Advances in the Analysis of Irregular Longitudinal Data Using Inverse Intensity Weighting

by

Grace Elizabeth Tompkins

A thesis
presented to the University of Waterloo
in fulfillment of the thesis
requirement for the degree of
Doctor of Philosophy
in
Statistics (Biostatistics)

Waterloo, Ontario, Canada, 2025

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

Supervisor(s):

Michael Wallace
Associate Professor
Depertment of Statistics and Actuarial Sciences
University of Waterloo

Joel Dubin
Professor
Department of Statistics and Actuarial Sciences
School of Public Health Sciences
University of Waterloo

Internal Member:

Richard Cook
Professor
Department of Statistics and Actuarial Sciences
University of Waterloo

Internal Member:

Lan Wen
Assistant Professor
Department of Statistics and Actuarial Sciences
University of Waterloo

Internal-External Member: **Shannon Majowicz**
Associate Professor
School of Public Health Sciences
University of Waterloo

External Member: **Yingwei (Paul) Peng**
Professor
Department of Public Health Sciences
Department of Mathematics and Statistics
Queen's University

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Grace Tompkins was the sole author of Chapters 1, 2, and 6 which were written under the supervision of Dr. Michael Wallace and Dr. Joel Dubin. These chapters were not written for publication.

A version of Chapter 3 was accepted for publication in *Statistical Methods in Medical Research* on December 21, 2024, and is titled "On Flexible Inverse Probability of Treatment and Intensity Weighting: Informative Censoring, Variable Selection, and Weight Trimming". This paper was co-authored by supervisors Dr. Joel Dubin and Dr. Michael Wallace.

Chapter 4 has been written in preparation for submission to a peer-reviewed journal. This paper was co-authored by Dr. Joel Dubin and Dr. Michael Wallace. Chapter 5 has also been prepared for submission to a peer-reviewed journal, and was co-authored by Dr. Lan Wen alongside Dr. Joel Dubin and Dr. Michael Wallace.

As lead author for all chapters presented in this thesis, Grace Tompkins was responsible for proposing the research projects, conducting, literature reviews, performing simulation studies, implementing the real data analyses, and writing each chapter. Supervisors Dr. Joel Dubin and Dr. Michael Wallace provided guidance during each step of the research and provided feedback on drafts of each chapter. Dr. Lan Wen provided feedback and guidance for Chapter 5 throughout the writing process.

Abstract

The analysis of irregular longitudinal data can be complicated by the fact that the timing at which individuals are observed in the data are related to the longitudinal outcome. For example, this can occur when patients are more likely to visit a clinician when their symptoms are worse. In such settings, the observation process is referred to as informative, and any analysis that ignores the observation process can be biased. Inverse intensity weighting (IIW) is a method that has been developed to handle specific cases of informative observation processes. IIW weights observations by the inverse probability of being observed at any given time, and creates a pseudopopulation where the observation process is subsequently ignorable. IIW can also be easily combined with inverse probability of treatment weighting (IPTW) to handle non-ignorable treatment assignment processes. While IIW is relatively intuitive and easy to implement compared to other existing methods, there are few peer-reviewed papers examining IIW and its underlying assumptions.

In this thesis, we begin by evaluating a flexible weighting method which combines IIW and IPTW through multiplication to handle informative observation processes and non-randomized treatment assignment processes. We show that the flexible inverse probability of treatment and intensity weighting (FIPTIW) weighting method is sensitive to violations of the noninformative censoring assumption and show that a previously proposed extension fails under such violations. We also show that variables confounding the observation and outcome processes should always be included in the observation intensity model. Finally, we show scenarios where weight trimming should and should not be used, and highlight sensitivities of the FIPTIW method to extreme weights. We also include an application of the methodology to a real data set to examine the impacts of household water sources on malaria diagnoses of children in Uganda.

Next, we investigate the impact of missing data on the estimation of IIW weights, and evaluate the performance of existing missing data methods through empirical simulation. We show that there is no "one-size-fits-all" approach to handling missing data in the IIW model, and show that the results are highly dependent on the type of covariates that are missing in the observation times model. We then apply the missing data methods to a real data set to estimate the association between sex assigned at birth and malaria diagnoses in children living in Uganda.

Finally, we provide an in-depth evaluation on the assumptions made on IIW across various peer-reviewed papers published in the literature. For each set of assumptions, we construct directed acyclic graphs (DAGs) to visualize the assumptions made on the observation and censoring processes which we use to highlight inconsistencies and potential

ambiguity among the assumptions presented in existing works involving IIW. We also discuss when causal estimates of the marginal outcome model can be obtained, and propose a general set of assumptions for IIW.

Acknowledgments

First, I'd like to give a heartfelt thank you to my supervisors, Michael and Joel for their support, knowledge, and patience over the past five years. Thank you for making this experience more enjoyable - I don't think I've ever attended a meeting where no laughter and jokes were shared. I truly could not have done it without either of you. I've learned so much from each of you, in statistics and beyond. I am going to miss you!

To Lan - thank you so much for working through my third project so closely with me. This project would never have come to fruition without your feedback from my thesis proposal, and all of the meetings over the past two years. Your students are very lucky to have such a knowledgeable, curious, and kind supervisor. I hope we will stay in touch.

To Richard Cook, Shannon Majowicz, and Paul (Yiweng) Peng - thank you so much for sitting on my committee. Your comments and questions have been so valuable, and I greatly appreciate all of the time and effort you have put into this process.

To Derrick Lee at St.FX - thank you for encouraging me to take on this challenge. You have been a mentor to me like no other, and even from 2000km away I have felt your unwavering support and encouragement.

Augustine and Tatiana, there are no two other people in this world I would have handpicked to go through this tumultuous, stressful, difficult, and, in many ways, amazing experience with. From meeting on Zoom to living on Amos Ave together and being office mates in M3, there are so many happy memories that I know I will always look back on. Thank you for always being there for me through it all. We are bonded for life.

Mom, Dad, and Emma - pieces of this dissertation have been written at all of your kitchen tables. Thank you for all of the calls, airport pickups, and (some very extended) visits over the past five years - you made home feel not so far away. Mom and Dad - I know you will brush off any acknowledgment of the parts you played in this achievement, but believe me when I say that both Emma and I would not be the women we are today without each of you. You provided us with so many opportunities and never doubted our dreams. I have never seen support and love like yours. Whether it's a drawing, job promotion, knitted sweater, a freshly caught striped bass in Grand Etang, or Ph.D. dissertation, we feel how proud you truly are of us every single day. So understand that when I say I couldn't have done this without you, I mean it. I love you so much!

And, Bruce. You made the bad days good. You made me see that there is more to life than a dissertation. Thank you for all of the love and support, laughs, and treats over the past four years. It has been so special to be able to share our accomplishments (and a single desk) together, but I can't wait to start this next chapter together.

Dedication

To my Mom and Dad

Table of Contents

List of Figures

List of Tables

Acronyms

ANOVA

analysis of variance

ATE

average treatment effect

CART

classification and regression trees

CCA

complete case analysis

CLT

central limit theorem

DAG

directed acyclic graph

FIPTICW

flexible inverse probability of treatment, intensity, and censoring weighting

FIPTIW

flexible inverse probability of treatment and intensity weighting

FIPTIW-GEE

flexible inverse probability of treatment and intensity weighted generalized estimating equation

GBM

generalized boosted model

GEE

generalized estimating equation

GLMM

generalized linear mixed effects model

HPP

homogeneous Poisson process

IIW

inverse intensity weighting

IIW-GEE

inverse intensity weighted generalized estimating equation

IPCW

inverse probability of censoring weighting

IPTW

inverse probability of treatment weighting

IPTW-GEE

inverse probability of treatment weighted generalized estimating equation

LOCF

last observation carried forward

MAR

missing at random

MCAR

missing completely at random

MI

multiple imputation

MICE

multiple imputation by chained equations

MICE-PPM

multiple imputation by chained equations using predictive mean matching

MNAR

missing not at random

MSE

mean squared error

NHPP

non-homogeneous Poisson process

OR

odds ratio

PH

proportional hazards

PMM

predictive mean matching

SI

single imputation

SI-PPM

single imputation with predictive mean matching

SITA

strongly ignorable treatment assignment

SUTVA

stable unit treatment value assumption

VAR

visiting at random

Chapter 1

Introduction

Longitudinal data consist of repeated measurements taken on individuals, which allow researchers to assess changes in health outcomes over some time period as they relate to various factors. For example, a clinical trial may follow patients to evaluate the use of an intervention or treatment where its effect may change over time. The results of studies of this nature may inform dose schedules for the treatment and capture changes in the efficacy of a drug during the study period, which can influence if and how a treatment is used or prescribed. The average treatment effect (ATE), which compares the average outcome in those treated versus not treated, is often of interest to researchers (Hernan and Robins, 2025).

Unique challenges arise when analyzing longitudinal data as repeated observations within an individual tend to be correlated over time (Fitzmaurice et al., 2011). Many methods for statistical inference require satisfying the assumption of independence between observations and are thus not appropriate for longitudinal data analysis where correlated observations are present. To overcome this issue, generalized estimating equations (GEEs), generalized linear mixed effects models (GLMMs), and, earlier, analysis of variance (ANOVA) techniques, such as repeated measures ANOVA and multivariate ANOVA for repeated measures have been developed as extensions to existing statistical methodology to account for the correlation inherently present between observations of the same individual. We refer to Fitzmaurice et al. (2011) for an introduction to these methods.

In many longitudinal data sets, particularly those involving observational studies, we may encounter irregular longitudinal data where the observation timings vary between individuals. That is, individuals may not have a common set of observation times and, in extreme cases, may be observed at completely unique times. We refer to the process

driving the times at which an individual has an observation recorded in the data as the observation process.

In many cases, the times at which individuals are observed in the data are related to the longitudinal outcome. When this occurs, the observation process is informative, and statistical methods that do not account for the observation process can provide biased estimates of the outcome model parameters. For example, consider a hypothetical study using historical data from a clinical database. It is expected that individuals with more severe symptoms receive more frequent medical consultations and thus may have more observations in the data. Sicker patients are then over-represented in the data, and this over-representation may bias estimates of quantities of interest, such as the ATE.

Many methods for handling informative observation processes exist, including parametric joint models (Garrett et al., 2024; Zhu et al., 2013), semiparametric joint models (Cai et al., 2012; Liang et al., 2009; Song et al., 2012; Sun et al., 2011; Sun et al., 2012), and inverse intensity weighting (IIW) (Buzkova and Lumley, 2007, 2009; Lin et al., 2004). Each method depends on a number of assumptions, including how the observation process is related to the longitudinal outcome, and what covariates are available in the analysis. An overview of methods developed for handling irregular longitudinal data can be found in Pullenayegum and Lim (2016).

The method that we focus on throughout this thesis is IIW. IIW is a weighting method that can be used in conjunction with an estimating equation, such as the GEE, to account for informative observation processes. Simply put, IIW weights each observation by the inverse probability of being observed, which creates a pseudopopulation where the observation process is ignorable. This method is desirable to analysts as it is relatively intuitive and simple to implement using standard statistical software. The underlying assumptions and an in-depth review of this method is presented in Chapter 2.

While the theoretical foundation of IIW is well developed in the literature, there exist numerous gaps in the literature involving IIW. The aim of this thesis is to provide further investigations into the sensitivity of IIW to violations of its assumptions, discuss variable selection and weight trimming, provide recommendations on handling missing data for IIW models, and finally clarify and propose a unified set of necessary assumptions for IIW.

We begin by providing an overview of the IIW method in Chapter 2. We then investigate the sensitivity of IIW to violations of the noninformative censoring assumption and discuss variable selection and weight trimming in Chapter 3. We follow with a study investigating methods for handling missing data in IIW models in Chapter 4. We discuss the various assumptions made on IIW in various published works, and propose a unified set of assumptions in Chapter 5. We conclude with a discussion of the thesis and potential

future projects in Chapter 6.

Chapter 2

Notation, Assumptions, and Review of Inverse Intensity Weighting

IIW was first proposed by Lin et al. (2004), and later extended by Buzkova and Lumley (2007). IIW involves fitting two models. The first is an observation times/intensity model which we use to estimate IIW weights. The second is the outcome model relating a set of covariates to a longitudinal outcome whose parameters are of primary interest. We typically use IIW to create a weighted data set, to which we fit the outcome model. Throughout this thesis, we present the IIW method developed by Buzkova and Lumley (2007) as it is more general and allows time-varying covariates to be included in the outcome model, and allows the use of a discontinuous hazard function in the observation intensity model.

We begin by discussing the notation and assumptions used throughout this thesis in Sections 2.1 and 2.2, and then review IIW in detail in Section 2.3.

2.1 Notation

We begin by defining general notation. Quantities that are **bolded** refer to vectors and matrices (i.e., $\boldsymbol{A}$) while unbolded quantities refer to univariate quantities (i.e., B). Time-varying covariates are denoted as a function of time (*i.e.*, $\boldsymbol{A}(t)$ or $B(t)$). Subscripts i refer to the individual or study unit (i.e., $\boldsymbol{A}_i(t)$ or B_i). Subscripts i, j refer to the jth observation of vector i (i.e., t_{ij} below). For any vector $\boldsymbol{A}$, let $\boldsymbol{A}^{obs}$ refer to the observed components of $\boldsymbol{A}$, and let $\boldsymbol{A}^{mis}$ refer to the unobserved/missing components of $\boldsymbol{A}$.

Consider a study from $t = 0$ to the study end time τ where $t_{i1}, t_{i2}, ..., t_{iK_i}$ are the times at which individual i is observed for $i = 1, 2, ..., n$ and $0 \leq t_{i1} < t_{i2} < ... < t_{iK_i} \leq \tau$. That is, each individual in the study has a potentially unique set of K_i observation times.

Throughout this thesis, we will focus on the following semiparametric marginal outcome model

$$g(E(Y_i(t)|\boldsymbol{X}_i(t))) = m(t) + \boldsymbol{\beta}^T \boldsymbol{X}_i(t), \tag{2.1}$$

with primary interest in the estimation of the parameter vector $\boldsymbol{\beta}$. In Equation (2.1), $\boldsymbol{X}_i(t) = (X_{i1}(t), X_{i2}(t), ..., X_{ip}(t))^T$ is a vector of p observed covariates for individual i at time t, and $Y_i(t)$ is the longitudinal outcome of individual i at time t, for $i = 1, 2, ..., n$ and $t = t_{i1}, t_{i2}, ..., t_{iK_i}$. Further, $g(\cdot)$ is a known, monotonic, and differentiable link function, and $m(t)$ is a time-varying intercept term. We assume the set of covariates $\boldsymbol{X}(t)$ contains a possibly time-varying, binary treatment $D(t)$ for which we'd like to estimate the treatment effect.

We explicitly define the observation process to be the underlying process that determines when individuals have data recorded. In other articles, this has been referred to as the "visit" (Aghababaei Jazi and Pullenayegum, 2022; Coulombe et al., 2021a; Lin et al., 2004; Pullenayegum, 2013; Pullenayegum et al., 2021; Pullenayegum and Lim, 2016), "participation" (Schmidt and Woll, 2017), "assessment" (Pullenayegum et al., 2023; Smith et al., 2022), or "monitoring" (Coulombe et al., 2021b) process. We opt to use a more general term as longitudinal data may be used in a variety of settings, such as for individuals repeatedly visiting a clinic, or physicians repeatedly monitoring/observing patients within a single hospital stay. This nomenclature also connects irregular longitudinal data analysis to the missing data literature. This is natural as the analysis of most longitudinal data sets is essentially a missing data problem as we typically cannot observe the covariates and outcome in continuous time (Lokku et al., 2020). However, there are some instances where data can be collected in nearly continuous time, including when data is obtained from wearable tracking technology or from continuous monitoring systems in intensity care units.

We denote the counting process for the number of observations individual i has by time t as $N_i(t) = \sum_{k=1}^{K_i} \mathbb{1}_{(t_{ik} \leq t)}$, where $\mathbb{1}_{(E)}$ is the indicator function for event E. We let C_i denote the censoring time at which follow-up ceases for individual i, such that $C_i \leq \tau$. Although individuals may be censored prior to end of the study, we consider the counting process for the counterfactual observation times, denoted $N_i^*(t)$. We relate the censored and uncensored counting processes as $N_i(t) = N_i^*(t \wedge C_i)$, where we define $a \wedge b = min(a, b)$. Let $\boldsymbol{V}_i(t)$ be a set of auxiliary covariates related to the observation process but omitted from the outcome model.

For any arbitrary process $A(t)$, we define $\overline{A}(t) = \{A(s) : 0 \leq s \leq t\}$ as the entire (and potentially counterfactual) history of the process up to and including time t. We denote $\overline{A}^{obs}(t)$ to include only the observed history of A up to and including time t. We let $\overline{A}(\infty) = \{A(s) : a > 0\}$ be the entire (counterfactual) process that includes times beyond the study end time. Further, we let $dN(t) = N(t) - N(t^-)$ where $N(t^-) = \lim_{s \to t} N(s)$. That is, $dN_i(t) = 1$ if individual i is observed at time t, and is zero otherwise. $dN_i^*(t)$ is similarly defined, where $dN_i^*(t) = 1$ if individual i is observed at time t in the counterfactual observation times, and is zero otherwise.

2.2 Assumptions

As in Buzkova and Lumley (2007), we assume that the outcome model covariates $\boldsymbol{X}(t)$ are known at all possible time points unless otherwise indicated. We discuss cases where $\boldsymbol{X}(t)$ is potentially not observed at all possible time points in Chapter 5.

Throughout this thesis, we will refer to five assumptions on the observation process. The first assumption states that the intensity or probability of being observed is conditionally independent of the longitudinal outcome and censoring time given the set of covariates $\boldsymbol{Z}(t)$, which can include the observed history of the outcome covariates $\overline{\boldsymbol{X}}(t)$, the observed history of the auxiliary covariates $\overline{\boldsymbol{V}}_i(t)$, previous observation times $\overline{\boldsymbol{N}}_i(t^-)$, and previous observed outcomes $\overline{\boldsymbol{Y}}_i^{obs}(t^-)$. That is, we assume

Assumption O1. *Conditional Independence of the Observation Process:*

$$E(dN_i^*(t)|\boldsymbol{Z}_i(t), \boldsymbol{X}_i(t), Y_i(t), C_i \geq t) = E(dN_i^*(t)|\boldsymbol{Z}_i(t)).$$

As $\boldsymbol{Z}_i(t)$ can contain the outcome model covariates ($\boldsymbol{X}_i(t)$), the observed history of the outcome model covariates ($\overline{\boldsymbol{X}}_i^{obs}(t)$), the observed history of auxiliary covariates related to the probability of being observed but omitted from the outcome model ($\overline{\boldsymbol{V}}_i^{obs}(t)$), information on past observation times ($\overline{\boldsymbol{N}}_i(t^-)$), and past observed outcomes prior to time t ($\overline{\boldsymbol{Y}}_i^{obs}(t^-)$), we can re-write Assumption O1 as

$$\begin{aligned} E(dN_i^*(t)|\overline{\boldsymbol{X}}_i^{obs}(t), \overline{\boldsymbol{V}}_i^{obs}(t), &\overline{\boldsymbol{N}}_i(t^-), \overline{\boldsymbol{Y}}_i^{obs}(t^-), \boldsymbol{X}_i(t), Y_i(t), C_i \geq t) \\ &= E(dN_i^*(t)|\boldsymbol{X}_i(t), \overline{\boldsymbol{X}}_i^{obs}(t), \overline{\boldsymbol{V}}_i^{obs}(t), \overline{\boldsymbol{N}}_i(t^-), \overline{\boldsymbol{Y}}_i^{obs}(t^-)). \end{aligned}$$

This implies $N_i^*(t) \perp Y_i(t), C_i \geq t | \overline{\boldsymbol{X}}_i^{obs}(t), \overline{\boldsymbol{V}}_i^{obs}(t), \overline{\boldsymbol{N}}_i(t^-),$ and $\overline{\boldsymbol{Y}}_i^{obs}(t^-)$.

Next, we make an assumption on the censoring times such that

Assumption O2. *Noninformative censoring:*

$$E(Y_i(t)|\boldsymbol{X}_i(t), C_i \geq t) = E(Y_i(t)|\boldsymbol{X}_i(t)).$$

Assumption O2 may not be met in some applications. For example, when analyzing the relationship between a treatment and a health outcome in an observational study, sicker (or in some cases, healthier) patients may drop out prior to the end of the study. In this setting, the censoring times are likely related to longitudinal outcome, which is a violation of Assumption O2. We will investigate the implications of violations of this assumption in Chapter 3.

We also assume

Assumption O3. *Separability: The outcome and observation times model parameters are separable (i.e., the models do not share parameters).*

We also assume

Assumption O4. *Correct specification: The observation intensity model is correctly specified.*

And finally, we assume

Assumption O5. *Completely observed observation-level covariates: The covariates related to the observation process are known at all possible observation times.*

This assumption means that at any possible observation time (including those not observed), we can obtain the value of $\boldsymbol{Z}_i(t)$ for all individuals. When $\boldsymbol{Z}_i(t)$ does not contain any time-varying covariates, this assumption is automatically met so long as $\boldsymbol{Z}_i(t)$ is observed at least once. However, the use of time-varying covariates in $\boldsymbol{Z}_i(t)$ can complicate analysis when these covariates are not observed in continuous time. When $\boldsymbol{Z}_i(t)$ is not observed completely, Buzkova and Lumley (2007) recommend carrying the last observation carried forward (LOCF) for unobserved $\boldsymbol{Z}_i(t)$. However, one must be cautious when employing this method when a non-trivial proportion of $\boldsymbol{Z}_i(t)$ are missing, as it can bias the parameters estimated in proportional hazards (PH) models (Andersen and Liestøl, 2003; Cao and Fine, 2021; Molenberghs et al., 2002; Molnar et al., 2008) and longitudinal models (Lachin, 2016; Lane, 2008; Saha and Jones, 2009). We discuss methods for handling missingness in the observation process covariates in Chapter 4.

When assumptions O1 to O5 are met, we will refer to the observation process as conditionally ignorable. That is, conditional on the observed history of the covariates, the observation process can adequately be accounted for in the analysis to obtain unbiased and consistent estimates of the outcome model parameters. We provide an in-depth examination of the underlying assumptions for IIW in Chapter 5.

2.3 Inverse Intensity Weighting

IIW is often used with a GEE to estimate measures of association between a longitudinal outcome and a set of potentially time-varying covariates. When a GEE is weighted by IIW, we refer to it as the inverse intensity weighted generalized estimating equation (IIW-GEE).

Estimating an IIW-GEE involves a two-step process where we first estimate the IIW weights in an observation times model and then use the estimated weights to obtain weighted estimates of the quantities of interest from the outcome model.

With IIW, we model the uncensored observation times as

$$E\left\{dN_i^*(t)|\boldsymbol{Z}_i(t)\right\} = \lambda_0(t)\exp\left\{\boldsymbol{\gamma}^T\boldsymbol{Z}_i(t)\right\}, \tag{2.2}$$

where $\lambda_0(\cdot)$ is an unspecified non-decreasing function, $\boldsymbol{Z}_i(t)$ is a vector of covariates for the observation times model, and $\boldsymbol{\gamma}$ is the corresponding parameter vector. Recall $\boldsymbol{Z}_i(t)$ may include outcome model covariates ($\boldsymbol{X}(t)$), past observed outcomes ($\overline{\boldsymbol{Y}}^{obs}(t^-)$), information about previous observation times ($\overline{\boldsymbol{N}}_i(t^-)$), and auxiliary covariates that are not included in the outcome model but related to the observation times ($\boldsymbol{V}_i(t)$). We emphasize that $\boldsymbol{Z}_i(t)$ must be known at all times (Assumption O5).

We define the IIW weight for the ith individual at time t, as

$$w_i^{IIW}(t;\boldsymbol{\gamma},h) = \frac{h(\boldsymbol{X}_i(t))}{\exp\left\{\boldsymbol{\gamma}^T\boldsymbol{Z}_i(t)\right\}}, \tag{2.3}$$

where $h(\cdot)$ can be any positive function of the outcome model covariates $\boldsymbol{X}_i(t)$. In the observation times model we can include (and must specify) various functional forms of the covariates $\boldsymbol{Z}_i(t)$ including interactions and higher-order terms (Lin et al., 2004). These IIW weights are proportional to the probability of individual i having an observation at time t relative to the other individuals, under the model in Equation (2.2) (Buzkova and Lumley, 2007).

In practice, these weights must be estimated from the data as the parameter vector $\boldsymbol{\gamma}$ is unknown in the observation times model. To estimate the weights, we use semiparametric

models (Lin et al., 2004). Buzkova and Lumley (2007) showed that the following estimation function can be used to estimate the parameter vector $\boldsymbol{\gamma}$:

$$U^{\dagger}(\boldsymbol{\gamma}) = \sum_{i=1}^{n} \int_0^{\tau} \{\boldsymbol{Z}_i(t) - \overline{\boldsymbol{Z}}(t, \boldsymbol{\gamma})\} dN_i(t),$$

where $\overline{\boldsymbol{Z}}(t, \gamma)$ is a weighted average of $\boldsymbol{Z}$ at time t, such that

$$\overline{\boldsymbol{Z}}(t, \boldsymbol{\gamma}) = \sum_{i=1}^{n} \boldsymbol{Z}_i(t) \frac{\exp\left\{\boldsymbol{\gamma}^T \boldsymbol{Z}_i(t)\right\} I(C_i \geq t)}{\sum_{j=1}^{n} \exp\left\{\boldsymbol{\gamma}^T \boldsymbol{Z}_j(t)\right\} I(C_j \geq t)}.$$

The solution to $U^{\dagger}(\boldsymbol{\gamma}) = 0$, $\hat{\boldsymbol{\gamma}}$, is a consistent estimator of the parameter vector $\boldsymbol{\gamma}$ in the observation times model (Buzkova and Lumley, 2007). Because of its form, the Cox PH model can be used to estimate the intensity as in Equation (2.2), however we note that the estimator for the asymptotic variance of $\hat{\boldsymbol{\gamma}}$ will differ from the asymptotic variance of parameters estimated using a Cox PH model (Buzkova and Lumley, 2007).

Recall that in the model for the weights in Equation (2.3) we also require the specification of a function $h(\cdot)$. A convenient choice of the numerator of the weights is $h(\boldsymbol{X}_i(t)) = 1$. We refer to such weights as non-stabilized weights. We can also use stabilized weights, as proposed by Buzkova and Lumley (2007), where $h(\boldsymbol{X}_i(t)) = \exp\{\boldsymbol{\delta}^T \boldsymbol{X}_i(t)\}$ and $\boldsymbol{\delta}$ is the parameter vector in the observation times model using only the outcome model covariates $\boldsymbol{X}_i(t)$. The stabilized weights are typically preferred as they can achieve a smaller variance than the non-stabilized weights (Buzkova and Lumley, 2007). To estimate the outcome model in Equation (3.1) when the observation process is conditionally ignorable we can use a weighted GEE, which involves specifying a number of components. First, we need to specify an appropriate link function $g(\cdot)$ for the semiparametric marginal model in Equation (3.1). Second, we require the specification of the conditional variance of each observation, given the covariates. We assume this quantity is dependent on the mean through

$$Var\left\{Y_i(t) | \boldsymbol{X}_i(t)\right\} = \phi v(\mu_i(t)), \tag{2.4}$$

where $v(\cdot)$ is a known variance function and ϕ is a positive scale parameter, which may need to be estimated.

