - \sqrt{n}\boldsymbol{\Delta} a_m(m-1) \right]}{\sum_{m=1}^{\ell} a_m(m-1)} \right) \\
&= \left(\frac{\sqrt{n}\boldsymbol{\Delta} \sum_{m=1}^{\ell} \left[\frac{q_m}{n}(m-1) - a_m(m-1) \right]}{\sum_{m=1}^{\ell} a_m(m-1)} \right) \\
&= \left(\frac{\sqrt{n}\boldsymbol{\Delta} \sum_{m=1}^{\ell} (m-1) \left[\frac{q_m}{n} - \frac{1}{n} \sum_{i=1}^{n} a_m \right]}{\sum_{m=1}^{\ell} a_m(m-1)} \right) \\
&= \left(\frac{\boldsymbol{\Delta} \sum_{m=1}^{\ell} (m-1) \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} I(k_i = m) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} a_m \right]}{\sum_{m=1}^{\ell} a_m(m-1)} \right)
\end{aligned}$$

Now, each m^{th} term in $\boldsymbol{\gamma}_2$ is IID over i, because $E[I(k_i = m)] = a_m$.

$$
\begin{aligned}
\sqrt{n}(\boldsymbol{\nu}_1-\boldsymbol{\nu}_2) =& \sqrt{n}\Bigg(\frac{1}{\frac{1}{n}\sum_{i=1}^{n}(k_i-1)}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)^T \\
&- \frac{1}{\sum_{m=1} a_m(m-1)}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)^T\Bigg) \\
=& \sqrt{n}\Bigg(\Bigg[\frac{1}{\frac{1}{n}\sum_{i=1}^{n}(k_i-1)} - \frac{1}{\sum_{m=1}^{\ell} a_m(m-1)}\Bigg]\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)^T\Bigg) \\
=& \sqrt{n}\Bigg(\Bigg[\frac{\sum_{m=1}^{\ell} a_m(m-1) - \frac{1}{n}\sum_{i=1}^{n}(k_i-1)}{\frac{1}{n}\sum_{i=1}^{n}(k_i-1)\sum_{m=1}^{\ell} a_m(m-1)}\Bigg]\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)^T\Bigg) \\
=& \sqrt{n}\Bigg(\Bigg[\frac{\sum_{m=1}^{\ell} a_m(m-1) - \frac{1}{n}\sum_{m=1}^{\ell}\sum_{i\in Q_m}(m-1)}{\frac{1}{n}\sum_{i=1}^{n}(k_i-1)\sum_{m=1}^{\ell} a_m(m-1)}\Bigg] \\
&\times \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)^T\Bigg) \\
=& \sqrt{n}\Bigg(\Bigg[\frac{\sum_{m=1}^{\ell} a_m(m-1) - \frac{1}{n}\sum_{m=1}^{\ell} q_m(m-1)}{\frac{1}{n}\sum_{i=1}^{n}(k_i-1)\sum_{m=1}^{\ell} a_m(m-1)}\Bigg] \\
&\times \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)^T\Bigg) \\
=& \sqrt{n}\Bigg(\Bigg[\frac{\sum_{m=1}^{\ell} a_m(m-1) - \sum_{m=1}^{\ell}\frac{(m-1)}{n}\sum_{i=1}^{n} I(k_i=m)}{\frac{1}{n}\sum_{i=1}^{n}(k_i-1)\sum_{m=1}^{\ell} a_m(m-1)}\Bigg] \\
&\times \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)^T\Bigg) \\
=& \sqrt{n}\Bigg(\frac{\frac{1}{n}\sum_{i=1}^{n}\sum_{m=1}^{\ell}\Big[(m-1)(a_m - I(k_i=m))\Big]}{\frac{1}{n}\sum_{i=1}^{n}(k_i-1)\sum_{m=1}^{\ell} a_m(m-1)} \\
&\times \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)^T\Bigg)
\end{aligned}
$$

Note that because $\frac{q_m}{n} \xrightarrow{p} a_m$,

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}(k_i-1) &= \sum_{i=1}^{n}\left(\frac{k_i}{n}-\frac{1}{n}\right) \\
&= \sum_{m=1}^{\ell}\sum_{i\in Q_m}\left(\frac{q_m}{n}-\frac{1}{n}\right) \\
&= \sum_{m=1}^{\ell}\frac{q_m}{n}m-1 \\
&= \sum_{m=1}^{\ell} a_m(m-1)+o_p(1).
\end{aligned}
$$

Also, since $\widehat{\boldsymbol{\Delta}} \xrightarrow{p} \boldsymbol{\Delta}$,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)^T &= \widehat{\boldsymbol{\Delta}}\sum_{i=1}^{n}(k_i-1)\\ &= \boldsymbol{\Delta}\sum_{m=1}^{\ell}a_m(m-1)+o_p(1).\end{aligned}$$

Therefore,

$$\begin{aligned}\sqrt{n}(\boldsymbol{\nu}_1-\boldsymbol{\nu}_2) &= \sqrt{n}\left(\frac{\frac{1}{n}\sum_{i=1}^{n}\sum_{m=1}^{\ell}\Big[(m-1)(a_m-I(k_i=m))\Big]}{\big[\sum_{m=1}^{\ell}a_m(m-1)\big]^2}\boldsymbol{\Delta}\sum_{m=1}^{\ell}a_m(m-1)\right)+o_p(1)\\ &= \left(\frac{\boldsymbol{\Delta}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{m=1}^{\ell}(m-1)\big[a_m-I(k_i=m)\big]}{\sum_{m=1}^{\ell}a_m(m-1)}\right)+o_p(1)\\ &= -\boldsymbol{\gamma}_2+o_p(1).\end{aligned}$$

Thus, $\sqrt{n}(\widehat{\boldsymbol{\Delta}}-\boldsymbol{\Delta})=\boldsymbol{\gamma}_1$, where $\boldsymbol{\gamma}_1$ can also be written as

$$\left(\frac{\sum_{i=1}^{n}\frac{1}{\sqrt{n}}\big[\sum_{j=1}^{k_i}(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)(\boldsymbol{W}_{ij}-\overline{\boldsymbol{W}}_i)^T-\boldsymbol{\Delta}(k_i-1)\big]}{\sum_{m=1}^{\ell}a_m(m-1)}\right)$$

$$
\begin{aligned}
\sqrt{n}\Big(\widehat{\boldsymbol{\Sigma}}(t)-\boldsymbol{\Sigma}(t)\Big)=&\sqrt{n}\bigg(\frac{1}{\sum_{i=1}^n Y_i(t)-1}\sum_{i=1}^n Y_i(t)(\overline{\boldsymbol{W}}_i-\widehat{\boldsymbol{\mu}}(t))(\overline{\boldsymbol{W}}_i-\widehat{\boldsymbol{\mu}}(t))^T\\
&-\frac{\sum_{i=1}^n Y_i(t)\widehat{\boldsymbol{\Delta}}/k_i}{\sum_{i=1}^n Y_i(t)}-\boldsymbol{\Sigma}(t)\bigg)\\
=&\sqrt{n}\bigg(\frac{1}{\sum_{i=1}^n Y_i(t)-1}\sum_{i=1}^n Y_i(t)(\overline{\boldsymbol{W}}_i-\boldsymbol{\mu}(t))(\overline{\boldsymbol{W}}_i-\boldsymbol{\mu}(t))^T\\
&-\frac{\sum_{i=1}^n Y_i(t)\widehat{\boldsymbol{\Delta}}/k_i}{\sum_{i=1}^n Y_i(t)}-\boldsymbol{\Sigma}(t)\bigg)+o_p(1)\\
=&\sqrt{n}\bigg(\frac{\sum_{i=1}^n Y_i(t)}{\big[\sum_{i=1}^n Y_i(t)-1\big]\sum_{i=1}^n Y_i(t)}\sum_{i=1}^n Y_i(t)(\overline{\boldsymbol{W}}_i-\boldsymbol{\mu}(t))(\overline{\boldsymbol{W}}_i-\boldsymbol{\mu}(t))^T\\
&-\frac{\sum_{i=1}^n Y_i(t)\widehat{\boldsymbol{\Delta}}/k_i}{\sum_{i=1}^n Y_i(t)}-\boldsymbol{\Sigma}(t)\frac{\sum_{i=1}^n Y_i(t)}{\sum_{i=1}^n Y_i(t)}\bigg)+o_p(1)\\
=&\frac{\sqrt{n}}{\frac{1}{n}\sum_{i=1}^n Y_i(t)}\bigg(\frac{\frac{1}{n}\sum_{i=1}^n Y_i(t)}{\frac{1}{n}\sum_{i=1}^n Y_i(t)-\frac{1}{n}}\frac{1}{n}\sum_{i=1}^n Y_i(t)(\overline{\boldsymbol{W}}_i-\boldsymbol{\mu}(t))(\overline{\boldsymbol{W}}_i-\boldsymbol{\mu}(t))^T\\
&-\frac{\widehat{\boldsymbol{\Delta}}}{n}\sum_{i=1}^n\frac{Y_i(t)}{k_i}-\boldsymbol{\Sigma}(t)\frac{1}{n}\sum_{i=1}^n Y_i(t)\bigg)+o_p(1)\\
=&\frac{\sqrt{n}}{E(Y_1(t))}\bigg(\frac{\frac{1}{n}\sum_{i=1}^n Y_i(t)}{\frac{1}{n}\sum_{i=1}^n Y_i(t)-\frac{1}{n}}\frac{1}{n}\sum_{i=1}^n Y_i(t)(\overline{\boldsymbol{W}}_i-\boldsymbol{\mu}(t))(\overline{\boldsymbol{W}}_i-\boldsymbol{\mu}(t))^T\\
&-\frac{\widehat{\boldsymbol{\Delta}}}{n}\sum_{i=1}^n\frac{Y_i(t)}{k_i}-\boldsymbol{\Sigma}(t)\frac{1}{n}\sum_{i=1}^n Y_i(t)\bigg)+o_p(1)\\
=&\frac{\sqrt{n}}{E(Y_1(t))}\bigg(\frac{\frac{1}{n}\sum_{i=1}^n Y_i(t)}{\frac{1}{n}\sum_{i=1}^n Y_i(t)-\frac{1}{n}}\frac{1}{n}\sum_{i=1}^n Y_i(t)(\overline{\boldsymbol{W}}_i-\boldsymbol{\mu}(t))(\overline{\boldsymbol{W}}_i-\boldsymbol{\mu}(t))^T\\
&-\frac{\widehat{\boldsymbol{\Delta}}}{n}\sum_{i=1}^n\frac{Y_i(t)}{k_i}-\boldsymbol{\Sigma}(t)\frac{1}{n}\sum_{i=1}^n Y_i(t)\pm\frac{\boldsymbol{\Delta}}{n}\sum_{i=1}^n Y_i(t)\bigg)+o_p(1)\\
=&\frac{\sqrt{n}}{E(Y_1(t))}\bigg(\frac{\frac{1}{n}\sum_{i=1}^n Y_i(t)}{\frac{1}{n}\sum_{i=1}^n Y_i(t)-\frac{1}{n}}\frac{1}{n}\sum_{i=1}^n Y_i(t)(\overline{\boldsymbol{W}}_i-\boldsymbol{\mu}(t))(\overline{\boldsymbol{W}}_i-\boldsymbol{\mu}(t))^T\\
&-\frac{1}{n}(\widehat{\boldsymbol{\Delta}}-\boldsymbol{\Delta})\sum_{i=1}^n\frac{Y_i(t)}{k_i}-\boldsymbol{\Sigma}(t)\frac{1}{n}\sum_{i=1}^n Y_i(t)+\frac{\boldsymbol{\Delta}}{n}\sum_{i=1}^n Y_i(t)\bigg)+o_p(1)\\
=&\frac{\sqrt{n}}{E(Y_1(t))}\bigg(\frac{\frac{1}{n}\sum_{i=1}^n Y_i(t)}{\frac{1}{n}\sum_{i=1}^n Y_i(t)-\frac{1}{n}}\frac{1}{n}\sum_{i=1}^n Y_i(t)(\overline{\boldsymbol{W}}_i-\boldsymbol{\mu}(t))(\overline{\boldsymbol{W}}_i-\boldsymbol{\mu}(t))^T\\
&-\frac{1}{n}\frac{\boldsymbol{\gamma}_1}{\sqrt{n}}\sum_{i=1}^n\frac{Y_i(t)}{k_i}-\boldsymbol{\Sigma}(t)\frac{1}{n}\sum_{i=1}^n Y_i(t)+\frac{\boldsymbol{\Delta}}{n}\sum_{i=1}^n Y_i(t)\bigg)+o_p(1),
\end{aligned}
$$

$$
\begin{aligned}
\sqrt{n}(\widehat{\boldsymbol{\tau}}_j(t)-\boldsymbol{\tau}_j(t)) &= \sqrt{n}\Big(\widehat{\boldsymbol{\Sigma}}(t)\Big[\widehat{\boldsymbol{\Sigma}}(t)+\frac{\widehat{\boldsymbol{\Delta}}}{j}\Big]^{-1}-\boldsymbol{\Sigma}(t)\Big[\boldsymbol{\Sigma}(t)+\frac{\boldsymbol{\Delta}}{j}\Big]^{-1}\Big)\\
&= \sqrt{n}\Big(\widehat{\boldsymbol{\Sigma}}(t)-\boldsymbol{\Sigma}(t)\Big[\boldsymbol{\Sigma}(t)+\frac{\boldsymbol{\Delta}}{j}\Big]^{-1}\Big[\widehat{\boldsymbol{\Sigma}}(t)+\frac{\widehat{\boldsymbol{\Delta}}}{j}\Big]\Big)\Big[\widehat{\boldsymbol{\Sigma}}(t)+\frac{\widehat{\boldsymbol{\Delta}}}{j}\Big]^{-1}\\
&= \sqrt{n}\Big(\widehat{\boldsymbol{\Sigma}}(t)-\boldsymbol{\Sigma}(t)\Big[\boldsymbol{\Sigma}(t)+\frac{\boldsymbol{\Delta}}{j}\Big]^{-1}\boldsymbol{\Sigma}(t)^{-1}\boldsymbol{\Sigma}(t)\Big[\widehat{\boldsymbol{\Sigma}}(t)+\frac{\widehat{\boldsymbol{\Delta}}}{j}\Big]\Big)\Big[\widehat{\boldsymbol{\Sigma}}(t)+\frac{\widehat{\boldsymbol{\Delta}}}{j}\Big]^{-1}\\
&= \sqrt{n}\Big(\Big\{\boldsymbol{\Sigma}(t)\Big[\boldsymbol{\Sigma}(t)+\frac{\boldsymbol{\Delta}}{j}\Big]^{-1}\boldsymbol{\Sigma}(t)^{-1}\Big\}\Big\{\boldsymbol{\Sigma}(t)\Big[\boldsymbol{\Sigma}(t)+\frac{\boldsymbol{\Delta}}{j}\Big]^{-1}\boldsymbol{\Sigma}(t)^{-1}\Big\}^{-1}\widehat{\boldsymbol{\Sigma}}(t)-\\
&\qquad \boldsymbol{\Sigma}(t)\Big[\boldsymbol{\Sigma}(t)+\frac{\boldsymbol{\Delta}}{j}\Big]^{-1}\boldsymbol{\Sigma}(t)^{-1}\boldsymbol{\Sigma}(t)\Big[\widehat{\boldsymbol{\Sigma}}(t)+\frac{\widehat{\boldsymbol{\Delta}}}{j}\Big]\Big)\Big[\widehat{\boldsymbol{\Sigma}}(t)+\frac{\widehat{\boldsymbol{\Delta}}}{j}\Big]^{-1}\\
&= \sqrt{n}\Big\{\boldsymbol{\Sigma}(t)\Big[\boldsymbol{\Sigma}(t)+\frac{\boldsymbol{\Delta}}{j}\Big]^{-1}\boldsymbol{\Sigma}(t)^{-1}\Big\}\Big(\Big\{\boldsymbol{\Sigma}(t)\Big[\boldsymbol{\Sigma}(t)+\frac{\boldsymbol{\Delta}}{j}\Big]^{-1}\boldsymbol{\Sigma}(t)^{-1}\Big\}^{-1}\widehat{\boldsymbol{\Sigma}}(t)-\\
&\qquad \boldsymbol{\Sigma}(t)\Big[\widehat{\boldsymbol{\Sigma}}(t)+\frac{\widehat{\boldsymbol{\Delta}}}{j}\Big]\Big)\Big[\widehat{\boldsymbol{\Sigma}}(t)+\frac{\widehat{\boldsymbol{\Delta}}}{j}\Big]^{-1}\\
&= \sqrt{n}\Big\{\boldsymbol{\Sigma}(t)\Big[\boldsymbol{\Sigma}(t)+\frac{\boldsymbol{\Delta}}{j}\Big]^{-1}\boldsymbol{\Sigma}(t)^{-1}\Big\}\Big(\Big\{\boldsymbol{\Sigma}(t)\Big[\boldsymbol{\Sigma}(t)+\frac{\boldsymbol{\Delta}}{j}\Big]\boldsymbol{\Sigma}(t)^{-1}\Big\}\widehat{\boldsymbol{\Sigma}}(t)-\\
&\qquad \boldsymbol{\Sigma}(t)\Big[\widehat{\boldsymbol{\Sigma}}(t)+\frac{\widehat{\boldsymbol{\Delta}}}{j}\Big]\Big)\Big[\widehat{\boldsymbol{\Sigma}}(t)+\frac{\widehat{\boldsymbol{\Delta}}}{j}\Big]^{-1}\\
&= \sqrt{n}\Big\{\boldsymbol{\Sigma}(t)\Big[\boldsymbol{\Sigma}(t)+\frac{\boldsymbol{\Delta}}{j}\Big]^{-1}\boldsymbol{\Sigma}(t)^{-1}\Big\}\Big(\Big\{\Big[\boldsymbol{\Sigma}(t)\boldsymbol{\Sigma}(t)+\boldsymbol{\Sigma}(t)\frac{\boldsymbol{\Delta}}{j}\Big]\boldsymbol{\Sigma}(t)^{-1}\Big\}\widehat{\boldsymbol{\Sigma}}(t)-\\
&\qquad \boldsymbol{\Sigma}(t)\Big[\widehat{\boldsymbol{\Sigma}}(t)+\frac{\widehat{\boldsymbol{\Delta}}}{j}\Big]\Big)\Big[\widehat{\boldsymbol{\Sigma}}(t)+\frac{\widehat{\boldsymbol{\Delta}}}{j}\Big]^{-1}\\
&= \sqrt{n}\Big\{\boldsymbol{\Sigma}(t)\Big[\boldsymbol{\Sigma}(t)+\frac{\boldsymbol{\Delta}}{j}\Big]^{-1}\boldsymbol{\Sigma}(t)^{-1}\Big\}\Big(\Big[\boldsymbol{\Sigma}(t)+\boldsymbol{\Sigma}(t)\frac{\boldsymbol{\Delta}}{j}\boldsymbol{\Sigma}(t)^{-1}\Big]\widehat{\boldsymbol{\Sigma}}(t)-\\
&\qquad \boldsymbol{\Sigma}(t)\Big[\widehat{\boldsymbol{\Sigma}}(t)+\frac{\widehat{\boldsymbol{\Delta}}}{j}\Big]\Big)\Big[\widehat{\boldsymbol{\Sigma}}(t)+\frac{\widehat{\boldsymbol{\Delta}}}{j}\Big]^{-1}\\
&= \sqrt{n}\Big\{\boldsymbol{\Sigma}(t)\Big[\boldsymbol{\Sigma}(t)+\frac{\boldsymbol{\Delta}}{j}\Big]^{-1}\boldsymbol{\Sigma}(t)^{-1}\Big\}\Big(\boldsymbol{\Sigma}(t)\widehat{\boldsymbol{\Sigma}}(t)+\boldsymbol{\Sigma}(t)\frac{\boldsymbol{\Delta}}{j}\boldsymbol{\Sigma}(t)^{-1}\widehat{\boldsymbol{\Sigma}}(t)-\\
&\qquad \Big[\boldsymbol{\Sigma}(t)\widehat{\boldsymbol{\Sigma}}(t)+\boldsymbol{\Sigma}(t)\frac{\widehat{\boldsymbol{\Delta}}}{j}\Big]\Big)\Big[\widehat{\boldsymbol{\Sigma}}(t)+\frac{\widehat{\boldsymbol{\Delta}}}{j}\Big]^{-1}\\
&= \sqrt{n}\Big\{\boldsymbol{\Sigma}(t)\Big[\boldsymbol{\Sigma}(t)+\frac{\boldsymbol{\Delta}}{j}\Big]^{-1}\boldsymbol{\Sigma}(t)^{-1}\Big\}\Big(\boldsymbol{\Sigma}(t)\frac{\boldsymbol{\Delta}}{j}\boldsymbol{\Sigma}(t)^{-1}\widehat{\boldsymbol{\Sigma}}(t)-\boldsymbol{\Sigma}(t)\frac{\widehat{\boldsymbol{\Delta}}}{j}\Big)\\
&\qquad \times\Big[\widehat{\boldsymbol{\Sigma}}(t)+\frac{\widehat{\boldsymbol{\Delta}}}{j}\Big]^{-1}
\end{aligned}
$$