The p-dimensional estimating function for the outcome model, as motivated by the GEE, is

$$U(\boldsymbol{\beta}; \hat{\boldsymbol{\gamma}}, h) = \sum_{i=1}^{n} \int_0^{\tau} \boldsymbol{X}_i(t) \left\{ \frac{dg(\mu)}{d\mu}\bigg|_{\mu_i(t;\boldsymbol{\beta})} \right\}^{-1} v(\mu_i(t;\boldsymbol{\beta}))^{-1} \times \{Y_i(t) - \mu_i(t;\boldsymbol{\beta})\} \frac{h(\boldsymbol{X}_i(t))}{\exp\left\{\boldsymbol{\gamma}^T \boldsymbol{Z}_i(t)\right\}} dN_i(t),$$

which resembles a GEE with an independent working correlation structure and subject-specific IIW weights. The use of other dependent correlation structures has, to our knowledge, yet to be explored in the literature and is beyond the scope of this thesis. Correct specification of the variance function $v(\mu_i(t;\boldsymbol{\beta}))$ increases the efficiency of the estimator for $\boldsymbol{\beta}$, however is not required to obtain consistent estimates, asymptotic normality, or the validity of its covariance estimator (Buzkova and Lumley, 2007). If the observation process is noninformative, the IIW-GEE simplifies to an independent unweighted GEE when using the stabilized weights (Buzkova and Lumley, 2007).

Chapter 3

On Flexible Inverse Probability of Treatment and Intensity Weighting

3.1 Introduction

Researchers in healthcare often have access to observational longitudinal data in the form of medical records or clinical data. Such data are often irregular with informative observation processes, which can introduce bias into outcome model parameter estimates if the observation process is not accounted for in the analysis. IIW (Buzkova and Lumley, 2007; Lin et al., 2004) has been developed to account for informative observation processes in the analysis of irregular longitudinal data.

With observational data, one must also consider how the treatment is assigned if causal quantities are to be estimated when describing the relationship between a treatment and a longitudinal outcome. When treatments are not assigned at random, treatment groups may have systematic differences which do not make them directly comparable. That is, estimates of causal quantities such as the ATE can be biased in this setting. Inverse probability of treatment weighting (IPTW) (Rosenbaum and Rubin, 1983) is one method that can be used to account for non-randomized treatment assignment processes. Like IIW, IPTW weights observations by the inverse probability of being treated, which creates a pseudopopulation where the treatment assignment process is subsequently ignorable.

As such, when analyzing observational longitudinal data, not only do we need to consider the type of observation process, but also consider how treatments are assigned. For example, if we consider the analysis of data from a clinical database, sicker patients may

be more likely to visit the clinic and have an observation recorded, and also may be more likely to be prescribed a treatment. In this setting, we have two potential sources of bias; one from the informative observation process, and one from the non-randomized treatment assignment process. To simultaneously account for both sources of bias, we can employ a weighting method proposed by Coulombe et al. (2021b) which we refer to as flexible inverse probability of treatment and intensity weighting (FIPTIW). This method combines IIW and IPTW in an intuitive way that is simple to implement using standard statistical software.

The aim of this chapter is to provide an overview and practical guidance for fitting models using FIPTIW by providing novel investigations on the sensitivity of the FIPTIW method, considering both binary and continuous outcomes. We begin with a discussion of the assumptions on the observation and treatment assignment processes in Section 3.2. We then review the existing methods for handling informative observation and treatment assignment processes in Section 3.3. We follow with three simulation studies in Section 3.4. The first investigates the impact of violations of the noninformative censoring assumption on the FIPTIW method, and also investigate if inverse probability of censoring weighting (IPCW) (Robins et al., 2000) can be further included to account for the bias introduced by such violations. Next, we investigate variable inclusion for IIW models to provide practical guidance on which variables to include in IIW (and hence FIPTIW) models. Finally, we evaluate weight trimming for FIPTIW, and investigate how the use of cubic splines to estimate the time-varying intercept may impact estimation under extreme weights. In Section 3.5, we implement a data analysis of a malaria data set. We conclude with a discussion in Section 3.6.

3.2 Assumptions

In this chapter, we will focus on the following marginal outcome model

$$g(\mu_i(t)) = g(E(Y_i(t)|\boldsymbol{X}_i(t))) = m(t) + \boldsymbol{\beta}^T \boldsymbol{X}_i(t), \tag{3.1}$$

with primary interest in the estimation of the parameter vector $\boldsymbol{\beta}$. In Equation (3.1), $\boldsymbol{X}_i(t) = (X_{i1}(t), X_{i2}(t), ..., X_{ip}(t))^T$ is a vector of p observed covariates for individual i at time t, and $Y_i(t)$ is the longitudinal outcome of individual i at time t, for $i = 1, 2, ..., n$ and $t = t_{i1}, t_{i2}, ..., t_{iK_i}$. Further, $g(\cdot)$ is a known, monotonic, and differentiable link function, and $m(t)$ is a time-varying intercept term. We assume the set of covariates $\boldsymbol{X}(t)$ contains a possibly time-varying, binary treatment $D(t)$ whose estimated effect is of primary interest,

along with other covariates of interest. We also assume the distribution of $\boldsymbol{Y}$ comes from an exponential family distribution.

To avoid repetition, readers are directed to Chapter 2.1 and 2.2 for the notation and assumptions for the observation process.

3.2.1 Assumptions on the Treatment Assignment Process

We explicitly define the treatment assignment process as the underlying mechanism that determines whether individuals are treated or not over time. As previously discussed, most observational studies do not have treatments randomly assigned to groups of individuals. In this setting, the treatment and control groups may systematically differ from each other and estimates of the ATE and other causal quantities may be biased. We assume that the probability of being assigned to the treatment group at any time is related to a set of possibly time-varying covariates $\boldsymbol{W}(t)$.

To discuss the assumptions we will make on the treatment assignment process, we first introduce the potential outcomes framework as developed by Rosenbaum and Rubin (1983). Recall $Y_i(t)$ is the observed outcome for individual i at time t. We also consider the potential outcome that individual i would have experienced had they been assigned to a specific treatment. We denote $Y_i^{(1)}(t)$ and $Y_i^{(0)}(t)$ to be the potential outcome under treatment and control at time t, respectively. We note that these quantities are counterfactual as we cannot observe both potential outcomes simultaneously in practice.

Regardless of the study design, the estimation of causal effects is a comparison of the potential outcomes. We often are interested in estimating the ATE, which we define as $E\{Y_i^{(1)}(t) - Y_i^{(0)}(t)\}$. To use IPTW, we require a number of assumptions in this framework. The first assumption is

Assumption T1. *Strongly ignorable treatment assignment (SITA): the treatment assignment is conditionally independent of the potential outcomes given the history of observed covariates $\boldsymbol{W}(t)$ and past treatment history, such that*

$$Y_i^{(0)}(t), Y_i^{(1)}(t) \perp D_i(t) | \overline{\boldsymbol{W}}_i(t), \overline{D}_i(t^-).$$

The SITA assumption is also known as *conditional exchangeability*. Implicit in Assumption T1 is the assumption that there are no unobserved confounders.

Next, we assume

Assumption T2. *Stable unit treatment value assumption (SUTVA): the treatment assignment of a given individual does not affect the outcome of another individual.*

The SUTVA assumption may not hold if, for example, the vaccination status of an individual impacts the probability of transmitting the disease to someone they are in contact with.

We also assume

Assumption T3. *Consistency: the observed outcome under a specific treatment is equal to that treatment's potential outcome, and we can only observe one potential outcome at a time.*

Under Assumption T3, under a binary treatment we can write the observed outcome as a function of the potential outcomes as

$$Y_i(t) = D_i(t)Y_i^{(1)}(t) + (1 - D_i(t))Y_i^{(0)}(t).$$

We further assume

Assumption T4. *Positivity: the probability of receiving a given treatment is greater than zero for every individual.*

We finally assume

Assumption T5. *Correct specification of the propensity score model: the model for the propensity score/treatment assignment process is correctly specified.*

When Assumptions T1 to T5 hold, we refer to the treatment assignment process as conditionally ignorable. That is, conditional on the observed history of the covariates, the treatment assignment process can be adequately accounted for in the analysis through IPTW to provide consistent and unbiased estimates of causal quantities like the ATE Coulombe et al. (2021b).

3.3 Methods

As previously discussed, we may need to simultaneously adjust for the observation process and treatment assignment process when observational longitudinal data is being analyzed.

When the observation process is conditionally ignorable (assumptions O1 to O5 hold) and the treatment assignment process is conditionally ignorable (assumptions T1 to T5 hold), we can employ the flexible weighting method proposed by Coulombe et al. (2021b) which we refer to as FIPTIW. FIPTIW combines IIW and IPTW to create a pseudopopulation in which both the observation and treatment assignment processes can be ignored.

We first define the IPTW weight for individual i at time t as

$$w_i^{IPTW}(t;\boldsymbol{\alpha},\pi,g) = \frac{g(\boldsymbol{W}_i(t))}{\mathbb{1}_{(D_i(t)=1)}\pi(\boldsymbol{W}_i(t);\boldsymbol{\alpha}) + \mathbb{1}_{(D_i(t)=0)}(1-\pi(\boldsymbol{W}_i(t);\boldsymbol{\alpha}))},$$

where $\pi(\boldsymbol{W}_i(t);\boldsymbol{\alpha}) = \Pr(D_i(t)=1|\boldsymbol{W}_i(t))$ is the probability of being in the treatment group at time t (conditional on covariates), which is also known as the propensity score, and $g(\cdot)$ can be any positive function of the outcome covariates $\boldsymbol{W}_i(t)$. A convenient choice of function for the numerator is $g(\boldsymbol{W}_i(t)) = 1$. For time-invariant treatments, stabilized weights weight may also be used, where we use the marginal probability of being treated in the numerator (i.e., $g(\boldsymbol{W}_i(t)) = \mathbb{1}_{(D_i(t)=1)}P(D_i(t)=1) + \mathbb{1}_{(D_i(t)=0)}(1-P(D_i(t)=1))$).

The propensity scores (and thus the IPTW weights) are unknown in practice and must be estimated from the available data. While many choices of model exist for estimating the probability of receiving a binary treatment, common methods include logistic regression or tree-based methods such as generalized boosted models (GBM) (Austin, 2011). We can model the probability of treatment assignment and obtain estimates of $\widehat{\boldsymbol{\alpha}}$ and then estimate the propensity scores as $\pi(\boldsymbol{W}_i(t);\widehat{\boldsymbol{\alpha}},g)$.

The FIPTIW weights are calculated by multiplying the IIW and IPTW weights together. That is, the FIPTIW weight for individual i at time t is

$$w_i^{FIPTIW}(t;\boldsymbol{\alpha},\boldsymbol{\gamma},\pi,h,g) = w_i^{IPTW}(t;\boldsymbol{\alpha},\pi,g) \times w_i^{IIW}(t;\boldsymbol{\gamma},h). \tag{3.2}$$

However, as each of the individual weights in the FIPTIW are unknown in practice, the FIPTIW weight must also be estimated from the data. We estimate it by

$$\widehat{w}_i^{FIPTIW}(t;\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\gamma}},\pi,h,g) = \widehat{w}_i^{IIW}(t;\widehat{\boldsymbol{\gamma}},h) \times \widehat{w}_i^{IPTW}(t;\widehat{\boldsymbol{\alpha}},\pi,g),$$

where the IIW weights, $\widehat{w}_i^{IIW}(t;\widehat{\boldsymbol{\gamma}},h)$, can be estimated as in Section 2.3.

The time-varying intercept $m(t)$ must also be specified or estimated in the outcome model. One can specify the parametric form of the time-varying intercept $m(t)$ as Buzkova and Lumley (2007) or use cubic splines along with a constant intercept as in Coulombe et al. (2021b) to estimate $m(t)$. The use of splines in the binary case has yet to be examined

in the literature. We compare the use of splines to specifying the time-varying intercept through simulation in Appendix A.2.2 for both continuous and binary outcomes.

We explicitly define a weighted GEE (where the working correlation structure is set to independent) which incorporates the FIPTIW weights into the model. We refer to this GEE as the flexible inverse probability of treatment and intensity weighted generalized estimating equation (FIPTIW-GEE), with the estimating equation specified as

$$U(\boldsymbol{\beta};\boldsymbol{\alpha},\boldsymbol{\gamma},\pi,h,g) = \sum_{i=1}^{n}\int_0^{\tau} \boldsymbol{X}_i(t)\left\{\frac{dg(\mu)}{d\mu}\bigg|_{\mu_i(t;\boldsymbol{\beta})}\right\}^{-1} v(\mu_i(t;\boldsymbol{\beta}))^{-1} \times \{Y_i(t)-\mu_i(t;\boldsymbol{\beta})\} w_i^{FIPTIW}(t;\boldsymbol{\alpha},\boldsymbol{\gamma},\pi,h,g)dN_i(t).$$

As the FIPTIW weights are unknown, we can then use the estimated weights in the estimating equation as

$$U(\boldsymbol{\beta};\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\gamma}},\pi,h,g) = \sum_{i=1}^{n}\int_0^{\tau} \boldsymbol{X}_i(t)\left\{\frac{dg(\mu)}{d\mu}\bigg|_{\mu_i(t;\boldsymbol{\beta})}\right\}^{-1} v(\mu_i(t;\boldsymbol{\beta}))^{-1} \times \{Y_i(t)-\mu_i(t;\boldsymbol{\beta})\} \widehat{w}_i^{FIPTIW}(t;\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\gamma}},\pi,h,g)dN_i(t). \tag{3.3}$$

The asymptotic variance for the FIPTIW estimator can be estimated through a two-step sandwich estimator (Coulombe et al., 2021b). Robust standard errors developed for inverse probability weighting by Robins et al. (2000) can also be used for a more conservative standard error and variance estimate.

Coulombe et al. (2021b) did not make any assumptions about the dependence between the observation and treatment assignment processes. However, when the treatment assignment and observation processes are conditionally independent given $\boldsymbol{X}_i(t), \boldsymbol{Z}_i(t), \boldsymbol{W}_i(t), Y_i(t)$, and the censoring time C_i, the weights intuitively reflect the inverse probability of individual i having an observation at time t and receiving the treatment they were assigned, relative to the other individuals. Further, under this assumption, we can show that the estimating equation in Equation (3.3) is unbiased by the following theorem:

Theorem 3.3.1. *The FIPTIW estimating equation in Equation (3.3) has zero mean at the point of true parameters* $\{\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\gamma},\pi,h,g\}$ *for any function* $h(\cdot)$ *and* $g(\cdot)$ *of covariates* $\boldsymbol{X}_i(t)$.

Proof. See Appendix A.1. □

We denote $\widehat{\boldsymbol{\beta}}$ to be the estimator of parameter vector $\boldsymbol{\beta}$ in Equation (3.1) where $\widehat{\boldsymbol{\beta}}$ is a solution to $U(\boldsymbol{\beta}; \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\gamma}}, \pi, h) = 0$. As we assume that the observations are independent and identically distributed and have shown that the estimating equation is unbiased in Theorem 3.3.1, then it follows that the solution $\widehat{\boldsymbol{\beta}}$ from $U(\widehat{\beta}; \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\gamma}}, \pi, h, g) = 0$ is a consistent estimator of $\boldsymbol{\beta}$, as the estimators for the parameters in $w_i^{FIPTIW}(t; \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\gamma}}, \pi, h, g)$ are each consistent (Tsiatis, 2006) That is, the estimator $\widehat{\boldsymbol{\beta}}$ satisfies

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} \boldsymbol{X}_i(t)\left\{\frac{dg(\mu)}{d\mu}\bigg|_{\mu_i(t;\widehat{\boldsymbol{\beta}})}\right\}^{-1} v(\mu_i(t;\widehat{\boldsymbol{\beta}}))^{-1} \\ \times \{Y_i(t) - \mu_i(t;\widehat{\boldsymbol{\beta}})\}\widehat{w}_i^{FIPTIW}(t; \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\gamma}}, \pi, h, g)dN_i(t) = 0.$$

As this is a weighted average with mean zero, by the Law of Large Numbers it follows that $\frac{1}{n}U(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\gamma}}, \pi, h) \xrightarrow{p} U(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \pi, h)$ and thus the estimator is consistent. The central limit theorem (CLT) can also be employed to prove the asymptotic normality of $\widehat{\boldsymbol{\beta}}$. Coulombe et al. (2021b) derived the asymptotic variance of the $h(\cdot)$ estimator.

Coulombe et al. (2021b) also performed sensitivity analyses to examine the impact of having treatment assignment confounders that were correlated with each other, having the same confounders in both the observation model and treatment assignment model, and model misspecification. In these simulations, the performance of the FIPTIW-GEE was not largely impacted by correlated treatment assignment confounders or having the same set of confounders affecting both the observation and treatment assignment processes. The FIPTIW method was also shown to be relatively insensitive to misspecification of the outcome model. However, the FIPTIW method was shown to be extremely sensitive to model misspecification when the observation process depended on non-linear functions of the covariates.

There are still various questions surrounding the FIPTIW method which are yet to be answered in the literature. First, it is known that censoring/dropout can bias model parameters when it is related to the longitudinal outcome (Ma et al., 2005), however the impacts of violating the noninformative censoring assumption have yet to be explored for the FIPTIW method. Coulombe et al. (2021b) noted that it could be possible to incorporate IPCW weights to account for violations of the noninformative censoring assumption. IPCW weights can be estimated by first fitting a Cox PH model to estimate the censoring hazard. Including IPCW weights into the model may remove some of the bias introduced by the censoring mechanism, however it will not account for the poor estimation of the IIW weights due to the violation of this assumption as the noninformative censoring assumption is necessary for sufficient estimation of the IIW weights. To the best of our knowledge,

there have been no papers investigating the inclusion of IPCW weights into the FIPTIW model.

Second, it has been shown in the causal inference literature that propensity score models for IPTW weights should include true treatment confounders (covariates related to both the treatment assignment and outcome) and covariates related only to the outcome (Brookhart et al., 2006). Including both of these types of covariates in propensity score models has been shown to minimize the mean squared error (MSE) of the outcome model parameter estimates (Brookhart et al., 2006). Further, including covariates that are only predictive of treatment assignment may inflate the variance of the estimator (Brookhart et al., 2006). However, variable inclusion for IIW (and hence FIPTIW) models has not yet been investigated in the literature.

Finally, extreme IPTW weights can occur when the estimated propensity scores are close to zero or one (near violations of Assumption T4). It has been shown that these extreme weights can lead to an increase in variance of the estimated ATE (Stuart, 2010). One solution to handling extreme weights when using IPTW is weight trimming (also called truncation) where weights above a certain threshold are set to a maximum value (Potter, 1993; Scharfstein et al., 1999). The threshold is often determined using percentile cut points where weights above the p_ath percentile are set to the value of the percentile p_a (Cole and Hernán, 2008). Similarly, we can set weights below the $(1 - p_b)$th percentile to the value of the $(1 - p_b)$th percentile. Weight trimming has been shown to improve the estimation of IPTW weights using logistic regression, but not when using classification and regression trees (CART) or random forests (Lee et al., 2011). Although the FIPTIW method can produce extreme weights from either process, the impacts of extreme weights and weight trimming have yet to be examined for FIPTIW.

In the following section, we aim to fill in the gaps in the existing literature by providing preliminary investigations on the impacts of violations of the noninformative censoring assumption and the inclusion of IPCW weights into the FIPTIW model in Section 3.4.2, variable inclusion in intensity models in Section 3.4.3, and weight trimming in Section 3.4.4.

3.4 Simulation Studies

3.4.1 Data Generating Mechanisms

The data generating mechanisms presented in this section are based on the simulation studies presented in Buzkova and Lumley (2007). We consider the scenario where we wish

to determine the ATE of a time-invariant treatment D on a binary longitudinal outcome $\boldsymbol{Y}(t)$. We note that in many scenarios, the estimation of a time-varying treatment may be of interest, however we limit our simulations to the setting where the treatment is time-invariant. Coulombe et al. (2021b) performed various studies using time-varying treatment and confounders, with similar results to the time-invariant setting. We let $D_i = 1$ indicate that individual i is in the treatment group and $D_i = 0$ indicate that individual i is in the control for the duration of the study. We also consider three other covariates, W, $G(t)$, and Z, which are each specified below, that may be related to the observation and/or treatment assignment processes, where $0 \leq t \leq \tau$.

For each of the n individuals, we simulate covariates such that $G(t) = W\log(t)$ where $W \sim Unif(0,1)$, D is a binary covariate with probability π, and $Z \sim N(\mu_{z,0}, \sigma^2_{z,0})$ if $D = 0$ or $Z \sim N(\mu_{z,1}, \sigma^2_{z,1})$ if $D = 1$. We let $\mu_{z,0} = 2$, $\sigma^2_{z,0} = 1$, $\mu_{z,1} = 0$, and $\sigma^2_{z,1} = 0.5$. The probability of being in the treated group π will vary in each simulation study.

To fit the data, a logistic model will be used as estimating odds ratios are of primary interest. However, to simulate the data, we will approximate the logit link using the probit link where $\Phi(x/1.7) \approx \exp(x)/(1+\exp(x))$ to make marginalization more straightforward, as in Buzkova and Lumley (2007).

We generate the outcome as

$$Y_i(t) = \mathbb{1}\left[f_0^*(t) + \beta_1^* D_i + \beta_2 G_i(t) + \beta_3 Z_i + \phi_i + \epsilon_i(t) > 0\right], \tag{3.4}$$

where $\epsilon_i(t) \sim N(0, \sigma^2_\epsilon)$ is the random error term (for each individual and time) and $\phi_i \sim N(0, \sigma^2_\phi)$ is an individual random effect that allows outcomes to be correlated within the same individual. From Equation (3.4), we have

$$\begin{aligned} E\{Y_i(t)|D_i\} &= P\{f_0^*(t) + \beta_1^* D_i + \beta_2 G_i(t) + \beta_3 Z_i + \phi_i + \epsilon_i(t) > 0|D_i\} \\ &= \Phi\left[\frac{f_0^*(t) + \beta_1^* D_i + \beta_2 E\{G_i(t)|D_i\} + \beta_3 E\{Z_i|D_i\}}{\sqrt{\beta_2^2 VAR\{G_i(t)|D_i\} + \beta_3^2 VAR\{Z_i|D_i\} + \sigma^2_\epsilon + \sigma^2_\phi}}\right]. \end{aligned} \tag{3.5}$$

Thus, to ensure proper marginalization, we can let $\beta_1^* = \beta_1 M$ and $f_0^*(t) = (2-t)M - \beta_2 E\{G_i(t)|D_i\} - \beta_3 E\{Z_i|D_i\}$ where

$$M = \frac{\sqrt{\beta_2^2 VAR\{G_i(t)|D_i\} + \beta_3^2 VAR\{Z_i|D_i\} + \sigma^2_\epsilon + \sigma^2_\phi}}{1.7} \tag{3.6}$$

so that

$$E\{Y_i(t)|D_i\} = \Phi((2-t)+\beta_1 D_i)/1.7) \approx \exp((2-t)+\beta_1 D_i)/(1+\exp(2-t)+\beta_1 D_i)). \tag{3.7}$$

We also compare the results to the case where a continuous outcome is simulated. For details on this simulation, refer to Appendix A.2.1.

For the simulations, we let $\beta_1 = 0.5$, $\beta_2 = 2$, $\beta_3 = 1$, $\sigma^2_\phi = 1.25$, and $\sigma^2_\epsilon = 1$. We simulate a random effects model to create a more realistic scenario for longitudinal data, however the interest lies in the estimation of the marginal model.

The estimation of β_1, the average treatment effect, is of primary interest for each simulation study. To estimate the outcome model, the time-varying intercept $m(t)$ must also be estimated or specified. Unless otherwise indicated, we estimate $m(t)$ using cubic splines along with a constant intercept as in Coulombe et al. (2021b). The knots of the splines are chosen by finding the tertiles of the observation times to estimate $m(t)$. In some settings, $m(t)$ will be assumed to be known, and will be modeled in the outcome model as an offset term.

We also consider a model for the observation times, which has an intensity specified as

$$\lambda_{i,obs}(t) = \nu_i \lambda_0(t) \exp\{\gamma_1 D_i + \gamma_2 G_i(t) + \gamma_3 Z_i\}, \tag{3.8}$$

where ν_i is Gamma distributed with mean one and variance $\sigma^2_\eta = 0.1$, which makes the observation times positively correlated. We let $\lambda_0(t) = \frac{\sqrt{t}}{2}$, and consider various values for $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3)$, which will vary across the simulations. To simulate the observation times, We use the thinning method presented by Lewis and Shedler (1979). The thinning method is one of the most popular methods for generating a non-homogeneous Poisson process (NHPP) and is based on finding a constant intensity function $\bar{\lambda}$ for a homogeneous Poisson process (HPP) that dominates the desired intensity $\lambda(t)$. A rejection-acceptance algorithm is used to reject a portion of the generated observation times until the desired intensity or rate is achieved. Algorithm 1 shows the thinning algorithm for an intensity function $\lambda(t)$ over $(0, \tau]$. We also generate the censoring time as $C_i \sim Unif(\tau/2, \tau)$ where τ is the study end time, unless otherwise indicated. We let $\tau = 7$ in all simulations. Directed acyclic graphs (DAGs) will be constructed for all data generating mechanisms.

For all simulation studies, the Robins (2000) type standard error estimates are used. These estimates are more conservative than those developed specifically for the FIPTIW method in Coulombe et al. (2021b). These standard errors tend to be larger than those developed in Coulombe et al. (2021b), but still undercover the 95% confidence interval. As such, we omit the Coulombe et al. (2021b) standard error estimates from the analysis.

All computations are performed using R version 4.4.0 (R Core Team, 2020). Code for all of the simulation studies are publicly available on GitHub at `https://github.com/grcetmpk/FIPTIW`.

Algorithm 1: Thinning Algorithm

Input: Intensity function $\lambda(t)$, maximum follow-up τ

Initialize $n = 0, m = 0, t_0 = 0, s_0 = 0, \bar{\lambda} = \sup_{0 \leq t \leq \tau} \lambda(t)$;
while $s_m < \tau$ **do**
    Generate $u \sim Unif(0, 1)$;
    Let $b = -\log(u)/\bar{\lambda}$;
    Set $s_{m+1} \leftarrow s_m + b$;
    Generate $R \sim Unif(0, 1)$;
    if $R \leq \lambda(s_{m+1})/\bar{\lambda}$ **then**
        $t_{n+1} \leftarrow s_{m+1}$;
        $n \leftarrow n + 1$;
    $m \leftarrow m + 1$;
if $t_n \leq \tau$ **then**
    return $\{t_k\}_{k=1,2,\ldots,n}$;
else
    return $\{t_k\}_{k=1,2,\ldots,n-1}$;

3.4.2 Simulation I: Violations of the Noninformative Censoring Assumption

As it is known that censoring/dropout can bias model parameters when it is related to the longitudinal outcome (Ma et al., 2005), we investigate the performance of the FIPTIW-GEE under various violations of the noninformative censoring assumption in this simulation study. We also examine how further incorporating IPCW weights into the FIPTIW method affects estimation of the outcome model parameters. We denote the FIPTIW method that also incorporates IPCW weights as the FIPTICW method.

We simulate covariates as in Section 3.4.1, and we simulate the probability of being treated as $\pi_i = \text{expit}(\alpha_0 + \alpha_1 W)$ where $\alpha_0 = -1$ and $\alpha_1 = 1.5$. The outcome is simulated as in Equation (3.4) and we simulate observation times according to Equation (3.8), where we let $\gamma_1 = 0.5, \gamma_2 = 0.3$, and $\gamma_3 = 0.6$.

In this simulation, we allow the covariates driving the observation times to also be related to the censoring time. We specify the censoring hazard as

$$\lambda_{i,c}(t) = \lambda_{0,c} \exp\{\eta_1 D_i + \eta_2 W_i + \eta_3 Z_i\}, \tag{3.9}$$

where $\lambda_{0,c} = 0.1 \times t$ is the baseline censoring hazard. In this simulation, we let $\eta_1 =$

0, $\eta_2 = (0, 0.2, 0.5)$, and $\eta_3 = (0, 0.4, 0.6)$ to see how the strength of the relationship between various covariates and the censoring times affects the estimation of the ATE, β_1, in Equation (3.7). Under this hazard function, when $\eta_2 = \eta_3 = 0$, the noninformative censoring assumption is satisfied. Otherwise, the assumption is violated.

To simulate the censoring time, we can use the inverse probability method proposed by Bender et al. (2005). We can use the PH model in Equation (3.9) to construct a survival function as

$$S(t) = \exp\{H_0(t) \times \exp(\eta_1 D_i + \eta_2 W_i + \eta_3 Z_i)\},$$

where $H_0(t) = \int_0^t \lambda_{0,c}(u)du$. So long as $\lambda_{0,c} > 0$ at all t, then the survival/censoring time can be expressed as

$$C_i = H_0^{-1}\left[-\log(U) \times \exp(-(\eta_1 D_i + \eta_2 W_i + \eta_3 Z_i))\right],$$

where $U \sim Unif(0, 1)$ (Bender et al., 2005). As such, we simulate the censoring time for each individual by randomly drawing U_i from a uniform distribution on (0,1) and calculate

$$C_i = \sqrt{\frac{2}{0.1}(-\log(U_i) \times \exp(-\eta_1 D_i + -\eta_2 W_i + -\eta_3 Z_i))}$$

for each individual.

Under this data generating mechanism, Assumptions O1, O3, O4, and O5 hold for the observation process, and T1 to T5 hold for the treatment assignment process. Assumption O2 will only hold when η_2 and η_3 are both zero, as W and $Z(t)$ are not included in the outcome model. The relationship between covariates at a single time point are shown in the DAG in Figure 3.1.

For each combination of (η_1, η_2, η_3), we simulate 1000 data sets for varying sample sizes of $n = 50$, 100, and 500. For each data set, we estimate the ATE, β_1, from an unweighted independent GEE and independent GEEs weighted by IPTW, IIW, FIPTIW, and FIPTIW including the IPCW weights (FIPTICW), where the IPCW, IIW, and IPTW weights were multiplied together. Stabilized IIW weights are estimated using a Cox PH model, stabilized IPTW weights are estimated by logistic regression, and the IPCW weights are estimated by a second Cox PH model. From each of these models over the 1000 simulated data sets, we calculate the empirical bias, MSE, and 95% confidence interval coverage (using the Robins et al. (2000) type standard error estimates) under each simulation scheme. All models are correctly specified.

The results for $n = 100$ are shown in Figure 3.2 and Table 3.1. The results for $n = 50$ and $n = 500$ when a binary outcome is simulated, along with the results for $n = 100$ when a normally-distributed outcome is simulated, can be found in Appendix A.2.3, and are summarized below.

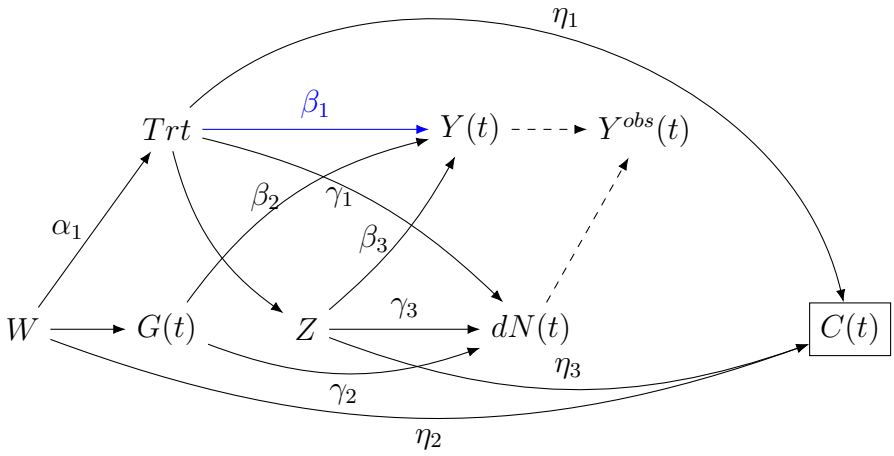


Figure 3.1: A DAG representing the data generation mechanism at a single time point for Simulation I. The estimand of interest is the ATE, β_1.

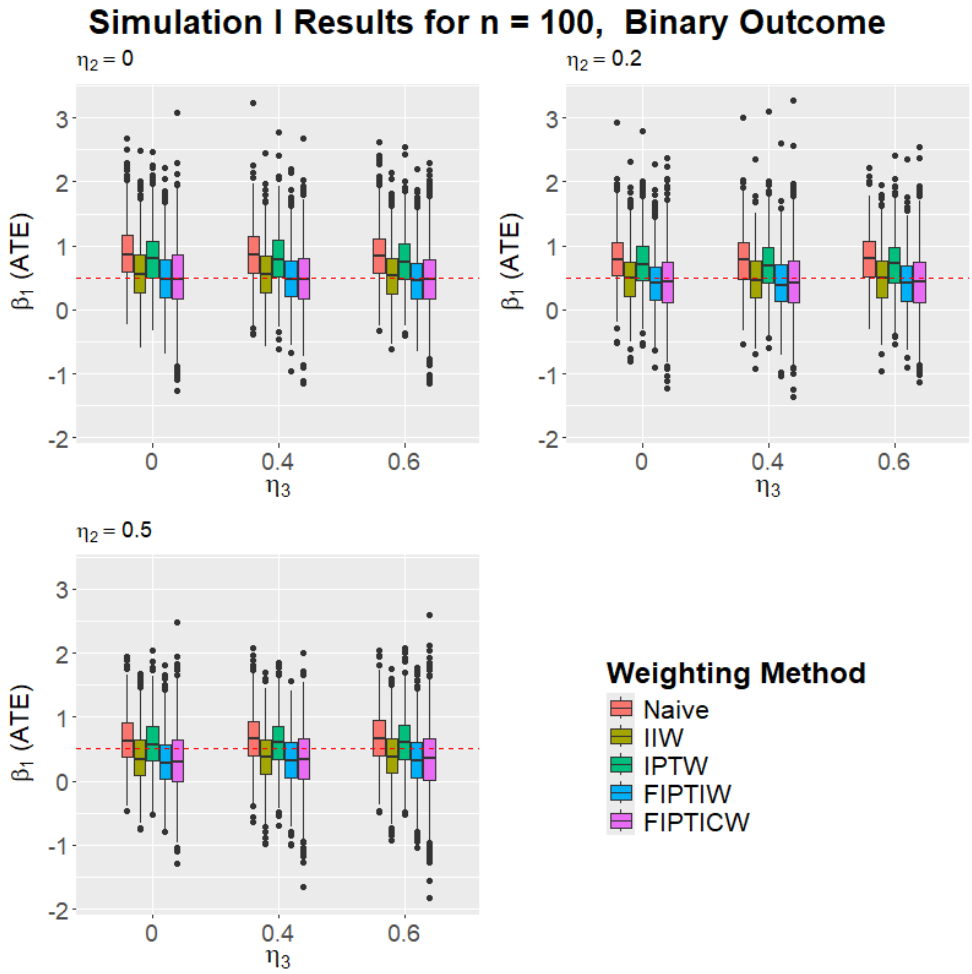


Figure 3.2: Results of Simulation I for $n = 100$ where a binary outcome is simulated. Average treatment effect (ATE, β_1) is calculated by fitting an independent GEE with various weights for each simulation scheme over the 1000 generated data sets. The weighting methods include no weighting (unweighted), inverse intensity weighting (IIW), inverse probability of treatment weighting (IPTW), flexible inverse probability of treatment and intensity weighting (FIPTIW) and flexible inverse probability of treatment weighting with inverse probability of censoring weights included (FIPTICW). The true value of the ATE is 0.5, and is denoted by the red horizontal line.

		Unweighted			**IIW**			**IPTW**		
η_2	η_3	Bias	AvgSE	Coverage	Bias	AvgSE	Coverage	Bias	AvgSE	Coverage
0.0	0.0	0.396	0.405	0.839	0.072	0.400	0.927	0.308	0.401	0.865
	0.4	0.374	0.408	0.848	0.055	0.404	0.937	0.292	0.404	0.879
	0.6	0.354	0.401	0.859	0.037	0.396	0.932	0.264	0.398	0.902
0.2	0.0	0.299	0.396	0.890	-0.011	0.395	0.935	0.224	0.393	0.908
	0.4	0.277	0.392	0.889	-0.026	0.395	0.933	0.203	0.387	0.904
	0.6	0.296	0.396	0.891	-0.012	0.394	0.940	0.212	0.391	0.906
0.5	0.0	0.143	0.377	0.925	-0.143	0.386	0.907	0.081	0.375	0.924
	0.4	0.167	0.384	0.908	-0.119	0.393	0.907	0.097	0.380	0.920
	0.6	0.174	0.382	0.924	-0.110	0.390	0.914	0.109	0.380	0.937

		FIPTIW			**FIPTICW**		
η_2	η_3	Bias	AvgSE	Coverage	Bias	AvgSE	Coverage
0.0	0.0	0.000	0.400	0.922	0.011	0.445	0.899
	0.4	-0.013	0.404	0.928	-0.016	0.445	0.911
	0.6	-0.038	0.396	0.938	-0.019	0.438	0.918
0.2	0.0	-0.073	0.395	0.929	-0.063	0.441	0.899
	0.4	-0.085	0.392	0.926	-0.067	0.432	0.891
	0.6	-0.082	0.393	0.918	-0.065	0.435	0.881
0.5	0.0	-0.195	0.385	0.890	-0.173	0.425	0.877
	0.4	-0.181	0.390	0.902	-0.159	0.425	0.886
	0.6	-0.165	0.389	0.905	-0.150	0.436	0.886

Table 3.1: Empirical bias, Average standard error ("AvgSE") and 95% confidence interval coverage estimates for Simulation I for $n = 100$. Average treatment effect (ATE, β_1) is calculated by fitting an independent GEE with various weights for each simulation scheme over the 1000 generated data sets. The weighting methods include no weighting (unweighted), inverse intensity weighting (IIW), inverse probability of treatment weighting (IPTW), flexible inverse probability of treatment and intensity weighting (FIPTIW) and flexible inverse probability of treatment weighting with inverse probability of censoring weights included (FIPTICW). Standard errors are calculated using the Robins et al. (2000) type standard error estimates.

When $n = 100$ and η_1, η_2 and η_3 are all zero, the noninformative censoring assumption (Assumption O2) is satisfied. In this setting, the FIPTIW and FIPTICW methods have negligible empirical biases while the other methods produce biased estimates. The FIPTICW method, though showing the smallest empirical bias, has a large variance. As the magnitude of η_2 increases, more emprical bias is introduced into our estimated ATEs under the FIPTIW and FIPTICW methods. Increasing η_3 did not impact the empirical bias as much as increasing η_2.

There are many settings where the FIPTIW and FIPTICW methods had worse performance than the IIW and IPTW methods in terms of both empirical bias and variance. Aside from when η_2 and η_3 are zero, the IIW and/or the IPTW methods outperform the FIPTIW and FIPTICW methods in most settings. In some cases, the unweighted method outperforms the FIPTIW and FIPTICW methods. These results are surprising as the IIW and IPTW methods only account for one of the three sources of bias in the data. The unweighted method accounts for none of the three sources of bias. We also note that often the best performing method in terms of bias and variance still results in biased estimation.

In general, the FIPTICW method tends to have a slightly smaller empirical bias than the FIPTIW method. However, it also tends to have a higher variance and is still biased in many settings as the IIW weights are not adequately estimated. In all settings the coverage is below 95%. The FIPTICW method tends to see the highest average standard error.

We see similar results for $n = 50$ (Figure A.1 in Appendix A.2.3) with larger variances for the estimated ATEs. We again see that the FIPTIW and FIPTICW methods are often outperformed by the IIW and IPTW methods, and provide empirically biased results for the estimated ATEs. When $n = 500$ (Figure A.2 in Appendix A.2.3), we see a decrease in the variances of the estimated ATEs. We did not see any extremity in the estimated ATEs as seen in the $n = 50$ case, however we did see a slight empirical bias introduced when $\eta = \{0, 0, 0\}$, which was surprising. Again, the FIPTIW and FIPTICW methods are often empirically biased and outperformed by the IIW and IPTW methods. Similar trends are seen for the case where a continuous, normally-distributed outcome is simulated with $n = 100$ (Figure A.3 in Appendix A.2.3).

From these results, we conclude that there is moderate sensitivity to the noninformative censoring assumption for the FIPTIW method. That is, the FIPTIW may result in spurious estimates of the ATE when the noninformative censoring assumption is not satisfied. Further including the IPCW weights into the FIPTIW model does not fully adjust for the empirical bias introduced by the noninformative censoring assumption, not to mention it results in increased variance in most settings. As such, work is needed to extend existing methodology to allow for the consideration of informative censoring in the analysis.

3.4.3 Simulation II: Variable Inclusion in Inverse Intensity Weighting Models

To examine which variables should be included in IIW models, we perform a simulation study similar to Brookhart et al. (2006).

In this simulation, we simulate the covariates as in Section 3.4.1 where the probability of being treated is $\pi = 0.5$. That is, the treatment assignment is randomized with equal likelihood of treatment allocation. We choose to simulate treatment assignments that are completely randomized to focus only on the variable inclusion for IIW, as variable inclusion for IPTW has previously been investigated (Brookhart et al., 2006; Zhu et al., 2015b). We simulate the outcome as in Equation (3.4). We simulate the observation times using the intensity model in Equation (3.8) where we let $\gamma_1 = 0.5$, $\gamma_2 = \{0, 0.3, 0.6\}$, and $\gamma_3 = 0.6$. We additionally let $\beta_2 = \{0, 2\}$. Under this mechanism, assumptions O1, O2, O3, and O5 hold. Assumption O4 will hold if all covariates are included in the intensity model (i.e. if the model is correctly specified). As the treatment assignment was randomized with equal likelihood, T1 to T5 implicitly hold. The DAG is shown for a single time point in this simulation study in Figure 3.3.

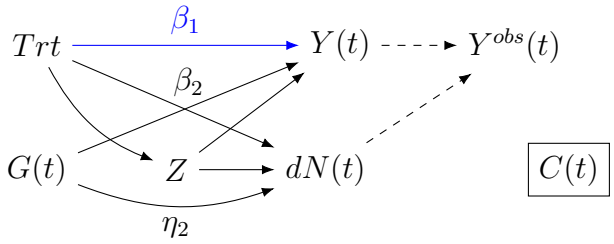


Figure 3.3: A DAG representing the data generation mechanism at a single time point for Simulation II. The estimand of interest is the ATE, β_1.

Under this data generating mechanism, we only need to fit an IIW model to adjust for the conditionally ignorable observation process as the treatment assignment is fully randomized. The covariate Z is related to the observation intensity, treatment, and the longitudinal outcome, making it a confounder for the estimation of the ATE. When $\beta_2 = 0$ and $\gamma_2 = 0$, the covariate $G(t)$ is not related to the observation intensity or longitudinal outcome. When $\beta_2 = 2$ and $\gamma_2 = 0$, the covariate $G(t)$ is only related to the longitudinal outcome. When $\beta_2 = 0$ and $\gamma_2 = 0.3$, the covariate $G(t)$ is only related to the observation intensity. When $\beta_2 = 2$ and $\gamma_2 = 0.3$, $G(t)$ is related to both processes but is not a confounder as it is not related to the treatment assignment.

We consider sample sizes of $n = 50, 100$, and 500. For each sample size, we generate 1000 data sets and fit seven models for the observation intensity using all possible combinations of D, $G(t)$, and Z as covariates. For each of these models, we obtain estimated stabilized IIW weights by fitting a Cox PH model. We then use these estimated weights in an independent GEE to obtain an estimate of β_1, the ATE. We also fit an unweighted model for comparison. From the 1000 data sets, we calculate the empirical bias, variance, MSE, and 95% confidence interval coverage (using the Robins et al. (2000) type standard error estimates) of β_1 for each possible combination of covariates in the weighting model. The results for $n = 100$ are shown in Table 3.2.

For a sample size of 100, when $\gamma_2 = 0$ and $\beta_2 = 0$, the covariate $G(t)$ is not related to the observation intensity or longitudinal outcome. In this setting, any weighting model that does not include the confounder Z and treatment D are empirically biased. Further including an unrelated covariate $G(t)$ does not impact the empirical bias, MSE, average SE, or coverage in this setting.

When $\gamma_2 = 0$ and $\beta_2 \neq 0$, the covariate $G(t)$ is related only to the longitudinal outcome. Again, any weighting model that does not include the confounder Z and treatment D produces estimates that are empirically biased. Further including the covariate only related to the longitudinal outcome, $G(t)$, did not influence the empirical bias, MSE, average SE, or coverage of the ATE.

When $\gamma_2 \neq 0$ and $\beta_2 = 0$, the covariate $G(t)$ is only related to the observation intensity, but neither the treatment nor the outcome. Again, any weighting model that does not include the confounder Z and treatment D is empirically biased. For $\gamma_2 = 0.3$, including $G(t)$ in the model decreases the empirical bias and slightly decreases the MSE. When $\gamma_2 = 0.6$, we see a more substantial reduction in empirical bias and MSE when $G(t)$ is included in the model. The standards error and coverage were similar or slightly better when including $G(t)$ in the intensity model.

When both γ_2 and β_2 are non-zero, $G(t)$ is related to both the observation and outcome processes but is unrelated to the treatment assignment. In this setting, the only models that produce estimates where the empirical bias is negligible include both D and Z. When $\gamma_3 = 0.3$, further including $G(t)$ in the model slightly increased the empirical bias and reduced the MSE. When $\gamma_2 = 0.6$, the empirical bias, and MSE substantially decreased when including $G(t)$ in the model. The coverage and average standard errors were similar.

The results for $n = 50$ and $n = 500$ can be found in Tables A.6 and A.7 in Appendix A.2.4. For the case of $n = 50$, we observe similar trends and see larger variances of the parameter estimates. Similarly, for $n = 500$ we observe similar trends and see smaller variances of the parameter estimates.

The results of this simulation highlight the importance of including observation process confounders in the intensity model. Including covariates that do not confound the relationship between the observation times and the outcome improved the estimation in many cases, with no substantial increases in variance. That is, we do not see an increase in variance when including variables that are only related to the observation intensity, which is contrary to what has been shown in the IPTW literature. As such, it is recommended to be conservative (i.e., more inclusive) with the variables included in the intensity model if there is reason to believe they could potentially be related to the outcome and/or observation times.

3.4.4 Simulation III: Weight Trimming

In this simulation, we consider weight trimming under FIPTIW when the IPTW and IIW weights are extreme due to the underlying observation intensity and/or propensity score. We recognize extreme weights can also occur under model misspecification, but limit this analysis to the scenario where the observation intensity and propensity score models are correctly specified. We also examine the impact of using cubic splines to estimate the time-varying intercept $m(t)$ under extreme weights.

We simulate data as in Section 3.4.1 where the probability of being treated is $\pi = \text{expit}(\alpha_0 + \alpha_1 W)$ where $\alpha_0 = -1$ and α_1 will vary for each simulation scheme. In this simulation, we generate the outcome as in Equation (3.4). We simulate the observation times according to the intensity in Equation (3.8), where the parameters $\boldsymbol{\gamma}$ will also vary for each scenario. Under this data generating mechanism, Assumptions O1 to O5 and T1 to T5 all hold. That is, both the observation and treatment assignment processes are conditionally ignorable. The DAG representing the data generating mechanism is shown in Figure 3.4

We consider scenarios where we have varying degrees of informativeness in the treatment assignment and observation process. To simulate various strengths of informativeness in the propensity score model, we set (α_0, α_1) to $(0, 0.5)$, $(0, 4)$, or $(0, 5.5)$ which correspond to low, moderate, and high degrees of informativeness. To simulate various observation processes, we set $(\gamma_1, \gamma_2, \gamma_3)$ to $(0.5, 0.3, 0.6)$, $(0.5, 0.3, -0.4)$, or $(0.5, 0.3, -0.6)$, which correspond to low, moderate, and high degrees of informativeness. The degrees of informativeness in each process will affect the resultant IPTW, IIW, and FIPTIW weights. We consider the following combinations of the degrees of informativeness for the treatment assignment/observation processes: low/low, moderate/low, high/low, low/moderate, low/high, and moderate/moderate. Under the moderate/high, high/moderate, and high/high scenar-

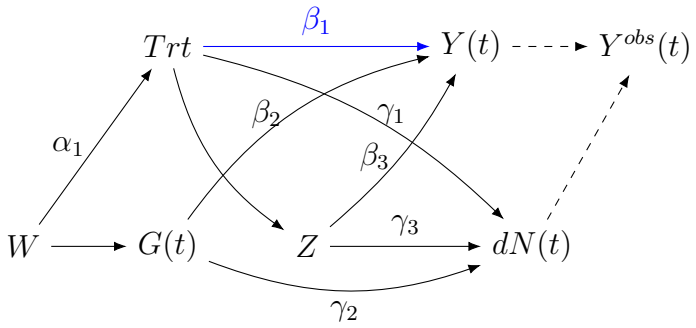


Figure 3.4: A DAG representing the data generation mechanism at a single time point for Simulation III. The estimand of interest is the ATE, β_1.

ios, we end up with propensity score estimates of zero or one which causes the estimation of the ATE to break down. We omit the results of such analyses.

For each simulation scheme, we simulate 1000 data sets with $n = 100$ and calculate the stabilized IIW weights and stabilized IPTW weights through Cox PH and logistic regression models, respectively. We obtain various estimates of the ATE (β_1) by trimming weights to the pth percentile from $p = 0.50$ to 1.00 in increments of 0.01 for each data set. We consider trimming the individual IIW and IPTW weights prior to multiplying (which we refer to as "trimming first") and also trimming the FIPTIW weights after multiplying the IIW and IPTW weights together (which we refer to as "trimming after"). For each set of weights, we use the FIPTIW-GEE to obtain estimates of the ATE, β_1. We consider the FIPTIW-GEE using cubic splines to estimate the intercept $m(t)$, and compare these results to the setting where the functional form of $m(t)$ is specified in the model as an offset term. We then aggregate the results and plot the relative (empirical) bias (RB), MSE, and and 95% confidence interval coverage (using the Robins et al. (2000) type standard error estimates) of the estimated ATE over the 1000 data sets for each cut point and trimming method considered.

Table 3.3 shows the average proportion of the estimated weights larger than 5, 10, and 20 prior to trimming across the 1000 simulated data sets. When the observation process is simulated to have a low degree of informativeness, we see the mean proportion of IPTW and FIPTIW weights larger than 5, 10, and 20 tends to increase as the treatment assignment process becomes more informative. When the treatment assignment process has low informativeness, we see the proportion of weights larger than 5, 10, and 20 monotonically increases as the observation process becomes more informative.

Figures 3.6 and 3.5 show the results of the estimated ATE for a binary outcome for

each of the six scenarios considered. We also simulated data using a continuous, normally-distributed outcome as described in Appendix A.2.1. These results are presented in Appendix A.2.5.

When the time-varying intercept is known and a binary outcome was simulated, the results show that weight trimming was beneficial in some cases. When the observation process is moderately or highly informative, weight trimming around the 95th percentile does slightly reduce the empirical bias of the estimated ATE. However, the differences in empirical bias, MSE, and coverage are small compared to when no trimming was used. That is, weight trimming does not substantially reduce the empirical bias or MSE of the estimated ATEs. We further see that the results are similar for trimming the IIW and IPTW weights before calculating the FIPTIW weights, and trimming the FIPTIW weights after multiplication.

When a cubic spline is used to estimate the time-varying intercept in the outcome model, the model is highly sensitive to extreme weights. In most settings, the outcome model produces biased estimates at all percentiles. The bias tends to be minimized at the 100th percentile (no trimming), which in some cases, is still producing empirically biased estimates. In some settings, weight trimming does help to minimize the MSE or maximize the coverage. However, we see substantial undercoverage in many settings where the treatment assignment process is moderately or highly informative. In the extreme case of a moderately informative treatment assignment and observation processes, the estimated ATEs were so poor that the empirical bias and MSE were in the thousands. We did not see this behaviour when the time-varying intercept was known.

For the case where a continuous, normally-distributed outcome is generated (see Appendix A.2.5), the same sensitivity is present when cubic splines are used under extreme weights. However, weight trimming appears to be more beneficial for the continuous outcome case when the time-varying intercept is known and is modeled using an offset. The bias and MSE are often minimized around the 95th percentile, although the results do not substantially differ when no trimming is used.

3.5 Application to the Program for Resistance, Immunology, Surveillance, and Modeling of Malaria in Uganda Data Set

Malaria is spread through infected female mosquitoes, which require aquatic habitats to breed (Shayo et al., 2021). Larval source management has been used as a malaria pre-

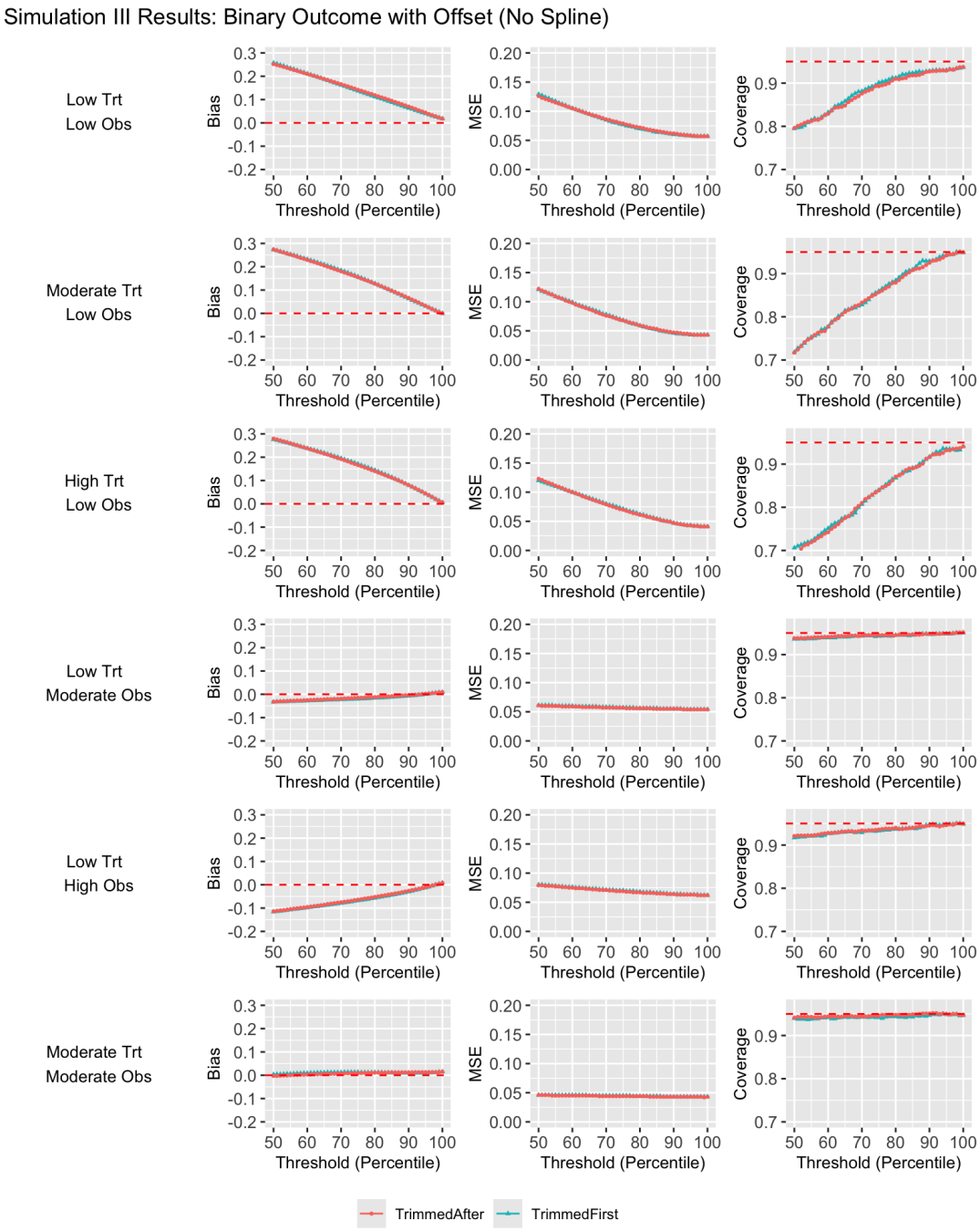


Figure 3.5: Results of the simulation under various degrees of informativeness (low, moderate, and high) in the treatment assignment ("Trt") and observation ("Obs") processes when the time-varying intercept is known. "Trimmed before" refers to trimming the individual IPTW and IIW weights prior to multiplying and calculating the FIPTIW weights. "Trimmed after" refers to trimming the FIPTIW weights after multiplying the IPTW and IIW weights together. The true average treatment effect β_1 is simulated to be 0.5. The parametric form of the time-varying intercept is known, and included as an offset term in the outcome model.

vention strategy to reduce both the densities of infected adult mosquitoes and prevalence of malaria in humans (Msellemu et al., 2016; Mwakalinga et al., 2018). A recent study has shown an association between residing in households with non-piped water sources and malaria diagnoses in a sample of 6707 children in Tarzania. Further, a recent meta-analysis found that unprotected household water sources such as unprotected wells, springs, rivers, dams, and streams were associated with higher prevalence of malaria in young children (Yang et al., 2020). We similarly wish to evaluate if there is an increase in the odds of malaria diagnosis for those residing in households with unprotected water sources compared to those with non-piped, protected water sources among children ages two to 11 living in Uganda. The code used to produce the results this data application is openly available in the GitHub repository `https://github.com/grcetmpk/MalariaFIPTIW`.