$$
\begin{aligned}
\sqrt{n}\bigg(\boldsymbol{\Sigma}(t)\frac{\boldsymbol{\Delta}}{j}\boldsymbol{\Sigma}(t)^{-1}\widehat{\boldsymbol{\Sigma}}(t)-\boldsymbol{\Sigma}(t)\frac{\widehat{\boldsymbol{\Delta}}}{j}\bigg) =&\sqrt{n}\bigg(\boldsymbol{\Sigma}(t)\frac{\boldsymbol{\Delta}}{j}\boldsymbol{\Sigma}(t)^{-1}\widehat{\boldsymbol{\Sigma}}+\boldsymbol{\Sigma}(t)\frac{\widehat{\boldsymbol{\Delta}}}{j}\pm\boldsymbol{\Sigma}(t)\frac{\boldsymbol{\Delta}}{j}\bigg)\\
=&\sqrt{n}\bigg(\boldsymbol{\Sigma}(t)\frac{\boldsymbol{\Delta}}{j}\boldsymbol{\Sigma}(t)^{-1}\widehat{\boldsymbol{\Sigma}}(t)-\boldsymbol{\Sigma}(t)\frac{\widehat{\boldsymbol{\Delta}}}{j}+\boldsymbol{\Sigma}(t)\frac{\boldsymbol{\Delta}}{j}\\
&-\boldsymbol{\Sigma}(t)\frac{\boldsymbol{\Delta}}{j}\boldsymbol{\Sigma}(t)^{-1}\boldsymbol{\Sigma}(t)\bigg)\\
=&\sqrt{n}\bigg(\boldsymbol{\Sigma}(t)\frac{\boldsymbol{\Delta}}{j}\boldsymbol{\Sigma}(t)^{-1}\big(\widehat{\boldsymbol{\Sigma}}(t)-\boldsymbol{\Sigma}(t)\big)+\boldsymbol{\Sigma}(t)\big(\boldsymbol{\Delta}-\widehat{\boldsymbol{\Delta}}\big)\bigg)
\end{aligned}
$$

Substituting $(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})$ as well as $\frac{\gamma_1}{\sqrt{n}}$ for $(\widehat{\boldsymbol{\Delta}}-\boldsymbol{\Delta})$ gives the result $\sqrt{n}(\widehat{\boldsymbol{\tau}}_j(t)-\boldsymbol{\tau}_j(t))=\boldsymbol{D}_{i(j)}^{k_i}(t)+o_p(1)$.

APPENDIX B

APPENDICES FOR CHAPTER 3

B.1. Additional Simulations

To support our claim that our proposed method will not perform worse than the complete case estimator even if a poor auxiliary outcome is chosen, we assessed these methods using an auxiliary outcome that is uncorrelated with the true outcome. Using the same simulation setup as Section 3.3, Tables B.1 and B.2 show that both $\widehat{\beta}_{HS}$ and $\widehat{\beta}_{VS}$ perform similarly to $\widehat{\beta}_{CC}$ in terms of both bias and efficiency when the auxiliary outcome is useless.

We also evaluated the proposed method when the auxiliary outcome is perfectly correlated with the true outcome. In Tables B.1 and B.2, the efficiency is close to optimal level for $\widehat{\beta}_{HS}$ and $\widehat{\beta}_{VS}$ when the correlation between $\boldsymbol{Y}$ and $\boldsymbol{S}$ is 1.0. As explained in Section 3.2 of the article and proved in the following appendix of this supplementary material, efficiency greater than 1 for these estimators is due to the use of a bandwidth matrix in the kernel density estimation. As the sample size approaches ∞, the efficiency will approach optimality.

We also compared the performance of our proposed method to the gold standard of multiple imputation. In Section 3.1, we argue that multiple imputation is subject to misspecification of the imputation model. To test this, we generated the auxiliary outcome with a quadratic relationship to the true outcome; we then used the auxiliary outcome, timepoints, and a continuous baseline covariate to generate imputed values for the true outcome. We accomplished this using the algorithm for multiple imputation of a linear mixed-effects model from the `mice` package version 3.6.0 in R. The imputation model incorrectly assumes a linear relationship for this model, and as indicated in Tables B.3 and B.4, this has an impact on the bias. We also compared these results to predictive mean matching (PMM), which draws from a suitable set of "donors" from the observed outcomes. We used two PMM estimators, one with 3 donors and one with 10 donors (Schenker and Taylor, 1996; Schenker and Taylor, 1996), and found that both versions are more robust to misspecification than the MI estimator. Table B.4 shows that PMM performs well in estimating the coefficient for the baseline covariate, with both bias and standard error remaining low in all situations. However, in Table B.3 we can see that PMM loses some efficiency in estimating the coefficient for timepoint.

Table B.1: Results for timepoint coefficient from 500 simulations with a useless auxiliary outcome and a perfect auxiliary outcome. SD = standard deviation; $\widehat{SE}$ = estimated standard error; MSE = mean squared error; RE = relative efficiency. $\widehat{\beta}_{HS}$ = estimator based on both hybrid and validation subsample; $\widehat{\beta}_{VS}$ = estimator based on internal validation sample; $\widehat{\beta}_{CC}$ = complete-case estimator.

Cor($\boldsymbol{Y}, \boldsymbol{S}$)	$(\rho_V, \rho_{\underline{V}}, \rho_{\overline{V}})$	Estimator	Bias (% Bias)	SD	$\widehat{\text{SE}}$	MSE	RE
0.0	(0.8, 0.1, 0.1)	$\widehat{\beta}_{HS}$	-0.004 (0.360)	0.083	0.090	0.007	1.117
		$\widehat{\beta}_{VS}$	-0.004 (0.360)	0.083	0.090	0.007	1.117
		$\widehat{\beta}_{CC}$	-0.004 (0.361)	0.083	0.090	0.007	1.116
	(0.6, 0.2, 0.2)	$\widehat{\beta}_{HS}$	-0.007 (0.694)	0.097	0.104	0.009	1.285
		$\widehat{\beta}_{VS}$	-0.007 (0.695)	0.097	0.104	0.009	1.285
		$\widehat{\beta}_{CC}$	-0.007 (0.694)	0.097	0.104	0.009	1.283
	(0.5, 0.25, 0.25)	$\widehat{\beta}_{HS}$	-0.006 (0.649)	0.111	0.114	0.012	1.405
		$\widehat{\beta}_{VS}$	-0.007 (0.652)	0.111	0.114	0.012	1.404
		$\widehat{\beta}_{CC}$	-0.007 (0.652)	0.111	0.114	0.012	1.402
	(0.3, 0.35, 0.35)	$\widehat{\beta}_{HS}$	-0.001 (0.087)	0.146	0.146	0.021	1.803
		$\widehat{\beta}_{VS}$	-0.001 (0.129)	0.146	0.146	0.021	1.797
		$\widehat{\beta}_{CC}$	-0.001 (0.125)	0.147	0.146	0.021	1.799
1.0	(0.8, 0.1, 0.1)	$\widehat{\beta}_{HS}$	-0.006 (0.589)	0.077	0.082	0.006	1.011
		$\widehat{\beta}_{VS}$	-0.006 (0.581)	0.076	0.082	0.006	1.010
		$\widehat{\beta}_{CC}$	-0.004 (0.361)	0.083	0.090	0.007	1.116
	(0.6, 0.2, 0.2)	$\widehat{\beta}_{HS}$	-0.006 (0.604)	0.079	0.083	0.006	1.020
		$\widehat{\beta}_{VS}$	-0.006 (0.634)	0.079	0.093	0.006	1.019
		$\widehat{\beta}_{CC}$	-0.006 (0.694)	0.097	0.104	0.009	1.283
	(0.5, 0.25, 0.25)	$\widehat{\beta}_{HS}$	-0.008 (0.757)	0.079	0.083	0.006	1.025
		$\widehat{\beta}_{VS}$	-0.008 (0.773)	0.079	0.083	0.006	1.025
		$\widehat{\beta}_{CC}$	-0.007 (0.652)	0.111	0.114	0.012	1.402
	(0.3, 0.35, 0.35)	$\widehat{\beta}_{HS}$	-0.005 (0.534)	0.082	0.084	0.007	1.036
		$\widehat{\beta}_{VS}$	-0.006 (0.602)	0.082	0.084	0.007	1.036
		$\widehat{\beta}_{CC}$	-0.001 (0.125)	0.147	0.146	0.022	1.799

Table B.2: Results for baseline covariate coefficient from 500 simulations with a useless auxiliary outcome and a perfect auxiliary outcome. SD = standard deviation; $\widehat{SE}$ = estimated standard error; MSE = mean squared error; RE = relative efficiency. $\widehat{\beta}_{HS}$ = estimator based on both hybrid and validation subsample; $\widehat{\beta}_{VS}$ = estimator based on internal validation sample; $\widehat{\beta}_{CC}$ = complete-case estimator.

Cor($\boldsymbol{Y}, \boldsymbol{S}$)	$(\rho_V, \rho_{\underline{V}}, \rho_{\overline{V}})$	Estimator	Bias (% Bias)	SD	$\widehat{\text{SE}}$	MSE	RE
0.0	(0.8, 0.1, 0.1)	$\widehat{\beta}_{HS}$	-0.011 (1.609)	0.215	0.217	0.046	1.067
		$\widehat{\beta}_{VS}$	-0.011 (1.606)	0.216	0.217	0.046	1.067
		$\widehat{\beta}_{CC}$	-0.011 (1.606)	0.216	0.215	0.046	1.054
	(0.6, 0.2, 0.2)	$\widehat{\beta}_{HS}$	-0.008 (1.140)	0.224	0.231	0.050	1.135
		$\widehat{\beta}_{VS}$	-0.008 (1.137)	0.224	0.231	0.050	1.135
		$\widehat{\beta}_{CC}$	-0.008 (1.135)	0.224	0.228	0.050	1.122
	(0.5, 0.25, 0.25)	$\widehat{\beta}_{HS}$	-0.009 (1.323)	0.236	0.240	0.056	1.177
		$\widehat{\beta}_{VS}$	-0.009 (1.316)	0.236	0.240	0.056	1.176
		$\widehat{\beta}_{CC}$	-0.009 (1.318)	0.236	0.237	0.056	1.163
	(0.3, 0.35, 0.35)	$\widehat{\beta}_{HS}$	-0.005 (0.654)	0.250	0.258	0.063	1.268
		$\widehat{\beta}_{VS}$	-0.005 (0.654)	0.250	0.255	0.063	1.267
		$\widehat{\beta}_{CC}$	-0.005 (0.687)	0.250	0.255	0.063	1.252
1.0	(0.8, 0.1, 0.1)	$\widehat{\beta}_{HS}$	-0.012 (1.689)	0.206	0.206	0.043	1.102
		$\widehat{\beta}_{VS}$	-0.012 (1.715)	0.206	0.206	0.043	1.012
		$\widehat{\beta}_{CC}$	-0.011 (1.606)	0.216	0.215	0.047	1.054
	(0.6, 0.2, 0.2)	$\widehat{\beta}_{HS}$	-0.009 (1.219)	0.206	0.207	0.043	1.015
		$\widehat{\beta}_{VS}$	-0.008 (1.177)	0.207	0.043	0.043	1.016
		$\widehat{\beta}_{CC}$	-0.008 (1.135)	0.224	0.228	0.050	1.112
	(0.5, 0.25, 0.25)	$\widehat{\beta}_{HS}$	-0.010 (1.377)	0.208	0.207	0.043	1.016
		$\widehat{\beta}_{VS}$	-0.010 (1.377)	0.208	0.207	0.043	1.016
		$\widehat{\beta}_{CC}$	-0.009 (1.318)	0.236	0.237	0.056	1.163
	(0.3, 0.35, 0.35)	$\widehat{\beta}_{HS}$	-0.006 (0.893)	0.209	0.207	0.044	1.016
		$\widehat{\beta}_{VS}$	-0.006 (0.917)	0.209	0.207	0.044	1.016
		$\widehat{\beta}_{CC}$	-0.005 (0.687)	0.250	0.255	0.063	1.252

This is likely due to the nature of the method; PMM imputes missing outcomes from a pool of observed outcomes, and does so by measuring distance between means conditional on covariates. The method does not account for correlated observations, and this results in efficiency loss in a longitudinal model.

Finally, as explained in Section 3.2, we evaluated the robustness of $\widehat{\beta}_{HS}$ to MAR outcomes in the hybrid set. For $\widehat{\beta}_{VS}$, we require that the validation set be a representative subsample of the population, in order to ensure that kernel density estimation provides a valid probability estimate. However, for $\widehat{\beta}_{HS}$, we also use hybrid set outcomes in the kernel density estimation, and we cannot guarantee that the hybrid set will have MCAR outcomes. Therefore, we simulated MAR data in the hybrid set and assessed the performance of $\widehat{\beta}_{HS}$. With two timepoints, we specified n_V, $n_{\underline{V}}$, and $n_{\overline{V}}$ first, and specified that each subject k miss exactly one outcome. Then, for each subject $k \in \underline{V}$, if $X_k \geq \mu_X$, the mean of the distribution of X, we set Y_{k1} missing with 80% probability. If $X_k < \mu_X$, we specified Y_{k2} missing with 80% probability. The results in Tables B.5 and B.6 affirm that the bias and efficiency are not affected by MAR data in the hybrid set, even when the hybrid set is larger than the validation set.