We analyze observational data from the Program for Resistance, Immunology, Surveillance, and Modeling of Malaria in Uganda (PRISM) (Kamya et al., 2015) using the FIPTIW method to quantify the association between unprotected water sources and malaria diagnoses. The data are publicly available at `https://clinepidb.org/`. In the PRISM study, households in three sub-counties of Uganda were randomly selected to be enrolled in a longitudinal study. All individuals in the household were enrolled in the study if they met the eligibility criteria outlined by Kamya et al. (2015). Data were collected on individuals enrolled in the study between August 2011 and June 2017 through routine clinical visits roughly every three months. Participants were also able to attend a study clinic (and thus had observations recorded) any time they became ill.

The PRISM study provided demographic information on each study participant, including the individual's age and sex at enrollment, and the individual's weight, height, body temperature, malaria diagnosis, and prescribed treatments at each scheduled and unscheduled assessment. The study also provided information on the household that each patient resides in, including the sub-county in Uganda, dwelling characteristics such as materials, number of sleeping places, waste disposal, household water sources, a categorical and numerical household wealth index, and food security measures such as the number of meals given per day and the number of food problems per week. Household characteristics were matched to the participant through a unique household identifier. Individuals were classified as residing in households with protected water sources if their dwelling sourced drinking water from a protected spring or well, public tap, or borehole. Individuals were classified as having unprotected water sources if their drinking water was sourced from open public wells, rivers and streams, ponds and lakes, or unprotected springs. This categorization of drinking water sources closely follows that of Yang et al. (2020). We note that drinking water source (along with other household characteristics) did not change during follow-up and are considered time-invariant covariates. Participants that had drinking wa-

ter piped into their yard, compound, or dwelling were excluded from the study to focus on the difference within unpiped water sources.

Individuals residing within the same household are likely to be related on biological and environmental factors. To restore the independence assumption between individuals in the study, we randomly selected one child per household to be in the final sample. The final sample size was 287 children, all of whom resided in separate houses. From this sample, 218 children resided in households with protected water sources while 69 resided in households with unprotected water sources. Baseline demographics for other covariates considered in the study are given in Table 3.4.

			Variables used to estimate intensity							
γ_2	β_2		**Naive**	D	$G(t)$	Z	$D, G(t)$	D, Z	$G(t), Z$	$D, G(t), Z$
0	0									
		Bias:	-0.226	-0.226	-0.216	-0.044	-0.220	0.005	-0.043	0.005
		Variance:	0.191	0.191	0.193	0.143	0.189	0.141	0.143	0.141
		MSE:	0.242	0.242	0.240	0.144	0.237	0.141	0.144	0.141
		Avg SE:	0.399	0.399	0.402	0.366	0.399	0.368	0.366	0.368
		Coverage:	0.882	0.882	0.884	0.939	0.888	0.944	0.938	0.944
	2									
		Bias:	-0.223	-0.223	-0.210	-0.052	-0.216	-0.004	-0.051	-0.004
		Variance:	0.180	0.180	0.185	0.144	0.179	0.146	0.144	0.146
		MSE:	0.229	0.229	0.229	0.147	0.226	0.146	0.146	0.145
		Avg SE:	0.402	0.402	0.406	0.370	0.402	0.371	0.370	0.371
		Coverage:	0.893	0.893	0.898	0.943	0.899	0.943	0.943	0.944
0.3	0									
		Bias:	-0.211	-0.211	-0.206	-0.026	-0.210	0.023	-0.030	0.019
		Variance:	0.171	0.171	0.168	0.131	0.164	0.133	0.129	0.131
		MSE:	0.215	0.215	0.210	0.132	0.208	0.133	0.130	0.131
		Avg SE:	0.397	0.397	0.396	0.362	0.392	0.363	0.360	0.361
		Coverage:	0.905	0.905	0.906	0.950	0.908	0.944	0.952	0.947
	2									
		Bias:	-0.227	-0.227	-0.221	-0.050	-0.228	-0.002	-0.057	-0.009
		Variance:	0.204	0.204	0.194	0.154	0.189	0.155	0.147	0.148
		MSE:	0.255	0.255	0.243	0.157	0.240	0.154	0.150	0.148
		Avg SE:	0.405	0.405	0.399	0.368	0.395	0.368	0.362	0.363
		Coverage:	0.873	0.873	0.871	0.923	0.876	0.935	0.930	0.941
0.6	0									
		Bias:	-0.220	-0.220	-0.220	-0.032	-0.224	0.018	-0.039	0.010
		Variance:	0.186	0.186	0.176	0.134	0.172	0.133	0.130	0.129
		MSE:	0.234	0.234	0.225	0.135	0.223	0.133	0.131	0.129
		Avg SE:	0.399	0.399	0.391	0.360	0.388	0.359	0.353	0.354
		Coverage:	0.895	0.895	0.888	0.937	0.889	0.944	0.942	0.943
	2									
		Bias:	-0.185	-0.185	-0.189	-0.006	-0.196	0.044	-0.022	0.027
		Variance:	0.216	0.216	0.193	0.156	0.190	0.156	0.146	0.146
		MSE:	0.250	0.250	0.229	0.156	0.228	0.157	0.147	0.147
		Avg SE:	0.410	0.410	0.394	0.372	0.390	0.371	0.358	0.359
		Coverage:	0.893	0.893	0.888	0.932	0.888	0.927	0.927	0.934

Table 3.2: Simulation results for Simulation II for $n = 100$. Bias, variance, MSE, average standard error (Avg SE), and coverage for 95% confidence interval of the average treatment effect (ATE) is calculated by weighting the outcome model in Equation (3.7) by inverse intensity weighting (IIW) for each simulation scheme over the 1000 generated data sets. Variables included in the IIW model are listed in the table. The true value of the ATE is 0.5.

Degree of Informativeness		**Mean Proportion of IPTW Weights**			**Mean Proportion of IIW Weights**			**Mean Proportion of FIPTIW Weights**		
Treatment Assignment Process	Observation Process	>5	>10	>20	>5	>10	>20	>5	>10	>20
Low	Low	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Moderate	Low	0.52	0.10	0.01	0.00	0.00	0.00	0.02	0.00	0.00
High	Low	0.51	0.17	0.04	0.00	0.00	0.00	0.02	0.00	0.00
Low	Moderate	0.00	0.00	0.00	0.31	0.00	0.00	0.43	0.00	0.00
Low	High	0.00	0.00	0.00	4.66	0.20	0.01	5.11	0.22	0.01
Moderate	Moderate	0.09	0.01	0.00	0.43	0.00	0.00	1.37	0.10	0.01

Table 3.3: Distribution of weights under each simulation scheme for Simulation III when a binary outcome is simulated.

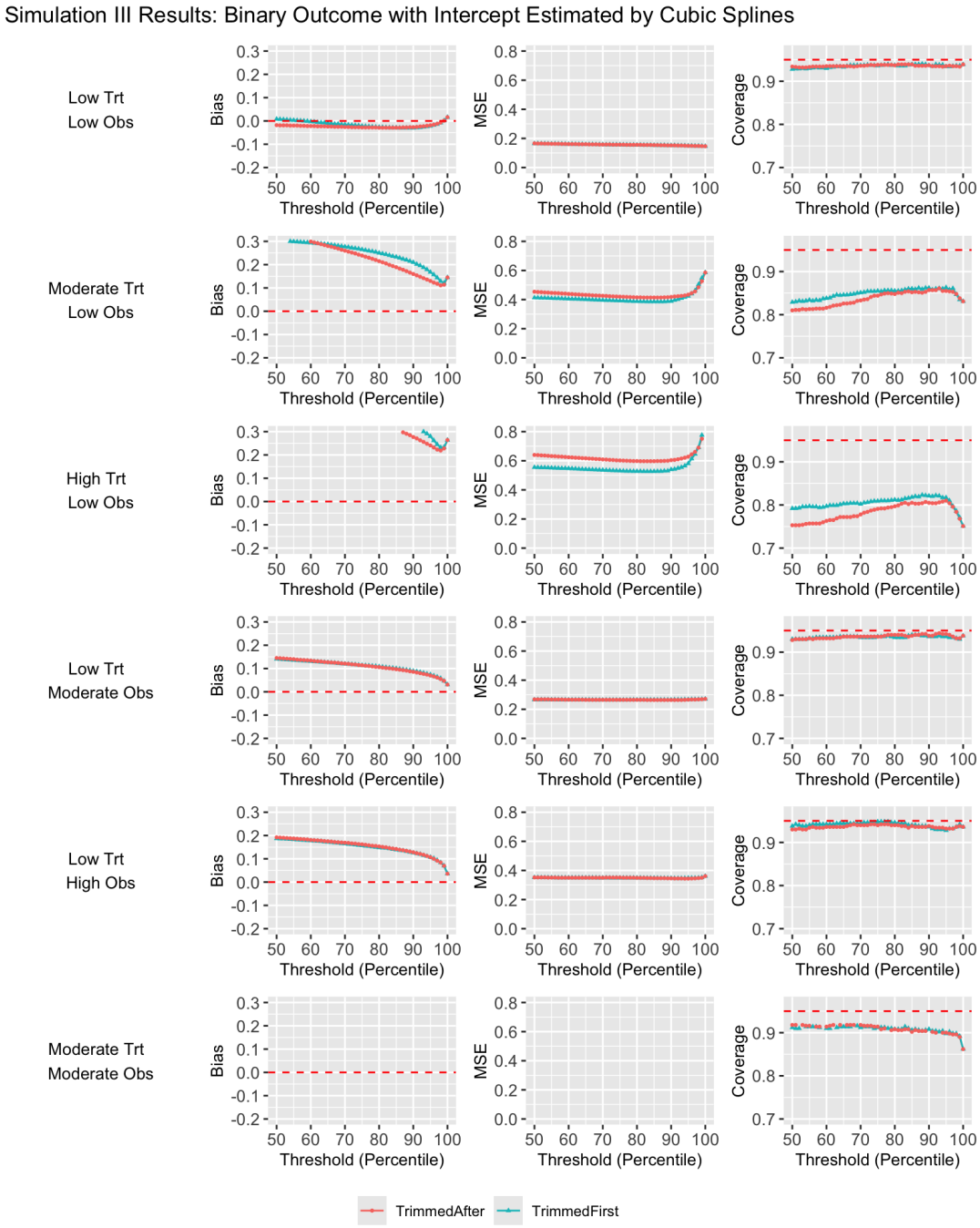


Figure 3.6: Results of the simulation under various degrees of informativeness (low, moderate, and high) in the treatment assignment ("Trt") and observation ("Obs") processes when the time-varying intercept is estimated by cubic splines. "Trimmed before" refers to trimming the individual IPTW and IIW weights prior to multiplying and calculating the FIPTIW weights. "Trimmed after" refers to trimming the FIPTIW weights after multiplying the IPTW and IIW weights together. The true average treatment effect β_1 is simulated to be 0.5. The parametric form of the time-varying intercept is unknown, and estimated using cubic splines.

Covariate	Mean (SD) or n (%)	Covariate	Mean (SD) or n (%)
Age at Enrollment	5.51 (2.03)	**Food Problems per Week**	
Sex		Never	47 (16.38%)
Female	145 (50.52%)	Seldom	50 (17.42%)
Male	142 (49.48%)	Sometimes	97 (33.8%)
Sub-County		Often	38 (13.24%)
Walukuba	85 (29.62%)	Always	55 (19.16%)
Kihihi	100 (34.84%)	**Waste Facilities**	
Nagongera	102 (35.54%)	Covered pit latrine, no slab	103 (35.89%)
Household Wealth Index		Covered pit latrine with slab	18 (6.27%)
Poorest	103 (35.89%)	Composting toilet	4 (1.39%)
Middle	98 (34.15%)	Uncovered pit latrine, no slab	112 (39.02%)
Least poor	86 (29.97%)	Flush toilet	8 (2.79%)
Drinking Water Source		Uncovered pit latrine with slab	8 (2.79%)
Public tap	105 (36.59%)	Vip latrine	3 (1.05%)
Protected public well	24 (8.36%)	No facility	31 (10.8%)
Protected spring	26 (9.06%)	**Number of Persons in House**	6.18 (2.6)
Borehole	63 (21.95%)	**Reason for Withdrawal**	
River/stream	25 (8.71%)	Moved out of area	46 (16.03%)
Open public well	21 (7.32%)	Reached 11 years of age	86 (29.97%)
Pond/lake	5 (1.74%)	Unable to be located for 120 days	15 (5.23%)
Unprotected spring	18 (6.27%)	Did not comply	7 (2.44%)
Unprotected Water Source		Withdrawal of informed consent	4 (1.39%)
No	218 (75.96%)	Completed study	35 (12.2%)
Yes	69 (24.04%)	Unknown	94 (32.75%)
Dwelling Type			
Modern	85 (29.62%)		
Traditional	202 (70.38%)		

Table 3.4: Baseline demographics for the 287 children included in the analysis.

As the FIPTIW method has been shown to be sensitive to violations of the noninformative censoring assumption, we employed artificial censoring (Joffe, 2001; Robins et al., 1992, 1995) to avoid such violations. Data on the reason for study withdrawal, but not the dates of withdrawal, were provided in the data. As such, the true censoring times were unknown. As a proxy for the true censoring time, we considered the last known observation times as the surrogate censoring times for each individual. The surrogate censoring times are used to estimate the unknown true censoring times, which are the earliest possible times at which an individual could be right-censored. Using this definition, surrogate censoring times varied from zero (baseline) to 1781 days (study completion), and 5.65% of individuals were potentially censored prior to six months. Of this set of individuals, 26.83% were considered to potentially violate the noninformative censoring assumption (i.e., dropped out of the study due to reaching 11 years of age, withdrawing informed consent, being unable to comply with the study protocol, being unable to be located for more than 120 days, or for unknown reasons). As we used the smallest possible censoring time for each individual, this means that a maximum of 1.52% of the the total study population could be considered censored prior to six months for reasons possibly related to the longitudinal outcome. We also performed a sensitivity analysis where individuals who had surrogate censoring times prior to six months were randomly censored between zero and six months (see Appendix A.3.3). There were negligible differences in the results of the analysis when random censoring was employed for this small proportion of individuals. As such, we consider this proportion of potentially informative censoring negligible. For this analysis, we use six months (182.5 days) as the point at which all individuals are artificially censored to restore the noninformative censoring assumption.

As patients could visit clinics between scheduled follow-ups, the number of observations per participant varied. From the 287 children we included in this sample, the number of observations ranged from one to 11, with an average of 4.83 observations per participant. Of the 1386 total observations, 287 were for enrollment, 525 were scheduled follow-up appointments, and 574 were unscheduled appointments. Of the 1386 visits, malaria was diagnosed 179 times (12.9%), and 111 of 287 patients (38.7%) had a malaria diagnosis at some point during the study.

We wish to estimate the average treatment effect of a time-invariant indicator of whether or not an individual resides in a household with a non-piped water source on malaria diagnoses. That is, we wish to estimate the marginal model

$$g(E(Y_i(t)|D_i) = m(t) + \beta_1 D_i \tag{3.10}$$

where $Y_i(t)$ is a binary, time-varying indicator of whether or not the patient was diagnosed with malaria at time t, $m(t)$ is a smooth function of time, D_i is a time-invariant indicator

of whether or not the individual resided in a household with unprotected water sources, and $g(\cdot)$ is the logit link function.

To estimate the average treatment effect, we considered adjusting for factors that may have confounded the relationship between household water sources and malaria diagnoses through the propensity score. To do so, we employed a logistic regression model. We considered age, sub-county, categorical wealth index, number of food problems per week, type of household waste disposal, dwelling type, and number of people living in the house as potential confounders. We note that we also have data on whether the patient was prescribed artmether-lumefantrine, quinine, artesunate, or no malaria medication for each observation. These medications are not preventive against malaria but used to treat it. However, in our sample only no observations had antimalarial medications prescribed within one week of the current observation time, so we omit this covariate from the analysis. After an initial model fit, the sub-county and human waste facility covariates caused inflated variances in the propensity score model parameter estimates. As such, these covariates were omitted from the final model. The resultant propensity score model was used to estimate stabilized IPTW weights for each individual at each time. No large IPTW weights were present, as the maximum IPTW weight was 2.52. This is similar to what was seen in the low informativeness category for the IPTW weights we defined in the simulation study in Section 3.4.4.

We also considered the factors driving the observation process. Based on the results of Simulation II (Section 3.4.3), we included covariates believed to be related to both the probability of being observed and the longitudinal outcome. As such, we allowed the probability of visiting a clinic at any given time to depend on whether or not the individual resided in a household with protected water sources, their age, sub-county, dwelling type, categorical household wealth index, degree of food problems per week, number of persons living in the house, and the individual's malaria status at the last clinic visit. A Cox PH model with these covariates was used to estimate stabilized IIW weights. From this model, the maximum weight was 3.66. This is similar to the low informativeness scenario as defined in Simulation II.

The resultant FIPTIW weights were calculated by multiplying the IPTW and IIW weights together for each individual at each time point. The FIPTIW weights were not extreme, as the maximum FIPTIW weight was 3.28.

To estimate β_1 in the outcome model in Equation (3.10), the smooth function of time $m(t)$ was estimated using a cubic spline with a constant intercept, as in Coulombe et al. (2021b). The knots of the spline were chosen by finding the tertiles of the time since enrollment.

Independent GEEs were fit using various weighting methods (none, IPTW only, IIW only, FIPTIW, and trimmed FIPTIW). When employing the trimmed FIPTIW weighting method, we trimmed the FIPTIW weights after multiplication to the 95th percentile. This resulted in FIPTIW weights above 2.20 being trimmed. Robust standard errors developed for inverse probability weighting by Robins et al. (2000) were used to calculate the standard errors of the ATE. The results of the analysis for each weighting method are shown in Table 3.6. Balance after weighting was assessed for the FIPTIW method at baseline by calculating standardized mean differences using the **cobalt** (Greifer, 2024) package in R (R Core Team, 2020). A covariate was considered balanced when the standardized mean difference was less than 0.1, as recommended by Zhu et al. (2015a). The assessment of balance is presented in Table 3.5.

Two sensitivity analyses were also performed for this study. Appendix A.3.1 presents a sensitivity analysis where we include individuals who are censored into the analysis. Appendix A.3.2 presents a sensitivity analysis where we include all children in each household.

Covariate	**Type**	**SMD**	**Balanced?**
Age at Baseline	Continuous	0.101	No
Sub-County (Kihihi)	Binary	0.381	No
Sub-county (Nagongera)	Binary	0.009	Yes
Sub-county (Walukuba)	Binary	-0.390	No
Household Wealth Index (Least poor)	Binary	-0.070	Yes
Household Wealth Index (Middle)	Binary	0.066	Yes
Household Wealth Index (Poorest)	Binary	0.004	Yes
Food Problems Per Week (Always)	Binary	0.015	Yes
Food Problems Per Week (Never)	Binary	0.032	Yes
Food Problems Per Week (Often)	Binary	-0.041	Yes
Food Problems Per Week (Seldom)	Binary	-0.058	Yes
Food Problems Per Week (Sometimes)	Binary	0.051	Yes
Dwelling Type (Traditional)	Binary	0.123	No
Number of Persons Living in House	Continuous	0.235	No
Last Observed Malaria Diagnosis	Binary	0.059	Yes

Table 3.5: Balance assessment for the PRISM study at baseline. Standardized mean differences (SMD) were calculated to assess balance on the FIPTIW weighted (untrimmed) data set for the covariates used in the IIW and IPTW models.

The results show that balance was not completely achieved at baseline for all covariates used in the FIPTIW model. However, as balance was only examined at baseline, it may noy fully capture the true extend of covariate balance across out entire modeling framework.

We see the results differed based on the weighting method that was employed in the analysis. Using no weighting method estimated that the odds of being diagnosed with

Weighting Method	β_1	SE(β_1)	95% CI for β_1	Odds Ratio (OR)	95% CI for OR
None	0.429	0.186	(0.065, 0.793)	1.536	(1.067, 2.211)
IPTW	0.327	0.168	(-0.002, 0.657)	1.387	(0.998, 1.928)
IIW	0.555	0.178	(0.206, 0.903)	1.742	(1.229, 2.468)
FIPTIW	0.424	0.168	(0.095, 0.753)	1.528	(1.099, 2.122)
FIPTIW (Trimmed)	0.409	0.166	(0.084, 0.734)	1.505	(1.087, 2.083)

Table 3.6: Results of the estimation of the odds ratio (OR) of malaria diagnoses for children residing in households with unprotected water sources versus those residing in households with protected water sources (β_1) by different weighting methods for an independent GEE in the PRISM cohort. Sample does not include those with piped water sources.

malaria was 1.54 times higher for individuals residing in a household with unprotected water sources (95% CI: (1.07, 2.21)). This estimate did not account for the non-randomized exposure nor the informative observation process. Employing only IPTW accounted for the non-randomized exposure, and reduced the estimated OR to 1.39, but this estimate was insignificant (95% CI: (1.00, 1.93)). Employing only IIW accounted for the informative observation process and resulted in a larger estimated OR of 1.74 (95% CI: (1.23, 2.47)). Weighting the independent GEE by FIPTIW accounted for both processes, and estimated the OR to be 1.53 (95% CI: (1.10, 2.12)), which was statistically significant.

As we saw in Simulation II in Section 3.4.4, weight trimming was not recommended when we had low informativeness in the treatment assignment and observation processes when using cubic splines to estimate the offset term. However, we estimate the OR using this method to highlight the differences of the estimated OR when weight trimming is employed in this scenario. When FIPTIW weights were trimmed above the 95th percentile, the estimated OR was 1.51 (95% CI: (1.09, 2.08). This result was also statistically significant.

In this example, the naive estimate is similar to the estimate seen when FIPTIW is employed. It is coincidental that in this data set, the observation process and treatment assignment processes are inversely affecting the bias of the ATE to similar degrees. For those without piped water sources, the results of this analysis showed a significant difference in the odds of being diagnosed with malaria for households with unprotected water sources when the observation and treatment assignment processes were accounted for in the analysis, given the available covariates. As such, we conclude from the analysis that efforts should be made to provide individuals with access to protected water sources, along with other protective measures, to reduce their individual risk of malaria.

3.6 Discussion

The analysis of irregular longitudinal data may be complicated by non-randomized treatment assignments and/or informative observation processes, particularly in observational data sets. The FIPTIW method can be employed in certain scenarios to account for these sources of bias. However, we have shown the existing methodology is sensitive to violations of the noninformative censoring assumption, as it may result in biased estimates of causal treatment effects when informative censoring is present. Further, the inclusion of IPCW weights into the FIPTIW model does not account for the bias introduced into the outcome model parameter estimates when informative censoring is present. This is because the estimation of the IIW weights rely on the assumption of noninformative censoring. As such, we have identified estimating IIW weights under informative censoring as an important area of research. Jackson et al. (2014) uses multiple imputation and bootstrapping to estimate the parameters of a Cox PH model under violations of the independent censoring assumption. It may be possible to extend this method to account for dependent censoring in the estimation of IIW weights. Multiple outputation (Pullenayegum et al., 2023) may also be a useful approach under such violations.

Variable inclusion for IIW (and thus FIPTIW) was also investigated. We have shown that omitting covariates that are related to both the observation and outcome processes in the observation intensity models can result in biased estimates of outcome model parameters. Further, we have shown that unlike in propensity score models, the inclusion of additional covariates related only to the observation intensity do not substantially increase the variance of the resultant estimates of outcome model parameters. The results of the simulations have shown that analysts should be conservative (i.e., more inclusive) with the covariates included in intensity models for IIW and FIPTIW, if there is any indication of possibly being related to both the observation process and longitudinal outcome. It may be useful to further increase the parameter γ_2 while leaving $\beta_2 = 0$ to see if there are cases where highly informative covariates increase the variance of the estimates. We identify this as a limitation to the analysis.

The FIPTIW-GEE was shown to be sensitive to extreme weights when the time-varying intercept was estimated by cubic splines. Weight trimming did not alleviate the bias introduced into the parameter estimates in this setting. When the time-varying intercept was assumed to be known and was treated as an offset in the outcome model, the sensitivity to extreme weights was not present. Weight trimming around the 95th percentile did reduce the bias in many cases, however the reduction in bias and MSE was not substantial when compared to the case where no trimming was performed. As in most data sets the time-varying intercept will not be known, we do not recommend using weight trimming.

Analysts should also be aware of the potential implications of extreme weights in their analyses.

In the real data analysis, household level clustering was present. To circumvent this issue, only one individual per household was included in the final sample. To perform a more appropriate analysis where all individuals in each household are included, one could extend the method presented in Pullenayegum et al. (2021) to incorporate within-household correlations in the FIPTIW method. Further, extending the FIPTIW methodology to allow for informative censoring (as previously discussed) would also allow us to use the full sample from the PRISM study. We identify both of these items are identified as important areas of future research.

Another limitation to our real data analysis was that balance was only assessed at baseline. To our knowledge, no specific methods for assessing balance after FIPTIW weighting have been developed for irregular longitudinal data. A more holistic approach to assessing balance over the entire data set would be more informative, however such development was beyond the scope of this thesis chapter.

The real data analysis was also restricted to using time-invariant covariates in the intensity model as covariates related to the observation intensity were required to be known at all possible time points (i.e., every day). Covariates such as whether other members in the household were infected with malaria or whether insecticide treated nets were used are important covariates that may impact both malaria risk and whether or not an individual decides to visit a clinic. We identify this as a limitation to the current analysis.

In most settings, the time-varying intercept will be unknown in the outcome model and must be estimated. Cubic splines have been used to estimate this quantity, however we have shown it is sensitive to extreme weights. If it were desirable to treat the time-varying intercept in the outcome model as a nuisance parameter, the Lin-Ying semiparametric approach (Lin and Ying, 2001) may be able to be extended to include FIPTIW weights to account for both sources of bias. It may also be beneficial to examine weight trimming in the simpler setting where a time-invariant intercept is present in the model.

Power analyses may also be of interest to researchers performing hypothesis testing. Although coverage was investigated in the simulation studies, more detailed power analyses using IIW and FIPTIW are warranted. We identify this as another area of future work.

Chapter 4

Methods for Handling Missing Data in the Estimation of Inverse Intensity Weights

4.1 Introduction

IIW is a weighting method that can be used to adjust for conditionally ignorable observation processes. In this method, which is explained in more detail in Chapter 2, there are two models that are fit. The first is the observation times model, which describes the intensity of observations for each individual as they relate to a set of covariates $\boldsymbol{Z}(t)$. In practice, the observation times model parameters are estimated through a PH model. These parameter estimates are then used to estimate the IIW weights. The second model is a marginal outcome model, which relates the longitudinal outcome $Y(t)$ to a set of covariates $\boldsymbol{X}(t)$. The IIW weights are used in the outcome model to obtain weighted estimates of the outcome model parameters, which (when the assumptions are met) will account for the informative observation process. The estimation of the parameters in the outcome model are typically of primary interest to researchers.

The IIW method may be desirable to analysts as it is relatively intuitive and is simple to implement using standard statistical software. However, IIW relies on a number of assumptions involving the underlying observation process, which are detailed in Section 2.2. One such assumption is that the covariates used in the observation times model are known at *all* possible time points. That is, the covariates related to the probability of having an observation ($\boldsymbol{Z}(t)$) are *always* observed when every possible time point over the duration of

the study is considered. This requires coarsening the (potentially continuous) time space to an appropriate set of possible discrete time points. For example, consider an observational clinical study where we are evaluating a measure of disease severity as the longitudinal outcome. Sicker patients will visit the clinic more often, and thus the longitudinal outcome is likely related to the observation times (i.e., the observation process is informative). If we assume that patients visit the clinic at most once per day, an appropriate set of possible observation times may be each day in the study, from baseline to the study end date. To avoid introducing bias into estimates obtained from IIW, the covariates related to the observation process will need to be known every day for every patient, whether or not the patient visited the clinic and had the longitudinal outcome recorded.

If time-invariant covariates are the only factors driving an individual to be observed, this assumption will be met so long as we have the data recorded at least once. However, time-varying factors may also influence an individual's decision to visit a clinic, such as changes in one's health. The assumptions that these factors are known at all possible observation times is often unreasonable to make in practice, unless data are collected through strict protocols or through, for example, wearable technology that can track symptoms such as heart rate and body temperature in nearly continuous time. However, for observational data, it is much more atypical that covariates are known in continuous time. As such, time-varying covariates may pose challenges for IIW if they are intermittently missing, which results in a missing data problem. Missing data has been shown to affect the asymptotic properties of PH models when the missing data are not accounted for in the analysis (Hsu and Yu, 2019). Ignoring the missing data (i.e., complete case analysis (CCA)) can bias the estimates of PH models, and warrants the use of methods developed specifically for handling missing data.

Marshall et al. (2010) compared various missing data methods for missing covariates in Cox PH models in the context of prognostic modeling for failure time models. However, methods for handling missing data for IIW models, and its effect on the resultant outcome model estimates has, to our knowledge, yet to be explored in the literature under informative observation processes. As such, the goal of this paper is to examine how missing data in the intensity model covariates can affect the estimation of the outcome model parameters, and compare existing missing data methods to provide recommendations on how the missing data can be handled in practice. In Section 4.2, we begin by discussing missing data mechanisms and existing methods for handling missing data. Then, we perform a simulation study in Section 4.3. A real data analysis follows in Section 4.4. We then conclude with a discussion in Section 4.5.

4.2 Methods

Two models involving two different sets of covariates are used when employing IIW, and as such, missingness can occur in one or both sets of covariates. As the focus on this paper is on missingness in the observation model covariates $\boldsymbol{Z}(t)$, we assume that the outcome model covariates $\boldsymbol{X}(t)$ have no missingness.

The degree of bias and the approaches used to account for such biases depend on the missingness mechanism and amount of missingness present in the data. Rubin (1976) proposed classifications of the data based on the missing data mechanisms, which include missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). If the data are MCAR, the probability of the data being observed/missing does not depend on the observed or missing data. If the probability of the data being observed/missing depends only on observed data (and not on any missing data), then we refer to the data as being MAR. When the probability of the data being observed/missing depends on the missing data, we say that the data are MNAR. The estimates of the PH model parameter estimates (and hence the IIW weights) may be biased depending on the missingness mechanism and degree of missingness (Marshall et al., 2010).

In most cases, one cannot simply ignore the missing data. CCA, where any observations with missing data are excluded from the analysis, is known to result in biased estimates of PH model parameters if the data are not missing completely at random (Buuren et al., 1999).

In the original IIW paper, Buzkova and Lumley (2007) recommend using LOCF for unobserved observation model covariates. This method assumes that the covariates in the observation model do not greatly change between two (observed) observation times, which may not be a realistic assumption in practice (Cao and Fine, 2021). Further, it is known that employing this method when a non-trivial proportion of covariates are missing can bias the parameters estimated in PH models used to model survival data (Andersen and Liestøl, 2003; Molenberghs et al., 2002). Lachin (2016) notes that the origins of LOCF are unclear in the literature, and despite its frequent use in applied work, there are no peer-reviewed articles where LOCF provides unbiased estimation.

Single imputation (SI) is a commonly used method to handle missing data, where the missing values are replaced with an estimate in the data set. Though many versions of SI exist, regression-based imputation is commonly used where the missing values are replaced using predictions from a regression model relating the missing covariate to the observed covariates.

Imputation can also be done using multiple imputation by chained equations (MICE). In

MICE, a single imputation is performed for each missing value in the data set. The imputed values are then regressed on other variables in the imputation model, then the missing values are replaced with the predictions from this model. The full imputation process is repeated m times, with imputations being updated at each iteration. Creating multiple imputed data sets can account for the statistical uncertainty associated with imputations (Azur et al., 2011). Further, the use of chained equations allows us to impute various data types, for example continuous, binary, categorical covariates (Azur et al., 2011). The ideal number of imputed data sets to create, m, depends on the amount of missingness in the data and the size of the data set. Authors have noted as few as three imputed data sets are sufficient for data with up to 20% missingness (Buuren et al., 1999), however more recent studies recommend larger numbers of imputations to increase statistical power and decrease the variation of the estimates where computationally possible (Graham et al., 2007; White et al., 2011). Their recommendations depend on the amount of missing data, amount of acceptable "power falloff", the width of the confidence interval, p-value, and desired "fraction of missingness information", which is the ratio of the between- and total-imputation variance. As such, as many as 100 imputations are recommended for some applications.

For imputation methods on Cox PH models, the binary indicator denoting whether or not the subject was observed and the time of the event should be included in the imputation model, along with any other covariates that may be related to the missingness (Austin et al., 2021). Both SI and MICE can be used with or without predictive mean matching (PMM), which incorporates a nonparametric component into the model to select which values are imputed. This may be beneficial when parametric assumptions are not fully met (Marshall et al., 2010). Single imputation with predictive mean matching (SI-PPM) and multiple imputation by chained equations using predictive mean matching (MICE-PPM) can be implemented using the **mice** (van Buuren and Groothuis-Oudshoorn, 2011) package in R.

Machine learning-based imputation methods have also been proposed for handling missing data in PH models. Guo et al. (2021) found that nonparametric imputation using random forests (commonly referred to as missForest) was robust for estimating type-I error using Cox PH models under all missingness mechanisms (Guo et al., 2021). For this method, nonparametric unsupervised random forest models are employed to perform the missing value imputation for mixed-type data (i.e., continuous and/or categorical data, and data with complex interactions and nonlinear relations) (Stekhoven and Buehlmann, 2012). As this method is unsupervised, the outcome of interest in the Cox PH model is excluded from the imputation model. However, Hong and Lynn (2020) found that missForest can produce biased regression coefficient estimates when data are highly skewed. Shah

et al. (2014) also reported that missForest produced biased estimates for regression model parameters under MAR, along with poor coverage of confidence intervals. The missForest method can be easily implemented using the **missForest** (Stekhoven and Buehlmann, 2012) package in R.

4.3 Simulation Studies

To investigate the impacts of missing data and compare methods for handling missingness in the observation model covariates $\boldsymbol{Z}(t)$, we will compare five existing methods for handling missing data in proportional hazards models (CCA, LOCF, SI-PPM, MICE-PPM, missForest) under various forms and degrees of missingness.

The primary objective of the analysis in the simulation study is to determine the ATE of a time-invariant treatment D on a continuous or binary longitudinal outcome $Y(t)$. Although time-varying treatments can be evaluated using IIW, we focus on the simpler case of a time-invariant treatment to focus on the missingness of the observation model covariates $\boldsymbol{Z}(t)$.

As the simulation results are highly dependent on the data generating mechanism, we present two simulation studies involving two types of data. The first study in Section 4.3.1 shows the how missing data methods behave when the covariates in $\boldsymbol{Z}(t)$ are highly variable and time-varying. The second study in Section 4.3.2 shows how missing data methods behave when the missing covariates in $\boldsymbol{Z}(t)$ are less variable, and linearly related with time.

For all simulation studies, the Robins (2000) type standard error estimates are used. These estimates are more conservative than those developed specifically for the FIPTIW method in Coulombe et al. (2021b). These standard errors tend to be larger than those developed in Coulombe et al. (2021b), but still undercover the 95% confidence interval. As such, we omit the Coulombe et al. (2021b) standard error estimates from the analysis.

All computations are performed using R version 4.4.2 (R Core Team, 2020). Codes for all of the simulation studies are publicly available on GitHub at `https://github.com/grcetmpk/IIW-Imputation`.

4.3.1 Simulation Study I: Missingness Imposed on Highly Variable Covariates

We first consider the analysis of a binary outcome $Y(t)$ on a time-invariant treatment D. We assume the marginal model

$$E\{Y_i(t)|D_i\} = \frac{\exp(m(t) + \beta_1 D_i)}{1 + \exp(m(t) + \beta_1 D_i)}, \tag{4.1}$$

where we let $m(t) = (2 - t)$.

We consider four independently generated, time-varying covariates related to the observation process, $\boldsymbol{V}_1(t), \boldsymbol{V}_2(t), \boldsymbol{V}_3(t)$, and $\boldsymbol{V}_4(t)$. These covariates are also related to the treatment D and longitudinal outcome $Y(t)$. We let $V_{i1}(t) \sim N(1, 1)$ if $D_i = 0$ or $V_{i1}(t) \sim N(3, 2)$ if $D_i = 1$. We then let $V_{i2}(t) \sim U(0, 1)$ if $D_i = 0$ or $V_{i2}(t) \sim U(1, 2)$ if $D_i = 1$ for every time t. We simulate $V_{i3}(t)$ as a time-varying binary covariate where the probability of being one is $\pi_{V_3} = 0.2$ if $D_i = 0$ and $\pi_{V_3} = 0.7$ if $D_i = 1$ for any time t. Finally, we let $V_{i4}(t) = V^*_{i4}(t) \times (t - 1)^2$ where $V^*_{i4}(t) \sim U(0.5, 2)$ if $D_i = 0$ and $V^*_{i4}(t) \sim U(1.5, 4)$ if $D_i = 1$ for any time t.

We also consider a time-varying covariate that is not included in the observation-times or outcome model, $R_i(t)$, which may be associated with the missingness probability. This covariate is assumed to be known at all times, and will be used to simulate MAR data. We simulate $R_i(t) = R^*_i(t) \times (-t/2)$ where $R^*_i(t) \sim N(0, 1)$ if $D_i = 0$ and $R^*_i(t) \sim N(2, 0.5)$ if $D_i = 1$ at any time t. $R_i(t)$ will also be included in any imputation models.

A logistic model will be used to estimate the marginal model in Equation (4.1). However, as marginalization over a logistic link is complex, we will approximate the logit link using the probit link as $\Phi(x/1.7) \approx \exp(x)/(1 + \exp(x))$ (Buzkova and Lumley, 2007).

We generate the outcome as

$$Y_i(t) = \mathbb{1}\left[f^*_0(t) + \beta^*_1 D_i + \beta_2 V_{i1}(t) + \beta_3 V_{i2}(t) + \beta_4 V_{i3}(t) + \beta_5 V_{i4}(t) + \phi_i + \epsilon_i(t) > 0\right] \tag{4.2}$$

where $\epsilon_i(t) \sim N(0, \sigma^2_\epsilon)$ is the random error term (for each individual and time), $\phi_i \sim N(0, \sigma^2_\phi)$ is an individual random effect that allows outcomes to be correlated within the same individual, and $f^*_0(t)$ is a term used to ensure proper marginalization, and is defined below. From Equation (4.1), we have

$$\begin{aligned} E\{Y_i(t)|D_i\} &= P\{f^*_0(t) + \beta^*_1 D_i + \beta_2 V_{i1}(t) + \beta_3 V_{i2}(t) + \beta_4 V_{i3}(t) + \beta_5 V_{i4}(t) + \phi_i + \epsilon_i(t) > 0\} \\ &= \Phi\left[\frac{f^*_0(t) + \beta^*_1 D_i + \beta_2 V_{i1}(t) + \beta_3 V_{i2}(t) + \beta_4 V_{i3}(t) + \beta_5 V_{i4}(t)}{\sqrt{\sigma^2_{total}}}\right], \end{aligned} \tag{4.3}$$

where $\sigma^2_{total} = \beta_2^2 \text{Var}\{V_{i1}(t)|D_i\} + \beta_3^2 \text{Var}\{V_{i2}(t)|D_i\} + \beta_4^2 \text{Var}\{V_{i3}(t)|D_i\} + \beta_5^2 \text{Var}\{V_{i4}(t)|D_i\} + \sigma^2_\epsilon + \sigma^2_\phi$.

Thus, to ensure proper marginalization, we can let $\beta_1^* = \beta_1 M$ and

$$f_0^*(t) = (2-t)M - \beta_2 E[V_{i1}(t)|D_i] - \beta_3 E[V_{i2}(t)|D_i] - \beta_4 E[V_{i3}(t)|D_i] - \beta_5 E[V_{i4}(t)|D_i]$$

where

$$M = \frac{\sqrt{\sigma^2_{total}}}{1.7}$$

so that

$$E\{Y_i(t)|D_i\} = \Phi((2-t)+\beta_1 D_i)/1.7) \approx \exp\{(2-t)+\beta_1 D_i\}/(1+\exp\{(2-t)+\beta_1 D_i\}). \quad (4.4)$$

We let $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5) = (0.5, 1, 0.4, 0.3, 0.4)$. The estimation of β_1, the ATE, is of primary interest for each simulation study. We estimate $m(t)$ using cubic splines, where the knots are chosen as the tertiles of the observation times, as in Coulombe et al. (2021b).

We also consider a model for the observation times, which has an intensity specified as

$$\lambda_{i,obs}(t) = 0.01 \times \nu_i \lambda_0(t) \exp\{\gamma_1 D_i + \gamma_2 V_{i1}(t) + \gamma_3 V_{i2}(t) + \gamma_4 V_{i3}(t) + \gamma_5 V_{i4}(t)\}, \quad (4.5)$$

where ν_i is Gamma distributed with mean 1 and variance $\sigma^2_\eta = 0.1$, which makes the observation times positively correlated. We emphasize that because D is time-invariant, it is thus known at all possible observation times and can be included in the intensity model. We let $\lambda_0(t) = \frac{\sqrt{t}}{2}$, and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5) = (0.5,\ 0.4,\ 0.5,\ 0.3,\ -0.3)$. To simulate the observation times, we discretize the study period over a grid of 0.01 units from 0 to τ. Bernoulli draws with probability proportional to the intensity in Equation (4.5) were used to determine when observation times occurred. We let $\tau = 2$ in all simulations. The relationship between the covariates is represented for a single time point in the directed acyclic graph (DAG) in Figure 4.1.

We allow $\boldsymbol{V}(t) = \{\boldsymbol{V}_1(t), \boldsymbol{V}_2(t), \boldsymbol{V}_3(t), \boldsymbol{V}_4(t)\}$ to be missing at some time points. We consider various missingness mechanisms and degrees (proportions) of missingness for these observation intensity covariates. We begin by simulating the intensity data set involving covariates $\boldsymbol{V}(t)$ at all possible observation times from 0 to τ, by increments of 0.01. Then, we let a specified proportion (p_{mis}) of observations have missing $V_1(t), V_2(t), V_3(t)$, and $V_4(t)$ based on the missingness mechanisms described below. We emphasize that this does not affect whether $Y(t)$, D, and $R(t)$ are observed.

We consider five missingness mechanisms: no missingness (fully observed $\boldsymbol{V}_i(t)$), a scenario where $\boldsymbol{V}_i(t)$ can only be observed when $Y_i(t)$ are observed (special case of MNAR),

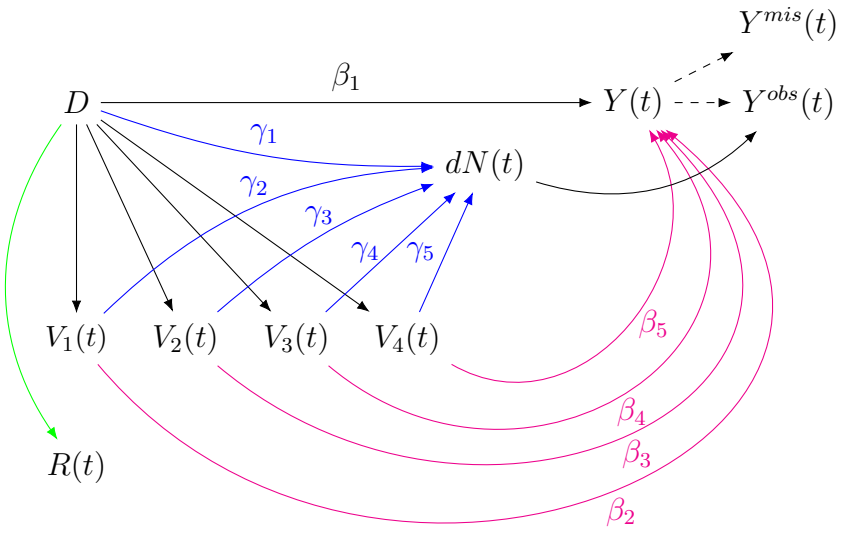


Figure 4.1: A directed acyclic graph (DAG) representing the data generating mechanism for a single time point in the simulation study.

along with MCAR, MAR, and MNAR for each possible $p_{mis} = (0.25, 0.50, 0.90)$, where p_{mis} will be the probability of all four observation level covariates will be missing.

To simulate data that are MCAR, we let the probability of any observation in the intensity data set to be missing to be p_{miss}. To simulate data that are MAR, we allow the probability of being missing to depend on D, which is always observed. That is, let $A_i(t)$ indicate that the observations for individual i are missing at time t. We let $A_i(t) \sim Bin(1, \pi_i(t))$ such that

$$\pi_i(t) = Pr(A_i(t) = 1 | D_i, R_i(t)]) = \frac{\exp(\eta_0 + \eta_1 D_i + \eta_2 R_i(t))}{1 + \exp(\eta_0 + \eta_1 D_i + \eta_2 R_i(t))}.$$

We choose η_0, η_1, and η_2 such that the overall proportion of missingness in the intensity data set is equal to p_{miss}. For this simulation, we let $\eta_1 = 1$ and $\eta_2 = 2$ and solve for η_0 for each data set to provide the appropriate amount of missingness.

To simulate data that are MNAR, we allow the probability of being missing to depend on $V_{i3}(t)$ and $V_{i4}(t)$, which are subject to missingness. For MNAR, we let

$$\pi_i(t) = Pr(A_i(t) = 1 | V_{i3}(t), V_{i4}(t)) = \frac{\exp(\eta_0 + \eta_1 V_{i3}(t) + \eta_2 V_{i4}(t))}{1 + \exp(\eta_0 + \eta_1 V_{i3}(t) + \eta_2 V_{i4}(t))}.$$

where we let $\eta_1 = 1$, $\eta_2 = 2$, and solve for η_0 for each data set to provide the appropriate amount of missingness.

For each missingness mechanism and proportion of missingness, we evaluate the use of the five missing data methods (CCA, LOCF, SI-PPM, MICE-PPM, missForest) for handling missing data for IIW models. In all cases, we include time, D, $V_1(t)$ to $V_4(t)$, $R(t)$, and an indicator of whether or not $Y(t)$ was observed as covariates in the imputation models, where applicable. For MICE-PPM, five imputations were performed. For missForest, the maximum number of iterations (*niter*) was set to five, and 15 trees were used for each imputation. All other values for the missForest function were kept at the default values, including the number of predictors used at each split, *mtry*, whose default value is $\sqrt{p}$ where p is the number of predictors. The computation time was too large to tune individual model parameters for this simulation study.

We generate 1000 data sets using a sample size of $n = 100$ for each combination of the missingness mechanism, degree of missingness, and outcome type and apply each of the five missing data methods to solve for the IIW weights. Then, we use an IIW-GEE to estimate the outcome model parameters in Equation (4.1). We also fit an unweighted (naive) GEE for comparison. We then aggregate the results over the 1000 data sets and

calculate the bias, MSE, and 95% confidence interval coverage using the Robins et al. (2000) type standard error estimates.

We also considered this simulation study where a continuous outcome was generated, which is presented in Appendix B.1.2.

When a binary outcome was simulated, the mean number of events per individual is 7.96 across the 1000 data sets. The mean minimum and maximum number of events per individual across the 1000 data sets are 1.00 and 22.01, respectively. As the intensity in Equation (4.5) determines whether observations occur or not, some individuals may have no observations within the study. Considering this, the mean sample size across the 1000 data sets was 92.43.

Extreme estimates of the ATE were produced in some cases when a binary outcome was simulated. As such, we flagged and removed any simulation runs that produced ATE estimates over 100 from the simulation study, and aggregated the results after the removal. Table 4.1 shows the number of estimates removed for each scenario. The empirical bias, MSE, and 95% confidence interval coverage after removing extreme estimates are shown in Tables 4.2, 4.3, and 4.4, respectively. The results where extreme estimates were not removed in the analysis are shown in Appendix B.1.1.

When IIW is not used and an independent GEE is fitted, the estimates are empirically biased with no missingness. When IIW weights are employed, the resultant estimates have negligible empirical bias with no missingness.

When $\boldsymbol{Z}(t)$ is only observed at the observation times, none of the existing methods provide estimates of the ATE with negligible empirical bias. All five methods perform similarly, with large empirical biases, large MSEs, and poor coverage.

When the data are MCAR, SI-PPM, and MICE-PPM produce estimates with negligible empirical bias at all degrees of missingness, with MICE-PPM tending to have sightly smaller MSE and better coverage. CCA also produces estimates of the ATE with negligible empirical bias up to 50% missingness. LOCF and missForest give empirically biased estimates in all cases. Similar results are seen when the data are MAR.

When the data are MNAR, SI-PPM and MICE-PPM produce empirically biased estimates in all cases, with MICE-PPM slightly out-performing SI-PPM in terms of MSE and coverage. CCA has negligible empirical bias at 25% missingness, while LOCF and missForest produce empirically biased estimates in all settings.

MICE-PPM appears to provide the best overall estimation of the ATE in this setting. In most cases, the estimates produced have negligible empirical bias. However, this method produced extreme estimates at 90% missingness for MNAR. Further, when $\boldsymbol{Z}(t)$ is known

only at the observation times, all methods, including MICE-PPM, have large empirical biases.

Simulation I (Binary Outcome): Number of Estimated ATEs > 100

Missingness Mechanism	p_{miss}	CCA	LOCF	SI-PMM	MICE-PMM	missForest
No IIW/Naive	-	0	-	-	-	-
No Missingness	0	0	-	-	-	-
Observation Times Only	-	0	0	0	0	0
MCAR	0.25	0	0	0	0	0
	0.50	0	0	**1**	0	0
	0.90	**4**	0	**2**	0	**13**
MAR	0.25	0	0	0	0	0
	0.50	0	0	0	0	0
	0.90	0	0	0	0	**6**
MNAR	0.25	0	0	0	0	0
	0.50	**16**	0	**1**	0	**26**
	0.90	**197**	0	**35**	**23**	**158**

Table 4.1: Number of instances where an extreme estimated ATE (> 100) was produced from the outcome model in Equation (4.1) using an independent GEE weighted by IIW (unless otherwise stated) where the missing observation model covariates were imputed using the indicated method in Simulation I. The ATE is considered extreme if the estimate $\widehat{\beta}_1$ was more than 100. The true value of the ATE is 0.5.

Simulation I (Binary Outcome): Bias (Extreme Estimates Omitted)

Missingness Mechanism	p_{miss}	CCA	LOCF	SI-PMM	MICE-PMM	missForest
No IIW/Naive	-	0.342	-	-	-	-
No Missingness	0	-0.006	-	-	-	-
Observation Times Only	-	0.361	0.342	0.343	0.342	0.336
MCAR	0.25	-0.001	0.074	-0.009	-0.010	-0.082
	0.50	0.005	0.158	**-0.007**	-0.009	-0.187
	0.90	**0.029**	0.305	**0.003**	-0.011	**-0.482**
MAR	0.25	-0.005	0.043	-0.009	-0.009	-0.049
	0.50	-0.003	0.112	-0.013	-0.015	-0.123
	0.90	0.012	0.283	-0.009	-0.009	-0.434
MNAR	0.25	0.016	0.082	-0.008	-0.010	-0.081
	0.50	**0.043**	0.175	**-0.003**	-0.006	**-0.174**
	0.90	**0.086**	0.311	**0.010**	**-0.010**	**-0.400**

Table 4.2: Bias of β_1 calculated across the 1000 simulation runs for each missing data method and missingness scheme when a binary outcome is simulated and extreme estimates ($\widehat{\beta_1} > 100$) are removed in Simulation I. $\widehat{\beta_1}$ was estimated by Equation (4.1) using an independent GEE weighted by IIW (unless otherwise stated) where the missing observation model covariates were imputed using the indicated method. The true value of β_1 is 0.5. **Bolded** quantities represent those which were previously extreme

.

Simulation I (Binary Outcome): MSE when Extreme Estimates Omitted

Missingness Mechanism	p_{miss}	CCA	LOCF	SI-PMM	MICE-PMM	missForest
No IIW/Naive	-	0.202	-	-	-	-
No Missingness	0	0.100	-	-	-	-
Observation Times Only	-	0.229	0.208	0.207	0.205	0.293
MCAR	0.25	0.100	0.097	0.103	0.102	0.117
	0.50	0.101	0.112	**0.109**	0.101	0.171
	0.90	**0.131**	0.179	**0.133**	0.123	**0.738**
MAR	0.25	0.100	0.097	0.100	0.099	0.109
	0.50	0.100	0.103	0.103	0.102	0.136
	0.90	0.113	0.166	0.125	0.115	0.512
MNAR	0.25	0.113	0.100	0.103	0.099	0.128
	0.50	**0.143**	0.119	**0.111**	0.109	**0.219**
	0.90	**0.259**	0.182	**0.191**	**0.141**	**0.944**

Table 4.3: MSE of β_1 calculated across the 1000 simulation runs for each missing data method and missingness scheme when a binary outcome is simulated and extreme estimates ($\widehat{\beta}_1 > 100$) are removed in Simulation I. $\widehat{\beta}_1$ was estimated by Equation (4.1) using an independent GEE weighted by IIW (unless otherwise stated) where the missing observation model covariates were imputed using the indicated method. **Bolded** quantities represent those which were previously extreme

.