B.2. Asymptotic Normality of Proposed Estimators

We outline the proof of asymptotic normality here using assumptions similar to Pepe (1992). First, the proportions $n_V/N = \rho_V$, $n_{\underline{V}}/N = \rho_{\underline{V}}$, $n_{\overline{V}}/N = \rho_{\overline{V}}$, and $n_{V_1}/N = \rho_{V_1}$ are all strictly greater than 0 as $N \to \infty$. Second, regularity conditions (Cox and Hinkley, 1974) hold for $P_{\boldsymbol{\beta}}(\boldsymbol{Y}|\boldsymbol{X})$, $P_{\boldsymbol{\beta}}(\boldsymbol{S}, \underline{\boldsymbol{Y}}|\boldsymbol{X})$, and $P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})$. We assume that these probabilities are bounded away from 0 uniformly in a neighborhood of the true $\boldsymbol{\beta}_0$, and that their first and second derivatives are bounded. Finally, for any bandwidth matrix $\boldsymbol{H}$ used to estimate kernel densities in the estimated likelihood, we assume that $N|\boldsymbol{H}| \to \boldsymbol{0}$. For more information on this assumption, see (Simonoff, 1998).

B.2.1. Asymptotic Normality of $\widehat{\boldsymbol{\beta}}_{VS}$

We start by separating the score equation into true and estimated components, with the shorthand $\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{Y}|\boldsymbol{X}) = \frac{\partial}{\partial \boldsymbol{\beta}} P_{\boldsymbol{\beta}}(\boldsymbol{Y}|\boldsymbol{X})$.

Table B.3: Results for timepoint coefficient from 500 simulations comparing proposed estimators with multiple imputation and predictive mean matching estimators. SD = standard deviation; $\widehat{SE}$ = estimated standard error; MSE = mean squared error; RE = relative efficiency. $\widehat{\beta}_{HS}$ = estimator based on both hybrid and validation subsample; $\widehat{\beta}_{VS}$ = estimator based on internal validation sample; $\widehat{\beta}_{PMM3}$ = predictive mean matching estimator using 3 donors; $\widehat{\beta}_{PMM10}$ = predictive mean matching estimator using 10 donors; $\widehat{\beta}_{MI}$ = multiple imputation estimator; $\widehat{\beta}_{CC}$ = complete-case estimator.

Cor($\boldsymbol{Y}, \boldsymbol{S}$)	$(\rho_V, \rho_{\underline{V}}, \rho_{\overline{V}})$	Estimator	Bias (% Bias)	SD	$\widehat{\text{SE}}$	MSE	RE
0.8	(0.8, 0.1, 0.1)	$\widehat{\beta}_{HS}$	0.006 (0.553)	0.079	0.096	0.006	1.063
		$\widehat{\beta}_{VS}$	0.003 (0.330)	0.078	0.086	0.006	1.063
		$\widehat{\beta}_{PMM3}$	0.002 (0.203)	0.087	0.1.03	0.008	1.274
		$\widehat{\beta}_{PMM10}$	0.000 (0.203)	0.087	0.102	0.008	1.261
		$\widehat{\beta}_{MI}$	0.020 (2.028)	0.086	0.123	0.008	1.519
		$\widehat{\beta}_{CC}$	-0.003 (0.345)	0.084	0.090	0.007	1.116
	(0.4, 0.3, 0.3)	$\widehat{\beta}_{HS}$	0.035 (3.481)	0.102	0.100	0.012	1.238
		$\widehat{\beta}_{VS}$	0.026 (2.581)	0.103	0.104	0.011	1.285
		$\widehat{\beta}_{PMM3}$	-0.003 (0.298)	0.139	0.159	0.019	1.969
		$\widehat{\beta}_{PMM10}$	-0.006 (0.635)	0.135	0.155	0.018	1.911
		$\widehat{\beta}_{MI}$	0.078 (7.812)	0.216	0.486	0.053	6.001
		$\widehat{\beta}_{CC}$	-0.006 (0.617)	0.125	0.127	0.016	1.564
0.4	(0.8, 0.1, 0.1)	$\widehat{\beta}_{HS}$	-0.003 (0.265)	0.083	0.090	0.007	1.112
		$\widehat{\beta}_{VS}$	-0.003 (0.266)	0.083	0.090	0.007	1.111
		$\widehat{\beta}_{PMM3}$	-0.002 (0.175)	0.090	0.111	0.008	1.372
		$\widehat{\beta}_{PMM10}$	-0.002 (0.207)	0.089	0.110	0.008	1.354
		$\widehat{\beta}_{MI}$	0.021 (2.071)	0.097	0.131	0.053	1.620
		$\widehat{\beta}_{CC}$	-0.003 (0.345)	0.084	0.090	0.016	1.116
	(0.4, 0.3, 0.3)	$\widehat{\beta}_{HS}$	0.000 (0.030)	0.121	0.122	0.015	1.511
		$\widehat{\beta}_{VS}$	0.001 (0.096)	0.120	0.123	0.014	1.515
		$\widehat{\beta}_{PMM3}$	-0.009 (0.920)	0.149	0.179	0.022	2.212
		$\widehat{\beta}_{PMM10}$	-0.008 (0.840)	0.148	0.174	0.022	2.145
		$\widehat{\beta}_{MI}$	0.077 (7.709)	0.227	0.227	0.057	2.805
		$\widehat{\beta}_{CC}$	-0.006 (0.617)	0.125	0.127	0.016	1.564

Table B.4: Results for baseline covariate coefficient from 500 simulations comparing proposed estimators with multiple imputation and predictive mean matching estimators. SD = standard deviation; $\widehat{SE}$ = estimated standard error; MSE = mean squared error; RE = relative efficiency. $\widehat{\beta}_{HS}$ = estimator based on both hybrid and validation subsample; $\widehat{\beta}_{VS}$ = estimator based on internal validation sample; $\widehat{\beta}_{PMM3}$ = predictive mean matching estimator using 3 donors; $\widehat{\beta}_{PMM10}$ = predictive mean matching estimator using 10 donors; $\widehat{\beta}_{MI}$ = multiple imputation estimator; $\widehat{\beta}_{CC}$ = complete-case estimator.

Cor($\boldsymbol{Y}, \boldsymbol{S}$)	$(\rho_V, \rho_{\underline{V}}, \rho_{\overline{V}})$	Estimator	Bias (% Bias)	SD	$\widehat{\text{SE}}$	MSE	RE
0.8	(0.8, 0.1, 0.1)	$\widehat{\beta}_{HS}$	-0.017 (2.385)	0.206	0.209	0.043	1.025
		$\widehat{\beta}_{VS}$	-0.012 (1.708)	0.207	0.209	0.043	1.026
		$\widehat{\beta}_{PMM3}$	-0.003 (0.395)	0.215	0.207	0.046	1.018
		$\widehat{\beta}_{PMM10}$	-0.006 (0.789)	0.216	0.207	0.047	1.018
		$\widehat{\beta}_{MI}$	-0.016 (2.344)	0.217	0.285	0.047	1.398
		$\widehat{\beta}_{CC}$	-0.012 (1.672)	0.215	0.215	0.046	1.054
	(0.4, 0.3, 0.3)	$\widehat{\beta}_{HS}$	-0.024 (3.478)	0.214	0.215	0.046	1.057
		$\widehat{\beta}_{VS}$	-0.014 (1.960)	0.218	0.218	0.046	1.070
		$\widehat{\beta}_{PMM3}$	0.007 (1.014)	0.251	0.231	0.047	1.134
		$\widehat{\beta}_{PMM10}$	0.002 (0.237)	0.250	0.225	0.047	1.106
		$\widehat{\beta}_{MI}$	-0.062 (8.821)	0.937	1.006	0.882	4.941
		$\widehat{\beta}_{CC}$	-0.008 (1.173)	0.239	0.245	0.057	1.202
0.4	(0.8, 0.1, 0.1)	$\widehat{\beta}_{HS}$	-0.012 (1.692)	0.214	0.216	0.046	1.062
		$\widehat{\beta}_{VS}$	-0.011 (1.592)	0.214	0.216	0.046	1.061
		$\widehat{\beta}_{PMM3}$	-0.009 (1.339)	0.216	0.209	0.047	1.026
		$\widehat{\beta}_{PMM10}$	-0.012 (1.706)	0.217	0.208	0.047	1.024
		$\widehat{\beta}_{MI}$	-0.013 (1.843)	0.218	0.265	0.064	1.299
		$\widehat{\beta}_{CC}$	-0.012 (1.672)	0.215	0.215	0.060	1.054
	(0.4, 0.3, 0.3)	$\widehat{\beta}_{HS}$	-0.010 (1.363)	0.235	0.243	0.055	1.192
		$\widehat{\beta}_{VS}$	-0.008 (1.213)	0.233	0.242	0.054	1.188
		$\widehat{\beta}_{PMM3}$	-0.011 (1.554)	0.257	0.237	0.066	1.164
		$\widehat{\beta}_{PMM10}$	-0.013 (1.818)	0.256	0.231	0.066	1.132
		$\widehat{\beta}_{MI}$	-0.015 (2.077)	0.285	0.432	0.081	2.119
		$\widehat{\beta}_{CC}$	-0.008 (1.173)	0.239	0.245	0.057	1.202

Table B.5: Results for timepoint coefficient from 500 simulations evaluating impact of MAR outcomes in the hybrid set. SD = standard deviation; $\widehat{SE}$ = estimated standard error; MSE = mean squared error; RE = relative efficiency. $\widehat{\beta}_{HS}$ = estimator based on both hybrid and validation subsample; $\widehat{\beta}_{CC}$ = complete-case estimator.

Cor($\boldsymbol{Y}, \boldsymbol{S}$)	$(\rho_V, \rho_{\underline{V}}, \rho_{\overline{V}})$	Estimator	Bias (% Bias)	SD	$\widehat{\text{SE}}$	MSE	RE
0.8	(0.8, 0.1, 0.1)	$\widehat{\beta}_{HS}$	-0.001 (0.143)	0.085	0.086	0.007	1.060
		$\widehat{\beta}_{CC}$	-0.002 (0.187)	0.088	0.089	0.008	1.102
	(0.7, 0.15, 0.15)	$\widehat{\beta}_{HS}$	-0.002 (0.167)	0.088	0.088	0.008	1.095
		$\widehat{\beta}_{CC}$	-0.003 (0.264)	0.094	0.094	0.009	1.168
	(0.6, 0.2, 0.2)	$\widehat{\beta}_{HS}$	-0.004 (0.382)	0.092	0.092	0.008	1.135
		$\widehat{\beta}_{CC}$	-0.005 (0.521)	0.100	0.101	0.010	1.245
	(0.5, 0.25, 0.25)	$\widehat{\beta}_{HS}$	-0.003 (0.268)	0.094	0.095	0.009	1.181
		$\widehat{\beta}_{CC}$	-0.006 (0.562)	0.106	0.108	0.011	1.341
	(0.4, 0.3, 0.3)	$\widehat{\beta}_{HS}$	-0.002 (0.230)	0.101	0.100	0.010	1.238
		$\widehat{\beta}_{CC}$	-0.008 (0.797)	0.113	0.119	0.013	1.471
	(0.3, 0.35, 0.35)	$\widehat{\beta}_{HS}$	-0.007 (0.745)	0.111	0.106	0.012	1.309
		$\widehat{\beta}_{CC}$	-0.008 (0.814)	0.131	0.133	0.017	1.647
0.4	(0.8, 0.1, 0.1)	$\widehat{\beta}_{HS}$	-0.002 (0.175)	0.088	0.089	0.008	1.100
		$\widehat{\beta}_{CC}$	-0.002 (0.187)	0.088	0.089	0.008	1.102
	(0.7, 0.15, 0.15)	$\widehat{\beta}_{HS}$	-0.005 (0.247)	0.093	0.094	0.009	1.163
		$\widehat{\beta}_{CC}$	-0.003 (0.264)	0.094	0.094	0.009	1.168
	(0.6, 0.2, 0.2)	$\widehat{\beta}_{HS}$	-0.005 (0.509)	0.099	0.100	0.010	1.237
		$\widehat{\beta}_{CC}$	-0.005 (0.521)	0.100	0.101	0.010	1.245
	(0.5, 0.25, 0.25)	$\widehat{\beta}_{HS}$	-0.005 (0.546)	0.105	0.107	0.107	1.327
		$\widehat{\beta}_{CC}$	-0.006 (0.562)	0.106	0.108	0.011	1.341
	(0.4, 0.3, 0.3)	$\widehat{\beta}_{HS}$	-0.008 (0.763)	0.111	0.117	0.011	1.447
		$\widehat{\beta}_{CC}$	-0.008 (0.797)	0.113	0.119	0.012	1.471
	(0.3, 0.35, 0.35)	$\widehat{\beta}_{HS}$	-0.010 (0.983)	0.122	0.130	0.013	1.608
		$\widehat{\beta}_{CC}$	-0.008 (0.814)	0.131	0.133	0.017	1.647

Table B.6: Results for baseline covariate coefficient from 500 simulations evaluating impact of MAR outcomes in hybrid set. SD = standard deviation; $\widehat{SE}$ = estimated standard error; MSE = mean squared error; RE = relative efficiency. $\widehat{\beta}_{HS}$ = estimator based on both hybrid and validation subsample; $\widehat{\beta}_{CC}$ = complete-case estimator.

Cor($\boldsymbol{Y}, \boldsymbol{S}$)	$(\rho_V, \rho_{\underline{V}}, \rho_{\overline{V}})$	Estimator	Bias (% Bias)	SD	$\widehat{\text{SE}}$	MSE	RE
0.8	(0.8, 0.1, 0.1)	$\widehat{\beta}_{HS}$	-0.020 (2.891)	0.212	0.211	0.050	1.039
		$\widehat{\beta}_{CC}$	-0.021 (3.071)	0.225	0.217	0.051	1.067
	(0.7, 0.15, 0.15)	$\widehat{\beta}_{HS}$	-0.021 (3.006)	0.214	0.215	0.053	1.055
		$\widehat{\beta}_{CC}$	-0.023 (3.321)	0.232	0.225	0.054	1.107
	(0.6, 0.2, 0.2)	$\widehat{\beta}_{HS}$	-0.028 (4.008)	0.216	0.217	0.056	1.069
		$\widehat{\beta}_{CC}$	-0.032 (4.632)	0.237	0.234	0.057	1.152
	(0.5, 0.25, 0.25)	$\widehat{\beta}_{HS}$	-0.023 (3.265)	0.223	0.221	0.062	1.087
		$\widehat{\beta}_{CC}$	-0.028 (3.974)	0.252	0.245	0.064	1.205
	(0.4, 0.3, 0.3)	$\widehat{\beta}_{HS}$	-0.023 (3.250)	0.232	0.225	0.066	1.107
		$\widehat{\beta}_{CC}$	-0.024 (3.393)	0.261	0.257	0.069	1.262
	(0.3, 0.35, 0.35)	$\widehat{\beta}_{HS}$	-0.024 (3.489)	0.236	0.229	0.070	1.125
		$\widehat{\beta}_{CC}$	-0.028 (4.019)	0.271	0.272	0.074	1.334
0.4	(0.8, 0.1, 0.1)	$\widehat{\beta}_{HS}$	-0.021 (3.032)	0.223	0.219	0.050	1.076
		$\widehat{\beta}_{CC}$	-0.021 (3.071)	0.225	0.217	0.051	1.067
	(0.7, 0.15, 0.15)	$\widehat{\beta}_{HS}$	-0.023 (3.262)	0.229	0.227	0.053	1.115
		$\widehat{\beta}_{CC}$	-0.023 (3.321)	0.232	0.225	0.054	1.107
	(0.6, 0.2, 0.2)	$\widehat{\beta}_{HS}$	-0.032 (4.587)	0.234	0.235	0.056	1.154
		$\widehat{\beta}_{CC}$	-0.032 (4.632)	0.237	0.234	0.057	1.152
	(0.5, 0.25, 0.25)	$\widehat{\beta}_{HS}$	-0.027 (3.911)	0.247	0.245	0.062	1.202
		$\widehat{\beta}_{CC}$	-0.028 (3.974)	0.252	0.245	0.064	1.205
	(0.4, 0.3, 0.3)	$\widehat{\beta}_{HS}$	-0.024 (3.402)	0.256	0.256	0.066	1.255
		$\widehat{\beta}_{CC}$	-0.024 (3.393)	0.261	0.257	0.069	1.262
	(0.3, 0.35, 0.35)	$\widehat{\beta}_{HS}$	-0.029 (4.160)	0.263	0.268	0.070	1.315
		$\widehat{\beta}_{CC}$	-0.024 (4.019)	0.271	0.272	0.074	1.334

$$\boldsymbol{U}_{VS}(\boldsymbol{\beta}) = \sum_{i\in V} \frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}{P_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)} + \sum_{k\in \underline{V}} \frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_k, \underline{\boldsymbol{Y}}_k|\boldsymbol{X}_k)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_k, \underline{\boldsymbol{Y}}_k|\boldsymbol{X}_k)} + \sum_{j\in \overline{V}} \frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)}$$

$$+ \sum_{k\in \underline{V}} \left\{ \frac{\widehat{\boldsymbol{D}}_{\boldsymbol{\beta}}(\boldsymbol{S}_k, \underline{\boldsymbol{Y}}_k|\boldsymbol{X}_k)}{\widehat{P}_{\boldsymbol{\beta}}(\boldsymbol{S}_k, \underline{\boldsymbol{Y}}_k|\boldsymbol{X}_k)} - \frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_k, \underline{\boldsymbol{Y}}_k|\boldsymbol{X}_k)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_k, \underline{\boldsymbol{Y}}_k|\boldsymbol{X}_k)} \right\} \tag{B.1}$$

$$+ \sum_{j\in \overline{V}} \left\{ \frac{\widehat{\boldsymbol{D}}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)}{\widehat{P}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)} - \frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)} \right\} \tag{B.2}$$