Simulation I (Binary Outcome): Coverage when Extreme Estimates Omitted

Missingness Mechanism	p_{miss}	CCA	LOCF	SI-PMM	MICE-PMM	missForest
No IIW/Naive	-	0.745	-	-	-	-
No Missingness	0	0.958	-	-	-	-
Observation Times Only	-	0.758	0.767	0.746	0.748	0.757
MCAR	0.25	0.953	0.946	0.964	0.956	0.947
	0.50	0.955	0.904	**0.946**	0.955	0.922
	0.90	**0.943**	0.782	**0.932**	0.943	**0.796**
MAR	0.25	0.958	0.952	0.952	0.958	0.950
	0.50	0.955	0.929	0.947	0.955	0.940
	0.90	0.947	0.817	0.934	0.934	0.802
MNAR	0.25	0.947	0.938	0.947	0.955	0.944
	0.50	**0.949**	0.892	**0.943**	0.940	**0.904**
	0.90	**0.931**	0.784	**0.919**	**0.926**	**0.812**

Table 4.4: Coverage of β_1 calculated across the 1000 simulation runs for each missing data method and missingness scheme when a binary outcome is simulated and extreme estimates ($\widehat{\beta}_1 > 100$) are removed in Simulation I. $\widehat{\beta}_1$ was estimated by Equation (4.1) using an independent GEE weighted by IIW (unless otherwise stated) where the missing observation model covariates were imputed using the indicated method. Standard errors for the 95% confidence intervals are calculated using the Robins et al. (2000) type standard error estimates. **Bolded** quantities represent those which were previously extreme

.

4.3.2 Simulation Study II: Missingness Imposed on Covariates Linearly Related with Time

To evaluate the performance of various missing data methods when the covariates included in the observation times model are linearly related with time (or at least monotonically increasing with time), we perform a second simulation study. We wish to estimate the ATE, β_1, in Equation (4.4). For each individual, the probability of being treated ($Pr(D = 1)$) is 0.5.

We consider four independently generated, potentially time-varying covariates related to the observation process, $\boldsymbol{V}_1(t), \boldsymbol{V}_2(t), \boldsymbol{V}_3(t)$, and $\boldsymbol{V}_4(t)$. These covariates are also related to the treatment D and longitudinal outcome $Y(t)$. We simulate $V_{i1}(t) = V_1 \times t/3$ where $V_{i1} \sim N(0, 0.5)$ if $D_i = 0$ or $V_{i1} \sim N(2, 0.5)$ if $D_i = 1$. We then simulate $V_{i2}(t) = V^*_{i2}(t) \times \log(t + 1)/2$ where $V^*_{i2}(t) \sim U(0, 1)$ if $D_i = 0$ or $V^*_{i2}(t) \sim U(1, 2)$ if $D_i = 1$ for any time t. We simulate $V_{i3}(t)$ as a time-varying binary covariate where the probability of being one is $\pi_{V_3} = 0.2$ if $D_i = 0$ and $\pi_{V_3} = 0.7$ if $D_i = 1$ for any time t. Finally, we simulate $V_{i4}(t) = V^*_{i4}(t) \times (t - 1)/3$ where $V^*_{i4}(t) \sim U(0.5, 2)$ if $D_i = 0$ and $V^*_{i4}(t)) \sim U(1.5, 3)$ if $D_i = 1$ at any time t.

We generate a fifth covariate that is unrelated to the observation intensity and longitudinal outcome. We let $R_i(t) \sim Bin(1, 0.8)$ if $D = 0$ and $R_i(t) \sim Bin(1, 0.4)$ if $D_i = 1$ at any time t. We assume $R_i(t)$ is known at all times, and will use this covariate to simulate MAR and in all imputation models.

The outcome is again generated as in Equation (4.2), where $\boldsymbol{\beta} = (0.5, 1, 0.4, 0.5, 0.6)$. The observation times model is the same as in Equation (4.5), where $\boldsymbol{\gamma} = (0.6, 0.6, 0.5, 0.7, 0.6)$. The same missingness mechanisms described in Section 4.3.1 are used for this simulation.

The same methods as described in Section 4.3.1 are also used in this simulation. We calculate the same metrics (empirical bias, MSE, and 95% confidence interval coverage) of 1000 simulated data sets for each missingness mechanism, degree of missingness, and method. The sample size is $n = 100$. The results of the simulation study where a continuous outcome are presented in Appendix B.2.2.

When a binary outcome was simulated, the mean number of events per individual is 4.30 across the 1000 data sets. The mean minimum and maximum number of events per individual across the 1000 data sets are 1.00 and 13.94, respectively. As the intensity in Equation (4.5) determines whether observations occur or not, some individuals may have no observations within the study. Considering this, the mean sample size across the 1000 data sets was 87.08.

In some scenarios, extreme estimates of the ATE were produced in the simulation. We investigate the results had the extreme estimates been flagged and removed. That is, we removed any estimated ATE that was above 100 (recall that the true value of the ATE is 0.5). Table 4.5 shows the number of extreme estimates per method. Tables 4.6, 4.7, and 4.8 show the empirical bias, MSE, and coverage, respectively, when extreme estimates are removed. We see 18 total extreme events produce by CCA, and 9 extreme events produced by missForest out of the 1000 simulation runs. The results including these extreme estimates can be found in B.2.1.

When IIW is not employed, the outcome model produces empirically biased estimates of the ATE. When IIW is employed and there is no missingness, the estimates of the ATE have negligible empirical bias.

When $\boldsymbol{Z}(t)$ is known only at the observation times (a special case of MNAR), LOCF provides the best estimation in terms of empirical bias, MSE, and coverage, with a relative empirical bias of 14.2%. CCA, SI-PPM, MICE-PPM and missForest are largely biased.

When the data are MCAR and the missingness is 25%, CCA, LOCF, and missForest produce estimates with negligible empirical bias while SI-PPM and MICE-PPM have slightly higher empirical biases (<6% relative empirical bias). At 50% missingness, the empirical biases of SI-PPM, MICE-PPM, and missForest increase. At 90% missingness, LOCF has negligible empirical bias and provides the best estimation in terms of empirical bias, MSE, and coverage. CCA produces a slightly higher empirical bias, while SI-PPM, MICE-PPM, and missForest are biased.

When the data are MAR, LOCF and CCA again provide estimates of the ATE where the empirical bias is negligible at all degrees of missingness. SI-PPM andMICE-PPM have small empirical biases at 25% missingness, but are biased above 50% missigingness. missForest produces estimates with negligible empirical bias at 25% and 50% missingness, but is empirically biased at 90% missingness. In all cases, LOCF produces the smallest MSE and best coverage.

When the data are MNAR and the missingness is 25%, LOCF produces estimates with negligible empirical bias while CCA, SI-PPM, MICE-PPM, and missForest are slightly biased. At 50% missingness, LOCF again has negligible empirical bias while the other methods have small to moderate relative empirical biases of around 10%. At 90% missingness, missForest has the smallest empirical bias but large MSE. LOCF has a slightly larger empirical bias, but substantially smaller MSE and better coverage. All other methods are biased.

Overall, when the missing covariates are linearly related with time, LOCF is the best performing method. LOCF never produces extreme estimates, and consistently provides

the lowest bias and MSE, and best coverage. For the setting where the covariates are only known at the observation times, LOCF still sees a moderate empirical bias (14.2% relative empirical bias), but performs substantially better than the other methods that were compared in this simulation study. While LOCF has been criticized in the literature, its performance here is likely due to the fact that the missing covariates do not change greatly between observation times.

Simulation II (Binary Outcome): Number of Estimated ATEs > 100

Missingness Mechanism	p_{miss}	**CCA**	**LOCF**	**SI-PMM**	**MICE-PMM**	**missForest**
No IIW/Naive	-	0	-	-	-	-
No Missingness	0	0	-	-	-	-
Observation Times Only	-	0	0	0	0	**4**
MCAR	0.25	0	0	0	0	0
	0.50	0	0	0	0	0
	0.90	0	0	0	0	**1**
MAR	0.25	0	0	0	0	0
	0.50	0	0	0	0	0
	0.90	0	0	0	0	**1**
MNAR	0.25	0	0	0	0	0
	0.50	0	0	0	0	0
	0.90	**18**	0	0	0	**3**

Table 4.5: Number of instances where an extreme estimated ATE ($\widehat{\beta}_1 > 100$) was produced from the outcome model in Equation (4.1) using an independent GEE weighted by IIW (unless otherwise stated) where the missing observation model covariates were imputed using the indicated method for Simulation II. The ATE is considered extreme if the estimate $\widehat{\beta}_1$ was more than 100. The true value of the ATE β_1 is 0.5.

Simulation II (Binary Outcome): Bias when Extreme ATEs removed

Missingness Mechanism	p_{miss}	CCA	LOCF	SI-PMM	MICE-PMM	missForest
No IIW/Naive	-	0.185	-	-	-	-
No Missingness	0	0.007	-	-	-	-
Observation Times Only	-	0.283	0.071	0.184	0.181	**0.229**
MCAR	0.25	0.003	0.007	0.026	0.028	-0.010
	0.50	-0.003	0.006	0.051	0.045	-0.031
	0.90	0.020	-0.005	0.081	0.063	**-0.059**
MAR	0.25	0.002	0.007	0.026	0.027	-0.012
	0.50	-0.005	0.005	0.048	0.042	-0.028
	0.90	0.019	-0.006	0.085	0.080	**-0.042**
MNAR	0.25	0.024	0.005	0.027	0.026	-0.026
	0.50	0.050	0.003	0.048	0.044	-0.053
	0.90	**0.139**	-0.025	0.105	0.090	**-0.015**

Table 4.6: Empirical bias of the ATE, β_1, calculated across the 1000 simulation runs for each missing data method and missingness scheme when a when a binary outcome is simulated and extreme estimates ($\widehat{\beta}_1 > 100$) are removed in Simulation II. $\widehat{\beta}_1$ was estimated by Equation (4.1) using an independent GEE weighted by IIW (unless otherwise stated) where the missing observation model covariates were imputed using the indicated method. The true value of β_1 is 0.5. **Bolded** quantities represent estimates that were previously infinite without exclusion of extreme ATEs.

Simulation II (Binary Outcome): MSE when Extreme ATEs are Removed

Missingness Mechanism	p_{miss}	CCA	LOCF	SI-PMM	MICE-PMM	missForest
No IIW/Naive	-	0.066	-	-	-	-
No Missingness	0	0.037	-	-	-	-
Observation Times Only	-	0.237	0.052	0.068	0.065	**0.132**
MCAR	0.25	0.037	0.037	0.044	0.042	0.039
	0.50	0.039	0.037	0.054	0.050	0.046
	0.90	0.106	0.036	0.117	0.090	**0.143**
MAR	0.25	0.037	0.037	0.043	0.041	0.039
	0.50	0.038	0.037	0.054	0.048	0.044
	0.90	0.102	0.036	0.130	0.102	**0.147**
MNAR	0.25	0.038	0.037	0.042	0.041	0.042
	0.50	0.046	0.037	0.057	0.051	0.057
	0.90	**0.219**	0.037	0.159	0.111	**0.299**

Table 4.7: MSE of β_1 calculated across the 1000 simulation runs for each missing data method and missingness scheme when a when a binary outcome is simulated and extreme estimates ($\widehat{\beta}_1 > 100$) are removed in Simulation II. $\widehat{\beta}_1$ was estimated by Equation (4.1) using an independent GEE weighted by IIW (unless otherwise stated) where the missing observation model covariates were imputed using the indicated method. The true value of β_1 is 0.5. **Bolded** quantities represent estimates that were previously infinite without exclusion of extreme ATEs.

Simulation II (Binary Outcome): Coverage when Extreme ATEs Removed

Missingness Mechanism	p_{miss}	CCA	LOCF	SI-PMM	MICE-PMM	missForest
No IIW/Naive	-	0.824	-	-	-	-
No Missingness	0	0.931	-	-	-	-
Observation Times Only	-	0.835	0.911	0.823	0.832	**0.745**
MCAR	0.25	0.929	0.932	0.930	0.938	0.928
	0.50	0.922	0.930	0.909	0.914	0.917
	0.90	0.900	0.933	0.872	0.881	0.854
MAR	0.25	0.927	0.932	0.920	0.929	0.924
	0.50	0.928	0.935	0.919	0.923	0.931
	0.90	0.892	0.935	0.860	0.863	0.858
MNAR	0.25	0.940	0.936	0.925	0.936	0.922
	0.50	0.923	0.931	0.910	0.922	0.912
	0.90	0.838	0.923	0.819	0.841	0.780

Table 4.8: 95% confidence interval coverage of β_1 calculated across the 1000 simulation runs for each missing data method and missingness scheme when a when a binary outcome is simulated and extreme estimates ($\widehat{\beta}_1 > 100$) are removed in Simulation II. $\widehat{\beta}_1$ was estimated by Equation (4.1) using an independent GEE weighted by IIW (unless otherwise stated) where the missing observation model covariates were imputed using the indicated method. The true value of β_1 is 0.5.

4.4 Application to the Program for Resistance, Immunology, Surveillance, and Modeling of Malaria in Uganda Data Set

We demonstrate the impact of missing data using the observational data from the Program for Resistance, Immunology, Surveillance, and Modeling of Malaria in Uganda (PRISM) (Kamya et al., 2015). In this data set, households in three sub-counties of Uganda were randomly selected to be enrolled in a longitudinal study. All individuals in the household were enrolled in the study if they met the eligibility criteria outlined by Kamya et al. (2015). Data were collected on individuals enrolled in the study between August 2011 and June 2017 through routine clinical visits roughly every three months. Participants also attended a study clinic (and thus had observations recorded) any time they became ill. The data collected on the 287 children included in this analysis and the distributions of their baseline covariates are described in Section 3.5. The number of observations ranged from one to 11, with a mean of 4.83 observations per child. Of the 1386 total observations, 287 were for enrollment, 525 were scheduled follow-up appointments, and 574 were unscheduled appointments. Of the 1386 visits, malaria was diagnosed 179 times (12.9%), and 111 of 287 patients (38.7%) had a malaria diagnosis at some point during the study. 145 children were female while 142 were male.

We wish to see if there is a relationship between sex assigned at birth and malaria diagnoses among children in Uganda using the IIW method. That is, we wish to estimate the marginal model

$$g(E(Y_i(t)|D_i) = m(t) + \beta_1 D_i, \tag{4.6}$$

where $Y_i(t)$ is a binary, time-varying indicator of whether or not the patient was diagnosed with malaria at time t, $m(t)$ is a smooth function of time, D_i is a time-invariant indicator of whether or not the child is male, and $g(\cdot)$ is the logit link function. β_1 will represent the log OR of malaria diagnoses for male versus female children in Uganda.

To use the IIW method, any time-varying covariates included in the observation intensity model must be imputed as no additional information was collected for time-varying covariates outside of these clinic visits. That is, $\boldsymbol{Z}(t)$ is only known at the observation times. While in the most general case none of the proposed methods were appropriate for the setting where $\boldsymbol{Z}(t)$ was known only at the observation times, LOCF was shown to be appropriate when the missing covariate was linearly related to time, while the other methods were shown to produce biased estimates of the ATE.

To demonstrate how missing data impacts the estimation of IIW weights, we will include

age as a time-varying covariate in the intensity model. As the exact age of each individual was given at the time of enrollment to the ten-thousands place (for example, 11.2010 years old), we can determine each individual's age at each possible time point given the individual's age at enrollment. As age is known at all time points, we can use it in the observation intensity model to estimate the IIW weights and hence the true ATE.

To demonstrate how missing data can affect estimation, we will perform an analysis assuming age is known only at the observation times, and examine how the five missing data methods presented in this paper (CCA, LOCF, SI-PPM, MICE-PPM, and missForest) perform under this setting. Along with age, we allow the probability of visiting a clinic at any given time to depend on their sub-county, dwelling type, categorical household wealth index, degree of food problems per week, number of persons living in the house, and the individual's malaria status at the last clinic visit, as in Section 3.5. These covariates, along with an indicator of whether or not an observation occurred, were also included in the imputation models for SI-PPM, MICE-PPM, and missForest.

Table 4.9 shows the results of the estimated ATE (β_1 in Equation (4.6)) when age is only known at the observation times and the data are imputed using the indicated method. When the child's current age is known at all possible time points, the OR of malaria diagnosis for males vs females is 1.12 (95% CI: (0.81, 1.55)). When CCA is performed, the OR is underestimated at 1.04. LOCF and missForest perform similarly, and slightly overestimate the OR. SI-PPM substantially overestimates the OR, while MICE-PPM underestimates. It is unsurprising that the single and multiple imputation methods do not perform well as it was shown in Simulation Study II (Section 4.3.2) that these methods were biased when $\boldsymbol{Z}(t)$ was only known at the observation times.

Overall, LOCF and missForest appear to estimate the ATE and subsequent OR accurately while CCA greatly underestimates the true ATE. This emphasizes the importance of Assumption O5 and choosing the appropriate method to handle missing data when employing IIW.

4.5 Discussion

IIW can be used to analyze irregular longitudinal data with informative observation processes when a number of assumptions are met. One such assumption requires the covariates in the intensity model to be known at all times, which may not be met in practice. Often, all data (including the time-varying covariates driving the observation process) may only be observed when the longitudinal outcome is recorded (i.e., a clinic visit), or may be intermittently missing for a variety of reasons.

Method	$\widehat{\beta_1}$	**SE**($\widehat{\beta_1}$)	**95% CI for** $\widehat{\beta_1}$	**OR**	**95% CI for OR**
No Missingness (Truth)	0.116	0.166	(-0.209, 0.441)	1.123	(0.811, 1.555)
CCA	0.040	0.169	(-0.291, 0.371)	1.041	(0.747, 1.450)
LOCF	0.123	0.165	(-0.200, 0.446)	1.131	(0.818, 1.563)
SI-PMM	0.131	0.165	(-0.192, 0.454)	1.140	(0.825, 1.575)
MICE-PMM	0.038	0.166	(0.750, 1.438)	1.039	(0.750, 1.438)
missForest	0.123	0.165	(-0.200, 0.446)	1.131	(0.818, 1.563)

Table 4.9: Results for the PRISM study data analysis where β_1 in Equation 4.6 was estimated using the indicated missing data method. OR refers to the odds ratio ($\exp(\beta_1)$).

.

We have shown through simulation that there is no one size fits all approach to handling missing data in the observation model covariates. The performance of each missing data method depends on the missingness mechanism, the proportion of missingness, and the type of data that are missing.

In the most general case where the missing time-varying covariates are not directly correlated with time, CCA and LOCF are typically biased. MICE-PPM tends to provide estimation with negligible empirical biases, but is subject to producing extreme estimates of the ATE when a binary outcome is simulated. MICE-PPM also tends to have the smallest MSE, and tends to produce coverage close to 95%. When the covariates are only known at the observation times, none of the existing methods are appropriate for handling the missingness. As such, we identify this as an important area of future research, as this type of missingness is often seen with irregular longitudinal data.

In the case where the missing covariates are linearly related with time, such as age, LOCF and missForest tend to provide estimates where the empirical bias is negligible, while CCA, SI-PPM and MICE-PPM are biased. However, missForest was shown to be biased when $\boldsymbol{Z}(t)$ was known only at the observation times through simulation when multiple covariates were missing, but performed well in the real data analysis. As the missing covariate in the real data analysis was a linear function of time, this scenario was overly simplified and it is unsurprising that the bias of the methods was low compared to the simulation studies.

Throughout this chapter, we assumed that $\boldsymbol{X}(t)$ was known at all times, which is a reasonable assumption to make when $\boldsymbol{X}(t)$ does not contain any time-varying covariates. If it were the case that $\boldsymbol{X}(t)$ was subject to missingness, imputation methods appropriate for GEEs should be explored. One must be cautious that if $\boldsymbol{X}(t)$ is missing and is also

included in the observation model (i.e., is $\boldsymbol{X}(t) \subset \boldsymbol{Z}(t)$), two levels of missingness will need to be accounted for; one in the outcome model, and one in the observation times model (which we examined in this chapter).

We acknowledge that extreme estimates were produced by the simulation studies where a binary outcome was simulated. With higher degrees of missingness, more extreme estimates were produced. It is possible that the extreme estimates were produced due to the approximation of the logit link in the simulation study. We identify this as a potential weakness of the simulation study.

Chapter 5

Assessing the Assumptions of Inverse Intensity Weighting using Directed Acyclic Graphs

5.1 Introduction

Throughout the literature, different assumptions are made on the observation process when employing IIW. As there is no universally adopted set of assumptions, the purpose of this chapter is to highlight inconsistencies and potential ambiguity among the assumptions presented in existing works involving IIW. We also aim to clarify and discuss when causal estimates of the marginal outcome model can be obtained. To do so, we will construct DAG (Glymour, 2006; Greenland et al., 1999) for each set of assumptions. DAGs can be used to directly communicate complex associations and establish causality, and can help to identify what variables need to be controlled for in an analysis to account for confounding or selection bias (Digitale et al., 2022). Using the DAGs, we will discuss and compare each set of assumptions. After careful consideration of the assumptions, we will also propose a unified set of assumptions for IIW.

We begin by formally defining the observation process and notation in Section 5.2, and then review DAGs in Section 5.3. We then discuss the various sets of assumptions made on IIW in existing works in Section 5.4. We then conclude by proposing a unified set of assumptions in Section 5.5, followed by a discussion in Section 5.6.

5.2 Notation

Let $t_{i1}, t_{i2}, ..., t_{iK_i}$ be the times at which individual i has the longitudinal outcome $Y_i(t)$ observed for $i = 1, 2, ..., n$ and $0 \leq t_{i1} < t_{i2} < ... < t_{iK_i} \leq \tau$ where τ is the study end time. We consider the scenario where we are primarily interested in estimating the association between $Y(t)$ and a set of covariates $\boldsymbol{X}(t)$ through the marginal model in

$$g(\mu_i(t)) = \boldsymbol{\beta}^T \boldsymbol{X}_i(t), \tag{5.1}$$

where the estimation of the parameter vector $\boldsymbol{\beta}$ is of interest. In Equation (5.1), $\boldsymbol{X}_i(t) = (X_{i1}(t), X_{i2}(t), ..., X_{ip}(t))^T$ is a vector of p covariates for individual i at time t, and $\mu_i(t) = E(Y_i(t)|\boldsymbol{X}_i(t))$ where $Y_i(t)$ is the longitudinal outcome of individual i at time t, for $i = 1, 2, ..., n$ and $t = t_{i1}, t_{i2}, ..., t_{iK_i}$. Further, $g(\cdot)$ is a known, monotonic, and differentiable link function. We emphasize that the observation times are irregular and uncommon between individuals.

We also consider a set of auxiliary covariates, $\boldsymbol{V}(t)$, that are related to the observation process but omitted from the outcome model. For simplicity, throughout this chapter we do not allow $\boldsymbol{X}(t)$ to be influenced by $\boldsymbol{Z}(t)$, the longitudinal outcome, or any other auxiliary covariates to avoid treatment-confounder feedback, where a confounder affects $\boldsymbol{X}(t)$ and $\boldsymbol{X}(t)$ affects the confounder. We discuss how treatment-confounder feedback can affect estimation in Section 5.6.

We formally define the observation process as the underlying process that determines the times at which individuals have the longitudinal outcome recorded, i.e., the observation times. In other sources, the observation process has been referred to as the "visit" (Aghababaei Jazi and Pullenayegum, 2022; Coulombe et al., 2021a; Lin et al., 2004; Pullenayegum, 2013; Pullenayegum et al., 2021; Pullenayegum and Lim, 2016), "participation" (Schmidt and Woll, 2017), "assessment" (Smith et al., 2022), or "monitoring" (Coulombe et al., 2021b) process. We opt to use "observation" over other terms as it is more general and also connects irregular longitudinal data to the missing data literature, as the analysis of irregular longitudinal data can often be viewed as a missing data problem (Buzkova and Lumley, 2007). We emphasize that the observation process, in general, only determines whether $Y_i(t)$ is observed. Unless explicitly stated, we will assume $\boldsymbol{X}_i(t)$ and $\boldsymbol{V}_i(t)$ are observed at all possible (discretized) observation times, regardless of whether or not $Y_i(t)$ is observed. For example, if we assume patients can visit a clinic and have observations recorded at most once per day, $\boldsymbol{X}(t)$ and $\boldsymbol{V}(t)$ will need to be known every day, regardless of whether or not a visit occurred. We note that this is a very strong assumption which may not hold in practice, and violations of this assumption are discussed in Chapter 4. We

denote the counting process for the number of times individual i is observed by time t as $N_i(t) = \sum_{k=1}^{K_i} \mathbb{1}_{(t_{ik} \leq t)}$, where $\mathbb{1}_{(\cdot)}$ is the indicator function.

As in Young et al. (2020), we define censoring as any event occurring in the study by time t that ensures the future values of all covariates of interest are unknown. By definition, loss to follow-up and administrative censoring at the study end date are a form of censoring. We let C_i denote the censoring time at which follow-up ceases for individual i such that $C_i \leq \tau$. We define the indicator $C_i(t) = \mathbb{1}_{(C_i \geq t)}$, such that $C_i(t) = 1$ if individual i is censored by time t. We assume $C_i(0) = 0$ for all i.

Although individuals may be censored prior to end of the study, we also consider the counting process for the counterfactual (uncensored) observation times, denoted $N_i^*(t)$. We relate the censored and uncensored counting processes as $N_i(t) = N_i^*(t \wedge C_i)$ where we define $a \wedge b = min(a, b)$. We assume a temporal ordering $(C_i(t), \boldsymbol{X}_i(t), \boldsymbol{V}_i(t), N_i(t), Y_i(t))$ so that if an individual is censored by time t, all future components of $\boldsymbol{X}_i(t), \boldsymbol{V}_i(t), N_i(t)$, and $Y_i(t)$ will be missing.

For any arbitrary process $A(t)$, let $A^{obs}(t)$ denote that the observed value of $A(t)$, and let $A^{mis}(t)$ denote the missing (counterfactual) value of $A(t)$. We define $\overline{A}(t) = \{A(s) : 0 \leq s \leq t\}$ as the entire (and potentially counterfactual) history of the process up to and including time t. We denote $\overline{A}(t)^{obs}$ to include only the observed history of A up to and including time t. We let $\overline{A}(\infty) = \{A(s) : s > 0\}$ be the entire (counterfactual) process that includes times beyond the study end time. Further, we let $dN(t) = N(t) - N(t^-)$, where $N(t^-) = \lim_{s \to t} N(s)$. That is, $dN_i(t) = 1$ if individual i has the longitudinal outcome observed at time t, and is zero otherwise (which could be due to censoring). $dN_i^*(t)$ is similarly defined, where $dN_i^*(t) = 1$ if individual i has the longitudinal outcome observed at time t in the counterfactual observation times, and is zero otherwise.

5.3 Directed Acyclic Graphs

A DAG is a graphical model that can be used to represent the assumptions made in a model through nodes and edges. Each variable is represented by a node, and relationships between variables are represented by edges going from one node to another. These edges are assumed to be unidirectional, such that an arrow from X to Y means that X causes Y. DAGs must correctly represent all confounding, selection bias, censoring, and measurement error to be able to estimate causal effects (Digitale et al., 2022).

To estimate the causal effect of an association between an intervention X and an outcome Y, we must control for, or condition on, other non-causal associations between X

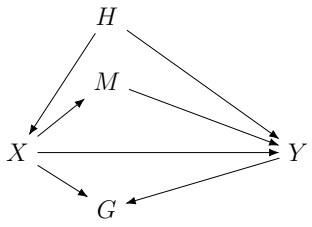


Figure 5.1: An example of a DAG at a single time point where M is a mediator, H is a confounder, and G is a collider for the association between X and Y.

and Y (Digitale et al., 2022). To control for other non-causal associations, only non-causal paths on the DAG connecting X to Y must be blocked. The path from X to Y is considered blocked by conditioning on a variable that is a common cause of both X and Y (for example, H in $X \leftarrow H \rightarrow Y$) or by conditioning on mediators along the path from X to Y (for example, M in $X \rightarrow M \rightarrow Y$). A path can also be blocked by uncontrolled-for colliders along the path (for example, G in $X \rightarrow G \leftarrow Y$).