Working with (B.2) first, by Slutsky's Theorem,

$$\begin{aligned}
(B.2) &= \sum_{j\in \overline{V}} \left\{ \frac{\widehat{\boldsymbol{D}}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)}{\widehat{P}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)} - \frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)} \right\} \times \frac{\widehat{P}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)} \\
&= \sum_{j\in \overline{V}} \left\{ \frac{\widehat{\boldsymbol{D}}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)} - \frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)\widehat{P}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)}{P^2_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)} \right\} + o_p(1) \\
&= \sum_{j\in \overline{V}} \left\{ \frac{\int \widehat{P}(\boldsymbol{S}_j|\boldsymbol{y}, \boldsymbol{X}_j)\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X}_j)d\boldsymbol{y}}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)} - \frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)\int \widehat{P}(\boldsymbol{S}_j|\boldsymbol{y}, \boldsymbol{X}_j)P_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X}_j)d\boldsymbol{y}}{P^2_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)} \right\} + o_p(1) \\
&= \sum_{j\in \overline{V}} \int \widehat{P}(\boldsymbol{S}_j|\boldsymbol{y}, \boldsymbol{X}_j) \left\{ \frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X}_j)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)} - \frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)P_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X}_j)}{P^2_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)} \right\} d\boldsymbol{y} + o_p(1).
\end{aligned}$$

Define

$$\boldsymbol{\psi}_j^{\boldsymbol{\beta}, \overline{V}}(\boldsymbol{y}) = \frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X}_j)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)} - \frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)P_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X}_j)}{P^2_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)},$$

Then

$$(B.2) = \sum_{j\in \overline{V}} \int \widehat{P}(\boldsymbol{S}_j|\boldsymbol{y}, \boldsymbol{X}_j)\boldsymbol{\psi}_j^{\boldsymbol{\beta}, \overline{V}}(\boldsymbol{y})d\boldsymbol{y} + o_p(1).$$

Now invoking Slutsky's Theorem again, and noting that $\int P(\boldsymbol{S}_j|\boldsymbol{y},\boldsymbol{X}_j)\boldsymbol{\psi}_j^{\boldsymbol{\beta},\overline{V}}(\boldsymbol{y})d\boldsymbol{y} = 0\ \forall j$,

$$(B.2) = \sum_{j\in\overline{V}} \int \left\{ \widehat{P}(\boldsymbol{S}_j|\boldsymbol{y},\boldsymbol{X}_j) - P(\boldsymbol{S}_j|\boldsymbol{y},\boldsymbol{X}_j) \right\} \boldsymbol{\psi}_j^{\boldsymbol{\beta},\overline{V}}(\boldsymbol{y})d\boldsymbol{y} + o_p(1)$$
$$= \sum_{j\in\overline{V}} \int \left\{ \frac{\widehat{P}(\boldsymbol{S}_j,\boldsymbol{y},\boldsymbol{X}_j)}{\widehat{P}(\boldsymbol{y},\boldsymbol{X}_j)} - \frac{P(\boldsymbol{S}_j,\boldsymbol{y},\boldsymbol{X}_j)}{P(\boldsymbol{y},\boldsymbol{X}_j)} \right\} \boldsymbol{\psi}_j^{\boldsymbol{\beta},\overline{V}}(\boldsymbol{y})d\boldsymbol{y} \times \frac{\widehat{P}(\boldsymbol{y},\boldsymbol{X}_j)}{P(\boldsymbol{y},\boldsymbol{X}_j)} + o_p(1)$$
$$= \sum_{j\in\overline{V}} \int \left\{ \frac{\widehat{P}(\boldsymbol{S}_j,\boldsymbol{y},\boldsymbol{X}_j)}{P(\boldsymbol{y},\boldsymbol{X}_j)} - \frac{P(\boldsymbol{S}_j,\boldsymbol{y},\boldsymbol{X}_j)\widehat{P}(\boldsymbol{y},\boldsymbol{X}_j)}{P^2(\boldsymbol{y},\boldsymbol{X}_j)} \right\} \boldsymbol{\psi}_j^{\boldsymbol{\beta},\overline{V}}(\boldsymbol{y})d\boldsymbol{y} + o_p(1).$$

At this point, we can express $\widehat{P}(\boldsymbol{S}_j,\boldsymbol{y},\boldsymbol{X}_j)$ and $\widehat{P}(\boldsymbol{y},\boldsymbol{X}_j)$ in terms of their respective constructions kernel density estimates,

$$\widehat{P}(\boldsymbol{S}_j,\boldsymbol{y},\boldsymbol{X}_j) = \frac{1}{n_V}\sum_{i\in V} \frac{1}{|\boldsymbol{H}_2|^{1/2}} \phi(\boldsymbol{L}_2(\boldsymbol{S}_j - \boldsymbol{S}_i, \boldsymbol{y} - \boldsymbol{Y}_i, \boldsymbol{X}_j - \boldsymbol{X}_i))$$
$$\widehat{P}(\boldsymbol{y},\boldsymbol{X}_j) = \frac{1}{n_V}\sum_{i\in V} \frac{1}{|\boldsymbol{H}_1|^{1/2}} \phi(\boldsymbol{L}_1(\boldsymbol{y} - \boldsymbol{Y}_i, \boldsymbol{X}_j - \boldsymbol{X}_j))$$

Let $\phi(\boldsymbol{L}_2(\boldsymbol{S}_j - \boldsymbol{S}_i, \boldsymbol{y} - \boldsymbol{Y}_i, \boldsymbol{X}_j - \boldsymbol{X}_i)) = \phi_{2j}$ and $\phi(\boldsymbol{L}_1(\boldsymbol{y} - \boldsymbol{Y}_i, \boldsymbol{X}_j)) = \phi_{1j}$, for clarity of presentation. Without loss of generality, if there are discrete covariates in the model, then let $\phi_{2j} = I(\boldsymbol{X}_{Dj} = \boldsymbol{X}_{Di})\phi(\boldsymbol{L}_2(\boldsymbol{S}_j - \boldsymbol{S}_i, \boldsymbol{y} - \boldsymbol{Y}_i, \boldsymbol{X}_j - \boldsymbol{X}_i))$, and likewise for ϕ_{1j}. Then

$$(B.2) = \frac{1}{n_V}\sum_{i\in V}\sum_{j\in\overline{V}} \int \left\{ \frac{|\boldsymbol{H}_2|^{-1/2}\phi_{2ij}}{P(\boldsymbol{y},\boldsymbol{X}_j)} - \frac{|\boldsymbol{H}_1|^{-1/2}\phi_{1ij}P(\boldsymbol{S}_j,\boldsymbol{y},\boldsymbol{X}_j)}{P^2(\boldsymbol{y},\boldsymbol{X}_j)} \right\} \boldsymbol{\psi}_j^{\boldsymbol{\beta},\overline{V}}(\boldsymbol{y})d\boldsymbol{y} + o_p(1)$$
$$= \frac{1}{n_V}\sum_{i\in v}\sum_{j\in\overline{V}} \boldsymbol{Q}_{ij} + o_p(1).$$

Finally, define

$$\zeta_i^{V,\overline{V}}(\boldsymbol{y}) = \frac{|\boldsymbol{H}_2|^{-1/2}\phi_{2ij}}{P(\boldsymbol{y},\boldsymbol{X}_j)} - \frac{|\boldsymbol{H}_1|^{-1/2}\phi_{1ij}P(\boldsymbol{S}_j,\boldsymbol{y},\boldsymbol{X}_j)}{P^2(\boldsymbol{y},\boldsymbol{X}_j)}$$

so that

$$\boldsymbol{Q}_{ij} = \int \zeta_i^{V,\overline{V}}(\boldsymbol{y})\boldsymbol{\psi}_j^{\boldsymbol{\beta},\overline{V}}(\boldsymbol{y})d\boldsymbol{y}.$$

Likewise, we can deconstruct (B.1) in a similar fashion. (B.1) can be expressed as

$$
\begin{aligned}
(B.1) &= \frac{1}{n_V}\sum_{i\in V}\sum_{k\in \underline{V}}\int \Big\{\zeta_i^{V,\underline{V}}(\boldsymbol{y}_{m_k})\Big\}\boldsymbol{\psi}_k^{\boldsymbol{\beta},\underline{V}}(\boldsymbol{y}_{m_k})d\boldsymbol{y}_{m_k} + o_p(1) \\
&= \frac{1}{n_V}\sum_{i\in V}\sum_{k\in \underline{V}}\boldsymbol{Q}_{ik} + o_p(1),
\end{aligned}
$$

where

$$
\begin{aligned}
\phi_{2ik} &= \phi(\boldsymbol{L}_2(\boldsymbol{S}_k - \boldsymbol{S}_i, \boldsymbol{y}_{m_k} - \boldsymbol{Y}_{m_k i}, \underline{\boldsymbol{Y}}_k - \boldsymbol{Y}_{T_k i}, \boldsymbol{X}_k - \boldsymbol{X}_k)), \\
\phi_{1ik} &= \phi(\boldsymbol{L}_1(\boldsymbol{y}_{m_k} - \boldsymbol{Y}_{m_k i}, \underline{\boldsymbol{Y}}_k - \boldsymbol{Y}_{T_k i}, \boldsymbol{X}_k - \boldsymbol{X}_k)), \\
\zeta_i^{V,\underline{V}}(\boldsymbol{y}_{m_k}) &= \frac{|\boldsymbol{H}_2|^{-1/2}\phi_{2ik}}{P(\boldsymbol{y}_{m_k}, \underline{\boldsymbol{Y}}_k, \boldsymbol{X}_k)} - \frac{|\boldsymbol{H}_1|^{-1/2}\phi_{1ik}P(\boldsymbol{S}_k, \boldsymbol{y}_{m_k}, \underline{\boldsymbol{Y}}_k, \boldsymbol{X}_k)}{P^2(\boldsymbol{y}_{m_k}, \underline{\boldsymbol{Y}}_k, \boldsymbol{X}_k)},
\end{aligned}
$$

and

$$
\boldsymbol{\psi}_k^{\boldsymbol{\beta},\underline{V}}(\boldsymbol{y}_{m_k}, \underline{\boldsymbol{Y}}_k) = \frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{y}_{m_k}, \underline{\boldsymbol{Y}}_k|\boldsymbol{X}_k)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_k, \underline{\boldsymbol{Y}}_k|\boldsymbol{X}_k)} - \frac{D_{\boldsymbol{\beta}}(\boldsymbol{S}_k, \underline{\boldsymbol{Y}}_k|\boldsymbol{X}_k)P(\boldsymbol{y}_{m_k}, \underline{\boldsymbol{Y}}_k|\boldsymbol{X}_k)}{P^2_{\boldsymbol{\beta}}(\boldsymbol{S}_k, \underline{\boldsymbol{Y}}_k|\boldsymbol{X}_k)}.
$$

Putting (B.1) and (B.2) together,

$$
\begin{aligned}
(B.1) + (B.2) &= \left(\frac{n_{\underline{V}} + n_{\overline{V}}}{\sqrt{n_V}}\right)\frac{1}{\sqrt{n_V}}\sum_{i\in V}\left[\frac{1}{n_{\underline{V}} + n_{\overline{V}}}\Big\{\sum_{k\in\underline{V}}\boldsymbol{Q}_{ik} + \sum_{j\in\overline{V}}\boldsymbol{Q}_{ij}\Big\}\right] + o_p(1) \\
&= \left(\frac{n_{\underline{V}} + n_{\overline{V}}}{\sqrt{n_V}}\right)\frac{1}{\sqrt{n_V}}\sum_{i\in V}\boldsymbol{Q}_i + o_p(1).
\end{aligned}
$$

Therefore, multiplying by $\frac{1}{\sqrt{N}}$, the entire score equation is

$$
\begin{aligned}
\frac{1}{\sqrt{N}}\boldsymbol{U}_{VS}(\boldsymbol{\beta}) =& \frac{1}{\sqrt{N}}\left[\sum_{i\in V}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}{P_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)} + \sum_{k\in\underline{V}}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_k, \underline{\boldsymbol{Y}}_k|\boldsymbol{X}_k)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_k, \underline{\boldsymbol{Y}}_k|\boldsymbol{X}_k)} + \sum_{j\in\overline{V}}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)}\right] \\
&+ \frac{1}{\sqrt{n_V}}\left[\frac{n_{\underline{V}} + n_{\overline{V}}}{\sqrt{N}\sqrt{n_V}}\sum_{i\in V}\sum_{k\in\underline{V}}\boldsymbol{Q}_i\right] + o_p(1).
\end{aligned}
$$

The next part of the proof is to establish the expected value of $\boldsymbol{Q}_i$ conditional on the validation data.

As before, we break this down into $\boldsymbol{Q}_{ik}$ and $\boldsymbol{Q}_{ij}$.

$$
\begin{aligned}
E\big[\boldsymbol{Q}_{ij}|\boldsymbol{S}_i,\boldsymbol{Y}_i,\boldsymbol{X}_i\big] &= E_{\boldsymbol{X}}\big[E_{\boldsymbol{S}}[\boldsymbol{Q}_{ij}|\boldsymbol{X}]\big|\boldsymbol{S}_i,\boldsymbol{Y}_i,\boldsymbol{X}_i\big] \\
&= E_{\boldsymbol{X}}\left[\int_{\boldsymbol{S}}\int_{\boldsymbol{y}} \zeta_i^{V,\overline{V}}(\boldsymbol{y})\psi_j^{\boldsymbol{\beta},\overline{V}}(\boldsymbol{y})P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})d\boldsymbol{y}d\boldsymbol{S}\middle|\boldsymbol{S}_i,\boldsymbol{Y}_i,\boldsymbol{X}_i\right] \\
&= E_{\boldsymbol{X}}\left[\int_{\boldsymbol{S}}\int_{\boldsymbol{y}} \zeta_i^{V,\overline{V}}(\boldsymbol{y})\left\{\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})}{P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})} - \frac{P_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}{P^2_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}\right\}P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})d\boldsymbol{y}d\boldsymbol{S}\middle|\boldsymbol{S}_i,\boldsymbol{Y}_i,\boldsymbol{X}_i\right] \\
&= E_{\boldsymbol{X}}\left[\int_{\boldsymbol{S}}\int_{\boldsymbol{y}} \zeta_i^{V,\overline{V}}(\boldsymbol{y})\left\{\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})\cancel{P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}}{\cancel{P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}} - \frac{P_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})D_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})\cancel{P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}}{P^{\cancel{2}}_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}\right\}d\boldsymbol{y}d\boldsymbol{S}\middle|\boldsymbol{S}_i,\boldsymbol{Y}_i,\boldsymbol{X}_i\right] \\
&= E_{\boldsymbol{X}}\left[\int_{\boldsymbol{S}}\int_{\boldsymbol{y}} \zeta_i^{V,\overline{V}}(\boldsymbol{y})\left\{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X}) - \frac{P_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})D_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}{P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}\right\}d\boldsymbol{y}d\boldsymbol{S}\middle|\boldsymbol{S}_i,\boldsymbol{Y}_i,\boldsymbol{X}_i\right] \qquad \text{(B.3)}
\end{aligned}
$$

(B.3) should now be split into two integrals.

$$
E_{\boldsymbol{X}}\left[\int_{\boldsymbol{S}}\int_{\boldsymbol{y}} \zeta_i^{V,\overline{V}}(\boldsymbol{y})\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})d\boldsymbol{y}d\boldsymbol{S}\middle|\boldsymbol{S}_i,\boldsymbol{Y}_i,\boldsymbol{X}_i\right] \qquad \text{(B.3a)}
$$

$$
-E_{\boldsymbol{X}}\left[\int_{\boldsymbol{S}}\int_{\boldsymbol{y}} \zeta_i^{V,\overline{V}}(\boldsymbol{y})\frac{P_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})D_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}{P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}d\boldsymbol{y}d\boldsymbol{S}\middle|\boldsymbol{S}_i,\boldsymbol{Y}_i,\boldsymbol{X}_i\right] \qquad \text{(B.3b)}
$$

$$
\begin{aligned}
\text{(B.3a)} = \;& E_{\boldsymbol{X}}\left[\int_{\boldsymbol{S}}\int_{\boldsymbol{y}}\left\{\frac{|\boldsymbol{H}_2|^{-1/2}\phi_{2ij}}{P(\boldsymbol{y},\boldsymbol{X})} - \frac{|\boldsymbol{H}_1|^{-1/2}\phi_{1ij}}{P^2(\boldsymbol{y},\boldsymbol{X})}\right\}\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})d\boldsymbol{y}d\boldsymbol{S}\middle|\boldsymbol{S}_i,\boldsymbol{Y}_i,\boldsymbol{X}_i\right] \\
= \;& E_{\boldsymbol{X}}\left[\int_{\boldsymbol{y}}\int_{\boldsymbol{S}}\left\{\frac{|\boldsymbol{H}_2|^{-1/2}\phi_{2ij}}{P(\boldsymbol{y},\boldsymbol{X})} - \frac{|\boldsymbol{H}_1|^{-1/2}\phi_{1ij}}{P^2(\boldsymbol{y},\boldsymbol{X})}\right\}\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})d\boldsymbol{S}d\boldsymbol{y}\middle|\boldsymbol{S}_i,\boldsymbol{Y}_i,\boldsymbol{X}_i\right] \\
= \;& \int_{\boldsymbol{X}}\int_{\boldsymbol{y}}\int_{\boldsymbol{S}}\left\{\frac{|\boldsymbol{H}_2|^{-1/2}\phi_{2ij}}{P(\boldsymbol{y},\boldsymbol{X})} - \frac{|\boldsymbol{H}_1|^{-1/2}\phi_{1ij}P(\boldsymbol{S},\boldsymbol{y},\boldsymbol{X})}{P^2(\boldsymbol{y},\boldsymbol{X})}\right\}\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})P(\boldsymbol{X})d\boldsymbol{S}d\boldsymbol{y}d\boldsymbol{X} \\
= \;& \int_{\boldsymbol{X}}\int_{\boldsymbol{y}}\int_{\boldsymbol{S}}\left\{|\boldsymbol{H}_2|^{-1/2}\phi_{2ij} - \frac{|\boldsymbol{H}_1|^{-1/2}\phi_{1ij}P(\boldsymbol{S},\boldsymbol{y},\boldsymbol{X})}{P(\boldsymbol{y},\boldsymbol{X})}\right\}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})}{P_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})}d\boldsymbol{S}d\boldsymbol{y}d\boldsymbol{X} \\
= \;& \int_{\boldsymbol{X}}\int_{\boldsymbol{y}}\int_{\boldsymbol{S}}|\boldsymbol{H}_2|^{-1/2}\phi_{2ij}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})}{P_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})}d\boldsymbol{S}d\boldsymbol{y}d\boldsymbol{X} \\
& -\int_{\boldsymbol{X}}\int_{\boldsymbol{y}}|\boldsymbol{H}_1|^{-1/2}\phi_{1ij}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})}{P_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})}\int_{\boldsymbol{S}}P(\boldsymbol{S}|\boldsymbol{y},\boldsymbol{X})d\boldsymbol{S}d\boldsymbol{y}d\boldsymbol{X}.
\end{aligned}
$$

We now use u-substitution on the first integral in (B.3a) with $(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}) = \boldsymbol{L}_2(\boldsymbol{S}-\boldsymbol{S}_i,\boldsymbol{y}-\boldsymbol{Y}_i,\boldsymbol{X}-\boldsymbol{X}_i)$.

This implies

$$d\boldsymbol{S}d\boldsymbol{y}d\boldsymbol{X} = \left|J(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})\right|d\boldsymbol{u}d\boldsymbol{v}d\boldsymbol{w} = \left|\boldsymbol{L}_2^{-1}\right|d\boldsymbol{u}d\boldsymbol{v}d\boldsymbol{w}.$$

We then make the substitution and note that $|\boldsymbol{L}_2^{-1}| = |\boldsymbol{H}_2|^{1/2}$. We also note here that

$$\begin{aligned}
(\boldsymbol{S},\boldsymbol{y},\boldsymbol{X}) &= \boldsymbol{L}_2^{-1}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}) + (\boldsymbol{S}_i,\boldsymbol{Y}_i,\boldsymbol{X}_i)\\
\boldsymbol{S} &= \boldsymbol{S}_i + (\boldsymbol{L}_2^{-1})^{\boldsymbol{S}}\boldsymbol{u}\\
\boldsymbol{y} &= \boldsymbol{Y}_i + (\boldsymbol{L}_2^{-1})^{\boldsymbol{y}}\boldsymbol{v}\\
\boldsymbol{X} &= \boldsymbol{X}_i + (\boldsymbol{L}_2^{-1})^{\boldsymbol{X}}\boldsymbol{w}
\end{aligned}$$

with the superscript denoting the respective columns of $\boldsymbol{L}_2^{-1}$ corresponding to that variable. Similarly, we can make a substitution on the second integral in (B.3a) with $(\boldsymbol{u},\boldsymbol{v}) = \boldsymbol{L}_1(\boldsymbol{y}-\boldsymbol{Y}_i,\boldsymbol{X}-\boldsymbol{X}_i)$, so that $d\boldsymbol{y}d\boldsymbol{X} = |\boldsymbol{L}_1^{-1}|d\boldsymbol{u}d\boldsymbol{v}$ and the analogous implications. The integrals then become

$$\begin{aligned}
\text{(B.3a)} = &\int_{\boldsymbol{w}}\int_{\boldsymbol{v}}\int_{\boldsymbol{u}}\left\{\frac{|\boldsymbol{L}_2^{-1}|}{|\boldsymbol{H}_2|^{1/2}}\phi(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})\right\}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{Y}_i + (\boldsymbol{L}_2^{-1})^{\boldsymbol{y}}\boldsymbol{v}|\boldsymbol{X}_i + (\boldsymbol{L}_2^{-1})^{\boldsymbol{X}}\boldsymbol{w})}{P_{\boldsymbol{\beta}}(\boldsymbol{Y}_i + (\boldsymbol{L}_2^{-1})^{\boldsymbol{y}}\boldsymbol{v}|\boldsymbol{X}_i + (\boldsymbol{L}_2^{-1})^{\boldsymbol{X}}\boldsymbol{w})}d\boldsymbol{u}d\boldsymbol{v}d\boldsymbol{w}\\
&-\int_{\boldsymbol{v}}\int_{\boldsymbol{u}}\left\{\frac{|\boldsymbol{L}_1^{-1}|}{|\boldsymbol{H}_1|^{1/2}}\phi(\boldsymbol{u},\boldsymbol{v})\right\}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{Y}_i + (\boldsymbol{L}_1^{-1})^{\boldsymbol{y}}\boldsymbol{u}|\boldsymbol{X}_i + (\boldsymbol{L}_1^{-1})^{\boldsymbol{X}}\boldsymbol{v})}{P_{\boldsymbol{\beta}}(\boldsymbol{Y}_i + (\boldsymbol{L}_1^{-1})^{\boldsymbol{y}}\boldsymbol{u}|\boldsymbol{X}_i + (\boldsymbol{L}_1^{-1})^{\boldsymbol{X}}\boldsymbol{v})}d\boldsymbol{u}d\boldsymbol{v}\\
= &\int_{\boldsymbol{w}}\int_{\boldsymbol{v}}\int_{\boldsymbol{u}}\left\{\phi(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})\right\}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}{P_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}d\boldsymbol{u}d\boldsymbol{v}d\boldsymbol{w} + O(\boldsymbol{H})\\
&-\int_{\boldsymbol{v}}\int_{\boldsymbol{u}}\left\{\phi(\boldsymbol{u},\boldsymbol{v})\right\}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}{P_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}d\boldsymbol{u}d\boldsymbol{v} + O(\boldsymbol{H}).
\end{aligned}$$

We write simply $O(\boldsymbol{H})$ to denote a function on the same order as the chosen bandwidth. Since, by assumption, the bandwidth matrix approaches $\boldsymbol{0}$ as N approaches ∞, these functions are also

$o_p(1)$. Finally, integrating over the specified variables,

$$\begin{aligned}(B.3a) = & \int_{\boldsymbol{w}}\int_{\boldsymbol{v}}\int_{\boldsymbol{u}}\left\{\phi(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})\right\}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}{P_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}d\boldsymbol{u}d\boldsymbol{v}d\boldsymbol{w}+O(\boldsymbol{H})\\ & -\int_{\boldsymbol{v}}\int_{\boldsymbol{u}}\left\{\phi(\boldsymbol{u},\boldsymbol{v})\right\}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}{P_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}d\boldsymbol{u}d\boldsymbol{v}+O(\boldsymbol{H})\\ & =\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}{P_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}-\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}{P_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}+O(\boldsymbol{H})\\ & =\boldsymbol{0}+O(\boldsymbol{H}).\end{aligned}$$

We can now find the form of the integral in (B.3b) with similar arguments.

$$\begin{aligned}(B.3b) = & E_{\boldsymbol{X}}\left[\int_{\boldsymbol{S}}\int_{\boldsymbol{y}}\left\{\frac{|\boldsymbol{H}_2|^{-1/2}\phi_{2ij}}{P(\boldsymbol{y},\boldsymbol{X})}-\frac{|\boldsymbol{H}_1|^{-1/2}\phi_{1ij}P(\boldsymbol{S},\boldsymbol{y},\boldsymbol{X})}{P^2(\boldsymbol{y},\boldsymbol{X})}\right\}\right.\\ & \left.\times\frac{P_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}{P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}d\boldsymbol{y}d\boldsymbol{S}\right|\boldsymbol{S}_i,\boldsymbol{Y}_i,\boldsymbol{X}_i\Bigg]\\ = & \int_{\boldsymbol{X}}\int_{\boldsymbol{S}}\int_{\boldsymbol{y}}\left\{\frac{|\boldsymbol{H}_2|^{-1/2}\phi_{1ij}}{P(\boldsymbol{y},\boldsymbol{X})}-\frac{|\boldsymbol{H}_1|^{-1/2}\phi_{1ij}P(\boldsymbol{S},\boldsymbol{y},\boldsymbol{X})}{P^2(\boldsymbol{y},\boldsymbol{X})}\right\}\\ & \times\frac{P_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})P(\boldsymbol{X})}{P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}d\boldsymbol{y}d\boldsymbol{S}d\boldsymbol{X}\\ = & \int_{\boldsymbol{X}}\int_{\boldsymbol{S}}\int_{\boldsymbol{y}}\left\{|\boldsymbol{H}_2|^{-1/2}\phi_{2ij}-\frac{|\boldsymbol{H}_1|^{-1/2}\phi_{1ij}P(\boldsymbol{S},\boldsymbol{y},\boldsymbol{X})}{P(\boldsymbol{y},\boldsymbol{X})}\right\}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}{P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}d\boldsymbol{y}d\boldsymbol{S}d\boldsymbol{X}\\ = & \int_{\boldsymbol{X}}\int_{\boldsymbol{S}}\int_{\boldsymbol{y}}|\boldsymbol{H}_2|^{-1/2}\phi_{2ij}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}{P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}d\boldsymbol{y}d\boldsymbol{S}d\boldsymbol{X}\\ - & \int_{\boldsymbol{X}}\int_{\boldsymbol{y}}|\boldsymbol{H}_1|^{-1/2}\phi_{1ij}\int_{\boldsymbol{S}}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}{P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}P(\boldsymbol{S}|\boldsymbol{y},\boldsymbol{X})d\boldsymbol{S}d\boldsymbol{y}d\boldsymbol{X}\end{aligned}$$

We now make the same u-substitutions as in (B.3a).

$$\begin{aligned}
(B.3b) = \ & \int_{\boldsymbol{w}}\int_{\boldsymbol{v}}\int_{\boldsymbol{u}}\left\{\frac{|\boldsymbol{L}_2|}{|\boldsymbol{H}_2|^{1/2}}\phi(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})\right\}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_i+(\boldsymbol{L}_2^{-1})^{\boldsymbol{S}}\boldsymbol{u}|\boldsymbol{X}_i+(\boldsymbol{L}_2^{-1})^{\boldsymbol{X}}\boldsymbol{w})}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_i+(\boldsymbol{L}_2^{-1})^{\boldsymbol{S}}\boldsymbol{u}|\boldsymbol{X}_i+(\boldsymbol{L}_2^{-1})^{\boldsymbol{X}}\boldsymbol{w})}d\boldsymbol{u}d\boldsymbol{v}d\boldsymbol{w} \\
& -\int_{\boldsymbol{v}}\int_{\boldsymbol{u}}\left\{\frac{|\boldsymbol{L}_1|}{|\boldsymbol{H}_1|^{1/2}}\phi(\boldsymbol{u},\boldsymbol{v})\right\}\int_{\boldsymbol{S}}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X}_i+(\boldsymbol{L}_1^{-1})^{\boldsymbol{X}}\boldsymbol{v})}{P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X}_i+(\boldsymbol{L}_1^{-1})^{\boldsymbol{X}}\boldsymbol{v})}\times \\
& P(\boldsymbol{S}|\boldsymbol{Y}_i+(\boldsymbol{L}_1^{-1})^{\boldsymbol{y}}\boldsymbol{u},\boldsymbol{X}_i+(\boldsymbol{L}_1^{-1})^{\boldsymbol{X}}\boldsymbol{v})d\boldsymbol{S}d\boldsymbol{u}d\boldsymbol{v} \\
= \ & \int_{\boldsymbol{w}}\int_{\boldsymbol{v}}\int_{\boldsymbol{u}}\left\{\phi(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})\right\}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_i|\boldsymbol{X}_i)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_i|\boldsymbol{X}_i)}d\boldsymbol{u}d\boldsymbol{v}d\boldsymbol{w} \\
& -\int_{\boldsymbol{v}}\int_{\boldsymbol{u}}\left\{\phi(\boldsymbol{u},\boldsymbol{v})\right\}\int_{\boldsymbol{S}}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X}_i)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X}_i)}P(\boldsymbol{S}|\boldsymbol{Y}_i,\boldsymbol{X}_i)d\boldsymbol{S}d\boldsymbol{u}d\boldsymbol{v}+O(\boldsymbol{H}) \\
= \ & \frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_i|\boldsymbol{X}_i)}{P_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}-E\left[\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X}_i)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X}_i)}\middle|\boldsymbol{Y}_i,\boldsymbol{X}_i\right]+O(\boldsymbol{H}).
\end{aligned}$$

Finally, the expected value is

$$E[\boldsymbol{Q}_{ij}|\boldsymbol{S}_i,\boldsymbol{Y}_i,\boldsymbol{X}_i] = E\left[\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X}_i)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X}_i)}\middle|\boldsymbol{Y}_i,\boldsymbol{X}_i\right]-\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_i|\boldsymbol{X}_i)}{P_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}+O(\boldsymbol{H}).$$

Without loss of generality in choice of $\boldsymbol{H}$, it can be shown through identical arguments that

$$E[\boldsymbol{Q}_{ik}|\boldsymbol{S}_i,\boldsymbol{Y}_i,\boldsymbol{X}_i] = E\left[\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S},\underline{\boldsymbol{Y}}|\boldsymbol{X}_i)}{P_{\boldsymbol{\beta}}(\boldsymbol{S},\underline{\boldsymbol{Y}}|\boldsymbol{X}_i)}\middle|\boldsymbol{Y}_i,\boldsymbol{X}_i\right]-\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_i,\boldsymbol{Y}_i|\boldsymbol{X}_i)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_i,\boldsymbol{Y}_i|\boldsymbol{X}_i)}+O(\boldsymbol{H}),$$

and therefore,

$$\begin{aligned}
E[\boldsymbol{Q}_i|\boldsymbol{S}_i,\boldsymbol{Y}_i,\boldsymbol{X}_i] = \ & E\left[\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X}_i)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X}_i)}\middle|\boldsymbol{Y}_i,\boldsymbol{X}_i\right]+E\left[\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S},\underline{\boldsymbol{Y}}|\boldsymbol{X}_i)}{P_{\boldsymbol{\beta}}(\boldsymbol{S},\underline{\boldsymbol{Y}}|\boldsymbol{X}_i)}\middle|\boldsymbol{Y}_i,\boldsymbol{X}_i\right] \\
& -\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_i|\boldsymbol{X}_i)}{P_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}-\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_i,\boldsymbol{Y}_i|\boldsymbol{X}_i)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_i,\boldsymbol{Y}_i|\boldsymbol{X}_i)}+O(\boldsymbol{H}),
\end{aligned}$$

Using arguments similar to those in Pepe (1992), conditional on the validation data $\boldsymbol{S}_i,\boldsymbol{Y}_i,\boldsymbol{X}_i$, each $\boldsymbol{Q}_i$ is a sum over $\underline{V}$ and $\overline{V}$ of independent, identically distributed random variables. Therefore, they converge in probability to their mean $E[\boldsymbol{Q}_i|\boldsymbol{S}_i,\boldsymbol{Y}_i,\boldsymbol{X}_i]$, and their variance is $Var(\boldsymbol{Q}_i|\boldsymbol{S}_i,\boldsymbol{Y}_i,\boldsymbol{X}_i)$, which converges in probability to $Var(E[\boldsymbol{Q}_i|\boldsymbol{S}_i, \boldsymbol{Y}_i, \boldsymbol{X}_i])$, defined in Section 3.2 as $\mathscr{K}_{VS}(\boldsymbol{\beta})$.

Conversely, if we condition on the observed data in $\underline{V}$ and $\overline{V}$, then it is trivial to show that $\boldsymbol{Q}_i$ is a sum over V of independent, identically distributed random variables with mean $\mathbf{0}$ (see equation (B.2)). Finally, by the Lyapunov central limit theorem (Billingsley, 1986), the conditional distribution of $\frac{1}{\sqrt{N}}\boldsymbol{U}_{VS}(\boldsymbol{\beta})$ is asymptotically normal with mean $\mathbf{0}$ and variance

$$\mathscr{I}(\boldsymbol{\beta}) + \frac{(\rho_{\underline{V}} + \rho_{\overline{V}})^2}{\rho_V}\mathscr{K}_{VS}(\boldsymbol{\beta})$$

with $\mathscr{I}(\boldsymbol{\beta})$ as defined in Section 3.2. By Taylor expansion, $\widehat{\boldsymbol{\beta}}_{VS}$ is then asymptotically normal with mean $\mathbf{0}$ and variance

$$\mathscr{I}(\boldsymbol{\beta})^{-1} + \frac{(\rho_{\underline{V}} + \rho_{\overline{V}})^2}{\rho_V}\mathscr{I}(\boldsymbol{\beta})^{-1}\mathscr{K}_{VS}(\boldsymbol{\beta})\mathscr{I}(\boldsymbol{\beta})^{-1}$$

Consistent estimates for these quantities are given in Section 3.2.1.