Consider the hypothetical DAG at a single time point in Figure 5.1. If we wanted to estimate the causal effect of X on Y, we would need to condition on the confounder H to block the non-causal path. We do not need to condition on M as it is a mediator on the causal path linking X and Y, and controlling for mediators may induce bias as it decomposes the total effect of X on Y into parts (Barrett, 2024). G is a collider on a non-causal path between X and Y, and as such we do not want to condition on it, or we will create a biased association between X and Y.

When longitudinal data are considered, the DAG should show the relationships of covariates across multiple time points. We also must consider if the study is prone to censoring. To represent censoring, we draw a box around a censoring indicator $C(t)$ which indicates that the analysis is restricted to those who were not censored by time t (Mansournia et al., 2017). Any arrows from a covariate to $C(t)$ indicate that censoring is influenced by those factors, which in many cases cause selection bias (Mansournia et al., 2017).

Consider the DAG in Figure 5.2 which shows the relationships between covariates at two time points. We now see that we have a non-causal path $X(1) \leftarrow H(0) \rightarrow Y(1)$, which means we also need to control for past values of $H(t)$. In this example, X is also related to the censoring mechanism. Without controlling for X, we will have selection bias due to differential loss to follow-up, which can bias the association between X and Y. If we were to condition on X, then C and Y will be conditionally independent, and the analysis can proceed without accounting for the censoring mechanism.

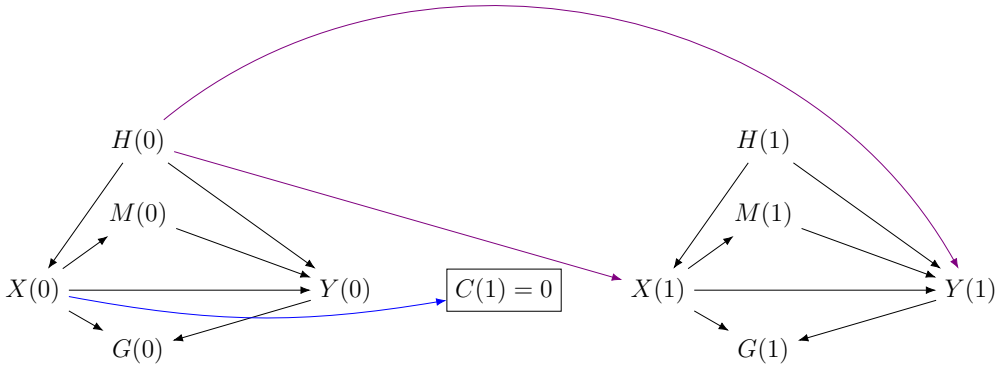


Figure 5.2: An example of a DAG at two time points where $M(t)$ is a mediator, $H(t)$ is a confounder, and $G(t)$ is a collider for the association between $X(t)$ and $Y(t)$ at $t = 0$ and $t = 1$. We also see $H(0)$ is confounding the association between $X(1)$ and $Y(1)$. X is also associated with the censoring time.

Throughout this chapter, colours will be used in the DAGs to help distinguish between lines. Unless otherwise indicated, the colours of the lines are not meaningful. Dashed arrows represent deterministic relationships.

5.4 Assumptions on IIW

To illustrate the differences in the sets of assumptions made on IIW, we compare the assumptions presented in one of the introductory IIW papers by Buzkova and Lumley (2007) to two other papers that use IIW, which are Coulombe et al. (2021b) and Pullenayegum and Lim (2016). DAGs will be used to visualize the relationships between the different types of covariates involved in IIW in various settings, starting with the simplest case where $\boldsymbol{X}(t)$ is fully observed and no censoring is present in Section 5.4.1. Then, in Section 5.4.2, we analyze the assumptions when noninformative censoring is present. Finally, in Section 5.4.3, we present a more general case where $\boldsymbol{X}(t)$ is only observed when $Y(t)$ is, and noninformative censoring is present.

5.4.1 Fully Observed $\boldsymbol{X(t)}$, No Censoring

We begin by comparing the sets of the assumptions in the setting where $\boldsymbol{X}(t)$ is observed at all possible time points, and no individuals are censored prior to the end of the study.

Buzkova and Lumley, 2007

In Buzkova and Lumley (2007), the authors make five assumptions on the observation and censoring processes. They first assume

Assumption B1. *Conditional Independence of the Observation Process:*

$$E(dN_i^*(t)|\boldsymbol{X}_i(t), \boldsymbol{Z}_i(t), Y_i(t), C_i \geq t) = E(dN_i^*(t)|\boldsymbol{Z}_i(t)).$$

As we can write $\boldsymbol{Z}_i(t) = \{\boldsymbol{X}_i(t), \overline{\boldsymbol{X}}_i(t^-), \overline{\boldsymbol{V}}_i(t), \overline{\boldsymbol{N}}_i(t^-), \overline{\boldsymbol{Y}}_i^{obs}(t^-)\}$ and $Y_i(t) = \{Y_i^{obs}(t), Y_i^{mis}(t)\}$, we can re-write Assumption B1 as

$$\begin{aligned} E(dN_i^*(t)|\boldsymbol{X}_i(t),&\overline{\boldsymbol{X}}_i(t^-), \overline{\boldsymbol{V}}_i(t), \overline{\boldsymbol{N}}_i(t^-), \overline{\boldsymbol{Y}}_i^{obs}(t^-), \boldsymbol{X}_i(t), Y_i^{obs}(t), Y_i^{mis}(t), C_i \geq t) \\ &= E(dN_i^*(t)|\boldsymbol{X}_i(t), \overline{\boldsymbol{X}}_i(t^-), \overline{\boldsymbol{V}}_i(t), \overline{\boldsymbol{N}}_i(t^-), \overline{\boldsymbol{Y}}_i^{obs}(t^-)), \end{aligned}$$

which implies

$$dN_i^*(t) \perp Y_i^{obs}(t), Y_i^{mis}(t), C_i \geq t|\boldsymbol{X}_i(t), \overline{\boldsymbol{X}}_i(t^-), \overline{\boldsymbol{V}}_i(t), \overline{\boldsymbol{N}}_i(t^-), \overline{\boldsymbol{Y}}_i^{obs}(t^-).$$

The second assumption by Buzkova and Lumley (2007) is a noninformative censoring assumption, such that

Assumption B2. *Noninformative censoring for the mean of the outcome:*

$$E(Y_i(t)|\boldsymbol{X}_i(t), C_i \geq t) = E(Y_i(t)|\boldsymbol{X}_i(t)),$$

which implies $Y_i(t) \perp C_i \geq t|\boldsymbol{X}_i(t)$. We will discuss the importance of this assumption when we consider censoring in Section 5.4.2.

The third assumption is

Assumption B3. *Separability: The outcome and observation times model parameters are separable (i.e., the models do not share parameters).*

The authors also implicitly assume

Assumption B4. *Correct specification of the observation times model: The observation intensity model is correctly specified, and the intensity and observation models do not share any parameters.*

and

Assumption B5. *Completely observed observation process covariates: $\boldsymbol{Z}(t)$ is known at all possible observation times.*

To visualize the assumptions made for IIW, we will first consider the scenario where there is no censoring present. In this scenario, $dN^*(t) = dN(t)$.

In this setting, the DAG can be constructed under Assumptions B1 to B5 as in Figure 5.3. We see that by conditioning on $X(1)$, $X(0)$, $V(1)$, $V(0)$, $dN(0)$, and $Y^{obs}(0)$, all non-causal paths from $dN(1)$ to $Y^{obs}(1)$ and $Y^{mis}(1)$ are blocked, as required by Assumptions B1 and B2.

If causal estimates were to be obtained for the association between $X(t)$ and $Y(t)$ through the marginal model in Equation (5.1), we also must ensure that there are no other causal paths from $X(1)$ to $Y(1)$, or that any confounding is adjusted for in the outcome model. In this setting, we cannot have $X(1) \leftarrow X(0) \rightarrow Y(1)$ if only IIW is to be used. However, one could also incorporate IPTW weights into the outcome model as presented in Chapter 3 to account for the confounding and obtain causal estimates of treatment effects if confounding between $X(t)$ and $Y(t)$ is present.

Coulombe et al., 2020

The IIW method was combined with IPTW in Coulombe et al. (2021b) to account for confounding factors. In this paper, the authors made a slightly different set of assumptions on the observation process than Buzkova and Lumley (2007).

Coulombe et al. (2021b) assumed that $N_i(t)$ was conditionally independent of $Y_i(t)$ given the set of covariates that contain all common predictors of the monitoring times and the outcome. We emphasize that other sets of assumptions were regarding $dN(t)$ or $dN^*(t)$, not $N(t)$. We can express this assumption as

Assumption C1.

$$N_i(t) \perp Y_i(t) | \boldsymbol{V}_i(t), \boldsymbol{X}_i(t)$$

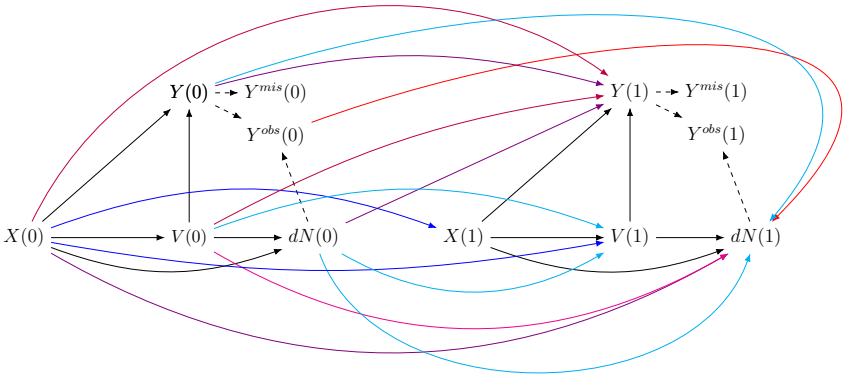


Figure 5.3: The DAG for the set of IIW assumptions presented by Buzkova and Lumley (2007) when $\boldsymbol{X}(t)$ and $\boldsymbol{V}(t)$ are known at all possible observation times. No censoring is present in this example.

as both $\boldsymbol{V}_i(t)$ and $\boldsymbol{X}_i(t)$ can be related to the longitudinal outcome and observation times. The authors do not discuss whether or not the set of common predictors can contain histories of the outcome or observation process covariates. As such, we assume no covariate histories are included in Assumption C1.

The authors also assumed noninformative censoring (Assumption B2), separability (Assumption B3), correct specification (Assumption B4), and assumed any covariates affecting the monitoring times were known at all times (Assumption B5).

From this set of assumptions, there are a number of possible DAGs that can be constructed under Assumption C1 in the absence of censoring. To reduce the number of potential DAGs that can be drawn, we will require an arrow between $Y(0)$ and $Y(1)$, such that past outcomes can influence future outcomes, which is a realistic assumption to make in practice.

As the histories of the observed outcome, outcome process covariates, auxiliary covariates, and past outcome times are not part of the conditioning set, $dN(0)$, $X(0)$, and/or $V(0)$ can be related to $dN(1)$ or $Y(1)$, but not both.

We first consider the scenario where $N(0)$, $X(0)$, and/or $V(0)$ can be related to $dN(1)$. However, as we are requiring an arrow from $Y(0)$ to $Y(1)$, this means $X(0)$ and $V(0)$ cannot be related to $dN(1)$ even if they are not related to $Y(1)$. This is because we would have non-causal pathways $Y(1) \leftarrow Y(0) \leftarrow X(0) \rightarrow dN(1)$, and $Y(1) \leftarrow Y(0) \leftarrow V(0) \rightarrow dN(1)$

that are not controlled for. As such, one possible DAG can be drawn as in Figure 5.4a.

In the setting where $dN(0)$, $X(0)$, and/or $V(0)$ can be related to $Y(1)$ but not $dN(1)$, we cannot allow $dN(0)$ to be related to $Y(1)$ or else we will have a path $N(1) \leftarrow dN(0) \rightarrow Y(1)$, which would not satisfy $N(1) \perp Y(1) \mid V(1), X(1)$. Therefore, the DAG in this setting can be drawn as in Figure 5.4b.

In either setting, if causal estimates are to be obtained from the marginal outcome model in Equation 5.1, we must also ensure all non-causal paths between $X(1)$ and $Y(1)$ are blocked. As such, we would either require for $X(0)$ to be unrelated to $X(1)$, or employ a method such as IPTW to account for the confounding.

From the DAGs, we can clearly see that this set of assumptions are more restrictive and less general than in Buzkova and Lumley (2007), as we do not condition on the histories $\bar{\boldsymbol{X}}(t)$, $\bar{\boldsymbol{V}}(t)$, or $\bar{Y}^{obs}(t)$.

Pullenayegum and Lim, 2016

Pullenayegum and Lim (2016) introduced the visiting at random (VAR) assumption, which states

Assumption P2. *Visiting at random (VAR):*

$$E(dN_i(t)|\overline{\boldsymbol{X}}_i(t), \overline{\boldsymbol{V}}_i(t), \overline{N}_i(t^-), \overline{Y}_i^{obs}(t^-), Y_i(t)) = E(dN_i(t)|\overline{\boldsymbol{X}}_i^{obs}(t), \overline{\boldsymbol{V}}_i(t), \overline{N}_i(t^-), \overline{Y}_i^{obs}(t^-)).$$

When $\boldsymbol{X}(t)$ is assumed to be known at all possible observation times, this assumption simplifies to

$$E(dN_i(t)|\overline{\boldsymbol{X}}_i(t), \overline{\boldsymbol{V}}_i(t), \overline{N}_i(t^-), \overline{Y}_i^{obs}(t^-), Y_i(t)) = E(dN_i(t)|\overline{\boldsymbol{X}}_i(t), \overline{\boldsymbol{V}}_i(t), \overline{N}_i(t^-), \overline{Y}_i^{obs}(t^-)).$$

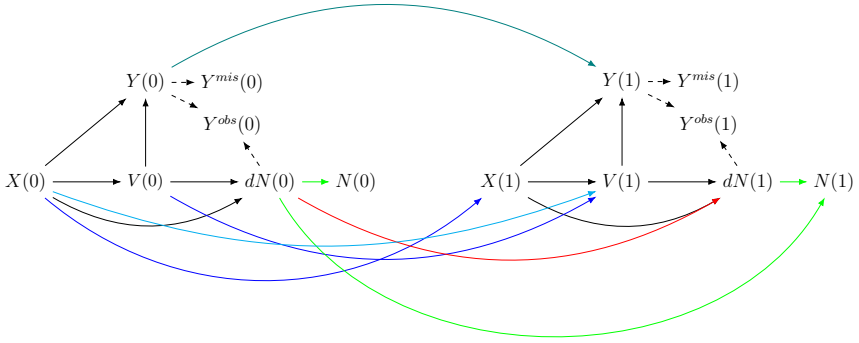
Further, as $\overline{\boldsymbol{A}}_i(t) = \{\boldsymbol{A}_i(t), \overline{\boldsymbol{A}}_i(t^-)\}$, $\overline{\boldsymbol{A}}_i(t^-) = \{\overline{\boldsymbol{A}}_i^{obs}(t^-), \overline{\boldsymbol{A}}_i^{mis}(t^-)\}$, and $\boldsymbol{A}_i(t) = \{\boldsymbol{A}_i^{obs}(t), \boldsymbol{A}_i^{mis}(t)\}$ for any $\boldsymbol{A}$, the VAR condition in Assumption P2 can be re-written as

$$\begin{aligned} E(dN_i(t)|\boldsymbol{X}_i(t), \overline{\boldsymbol{X}}_i(t^-), \overline{\boldsymbol{V}}_i(t), \overline{N}_i(t^-), \overline{Y}_i^{obs}(t^-), Y_i^{obs}(t), Y_i^{mis}(t)) \\ = E(dN_i(t)|\boldsymbol{X}_i(t), \overline{\boldsymbol{X}}_i(t^-), \overline{\boldsymbol{V}}_i(t), \overline{N}_i(t^-), \overline{Y}_i^{obs}(t^-)). \end{aligned}$$

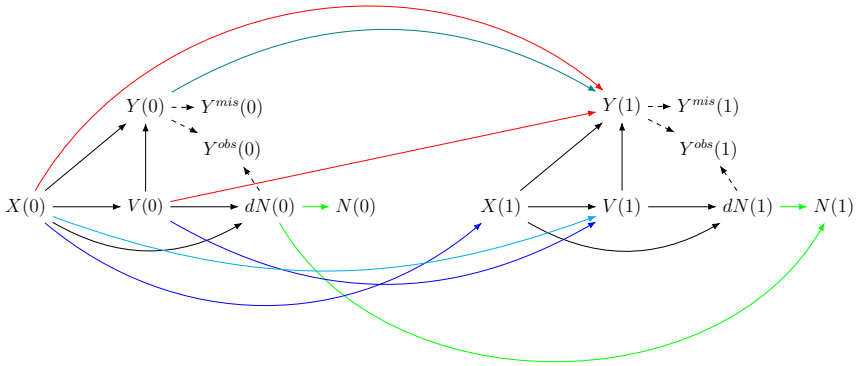
This then implies

$$dN_i(t) \perp Y_i^{obs}(t), Y_i^{mis}(t) | \boldsymbol{X}_i(t), \overline{\boldsymbol{X}}_i(t^-), \overline{\boldsymbol{V}}_i(t), \overline{N}_i(t^-), \overline{Y}_i^{obs}(t^-),$$

which is equivalent to Assumption B1 when no censoring is present as $dN_i(t)$ will be equivalent to $dN_i^*(t)$. As such, the DAG for the set of assumptions presented by Pullenayegum and Lim (2016) will be equivalent to the DAG shown in Figure 5.3.



(a) One possible DAG for the set of assumptions presented by Coulombe et al. (2021b), where the histories of the outcome model and auxiliary covariates are not related to future outcomes.



(b) One possible DAG for the set of assumptions presented by Coulombe et al. (2021b), where the histories of the outcome model and auxiliary covariates are related to future outcomes.

Figure 5.4: Two possible DAGs for the set of assumptions presented by Coulombe et al. (2021b) when $\boldsymbol{X}(t)$ and $\boldsymbol{V}(t)$ are known at all possible observation times. No censoring is present in this example.

Comparison of Assumptions

In the absence of censoring and when no covariates are missing, the sets of assumptions presented in Buzkova and Lumley (2007) and Pullenayegum and Lim (2016) are equivalent.

However, the set of assumptions presented in Coulombe et al. (2021b) have subtle but important differences.

The assumptions made by Coulombe et al. (2021b) regarding the observation process is on $N(t)$ (the number of observations by time t), while Buzkova and Lumley (2007) and Pullenayegum and Lim (2016) are on $dN(t)$ (the indicator of whether or not an observation occurred). However, $N(t)$ is a deterministic function of the past $dN(s)$ for $s \leq t$, which will still allow us to directly compare the assumptions.

A concern with the set of assumptions presented by Coulombe et al. (2021b) is that we only condition on current values of the covariates in the assumption made on the observation process. Because of this, we cannot allow $dN(t-1)$, $X(t-1)$, and/or $V(t-1)$ to be related to $Y(t)$ and not $dN(t)$. Further, if any $dN(s)$ for $s \leq t$ has an arrow going into any component of $Y(t)$, then the assumptions will be violated. In other words, it is assumed that whether or not a visit occurs cannot affect future outcomes. In practice, however, it is possible that a clinical visit can improve outcomes. For example, a doctor may prescribe medications or perform additional treatments that may improve the longitudinal outcome. As such, we do not believe this assumption is reasonable for many real-world settings, unless the total effect of $dN(t)$ on future outcomes is completely captured or mediated by $X(t)$ and/or $V(t)$.

Because of its limitations in the simplest setting with no censoring or missingness, we will omit the comparisons of the set of assumptions presented by Coulombe et al. (2021b) for the remaining sections in the paper.

5.4.2 Fully Observed $\boldsymbol{X(t)}$, Noninformative Censoring Present

We now consider the setting where again $\boldsymbol{X}(t)$ is known at all possible observation times, but allow for noninformative censoring in the sense of Assumption B2.

Buzkova and Lumley, 2007

When censoring is present, we must also consider the noninformative censoring assumption, which requires $Y_i(t) \perp C_i \geq t | \boldsymbol{X}_i(t)$. In this scenario, the DAG can be constructed under Assumptions B1 to B5 as in Figure 5.5.

Although in Assumption B1 the covariates $X(0)$, $V(0)$, and $dN(0)$ can be related to both $C(1)$ and $Y(1)$, by conditioning on $X(0)$, $V(0)$, and $dN(0)$ in Assumption B1, we must also consider the noninformative censoring assumption which states that $Y(t)$ and $C(t)$ are

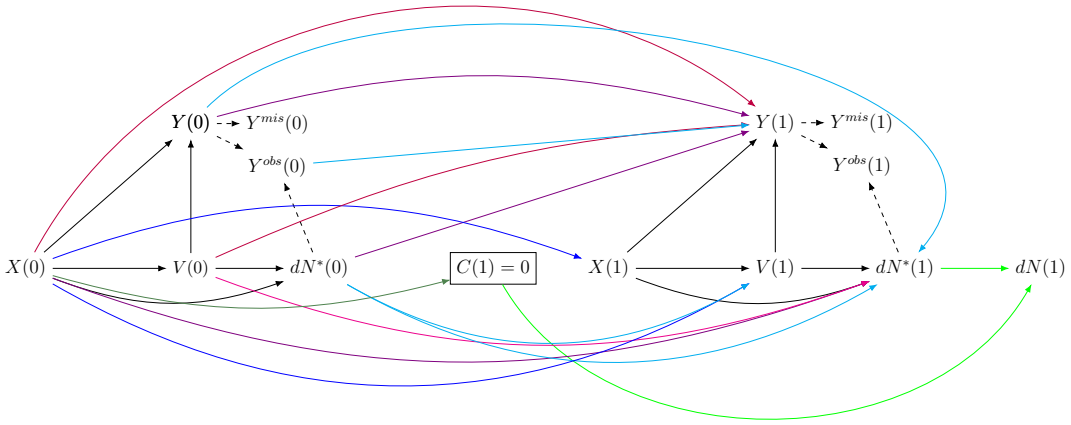


Figure 5.5: The DAG for the set of IIW assumptions presented by Buzkova and Lumley (2007) when $\boldsymbol{X}(t)$ is known at all possible observation times and noninformative censoring is present.

conditionally independent given only $X(t)$. As such, we only allow $X(0)$ to be related to $C(1)$. We also note that we do not allow future (possibly counterfactual) observation times depend on the history of the censoring process.

We again see that by conditioning on $X(1), X(0), V(1), V(0), dN(0)$ and $Y^{obs}(0)$, all non-causal paths from $dN(1)$ to $Y^{obs}(1)$ and $Y^{mis}(1)$, and $C(1)$ to $Y(1)$ are blocked, as required by Assumptions B1 and B2.

Further, if we wanted to estimate the casual effect of $X(t)$ on $Y(t)$ through the marginal model in Equation (5.1), we would have to ensure all non-causal paths between $X(1)$ and $Y(1)$ are blocked. This could be achieved by ensuring there are no common causes of $X(1)$ and $Y(1)$, or by employing a method like IPTW to account for the confounding that is present.

Pullenayegum and Lim, 2016

Recall that the VAR assumption (Assumption P2) implies

$$dN_i(t) \perp Y_i^{obs}(t), Y_i^{mis}(t) | \boldsymbol{X}_i(t), \overline{\boldsymbol{X}}_i(t^-), \overline{\boldsymbol{V}}_i(t), \overline{N}_i(t^-), \overline{Y}_i^{obs}(t^-).$$

This assumption is on the censored observation process ($dN_i(t)$), and does not say anything about the independence of $dN_i(t)$ and $C_i \geq t$, as in Assumption B1. As $dN(t)$ is a function of $dN^*(t)$ and $C(t)$, the assumption about visit process and censoring process are intertwined and cannot be separated. However, noninformative censoring in the sense of Assumption B2 was considered, which implies $Y_i(t) \perp C_i \geq t | \boldsymbol{X}_i(t)$. Considering this, the DAG will be the same as the one presented for the set of assumptions presented by Buzkova and Lumley (2007) in Figure 5.5.

5.4.3 *X(t)* Only Observed When *Y(t)* is, Noninformative Censoring Present

Pullenayegum and Lim (2016) was the only paper to present a set of assumptions for IIW which considered the setting where $\boldsymbol{X}_i(t)$ was not assumed to be completely observed. The authors state that the time-varying covariates $\boldsymbol{X}(t)$ are known at the visit times. As written, we assume that $dN_i^*(t)$ indicates if both $\boldsymbol{X}_i(t)$ and $Y_i(t)$ are observed. Otherwise, $\boldsymbol{X}(t)$ is missing. We assume that there is no partial missingness of $\boldsymbol{X}$ (i.e., if $dN_i^*(t) = 1$, then all covariates in $\boldsymbol{X}(t)$ are observed).

Pullenayegum and Lim (2016) state that under VAR (Assumption P2) and the noninformative censoring assumption (Assumption B2), along with the implicit assumptions B3, B4, and B5, IIW can be employed to adjust for the informative observation process, even in the presence of missing $\boldsymbol{X}(t)$. The auxiliary covariates $\boldsymbol{V}(t)$ are still assumed to be known at all possible observation times.

In its most general form, the VAR assumption (Assumption P2) is written as

$$E(dN_i(t)|\overline{\boldsymbol{X}}_i(t), \overline{\boldsymbol{V}}_i(t), \overline{N}_i(t^-), \overline{Y}_i^{obs}(t^-), Y_i(t)) = E(dN_i(t)|\overline{\boldsymbol{X}}_i^{obs}(t), \overline{\boldsymbol{V}}_i(t), \overline{N}_i(t^-), \overline{Y}_i^{obs}(t^-)),$$

which implies $dN_i(t) \perp \boldsymbol{X}_i^{mis}(t), \overline{\boldsymbol{X}}_i^{mis}(t^-), Y_i^{obs}(t), Y_i^{mis}(t), \overline{Y}_i^{mis}(t^-) | \boldsymbol{X}_i^{obs}(t), \overline{\boldsymbol{X}}_i^{obs}(t^-), \overline{\boldsymbol{V}}_i(t), \overline{N}_i(t^-), \overline{Y}_i^{obs}(t^-)$ in the presence of missing $\boldsymbol{X}(t)$.

Figure 5.6 shows a DAG highlighting a concern with the set of assumptions presented by Pullenayegum and Lim (2016) under VAR, when $\boldsymbol{X}(t)$ is also only observed when $Y(t)$ is, and noninformative censoring is present. While the authors claim that VAR in the sense of Assumption P2 can be employed when $dN^*(t)$ also determines when $\boldsymbol{X}(t)$ is observed, this scenario causes the VAR assumption to fail. To see this, consider how Pullenayegum and Lim (2016) conditions on $\overline{\boldsymbol{X}}^{obs}(t)$, which includes data up to and including time t.

In the DAG, we have $dN(1) \rightarrow X^{obs}(1) \leftarrow X(1) \rightarrow X^{mis}(1)$, and conditioning on $X^{obs}(1)$ would open a collider path which would cause the VAR assumption to fail, as $dN(1)$ would no longer be conditionally independent of $X^{mis}(1)$. Due to this, we cannot have an arrow from $dN^*(t)$ to $X^{obs}(t)$, which means that under Assumption P2, $dN^*(t)$ (and thus $dN(t)$), cannot determine when $\boldsymbol{X}(t)$ is observed.