B.2.2. Asymptotic Normality of $\widehat{\boldsymbol{\beta}}_{HS}$

Many steps in this proof are similar to those explained in Section B.2.1. The score equation can be written as

$$\begin{aligned}\frac{1}{\sqrt{N}}\boldsymbol{U}_{HS}(\boldsymbol{\beta}) =& \frac{1}{\sqrt{N}}\left\{\sum_{i\in V}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}{P_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)} + \sum_{k\in\underline{V}}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_k, \boldsymbol{Y}_k|\boldsymbol{X}_k)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_k, \boldsymbol{Y}_k|\boldsymbol{X}_k)} + \sum_{j\in\overline{V}}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)}\right\}\\ &+ \frac{1}{\sqrt{N}}\sum_{k\in\underline{V}}\left\{\frac{\widehat{\boldsymbol{D}'}_{\boldsymbol{\beta}}(\boldsymbol{S}_k, \boldsymbol{Y}_k|\boldsymbol{X}_k)}{\widehat{P'}_{\boldsymbol{\beta}}(\boldsymbol{S}_k, \boldsymbol{Y}_k|\boldsymbol{X}_k)} - \frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_k, \boldsymbol{Y}_k|\boldsymbol{X}_k)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_k, \boldsymbol{Y}_k|\boldsymbol{X}_k)}\right\} \quad \text{(B.4)}\\ &+ \frac{1}{\sqrt{N}}\sum_{j\in\overline{V}}\left\{\frac{\widehat{\boldsymbol{D}'}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)}{\widehat{P'}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)} - \frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)}\right\} \quad \text{(B.5)}\end{aligned}$$

$$(B.5) = \frac{1}{\sqrt{N}}\sum_{j\in\overline{V}}\int\left\{\frac{\widehat{P'}(\boldsymbol{S}_j, \boldsymbol{y}, \boldsymbol{X}_j)}{\widehat{P'}(\boldsymbol{y}, \boldsymbol{X}_j)} - \frac{P(\boldsymbol{S}_j, \boldsymbol{y}, \boldsymbol{X}_j)\widehat{P'}(\boldsymbol{y}, \boldsymbol{X}_j)}{P^2(\boldsymbol{y}, \boldsymbol{X}_j)}\right\}\psi_j^{\boldsymbol{\beta},\overline{V}}(\boldsymbol{y})d\boldsymbol{y} + o_p(1).$$

We now use $\widehat{P'}(\boldsymbol{S}, \boldsymbol{y}, \boldsymbol{X})$ and $\widehat{P'}(\boldsymbol{y}, \boldsymbol{X})$ as defined in Section 3.2.2.

$$
\begin{aligned}
\widehat{P'}(\boldsymbol{S}_j, \boldsymbol{y}, \boldsymbol{X}_j) &= \prod_{t=1}^{T} \widehat{P}(S_{jt}^{'}, y_t^{'}, \boldsymbol{X}_{jt}^{'}) \\
&= \prod_{t=1}^{T} \frac{1}{n_t} \sum_{\ell \in V_t} \frac{1}{|\boldsymbol{H}_2|^{1/2}} \phi(\boldsymbol{L}_2(S_{jt}^{'} - S_{\ell t}^{'}, y_t^{'} - Y_{\ell t}^{'}, \boldsymbol{X}_{jt}^{'} - \boldsymbol{X}_{\ell t}^{'})) \\
\widehat{P'}(\boldsymbol{y}, \boldsymbol{X}_j) &= \prod_{t=1}^{T} \widehat{P}(y_t^{'}, \boldsymbol{X}_{jt}^{'}) \\
&= \prod_{t=1}^{T} \frac{1}{n_t} \sum_{\ell \in V_t} \frac{1}{|\boldsymbol{H}_1|^{1/2}} \phi(\boldsymbol{L}_1(y_t^{'} - Y_{\ell t}^{'}, \boldsymbol{X}_{jt}^{'} - \boldsymbol{X}_{\ell t}^{'}))
\end{aligned}
$$

By the Continuous Mapping Theorem and Slutsky's Theorem, and using the consistency of these probability estimates, it can be shown through basic manipulation that

$$
\prod_{t=1}^{T} \widehat{P}(S_{jt}^{'}, y_t^{'}, \boldsymbol{X}_{jt}^{'}) = \widehat{P}(S_{j1}^{'}, y_1^{'}, \boldsymbol{X}_{j1}^{'}) \prod_{t=2}^{T} P(S_{jt}, y_t, \boldsymbol{X}_{jt})
$$
$$
\prod_{t=1}^{T} \widehat{P}(y_t^{'}, \boldsymbol{X}_{jt}^{'}) = \widehat{P}(y_1^{'}, \boldsymbol{X}_{j1}^{'}) \prod_{t=2}^{T} P(y_t, \boldsymbol{X}_{jt}).
$$

To illustrate an example of this, consider the situation with two timepoints, and let $\widehat{P'}_t = \widehat{P}(y_t^{'}, \boldsymbol{X}_{jt}^{'})$ and $P_t = P(y_t, \boldsymbol{X}_{jt})$ for $t = 1, 2$. We know that these are consistent estimates; i.e.,

$$\widehat{P'}_t \xrightarrow{p} P_t \text{ for } t = 1, 2.$$

Now specify a continuous, differentiable function such that

$$g(P_1, P_2) = P_1 P_2$$

Then, by the Continuous Mapping Theorem, we have

$$g(\widehat{P'}_1, \widehat{P'}_2) \xrightarrow{p} g(P_1, P_2)$$

which implies

$$\widehat{P'}_1\widehat{P'}_2 = P_1P_2.$$

Finally, by Slutsky's Theorem,

$$(\widehat{P'}_1 - P_1)(\widehat{P'}_2 - P_2) \xrightarrow{p} 0.$$

We can now use these properties to establish

$$\begin{aligned}
\widehat{P'}_1\widehat{P'}_2 =& \widehat{P'}_1\widehat{P'}_2 \pm \widehat{P'}_1P_2 \pm P_1\widehat{P'}_2 \pm P_1P_2 \\
=& (\widehat{P'}_1\widehat{P'}_2 - \widehat{P'}_1P_2 - P_1\widehat{P'}_2 + P_1P_2) + \widehat{P'}_1P_2 + P_1\widehat{P'}_2 - P_1P_2 \\
=& (\widehat{P'}_1 - P_1)(\widehat{P'}_2 - P_2) + (\widehat{P'}_1P_2 + P_1\widehat{P'}_2 - P_1P_2) \\
=& (\widehat{P'}_1 - P_1)(\widehat{P'}_2 - P_2) + (\widehat{P'}_1P_2 + P_1[\widehat{P'}_2 - P_2]) \\
=& \widehat{P'}_1P_2 + o_p(1).
\end{aligned}$$

These steps are easily extended to any number of timepoints $t > 2$. Therefore, we can write (B.5) as

$$(B.5) = \frac{1}{\sqrt{N}}\sum_{j\in\overline{V}}\frac{1}{n_{V_1}}\sum_{\ell\in V_1}\int \zeta_\ell^{V_1,\overline{V}}(\boldsymbol{y})\boldsymbol{\psi}_j^{\boldsymbol{\beta},\overline{V}}(\boldsymbol{y})d\boldsymbol{y} + o_p(1)$$

where

$$\zeta_\ell^{V_1,\overline{V}}(\boldsymbol{y}) = \frac{|\boldsymbol{H}_2|^{-1/2}\phi'_{2\ell j}\prod_{t=2}^T P(S_{jt}, y_t, \boldsymbol{X}_{jt})}{P(\boldsymbol{y}, \boldsymbol{X}_j)} - \frac{|\boldsymbol{H}_1|^{-1/2}\phi'_{1\ell j}P(\boldsymbol{S}_j, \boldsymbol{y}, \boldsymbol{X}_j)\prod_{t=2}^T P(y_t, \boldsymbol{X}_{jt})}{P^2(\boldsymbol{y}, \boldsymbol{X}_j)}$$

and

$$\begin{aligned}
\phi'_{2\ell j} =& \phi(\boldsymbol{L}_2(S'_{j1} - S'_{i1}, y'_1 - Y'_{i1}, \boldsymbol{X}'_{j1} - \boldsymbol{X}'_{i1})) \\
\phi'_{1\ell j} =& \phi(\boldsymbol{L}_1(y'_1 - Y'_{i1}, \boldsymbol{X}'_{j1} - \boldsymbol{X}'_{i1})).
\end{aligned}$$

Now we have that

$$(B.5) = \frac{1}{n_{V_1}\sqrt{N}} \sum_{\ell \in V_1} \sum_{j \in \overline{V}} \boldsymbol{Q}_{\ell j} + o_p(1)$$

where

$$\boldsymbol{Q}_{\ell j} = \int \zeta_{\ell}^{V_1,\overline{V}}(\boldsymbol{y}) \boldsymbol{\psi}_j^{\boldsymbol{\beta},\overline{V}}(\boldsymbol{y}) d\boldsymbol{y} + o_p(1).$$

We can use similar arguments to derive the formula for $\boldsymbol{Q}_{\ell k}$. As mentioned in Section 3.2.2, the important distinction here is not to use subject k in the kernel density estimation.

$$\begin{aligned}(B.4) =& \frac{1}{\sqrt{N}} \sum_{k \in \underline{V}} \left\{ \frac{\widehat{P'}(\boldsymbol{S}_k, \boldsymbol{y}_{m_k}, \underline{\boldsymbol{Y}}_k, \boldsymbol{X}_k)}{\widehat{P'}(\boldsymbol{y}_{m_k}, \underline{\boldsymbol{Y}}_k, \boldsymbol{X}_k)} - \frac{P(\boldsymbol{S}_k, \boldsymbol{y}_{m_k}, \underline{\boldsymbol{Y}}_k, \boldsymbol{X}_k)\widehat{P'}(\boldsymbol{y}, \boldsymbol{X}_k)}{P^2(\boldsymbol{y}_{m_k}, \boldsymbol{X}_k)} \right\} \boldsymbol{\psi}_k^{\boldsymbol{\beta},\underline{V}}(\boldsymbol{y}_{m_k}, \underline{\boldsymbol{Y}}_k) d\boldsymbol{y}_{m_k} \\ & \times + o_p(1) \\ =& \frac{1}{\sqrt{N}} \frac{1}{n_{V_1} - 1} \sum_{\substack{\ell \in V_1 \\ \ell \neq k}} \sum_{k \in \underline{V}} \int \zeta_{\ell}^{V_1,\underline{V}}(\boldsymbol{y}_{m_k}) \boldsymbol{\psi}_k^{\boldsymbol{\beta},\underline{V}}(\boldsymbol{y}_{m_k}) d\boldsymbol{y}_{m_k} + o_p(1) \\ =& \frac{1}{\sqrt{N}} \sum_{\substack{\ell \in V_1 \\ \ell \neq k}} \sum_{k \in \underline{V}} \boldsymbol{Q}_{\ell k} + o_p(1),\end{aligned}$$

where, as before,

$$\begin{aligned}\zeta^{V_1,\underline{V}}(\boldsymbol{y}_{m_k}) =& \frac{|\boldsymbol{H}_2|^{-1/2} \phi'_{2\ell k} \prod_{t=2}^{T} P(S_{kt}, y_t, \boldsymbol{X}_{kt})}{P(\boldsymbol{y}_{m_k}, \underline{\boldsymbol{Y}}_k, \boldsymbol{X}_k)} \\ & - \frac{|\boldsymbol{H}_1|^{-1/2} \phi'_{1\ell k} P(\boldsymbol{S}_k, \boldsymbol{y}_{m_k}, \underline{\boldsymbol{Y}}_k, \boldsymbol{X}_k) \prod_{t=2}^{N} P(y_t, \boldsymbol{X}_{kt})}{P^2(\boldsymbol{y}_{m_k}, \underline{\boldsymbol{Y}}_k, \boldsymbol{X}_k)} \\ \phi'_{2\ell k} =& \phi(\boldsymbol{L}_2(S'_{k1} - S'_{\ell 1}, y'_1 - Y'_{\ell 1}, \boldsymbol{X}_{k1} - \boldsymbol{X}_{\ell 1})) \\ \phi'_{1\ell k} =& \phi(\boldsymbol{L}_1(y'_1 - Y'_{\ell 1}, \boldsymbol{X}_{k1} - \boldsymbol{X}_{\ell 1})).\end{aligned}$$

Furthermore, if subject k is missing the true outcome, then replace $\frac{1}{n_{V_1}-1}$ with $\frac{1}{n_{V_1}}$, because then k is not a member of V_1. Since this will not affect more than one subject, this nuance has no impact on the variance as $N \to \infty$. To establish $E[\boldsymbol{Q}_{\ell} | \boldsymbol{S}_{\ell}, \boldsymbol{Y}_{\ell}, \boldsymbol{X}_{\ell}]$, repeat the same integration steps as with $\boldsymbol{Q}_i$ in Section B.2.1.

$$
\begin{aligned}
E[\ \ \boldsymbol{Q}_{\ell j}|\boldsymbol{S}_\ell,\boldsymbol{Y}_\ell,\boldsymbol{X}_\ell] =& E_{\boldsymbol{X}}\left[E_{\boldsymbol{S}}[\boldsymbol{Q}_{\ell j}|\boldsymbol{X}]\middle|\boldsymbol{S}_\ell,\boldsymbol{Y}_\ell,\boldsymbol{X}_\ell\right]\\
=&E_{\boldsymbol{X}}\left[\int_{\boldsymbol{S}}\int_{\boldsymbol{y}}\zeta_\ell^{V_1,\overline{V}}(\boldsymbol{y})\boldsymbol{\psi}_j^{\boldsymbol{\beta},\overline{V}}(\boldsymbol{y})P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})d\boldsymbol{y}d\boldsymbol{S}\middle|\boldsymbol{S}_\ell,\boldsymbol{Y}_\ell,\boldsymbol{X}_\ell\right]\\
=&E_{\boldsymbol{X}}\left[\int_{\boldsymbol{S}}\int_{\boldsymbol{y}}\zeta_\ell^{V_1,\overline{V}}(\boldsymbol{y})\left\{\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})}{P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}-\frac{P_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}{P^2_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}\right\}P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})d\boldsymbol{y}d\boldsymbol{S}\middle|\boldsymbol{S}_\ell,\boldsymbol{Y}_\ell,\boldsymbol{X}_\ell\right]\\
=&E_{\boldsymbol{X}}\left[\int_{\boldsymbol{S}}\int_{\boldsymbol{y}}\zeta_\ell^{V_1,\overline{V}}(\boldsymbol{y})\left\{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})-\frac{P_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}{P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}\right\}d\boldsymbol{y}d\boldsymbol{S}\middle|\boldsymbol{S}_\ell,\boldsymbol{Y}_\ell,\boldsymbol{X}_\ell\right]\\
=&E_{\boldsymbol{X}}\left[\int_{\boldsymbol{S}}\int_{\boldsymbol{y}}\zeta_\ell^{V_1,\overline{V}}(\boldsymbol{y})\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})d\boldsymbol{y}d\boldsymbol{S}\middle|\boldsymbol{S}_\ell,\boldsymbol{Y}_\ell,\boldsymbol{X}_\ell\right] \qquad \text{(B.6a)}\\
&-E_{\boldsymbol{X}}\left[\int_{\boldsymbol{S}}\int_{\boldsymbol{y}}\zeta_\ell^{V_1,\overline{V}}(\boldsymbol{y})\frac{P_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}{P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}d\boldsymbol{y}d\boldsymbol{S}\middle|\boldsymbol{S}_\ell,\boldsymbol{Y}_\ell,\boldsymbol{X}_\ell\right] \qquad \text{(B.6b)}
\end{aligned}
$$

Since choice of the bandwidth matrix $\boldsymbol{H}$ is flexible, we recommend accounting for $\boldsymbol{\Sigma}^{'}$ in the choice of $\boldsymbol{H}$, by multiplying the original choice of $\boldsymbol{H}$ by $\boldsymbol{\Sigma}^{'T}\boldsymbol{\Sigma}^{'}$. This will simplify the integral below, since we will not need to consider the additional matrices $\boldsymbol{L}_B^{-1}\boldsymbol{L}_G$ in the u-substitution. Again, we continue to write $\boldsymbol{H}$ to denote any choice of bandwidth matrix $\boldsymbol{H}$.

$$
\begin{aligned}
(B.6a) =&E_{\boldsymbol{X}}\left[\int_{\boldsymbol{S}}\int_{\boldsymbol{y}}\left\{\frac{|\boldsymbol{H}_2|^{-1/2}\phi^{'}_{2\ell j}\prod_{t=2}^T P(S_t,y_t,\boldsymbol{X}_t)}{P(\boldsymbol{y},\boldsymbol{X})}\right.\right.\\
&\left.\left.-\frac{|\boldsymbol{H}_1|^{-1/2}\phi^{'}_{1\ell j}P(\boldsymbol{S},\boldsymbol{y},\boldsymbol{X})\prod_{t=2}^T P(y_t,\boldsymbol{X}_t)}{P^2(\boldsymbol{y},\boldsymbol{X})}\right\}\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})d\boldsymbol{y}d\boldsymbol{S}\middle|\boldsymbol{S}_\ell,\boldsymbol{Y}_\ell,\boldsymbol{X}_\ell\right]\\
=&\int_{\boldsymbol{X}}\int_{\boldsymbol{S}}\int_{\boldsymbol{y}}\left\{|\boldsymbol{H}_2|^{-1/2}\phi^{'}_{2\ell j}\prod_{t=2}^T P(S_t,y_t,\boldsymbol{X}_t)\right.\\
&\left.-\frac{|\boldsymbol{H}_1|^{-1/2}\phi^{'}_{1\ell j}P(\boldsymbol{S},\boldsymbol{y},\boldsymbol{X})\prod_{t=2}^T P(y_t,\boldsymbol{X}_t)}{P(\boldsymbol{y},\boldsymbol{X})}\right\}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})}{P_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})}d\boldsymbol{y}d\boldsymbol{S}d\boldsymbol{X}\\
=&\int_{\boldsymbol{X}}\int_{\boldsymbol{S}}\int_{\boldsymbol{y}}\left\{|\boldsymbol{H}_2|^{-1/2}\phi^{'}_{2\ell j}\prod_{t=2}^T P(S_t,y_t,\boldsymbol{X}_t)\right.\\
&\left.-\frac{|\boldsymbol{H}_1|^{-1/2}\phi^{'}_{1\ell j}P(\boldsymbol{S},\boldsymbol{y},\boldsymbol{X})\prod_{t=2}^T P(y_t,\boldsymbol{X}_t)}{P(\boldsymbol{y},\boldsymbol{X})}\right\}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})}{P_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})}d\boldsymbol{y}d\boldsymbol{S}d\boldsymbol{X}.
\end{aligned}
$$

Making u-subsitutions as in Section B.2.1, we have

$$
\begin{aligned}
(B.6a) =& \int_{\boldsymbol{w}}\int_{\boldsymbol{v}}\int_{\boldsymbol{u}} \frac{|\boldsymbol{L}_2^{-1}|}{|\boldsymbol{H}_2|^{1/2}}\phi(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})\prod_{t=2}^{T} P(S_{\ell t}, Y_{\ell t}, \boldsymbol{X}_{\ell t})\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{Y}_\ell|\boldsymbol{X}_\ell)}{P_{\boldsymbol{\beta}}(\boldsymbol{Y}_\ell|\boldsymbol{X}_\ell)}d\boldsymbol{u}d\boldsymbol{v}d\boldsymbol{w}\\
&-\int_{\boldsymbol{v}}\int_{\boldsymbol{u}}|\boldsymbol{L}_1^{-1}||\boldsymbol{H}_1|^{-1/2}\phi(\boldsymbol{u},\boldsymbol{v})\prod_{t=2}^{T}P(Y_{\ell t},\boldsymbol{X}_{\ell t})\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{Y}_\ell|\boldsymbol{X}_\ell)}{P_{\boldsymbol{\beta}}(\boldsymbol{Y}_\ell|\boldsymbol{X}_\ell)}\\
&\times\int_{\boldsymbol{S}}P(\boldsymbol{S}|\boldsymbol{Y}_\ell,\boldsymbol{X}_\ell)d\boldsymbol{S}d\boldsymbol{u}d\boldsymbol{v}+O(\boldsymbol{H})\\
=&\Big\{\prod_{t=2}^{T}P(S_{\ell t},Y_{\ell t},\boldsymbol{X}_{\ell t})-\prod_{t=2}^{T}P(Y_{\ell t},\boldsymbol{X}_{\ell t})\Big\}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{Y}_\ell|\boldsymbol{X}_\ell)}{P_{\boldsymbol{\beta}}(\boldsymbol{Y}_\ell|\boldsymbol{X}_\ell)}+O(\boldsymbol{H})
\end{aligned}
$$

and

$$
\begin{aligned}
(B.6b) =& \int_{\boldsymbol{X}}\int_{\boldsymbol{y}}\int_{\boldsymbol{S}}\Big\{\frac{|\boldsymbol{H}_2|^{-1/2}\phi'_{2\ell j}\prod_{t=2}^{T}P(S_t,y_t,\boldsymbol{X}_t)}{P(\boldsymbol{y},\boldsymbol{X})}\\
&-\frac{|\boldsymbol{H}_1|^{-1/2}\phi'_{1\ell j}P(\boldsymbol{S},\boldsymbol{y},\boldsymbol{X})\prod_{t=2}^{T}P(y_t,\boldsymbol{X}_t)}{P^2(\boldsymbol{y},\boldsymbol{X})}\Big\}\frac{P_{\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{X})\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})P(\boldsymbol{X})}{P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}d\boldsymbol{S}d\boldsymbol{y}d\boldsymbol{X}\\
=&\int_{\boldsymbol{X}}\int_{\boldsymbol{y}}\int_{\boldsymbol{S}}\Big\{|\boldsymbol{H}_2|^{-1/2}\phi'_{2\ell j}\prod_{t=2}^{T}P(S_t,y_t,\boldsymbol{X}_t)\\
&-\frac{|\boldsymbol{H}_1|^{-1/2}\phi'_{1\ell j}P(\boldsymbol{S},\boldsymbol{y},\boldsymbol{X})\prod_{t=2}^{T}P(y_t,\boldsymbol{X}_t)}{P(\boldsymbol{y},\boldsymbol{X})}\Big\}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}{P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X})}d\boldsymbol{S}d\boldsymbol{y}d\boldsymbol{X}\\
=&\int_{\boldsymbol{w}}\int_{\boldsymbol{v}}\int_{\boldsymbol{u}}|\boldsymbol{H}_2|^{-1/2}|\boldsymbol{L}_2^{-1}|\phi(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})\prod_{t=2}^{T}P(S_{\ell t},Y_{\ell t},\boldsymbol{X}_{\ell t})\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_\ell|\boldsymbol{X}_\ell)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_\ell|\boldsymbol{X}_\ell)}d\boldsymbol{u}d\boldsymbol{v}d\boldsymbol{w}\\
&-\int_{\boldsymbol{v}}\int_{\boldsymbol{u}}|\boldsymbol{H}_1|^{-1/2}|\boldsymbol{L}_1^{-1}|\phi(\boldsymbol{u},\boldsymbol{v})\prod_{t=2}^{T}P(Y_{\ell t},\boldsymbol{X}_{\ell t})\int_{\boldsymbol{S}}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X}_\ell)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X}_\ell)}P(\boldsymbol{S}|Y_\ell,\boldsymbol{X}_\ell)d\boldsymbol{S}d\boldsymbol{u}d\boldsymbol{v}\\
&+O(\boldsymbol{H})\\
=&\prod_{t=2}^{T}P(S_{\ell t},Y_{\ell t},\boldsymbol{X}_{\ell t})\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_\ell|\boldsymbol{X}_\ell)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_\ell|\boldsymbol{X}_\ell)}-\prod_{t=2}^{T}P(Y_{\ell t},\boldsymbol{X}_{\ell t})\int_{\boldsymbol{S}}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X}_\ell)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}|\boldsymbol{X}_\ell)}P(\boldsymbol{S}|Y_\ell,\boldsymbol{X}_\ell)d\boldsymbol{S}\\
&+O(\boldsymbol{H})
\end{aligned}
$$

With similar steps for $E\big[\boldsymbol{Q}_{\ell k}\big|\boldsymbol{S}_\ell,\boldsymbol{Y}_\ell,\boldsymbol{X}_\ell\big]$, we have that $E\big[\boldsymbol{Q}_\ell\big|\boldsymbol{S}_\ell,\boldsymbol{Y}_\ell,\boldsymbol{X}_\ell\big] = E\big[\boldsymbol{Q}_{\ell k}\big|\boldsymbol{S}_\ell,\boldsymbol{Y}_\ell,\boldsymbol{X}_\ell\big] +$

$E\big[\boldsymbol{Q}_{\ell j}\big|\boldsymbol{S}_\ell,\boldsymbol{Y}_\ell,\boldsymbol{X}_\ell\big]$. Therefore, writing $\boldsymbol{Q}_\ell = \frac{1}{n_{\underline{V}}+n_{\overline{V}}}\big\{\sum_{k\in\underline{V}}\boldsymbol{Q}_{\ell k}+\sum_{j\in\overline{V}}\boldsymbol{Q}_{\ell j}\big\}$, the score equation is

$$\begin{aligned}\frac{1}{\sqrt{N}}\boldsymbol{U}_{HS}(\boldsymbol{\beta}) =&\frac{1}{\sqrt{N}}\left\{\sum_{i\in V}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}{P_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}+\sum_{k\in\underline{V}}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_k,\boldsymbol{Y}_k|\boldsymbol{X}_k)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_k,\boldsymbol{Y}_k|\boldsymbol{X}_k)}+\sum_{j\in\overline{V}}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)}\right\}\\ &+\frac{1}{\sqrt{N}(n_{V_1}-1)}\sum_{\substack{\ell\in V_1\\ \ell\neq k}}\frac{n_{\underline{V}}+n_{\overline{V}}}{n_{\underline{V}}+n_{\overline{V}}}\left\{\sum_{k\in\underline{V}}\boldsymbol{Q}_{\ell k}+\sum_{j\in\overline{V}}\boldsymbol{Q}_{\ell j}\right\}+o_p(1)\\ =&\frac{1}{\sqrt{N}}\left\{\sum_{i\in V}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}{P_{\boldsymbol{\beta}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}+\sum_{k\in\underline{V}}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_k,\boldsymbol{Y}_k|\boldsymbol{X}_k)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_k,\boldsymbol{Y}_k|\boldsymbol{X}_k)}+\sum_{j\in\overline{V}}\frac{\boldsymbol{D}_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)}{P_{\boldsymbol{\beta}}(\boldsymbol{S}_j|\boldsymbol{X}_j)}\right\}\\ &+\frac{1}{\sqrt{(n_{V_1}-1)}}\frac{n_{\underline{V}}+n_{\overline{V}}}{\sqrt{N}\sqrt{(n_{V_1}-1)}}\sum_{\substack{\ell\in V_1\\ \ell\neq k}}\boldsymbol{Q}_\ell+o_p(1).\end{aligned}$$

Using the same arguments as in Section B.2.1, the variance of the score equation is $\mathscr{I}(\boldsymbol{\beta})+\mathscr{K}_{HS}(\boldsymbol{\beta})$, where $\mathscr{K}_{HS}(\boldsymbol{\beta}) = \left(\frac{(\rho_{\underline{V}}+\rho_{\overline{V}})^2}{\rho_1}\right)\sum_{\ell\in V_1}\boldsymbol{Q}_\ell$. Finally, the asymptotic variance of $\widehat{\boldsymbol{\beta}}_{HS}$, the solution to $\boldsymbol{U}_{HS}(\boldsymbol{\beta})$, is

$$\mathscr{I}^{-1}(\boldsymbol{\beta})+\left(\frac{(\rho_{\underline{V}}+\rho_{\overline{V}})^2}{\rho_1}\right)\mathscr{I}^{-1}(\boldsymbol{\beta})\mathscr{K}_{HS}(\boldsymbol{\beta})\mathscr{I}^{-1}(\boldsymbol{\beta}).$$

B.3. Additional Details on Natural Cubic Splines

As there is no analytical integral for $\widehat{P}(\boldsymbol{S}_j|\boldsymbol{X}_j)$ or $\widehat{P}(\boldsymbol{S}_k|\underline{\boldsymbol{Y}}_k,\boldsymbol{X}_k)$, numerical integration techniques were required to calculate these quantities. The dimension of the integral for each subject j is equal to the number of missing observations for subject j. Due to the complexity of the integrand for these quantities, many numerical integration methods are not practical for more than one missing observation per subject, due to exponentially increased computing time. Therefore, the results presented in this dissertation were obtained not by approximating the value of the integral with numerical integration, but by approximating the integrand with a natural cubic spline. The cubic spline is constructed in the following way. All notation defined here is unique to this appendix.

Given $n+1$ pairs of data points (x_i, y_i), $i = 0, 1, ..., n$, the goal is to find a piecewise cubic polynomial

$S(x)$ such that

$$S(x) = \begin{cases} S_0(x) = a_0 + b_0(x - x_i) + c_0(x - x_i)^2 + d_0(x - x_i)^3 & x_0 \leq x \leq x_1 \\ S_1(x) = a_1 + b_1(x - x_i) + c_1(x - x_i)^2 + d_1(x - x_i)^3 & x_1 \leq x \leq x_2 \\ S_2(x) = a_2 + b_2(x - x_i) + c_2(x - x_i)^2 + d_2(x - x_i)^3 & x_2 \leq x \leq x_3 \\ \vdots \\ S_{n-1}(x) = a_{n-1} + b_{n-1}(x - x_i) + c_{n-1}(x - x_i)^2 + d_{n-1}(x - x_i)^3 & x_{n-1} \leq x \leq x_n \end{cases}$$

$S(x)$ is also subject to the following conditions:

1. Interpolating data: $S_i(x_i) = y_i, i = 0, 1, .., n$.
2. Continuity at interior points: $S_i(x_{i+1}) = S_{i+1}(x_{i+1}), i = 0, 1, ..., (n-2)$.
3. Continuity of slope at interior points: $S_i^{'}(x_{i+1}) = S_{i+1}^{'}(x_{i+1}), i = 0, 1, ..., (n-2)$.
4. Continuity of curvature at interior points: $S_i^{''}(x_{i+1}) = S_{i+1}^{''}(x_{i+1}), i = 0, 1, ..., (n-2)$.
5. Natural cubic spline condition: $S_0^{''}(x_0) = 0$ and $S_{n-1}^{''}(x_n) = 0$.

From here, a system of equations can be set up in order to solve for a_i, b_i, c_i, and d_i. Condition 1 implies $S_i(x_i) = a_i = y_i$, $i = 0, 1, ..., (n-1)$, and $S_{n-1}(x_n) = y_n$, as well as

$$y_{n-1} + b_{n-1}(x_n - x_{n-1}) + c_{n-1}(x_n - x_{n-1})^2 + d_{n-1}(x_n - x_{n-1})^3 = y_n.$$

Condition 2 implies

$$y_i + b_i(x_{i+1} - x_i) + c_i(x_{i+1} - x_i)^2 + d_i(x_{i+1} - x_i)^3 = y_{i+1}, \quad i = 0, 1, ..., (n-2).$$

From condition 3, we have

$$b_i + 2c_i(x_{i+1} - x_i) + 3d_i(x_{i+1} - x_i)^2 = b_{i+1}, \quad i = 0, 1, ..., (n-2).$$

Finally, from condition 4, we have

$$2c_i + 6d_i(x_{i+1} - x_i) = 2c_{i+1}, \quad i = 0, 1, ..., (n-2),$$

and the natural spline condition implies

$$2c_0 = 0 \text{ and } 6d_{n-1}(x_n - x_{n-1}) + 2c_{n-1} = 0.$$

Writing $h_i = x_{i+1} - x_i$, $i = 0, 1, ..., (n-1)$, and vectors $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, and $\boldsymbol{d}$, we can construct a system of equations to solve for the coefficients of $S(x)$. First, solve $\boldsymbol{Ac} = \boldsymbol{v}$ for $\boldsymbol{c}$, where

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ h_0 & 2(h_0 + h_1) & h_1 & 0 & \dots & 0 & 0 & 0 \\ 0 & h_1 & 2(h_1 + h_2) & h_2 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & h_{n-2} & 2(h_{n-2} + h_{n-1}) & h_{n-1} \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix}$$

and

$$\boldsymbol{v} = \begin{bmatrix} 0 \\ 3\left[\frac{1}{h_1}(y_2 - y_1) - \frac{1}{h_0}(y_1 - y_0)\right] \\ 3\left[\frac{1}{h_2}(y_3 - y_2) - \frac{1}{h_1}(y_2 - y_1)\right] \\ \vdots \\ 3\left[\frac{1}{h_{n-1}}(y_n - y_{n-1}) - \frac{1}{h_{n-2}}(y_{n-1} - y_{n-2})\right] \\ 0 \end{bmatrix}.$$

Next, calculate b_i and d_i as

$$b_i = \frac{1}{h_i}(a_{i+1} - a_i) - \frac{h_i}{3}(2c_i + c_{i+1}),$$
$$d_i = \frac{1}{3h_i}(c_{i+1} - c_i),$$
$$i = 0, 1, ..., (n-1).$$

Once all coefficients are obtained, $\widehat{P}(\boldsymbol{S}_j|\boldsymbol{X}_j)$ can be approximated as

$$\widehat{P}(\boldsymbol{S}_j|\boldsymbol{X}_j) = \int \widehat{P}(\boldsymbol{S}_j|y, \boldsymbol{X}_j)P_{\boldsymbol{\beta}}(y|\boldsymbol{X})dy \approx \int S(x)dx.$$

$S(x)$ has a closed-form integral, which is

$$\sum_{i=1}^{n-1} a_i(x_{i+1} - x_i) + \frac{1}{2}b_i(x_{i+1} - x_i)^2 + \frac{1}{3}c_i(x_{i+1} - x_i)^3 + \frac{1}{4}d_i(x_{i+1} - x_i)^4,$$

and now the likelihood can be calculated in full for one missing observation per subject. For a linear mixed-effects model, the missing outcome is assumed to be normally distributed, and thus has an infinite domain. This may seem to be problematic, as the cubic spline is defined over a finite set of intervals and thus the integral of $S(x)$ cannot be an indefinite integral. However, since the kernel function of $\widehat{P}(\boldsymbol{S}_j|y, \boldsymbol{X}_j)$ is 0 for large inputs, $\widehat{P}(\boldsymbol{S}_j|y, \boldsymbol{X}_j)P_{\boldsymbol{\beta}}(y|\boldsymbol{X}_j)$ is only nonzero over some finite domain. Therefore, in order to appropriately approximate $\widehat{P}(\boldsymbol{S}_j|y, \boldsymbol{X}_j)P_{\boldsymbol{\beta}}(y|\boldsymbol{X}_j)$ with $S(x)$, we need only to define $S(x)$ over a wide enough domain such that all inputs resulting in a nonzero function value are contained within the specified domain of $S(x)$.

In order to extend the method to two missing observations per subject, the cubic spline must be extended to two dimensions, a case often referred to as a thin plate spline. The extension is straightforward. The cubic spline is expressed as a plane in two dimensions, x and y, for data points (x_i, y_j, z_{ij}), $i = 0, 1, ..., n$, $j = 0, 1, ..., n$. The piecewise cubic thin plate spline $S(x, y)$ is now

$$\begin{aligned} S_{ij}(x, y) =& a_{ij} + b_{ij}(x - x_i) + c_{ij}(x - x_i)^2 + d_{ij}(x - x_i)^3 + e_{ij}(y - y_j) + f_{ij}(y - y_j)^2 \\ &+ g_{ij}(y - y_j)^3, \quad x_i \leq x \leq x_{i+1}, \quad y_j \leq y \leq y_{j+1} \end{aligned}$$

for $i = 0, 1, ...(n-1)$, $j = 0, 1, ..., (n-1)$.

Conditions 1 through 5 are also extended:

1. Interpolating data: $S_{ij}(x_i, y_j) = z_{ij}, \quad i = 0, 1, .., n, \quad j = 0, 1, ..., n$.

2. Continuity at interior points:

$$S_{ij}(x_{i+1}, y_j) = S_{i+1,j}(x_{i+1}, y_j)$$
$$S_{ij}(x_i, y_{j+1}) = S_{i,j+1}(x_i, y_{j+1})$$
$$i = 0, 1, ..., (n-2), \quad j = 0, 1, .., (n-2).$$

3. Continuity of slope at interior points:

$$S'_{ij}(x_{i+1}, y_j) = S'_{i+1,j}(x_{i+1}, y_j)$$
$$S'_{ij}(x_i, y_{j+1}) = S'_{i,j+1}(x_i, y_{j+1})$$
$$i = 0, 1, ..., (n-2), \quad j = 0, 1, .., (n-2).$$

4. Continuity of curvature at interior points:

$$S''_{ij}(x_{i+1}, y_j) = S''_{i+1,j}(x_{i+1}, y_j)$$
$$S''_{ij}(x_i, y_{j+1}) = S''_{i,j+1}(x_i, y_{j+1})$$
$$i = 0, 1, ..., (n-2), \quad j = 0, 1, .., (n-2).$$

5. Natural cubic spline condition:

$$S''_{00}(x_0, y_0) = 0$$
$$S''_{0j}(x_0, y_j) = 0$$
$$S''_{i0}(x_i, y_0) = 0$$
$$S''_{n-1,j}(x_n, y_j) = 0$$
$$S''_{i,n-1}(x_i, y_n) = 0$$
$$S''_{n-1,n-1}(x_n, y_n) = 0$$
$$i = 0, 1, ..., n, \quad j = 0, 1, .., n.$$

Now, define $h_i = (x_{i+1} - x_i)$ and $k_j = (y_{j+1} - y_j)$. Further define $\boldsymbol{c}_j$ as the vector of coefficients

c_{ij} for $i = 0, 1, ..., n$, and $\boldsymbol{f}_i$ as the vector of coefficients f_{ij} for $j = 0, 1, ..., n$. Define $\boldsymbol{b}_j$, $\boldsymbol{d}_j$, $\boldsymbol{e}_i$, and $\boldsymbol{g}_i$ likewise. We can solve for $\boldsymbol{c}_j$ using equations $\boldsymbol{A}\boldsymbol{c}_j = \boldsymbol{v}_j$, and additionally solve for $\boldsymbol{f}_i$ using equations $\boldsymbol{B}\boldsymbol{f}_i = \boldsymbol{w}_i$, where

$$\boldsymbol{B} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ k_0 & 2(k_0 + k_1) & k_1 & 0 & \dots & 0 & 0 & 0 \\ 0 & k_1 & 2(k_1 + k_2) & k_2 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & k_{n-2} & 2(k_{n-2} + k_{n-1}) & k_{n-1} \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix}$$

$$\boldsymbol{w}_i = \begin{bmatrix} 0 \\ 3\left[\frac{1}{k_1}(z_{i2} - z_{i1}) - \frac{1}{k_0}(z_{i1} - z_{i0})\right] \\ 3\left[\frac{1}{k_2}(z_{i3} - z_{i2}) - \frac{1}{k_1}(z_{i2} - z_{i1})\right] \\ \vdots \\ 3\left[\frac{1}{k_{n-1}}(z_{in} - z_{i,n-1}) - \frac{1}{k_{n-2}}(z_{i,n-1} - z_{i,n-2})\right] \\ 0 \end{bmatrix},$$

and

$$\boldsymbol{v}_j = \begin{bmatrix} 0 \\ 3\left[\frac{1}{h_1}(z_{2j} - z_{1j}) - \frac{1}{h_0}(z_{1j} - z_{0j})\right] \\ 3\left[\frac{1}{h_2}(z_{3j} - z_{2j}) - \frac{1}{h_1}(z_{2j} - z_{1j})\right] \\ \vdots \\ 3\left[\frac{1}{h_{n-1}}(z_{nj} - z_{n-1,j}) - \frac{1}{h_{n-2}}(z_{n-1,j} - z_{n-2,j})\right] \\ 0 \end{bmatrix}.$$

Finally, compute the remaining coefficients as

$$\begin{aligned}
b_{ij} &= \frac{1}{h_i}(a_{i+1,j} - a_{ij}) - \frac{h_i}{3}(2c_{ij} + c_{i+1,j}) \\
d_{ij} &= \frac{1}{3h_i}(c_{i+1,j} - c_{ij}) \\
e_{ij} &= \frac{1}{k_j}(a_{i+1,j} - a_i) - \frac{k_j}{3}(2f_{ij} + f_{i,j+1}) \\
g_{ij} &= \frac{1}{3k_j}(f_{i,j+1} - f_{ij}) \\
i &= 0, 1, ..., (n-1), \quad j = 0, 1, ..., (n-1).
\end{aligned}$$

Now, $\widehat{P}(\boldsymbol{S}_j | \boldsymbol{y}, \boldsymbol{X}_j)$ can be approximated for two missing observations per subject as the integral of $S(x, y)$, which is

$$\sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \Big[z_{ij}(x_{i+1} - x_i)(y_{j+1} - y_j) + \frac{1}{2} b_{ij}(x_{i+1} - x_i)^2 (y_{j+1} - y_j) + \frac{1}{3} c_{ij}(x_{i+1} - x_i)^3 (y_{j+1} - y_j)$$
$$+ \frac{1}{4} d_{ij}(x_{i+1} - x_i)^4 (y_{j+1} - y_j) + \frac{1}{2} e_{ij}(x_{i+1} - x_i)(y_{j+1} - y_j)^2 + \frac{1}{3} f_{ij}(x_{i+1} - x_i)(y_{j+1} - y_j)^3$$
$$+ \frac{1}{4} g_{ij}(x_{i+1} - x_i)(y_{j+1} - y_j)^4 \Big]$$

Extensions into more than two dimensions can be accomplished by repeating the steps of this algorithm.

BIBLIOGRAPHY

Austin, PC and Fine, JP (2017). Practical Recommendations for Reporting Fine-Gray Model Analyses for Competing Risk Data. *Statistics in Medicine* 36, 4391–4400.

Barrett, MJ, Sperling, SA, Blair, JC, Freeman, CS, Flanigan, JL, Smolkin, ME, Manning, CA, and Druzgal, JT (2019). Lower Volume, More Impairment: Reduced Cholinergic Basal Forebrain Grey Matter Density is Associated with Impaired Cognition in Parkinson Disease. *Journal of Neurology, Neurosurgery, and Psychiatry* 90, 1251–1256.

Beach, TG, Monsell, SE, Phillips, LE, and Kukull, W (2012). Accuracy of the Clinical Diagnosis of Alzheimer Disease at National Institute on Aging Alzheimer's Disease Centers, 2005-2010. *Journal of Neuropathology and Experimental Neurology* 71 (4), 266–273.

Billingsley, P (1986). *Probability and Measure*. New York: Wiley.

Birkhoff, G and Garabedian, HL (1960). Smooth Surface Interpolation. *Studies in Applied Mathematics* 39, 258–268.

Carpenter, JR, Kenward, MG, and Vansteelandt, S (2006). A Comparison of Multiple Imputation and Double Robust Estimation for Analyses with Missing Data. *Journal of the Royal Statistical Society, Series A* 169 (3), 571–584.

Casella, G and Berger, RL (2002). *Statistical Inference*. Pacific Grove, CA: Duxbury.

Clayton, D, Spiegelhalter, D, Dunn, G, and Pickles, A (1998). Analysis of Longitudinal Binary Data from Multiphase Sampling. *Journal of the Royal Statistical Society, Series B* 60 (1), 71–87.

Cox, DR (1972). Regressional Models and Life-Tables (with Discussion). *Journal of the Royal Statistical Society, Series B* 34, 187–220.

Cox, DR and Hinkley, DV (1974). *Theoretical Statistics*. New York: Chapman and Hall.

Dadgar, H, Khatoonabadi, AR, and Bakhtiyari, J (2013). Verbal Fluency Performance in Patients with Non-demented Parkinson's Disease. *Iran Journal of Psychiatry* 8 (1), 55–58.

Dignam, JJ, Zhang, Q, and Kocherginsky, MN (2013). The Use and Interpretation of Competing Risks Regression Models. *Clinical Cancer Research* 18, 2301–2308.

Erler, NS, Rizopoulos, D, Rosmalen, JV, Jaddoe, VWV, Franco, OH, and Lesaffre, EMEH (2016). Dealing with Missing Covariates in Epidemiologic Studies: A Comparison Between Multiple Imputation and a Full Bayesian Approach. *Statistics in Medicine* 35, 2955–2974.

Fine, JP and Gray, RJ (1999). A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association* 94, 496–509.

Fleming, TR and Harrington, DP (1991). *Counting Processes and Survival Analysis*. New York: Wiley.

Gill, A (1980). *Censoring and Stochastic Integrals: Mathmatical Centre Tracts 124*. Amsterdam: Mathematisch Centrum.

Gray, R (2014). *cmprsk: Subdistribution Analysis of Competing Risks*. R package version 2.2.7. URL: https://CRAN.R-project.org/package=cmprsk.

Ibarretxe-Bilbao, N, Tolosa, E, Junque, C, and M-J, M (2009). MRI and Cognitive Impairment in Parkinson's Disease. *Movement Disorders* 24, S748–S753.

Ibrahim, JG, Chen, M-H, Lipsitz, SR, and Herring, AH (2005). Missing-Data Methods for Generalized Linear Models. *Journal of the American Statistical Association* 100 (469), 332–346.

Jack, CR, Knopman, DS, Jagust, WJ, Shaw, LM, Aisen, PS, Weiner, MW, and al., et (2010). Hypothetical Model of Dynamic Biomarkers of the Alzheimer's Disease Pathological Cascade. *The Lancet Neurology* 9, 119–128.

Kalbfleisch, JD and Prentice, RL (1980). *The Statistical Analysis of Failure Time Data*. New York: Wiley.

Kang, JD and Schafer, JL (2007). Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data. *Statistical Science* 22, 523–539.

Kusnierczyk, W (2012). *rbenchmark: Benchmarking Routines for R*. R package version 1.0.0. URL: https://CRAN.R-project.org/package=rbenchmark.

Laird, NM and Ware, JH (1982). Random-Effects Models for Longitudinal Data. *Biometrics* 38 (4), 963–974.

Lin, DY and Wei, LJ (1989). The Robust Inference for the Cox Proportional Hazards Model. *Journal of the American Statistical Association* 84, 1074–1078.

Lu, T (2017). Simultaneous Inference for Semiparametric Mixed-Effects Joint Models with Skew Distribution and Covariate Measurement Error for Longitudinal Competing Risks Data Analysis. *Journal of Biopharmaceutical Statistics* 8, 1–19.

Moeschberger, ML, Tordoff, KP, and Kochar, N (2008). A Review of Statistical Analyses for Competing Risks. *Handbook of Statistics* 27, 321–341.

Nagano-Saito, A, Washimi, Y, Arahata, Y, Kachi, T, Lerch, JP, Evans, AC, Dagher, A, and Ito, K (2005). Cerebral Atrophy and its Relation to Cognitive Impairment in Parkinson Disease. *Neurology* 64 (2), 224–229.

Nazem, S, Siderowf, AD, Duda, JE, Ten Have, T, Colcher, A, Horn, SS, Moberg, PJ, Wilkinson, JR, Hurtig, HL, Stern, MB, and Weintraub, D (2009). Montreal Cognitive Assessment Performance in Patients with Parkinson's Diseasee with 'Normal' Global Cognition According to Mini-Mental State Examination Score. *Journal of the American Geriatric Society* 57 (2), 304–308.

Pepe, MS (1992). Inference Using Surrogate Outcome Data and a Validation Sample. *Biometrika* 79 (2), 355–365.

Pepe, MS, Reilly, M, and Fleming, TR (1994). Auxiliary Outcome Data and the Mean Score Method. *Journal of Statistical Planning and Inference* 42, 137–160.

Prentice, RL (1982). Covariate Measurement Errors and Parameter Estimation in a Failure Time Regression Model. *Biometrika* 69, 331–342.

Qi, L and Sun, Y (2014). Missing data approaches for probability regression models with missing outcomes with applications. *Journal of Statistical Distributions and Applications* 1 (23), 1–33.

Robins, JM, Rotnitzky, A, and Zhao, LP (1994). Estimation of Regression Coefficients when Some Regressors are not Always Observed. *Journal of the American Statistical Association* 89 (427), 846–866.

Rubin, DB (1987). *Multiple Imputation for Non-Response Surveys*. New York: Wiley.

Scharfstein, DO, Rotnitzky, A, and Robins, JM (1999). Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models. *Journal of the American Statistical Association* 94 (448), 1096–1120.

Schenker, N and Taylor, JMG (1996). Partially Parametric Techniques for Multiple Imputation. *Computational Statistics and Data Analysis* 22, 425–446.

Schoenberg, IJ (1946). Contributions to the Problem of Approximation of Equidistant Data by Analytic Functions: Part A. *Quarterly of Applied Mathematics* 4 (2), 45–99.

Schuff, N, Wu, IW, Buckley, S, Foster, ED, Coffey, CS, Gitelman, DR, Mendick, S, Seibyl, J, Simuni, T, Zhang, Y, Jankovic, J, Hunter, C, Tanner, CM, Rees, L, Factor, S, Berg, D, Wurster, I, Gauss, K, Sprenger, F, Seppi, K, Poewe, W, Mollenhauer, B, Knake, S, Mari, Z, McCoy, A, Ranola, M, and Marek, K (2015). Diffusion Imaging of Nigral Alterations in Early Parkinson's Disease with Dopaminergic Deficits. *Movement Disorders* 30 (14), 1885–1892.

Seaman, SR and Vansteelandt, S (2018). Introduction to Double Robust Methods for Incomplete Data. *Statistical Science* 33 (2), 184–187.

Shaw, LM, Vanderstichele, H, Knapik-Czajka, M, Figurski, M, Coart, E, Blennow, K, and al., et (2011). Qualification of the Analytical and Clinical Performance of CSF Biomarker Analyses in ADNI. *Acta Neuropathologica* 121, 597–609.

Simonoff, JS (1998). *Smoothing Methods in Statistics*. New York: Springer.

Spratt, M, Carpenter, J, Sterne, JAC, Carlin, JB, Heron, J, Henderson, J, and Tilling, K (2010). Strategies for Multiple Imputation in Longitudinal Studies. *American Journal of Epidemiology* 172 (4), 478–487.

Stone, C (1980). Optimal Rates of Convergence for Nonparametric Estimators. *The Annals of Statistics* 8 (6), 1348–1360.

Therneau, TM (2017). *survival: Survival Analysis Routines for R*. R package version 2.41.3. Ed. by T Lumley. URL: `https://CRAN.R-project.org/package=survival`.

Tomita, H, Fujisawa, H, and Henmi, M (2018). A Bias-Corrected Estimator in Multiple Imputation for Missing Data. *Statistics in Medicine* 37, 3373–3386.

Tsiatis, AA, DeGruttola, V, and Wulfsohn, MS (1995). Modeling the Relationship of Survival to Longitudinal Data Measured with Error: Applications to Survival and CD4 Counts in Patients with AIDS. *Journal of the American Statistical Association* 90, 27–37.

Weiner, MW, Veitch, DP, Aisen, PS, Beckett, LA, Cairns, NJ, Green, RC, and al., et (2012). The Alzheimer's Disease Neuroimaging Initiative: A Review of Papers Published Since its Inception. *Alzheimers and Dementia: The Journal of the Alzheimer's Association* 8, 51–68.

Weintraub, D, Mamikonyan, E, Papay, K, Shea, JA, Xie, SX, and Siderowf, A (2012). Questionnaire for Impulsive-Compulsive Disorders in Parkinson's Disease-Rating Scale. *Movement Disorders* 27 (2), 242–247.

White, MT, Shaw, LM, and Xie, SX (2016). Evaluation of Cerebrospinal Fluid Assay Variability in Alzheimer's Disease. *Journal of Alzheimer's Disease* 51, 463–470.

White, MT and Xie, SX (2013). Adjustment for Measurement Error in Evaluating Diagnostic Biomarkers by Using an Internal Reliability Sample. *Statistics in Medicine* 32, 4709–4725.

Xie, SX, Wang, CY, and Prentice, RL (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, 855–870.

Yang, H-J, Ahn, J-H, Lee, J, Lee, WK, Lee, J, and Kim, Y (2019). Measuring Anxiety in Patients with Early-Stage Parkinson's Disease: Rasch Analysis of the State-Trait Anxiety Inventory. *Frontiers in Neurology* 10 (49), 1–9.

Ye, W, Lin, X, and Taylor, J (2008). Semiparametric Modeling of Longitudinal Measurements and Time-to-Event Data: A Two-Stage Regression Calibration Approach. *Biometrics* 64, 1238–1246.