The VAR assumption has another subtle but important issue that involves conditioning on $\overline{\boldsymbol{X}}^{obs}(t)$. Consider again how $X^{obs}(t) = X(t)$ when $dN(t) = 1$ and is null otherwise, making $X^{obs}(t)$ a deterministic function of $dN(t)$. Thus, when $\overline{\boldsymbol{X}}^{obs}(t)$ is in the conditioning statement for VAR (which is for the conditional independence of $dN(t)$ and $Y(t)$), due to the temporal ordering of our covariates we are technically conditioning on $dN(t)$, which has not occurred yet. This makes this assumption unrealistic to make in practice.

Neither Buzkova and Lumley (2007) nor Coulombe et al. (2021b) accounted for the setting where the observation process may also determine whether or not $\boldsymbol{X}(t)$ is observed. As the assumptions on the observation process include $\boldsymbol{X}(t) = \{\boldsymbol{X}^{obs}(t), \boldsymbol{X}^{mis}(t)\}$, the observation process can depend on the missing components of $\boldsymbol{X}(t)$ which are unknown in practice. As such, biased estimates of the intensity model parameters, and hence the outcome model parameters, may be obtained. While the authors each discuss using LOCF, this method is known to result in biased estimates of PH model parameters (Andersen and Liestøl, 2003; Cao and Fine, 2021; Molenberghs et al., 2002; Molnar et al., 2008). If the missing components of $\boldsymbol{X}(t)$ are thought to be related to the observation intensity, imputation methods such as MICE may be useful, as shown in Chapter 4. However, we note that the way one should handle the missingness in the observation covariates depends on the type of missingness, amount of missingness, and type of covariate (for example, whether or not it is linearly related to time).

5.5 Proposed Assumptions

As we have shown the VAR condition is inappropriate for the setting where $dN^*(t)$ dictates when $\boldsymbol{X}(t)$ is observed, we propose a slight variation on the VAR assumption that more carefully considers the temporal ordering of the covariates by conditioning on the history of the observed outcome model covariates prior to time t ($\overline{\boldsymbol{X}}^{obs}(t^-)$). Further, we consider writing the assumption for the counterfactual observation process ($dN^*(t)$) as Pullenayegum and Lim (2016) discuss how considering the counterfactual observation process may allow for more critical evaluation of the censoring process, and how the observation process may be related to it. This may allow analysts to identify violations of the

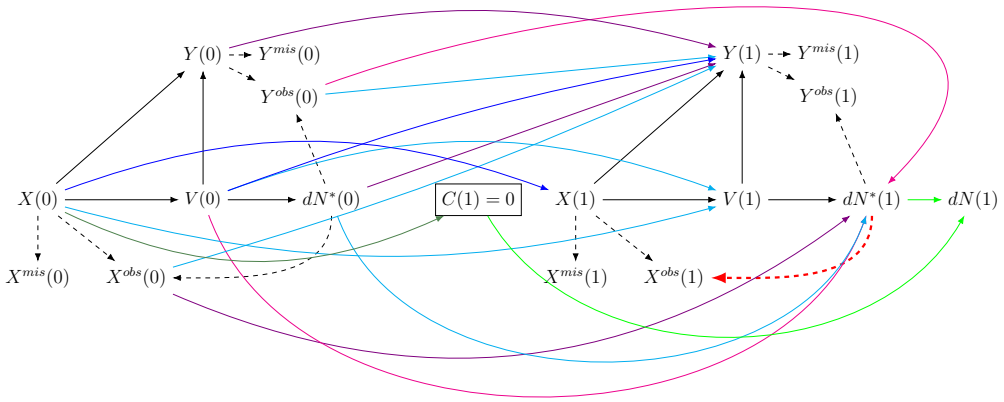


Figure 5.6: A DAG highlighting the **issue** with the VAR assumption presented by Pullenayegum and Lim (2016) in red, when $\boldsymbol{X}(t)$ and $Y(t)$ are only observed at observation times. **Noninformative censoring is present in this example**.

noninformative censoring assumption which may have been otherwise overlooked.

As such, our proposed variation of the VAR condition can be written as

Assumption P2(b). *Visiting at Random (Variation)*

$$E(dN_i^*(t)|\overline{\boldsymbol{X}}_i(t), \overline{\boldsymbol{V}}_i(t), \overline{N}_i(t^-), \overline{Y}_i(t)) = E(dN_i^*(t)|\overline{\boldsymbol{X}}_i^{obs}(t^-), \overline{\boldsymbol{V}}_i(t), \overline{N}_i(t^-), \overline{Y}_i^{obs}(t^-)),$$

which implies $dN_i^*(t) \perp \boldsymbol{X}_i^{obs}(t), \boldsymbol{X}_i^{mis}(t)\overline{\boldsymbol{X}}_i^{mis}(t^-), Y_i^{obs}(t), Y_i^{mis}(t), \overline{Y}_i^{mis}(t^-)|\overline{\boldsymbol{X}}_i^{obs}(t^-)$, $\overline{\boldsymbol{V}}_i(t), \overline{N}_i(t^-), \overline{Y}_i^{obs}(t^-)$.

Under this assumption, we require $\boldsymbol{X}(t)$ to be conditionally independent of $dN_i^*(t)$ given $\overline{\boldsymbol{X}}_i^{obs}(t^-), \overline{\boldsymbol{V}}_i(t), \overline{N}_i(t^-)$, and $\overline{Y}_i^{obs}(t^-)$. The DAG using the proposed variation on VAR is shown in Figure 5.7. We note that in this case, we can have an arrow from $dN^*(1)$ to $X^{obs}(1)$, as no collider paths will be opened as we are no longer conditioning on $X^{obs}(1)$. Further, we are not conditioning on any events that depend on $dN(t)$ at time t, and as such, no longer have the issue of conditioning on future covariates.

Again, if causal estimates are to be obtained from a marginal outcome model, we would either need to ensure that no arrows are going into $X(1)$, or include IPTW weights in our model to account for the confounding present.

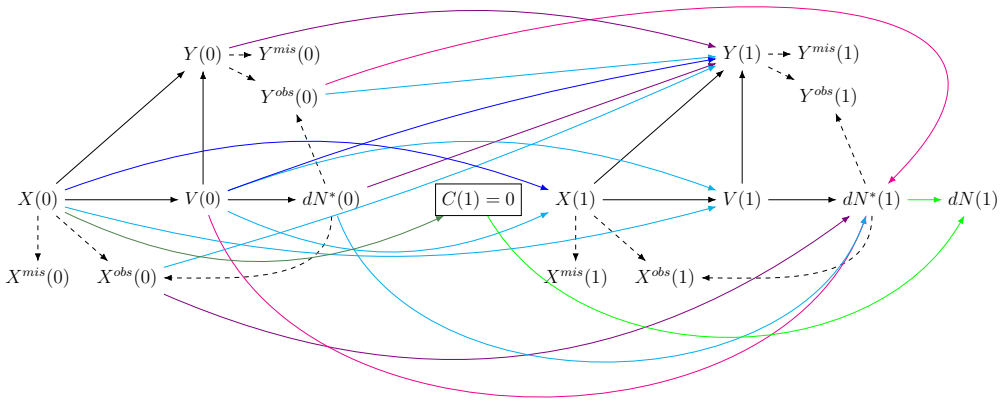


Figure 5.7: The DAG for the proposed variation of the VAR assumption presented by Pullenayegum and Lim (2016), when $\boldsymbol{X}(t)$ and $Y(t)$ are only observed at observation times. **Noninformative censoring is present in this example**.

To unify the sets of assumptions made on IIW, we propose explicitly referencing the following five assumptions when employing IIW; visiting at random (Assumption P2(b)), noninformative censoring (Assumption B2), separability (Assumption B3), correct specification (Assumption B4), and completely observed observation process covariates (Assumption B5). When all of these assumptions are met, we propose referring to the observation process as conditionally ignorable. That is, IIW can be used to adequately account for the bias introduced by the informative observation process.

5.6 Discussion

The current assumptions made on the observation and censoring processes are not universally agreed upon in the literature. To unify the assumptions and nomenclature associated with the observation and censoring processes for irregular longitudinal data analysis, we propose a slight variation to the most general observation process assumption, which is the VAR condition, and propose naming five explicit assumptions when employing IIW. This will allow practitioners employing IIW to think critically about each individual assumption. The DAGs provided will also allow analysts to visually assess the assumptions as a whole, and conceptualize the complex relationships between all of the possible covariates

in the models.

While the VAR assumption (as proposed by Pullenayegum and Lim (2016)) provided the most general assumption for the observation process, this assumption conditions on only the observed counterparts of the history of the outcome model covariates $\boldsymbol{X}(t)$, which allow them to be related to the observation intensity. For time-invariant covariates, this does not pose an issue as if you observe $\boldsymbol{X}(t)$ at least once, you know the value for all possible time points. However, if $\boldsymbol{X}(t)$ is thought to influence the observation intensity and does vary with time, the assumption that only the observed components of $\boldsymbol{X}$ are related to the observation intensity may not be met in practice. We acknowledge that requiring the covariates included in the observation times model to be known at all possible time points is a limitation to the IIW, and direct readers to Chapter 4 for a more in-depth analysis and discussion on this assumption.

Our proposed variation of the VAR assumption overcomes the issue of the collider bias and does not condition on future values. We also write the assumption in terms of the counterfactual observation process $dN^*(t)$ to more critically assess how the censoring process may impact the observation process.

In some settings, non-randomized time-varying treatments may be susceptible to treatment-confounder feedback where $X(t)$ (which causes $Y(t)$) is influenced by, for example, $Y(t-1)$. We emphasize that throughout this paper, we assume the covariates in the outcome model are randomized such that no treatment-confounder feedback is present. If treatment-confounder feedback is thought to be present in the data, traditional methods used to adjust for confounding may be biased, and g-estimation methods are necessary to obtain unbiased estimates of causal effects. For an introduction to g-estimation, see Chapter 21 of Hernan and Robins (2025).

Chapter 6

Discussion

The analysis of irregular longitudinal data is complex. One must consider the type of observation and censoring processes, whether confounding is present, and what covariates are available at various times. While IIW appears to be an intuitive method for handling informative observation processes, we have shown that the method has significant limitations which have yet to be addressed in the literature.

In Chapter 3, we examined an extension of the IIW method which incorporated IPTW weights into the outcome model. We found that this method, which we refer to as FIPTIW, is sensitive to violations of the noninformative censoring assumption. We showed that a previously proposed approach for handling informative censoring where IPCW weights are incorporated into the outcome model fails to account for the bias introduced by the censoring process. We also showed that the variables included in the intensity model should be conservative, as omitting any covariate related to both the longitudinal outcome and observation intensity can result in biased outcome model parameter estimates. Finally, we showed that FIPTIW is sensitive to extreme weights, and weight trimming cannot always be recommended. We also applied the FIPTIW method to a malaria data set, where we determined that children living in households with unprotected water sources had significantly increased odds of being diagnosed with malaria compared to those living in households with protected water sources.

Throughout this chapter, limitations of the FIPTIW method were apparent. First, in many cases estimating a time-varying intercept using cubic splines produced extreme estimates of the outcome model parameters, especially when extreme weights were present. As weight trimming did not adequately account for the bias in this setting, it may be useful to extend the FIPTIW method to employ the Lin and Ying (2001) semiparametric

approach where the time-varying intercept does not need to be specified. We also identified a need for FIPTIW to be extended further to account for violations of the noninformative censoring assumption, which often occurs in observational data sets, such as the PRISM study. We identify this as an important area of future work, and will be investigating if multiple imputation methods as presented in Jackson et al. (2014) or multiple outputation as presented in Pullenayegum et al. (2023) could be applied to the FIPTIW method to handle informative censoring. Further, we also hope to extend FIPTIW to handle clustered data, as in Pullenayegum et al. (2021). This would also allow for more children to be included in the real data analysis using the PRISM study, as children residing in the same household are likely to be correlated on a number of factors.

In Chapter 4, we emphasized the importance of the assumption that intensity model covariates are known at all possible observation times. We found that there is no "one-size-fits-all" approach to handling missing data in the observation intensity model. In the setting where the covariates included in the observation model were highly variable and not linearly related with time, MICE-PPM adequately accounted for the missing data in most cases. However, when the observation process covariates were only known at the observation times, none of the existing missing data methods we examined adequately handled the missingness. This type of missingness is a special case of MNAR, where the estimation of the PH model parameters is complicated by the fact that we never observe data when no visit/observation occurs. As we often only collect data at the observation times (including factors that may impact the probability of visiting at that time), it is necessary to develop methodology capable of handling this type of missingness in a general setting. We identify this as a future research project. Further, it is important to identify this type of missingness as a limitation of the IIW method as analysts may use standard missing data methods without evaluating if they are appropriate.

In Chapter 4, we also examined imputation in the setting where the missing observation intensity covariates were linearly related with time. Here, LOCF provided unbiased estimation in most settings. When the observation covariates were only known at the observation times (the special case of MNAR), LOCF had a small relative bias when a binary outcome was simulated, and was unbiased when a continuous outcome was simulated, while the other methods were largely biased. These results may be surprising to researchers as previous studies have criticized the use of LOCF to handle missing data (Lachin, 2016). However, we have identified a case where LOCF is sufficient for handling missing data in a relatively simplistic setting where the covariates are a monotonically increasing function of time. We emphasize that the results are highly dependent on how the data is generated and the types of covariates that are missing.

In this chapter, we also applied the IIW method and imputed missing data for a malaria

data set using the various methods presented in Chapter 4.2. The analysis found that when age was (a covariate that is linearly related with time) was only known at the observation times, LOCF and missForest were sufficient for handling the missing data. The results showed that there was no significant difference in the odds of malaria diagnoses by sex assigned at birth in children living in Uganda.

Chapter 5 provided an important evaluation of the assumptions that are made on IIW across three peer-reviewed papers. As the assumptions made on the observation process, censoring process, and outcome process are interconnected and complex, the DAGs provided a means to represent the relationships between the covariates for each set of assumptions. The DAGs also allowed us to compare and contrast existing sets of assumptions, and to determine which is the most general/flexible and appropriate. We found limitations with many of the existing sets of assumptions, and as such proposed a slight variation on the VAR assumption. We also suggested a unified set of assumptions, along with nomenclature, which we hope will be used in future projects using IIW. It is our hope that by clarifying and visualizing the assumptions, researchers will be able to more easily understand the assumptions and identify potential violations in their own applied work. Further, we hope that it will advance research in areas of irregular longitudinal data by serving as an in-depth review of IIW.

As in Chapter 3, Chapter 5 identified the importance of the noninformative censoring assumption. We again emphasize the need for the IIW method to be extended to handle violations of the noninformative censoring assumption, and identify this as a future research project.

References

Aghababaei Jazi, Omidali and Eleanor M. Pullenayegum (2022). "Variable selection in semiparametric regression models for longitudinal data with informative observation times". *Statistics in Medicine* 41.17, pp. 3281–3298.

Andersen, Per Kragh and Knut Liestøl (2003). "Attenuation caused by infrequently updated covariates in survival analysis". *Biostatistics* 4.4, pp. 633–649.

Austin, Peter C, Ian R White, Douglas S Lee, and Stef van Buuren (2021). "Missing data in clinical research: a tutorial on multiple imputation". *Canadian Journal of Cardiology* 37.9, pp. 1322–1331.

Austin, Peter C. (2011). "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies". *Multivariate Behavioral Research* 46, pp. 399–424.

Azur, Melissa J, Elizabeth A Stuart, Constantine Frangakis, and Philip J Leaf (2011). "Multiple imputation by chained equations: what is it and how does it work?" *International journal of methods in psychiatric research* 20.1, pp. 40–49.

Barrett, Malcolm (2024). *ggdag: Analyze and Create Elegant Directed Acyclic Graphs*. R package version 0.2.13.9000, https://r-causal.github.io/ggdag/. URL: https://github.com/r-causal/ggdag.

Bender, Ralf, Thomas Augustin, and Maria Blettner (2005). "Generating survival times to simulate Cox proportional hazards models". *Statistics in medicine* 24.11, pp. 1713–1723.

Brookhart, M. Alan, Sebastian Schneeweiss, Kenneth J. Rothman, Robert J. Glynn, Jerry Avorn, and Til Stürmer (2006). "Variable selection for propensity score models". *American Journal of Epidemiology* 163.12, pp. 1149–1156.

Buuren, Stef van, Hendriek C Boshuizen, and Dick L Knook (1999). "Multiple imputation of missing blood pressure covariates in survival analysis". *Statistics in medicine* 18.6, pp. 681–694.

Buzkova, Petra and Thomas Lumley (2007). "Longitudinal data analysis for generalized linear models with follow-up dependent on outcome-related variables". *Canadian Journal of Statistics* 35.4, pp. 485–500.

Buzkova, Petra and Thomas Lumley (2009). "Semiparametric modeling of repeated measurements under outcome-dependent follow-up". *Statistics in Medicine* 28.6, pp. 987–1003.

Cai, Na, Wenbin Lu, and Hao Helen Zhang (2012). "Time-varying latent effect model for longitudinal data with informative observation times". *Biometrics* 68.4, pp. 1093–1102.

Cao, Hongyuan and Jason P Fine (2021). "On the proportional hazards model with last observation carried forward covariates". *Annals of the Institute of Statistical Mathematics* 73, pp. 115–134.

Cole, Stephen R and Miguel A Hernán (2008). "Constructing inverse probability weights for marginal structural models". *American journal of epidemiology* 168.6, pp. 656–664.

Coulombe, Janie, Erica EM Moodie, and Robert W Platt (2021a). "Estimating the marginal effect of a continuous exposure on an ordinal outcome using data subject to covariate-driven treatment and visit processes". *Statistics in Medicine* 40.26, pp. 5746–5764.

Coulombe, Janie, Erica EM Moodie, and Robert W Platt (2021b). "Weighted regression analysis to correct for informative monitoring times and confounders in longitudinal studies". *Biometrics* 77.1, pp. 162–174.

Digitale, Jean C, Jeffrey N Martin, and Medellena Maria Glymour (2022). "Tutorial on directed acyclic graphs". *Journal of Clinical Epidemiology* 142, pp. 264–267.

Fitzmaurice, G.M., N.M. Laird, and J.H. Ware (2011). "Marginal Models: Introduction and Overview". In: *Applied Longitudinal Analysis*. John Wiley & Sons, Ltd, pp. 357–400. ISBN: 9781119513469. DOI: https://doi.org/10.1002/9781119513469.ch12. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119513469.ch12. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119513469.ch12.

Garrett, Rose, Brian Feldman, and Eleanor Pullenayegum (2024). "Bias in mixed models when analysing longitudinal data subject to irregular observation: when should we worry about it and how can recommended visit intervals help in specifying joint models when needed?" *arXiv preprint arXiv:2410.00662*.

Glymour, M Maria (2006). "Using causal diagrams to understand common problems in social epidemiology". *Methods in social epidemiology*, pp. 393–428.

Graham, John W, Allison E Olchowski, and Tamika D Gilreath (2007). "How many imputations are really needed? Some practical clarifications of multiple imputation theory". *Prevention science* 8, pp. 206–213.

Greenland, Sander, Judea Pearl, and James M. Robins (1999). "Causal diagrams for epidemiologic research". *Epidemiology* 10.1, pp. 37–48.

Greifer, Noah (2024). *cobalt: Covariate Balance Tables and Plots*. R package version 4.5.5. URL: https://CRAN.R-project.org/package=cobalt.

Guo, Chao-Yu, Ying-Chen Yang, and Yi-Hau Chen (2021). "The optimal machine learning-based missing data imputation for the cox proportional hazard model". *Frontiers in public health* 9, p. 680054.

Hernan, M.A. and J.M. Robins (2025). *Causal Inference: What If*. Chapman & Hall/CRC monographs on statistics & applied probability. Taylor & Francis. ISBN: 9781420076165.

Hong, Shangzhi and Henry S Lynn (2020). "Accuracy of random-forest-based imputation of missing data in the presence of non-normality, non-linearity, and interaction". *BMC medical research methodology* 20, pp. 1–12.

Hsu, Chiu-Hsieh and Mandi Yu (2019). "Cox regression analysis with missing covariates via nonparametric multiple imputation". *Statistical methods in medical research* 28.6, pp. 1676–1688.

Jackson, Dan, Ian R White, Shaun Seaman, Hannah Evans, Kathy Baisley, and James Carpenter (2014). "Relaxing the independent censoring assumption in the Cox proportional hazards model using multiple imputation". *Statistics in medicine* 33.27, pp. 4681–4694.

Joffe, Marshall M (2001). "Administrative and artificial censoring in censored regression models". *Statistics in medicine* 20.15, pp. 2287–2304.

Kamya, Moses R, Emmanuel Arinaitwe, Humphrey Wanzira, Agaba Katureebe, Chris Barusya, Simon P Kigozi, Maxwell Kilama, Andrew J Tatem, Philip J Rosenthal, Chris Drakeley, et al. (2015). "Malaria transmission, infection, and disease at three sites with varied transmission intensity in Uganda: implications for malaria control". *The American journal of tropical medicine and hygiene* 92.5, p. 903.

Lachin, John M (2016). "Fallacies of last observation carried forward analyses". *Clinical trials* 13.2, pp. 161–168.

Lane, Peter (2008). "Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches". *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry* 7.2, pp. 93–106.

Lee, Brian K, Justin Lessler, and Elizabeth A Stuart (2011). "Weight trimming and propensity score weighting". *PloS one* 6.3, e18174.

Lewis, P. A. W and G. S. Shedler (1979). "Simulation of nonhomogeneous poisson processes by thinning". *Naval Research Logistics Quarterly* 26.3, pp. 403–413. DOI: https://doi.org/10.1002/nav.3800260304. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/nav.3800260304. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/nav.3800260304.

Liang, Yu, Wenbin Lu, and Zhiliang Ying (2009). "Joint modeling and analysis of longitudinal data with informative observation times". *Biometrics* 65.2, pp. 377–384.

Lin, Danyu Y and Zhiliang Ying (2001). "Semiparametric and nonparametric regression analysis of longitudinal data". *Journal of the American Statistical Association* 96.453, pp. 103–126.

Lin, Haiqun, Daniel O. Scharfstein, and Robert A. Rosenheck (2004). "Analysis of longitudinal data with irregular, outcome-dependent follow-up". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 66.3, pp. 791–813. DOI: https://doi.org/10.1111/j.1467-9868.2004.b5543.x. eprint: https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9868.2004.b5543.x. URL: %5Bhttps://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2004.b5543.x%5D(https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2004.b5543.x).

Lokku, Armend, Lily S. Lim, Catherine S. Birken, and Eleanor M. Pullenayegum (2020). "Summarizing the extent of visit irregularity in longitudinal data". *BMC Medical Research Methodology* 20.135.

Ma, Guoguang, Andrea B Troxel, and Daniel F Heitjan (2005). "An index of local sensitivity to nonignorable drop-out in longitudinal modelling". *Statistics in Medicine* 24, pp. 2429–2150.

Mansournia, Mohammad Ali, Julian PT Higgins, Jonathan AC Sterne, and Miguel A Hernán (2017). "Biases in randomized trials: a conversation between trialists and epidemiologists". *Epidemiology* 28.1, pp. 54–59.

Marshall, Andrea, Douglas G Altman, and Roger L Holder (2010). "Comparison of imputation methods for handling missing covariate data when fitting a Cox proportional hazards model: a resampling study". *BMC medical research methodology* 10, pp. 1–10.

Molenberghs, Geert, Paige L Williams, and Stuart R Lipsitz (2002). "Prediction of survival and opportunistic infections in HIV-infected patients: a comparison of imputation methods of incomplete CD4 counts". *Statistics in medicine* 21.10, pp. 1387–1408.

Molnar, Frank J, Brian Hutton, and Dean Fergusson (2008). "Does analysis using "last observation carried forward" introduce bias in dementia research?" *Cmaj* 179.8, pp. 751–753.

Msellemu, Daniel, Hagai I Namango, Victoria M Mwakalinga, Alex J Ntamatungiro, Yeromin Mlacha, Zacharia J Mtema, Samson Kiware, Neil F Lobo, Silas Majambere, Stefan Dongus, et al. (2016). "The epidemiology of residual Plasmodium falciparum malaria transmission and infection burden in an African city with high coverage of multiple vector control measures". *Malaria journal* 15.1, pp. 1–19.

Mwakalinga, Victoria M, Benn KD Sartorius, Alex J Limwagu, Yeromin P Mlacha, Daniel F Msellemu, Prosper P Chaki, Nicodem J Govella, Maureen Coetzee, Stefan Dongus, and

Gerry F Killeen (2018). "Topographic mapping of the interfaces between human and aquatic mosquito habitats to enable barrier targeting of interventions against malaria vectors". *Royal Society Open Science* 5.5, p. 161055.

Potter, Frank J (1993). "The effect of weight trimming on nonlinear survey estimates". *Proceedings of the Section on Survey Research Methods*, pp. 758–763.

Pullenayegum, Eleanor M. (2013). "Doubly robust estimation, optimally truncated inverse-intensity weighting and increment-based methods for the analysis of irregularly observed longitudinal data". *Statistics in Medicine* 32, pp. 1054–1072.

Pullenayegum, Eleanor M., Catherine Birken, and Jonathon Maguire (Apr. 2023). "Causal inference with longitudinal data subject to irregular assessment times". *Statistics in Medicine*. DOI: 10.1002/sim.9727.

Pullenayegum, Eleanor M., Catherine Birken, Jonathon Maguire, and TARGet Kids! Collaboration (2021). "Clustered longitudinal data subject to irregular observation". *Statistical Methods in Medical Research* 30.4, pp. 1081–1100.

Pullenayegum, Eleanor M. and Lily SH Lim (2016). "Longitudinal data subject to irregular observation: A review of methods with a focus on visit processes, assumptions, and study design". *Statistical Methods in Medical Research* 25.6, pp. 2992–3014.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Robins, James M. (2000). "Robust estimation in sequentially ignorable missing data and causal inference models". In: *Proceedings of the American Statistical Association*. Vol. 1999. Indianapolis, IN, pp. 6–10.

Robins, James M., Donald Blevins, Grant Ritter, and Michael Wulfsohn (1992). "G-estimation of the effect of prophylaxis therapy for Pneumocystis carinii pneumonia on the survival of AIDS patients". *Epidemiology* 3.4, pp. 319–336.

Robins, James M., Miguel Angel Hernan, and Babette Brumback (2000). "Marginal structural models and causal inference in epidemiology". *Epidemiology*, pp. 550–560.

Robins, James M., Andrea Rotnitzky, and Lue Ping Zhao (1995). "Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data". *Journal of the American Statistical Association* 90.429, pp. 106–121.

Rosenbaum, Paul R. and Donald B. Rubin (1983). "The central role of the propensity score in observational studies for causal effects". *Biometrika* 70.1, pp. 41–55.

Rubin, Donald B (1976). "Inference and missing data". *Biometrika* 63.3, pp. 581–592.

Saha, Chandan and Michael P Jones (2009). "Bias in the last observation carried forward method under informative dropout". *Journal of Statistical Planning and Inference* 139.2, pp. 246–255.

Scharfstein, Daniel O, Andrea Rotnitzky, and James M. Robins (1999). "Adjusting for non-ignorable drop-out using semiparametric nonresponse models". *Journal of the American Statistical Association* 94.448, pp. 1096–1120.

Schmidt, S. C. E. and A. Woll (2017). "Longitudinal drop-out and weighting against its bias". *BMC Medical Research Methodology* 17.164.

Shah, Anoop D, Jonathan W Bartlett, James Carpenter, Owen Nicholas, and Harry Hemingway (2014). "Comparison of random forest and parametric imputation models for imputing missing data using MICE: a CALIBER study". *American journal of epidemiology* 179.6, pp. 764–774.

Shayo, Festo Kasmir, Keiko Nakamura, Saber Al-Sobaihi, and Kaoruko Seino (2021). "Is the source of domestic water associated with the risk of malaria infection? Spatial variability and a mixed-effects multilevel analysis". *International Journal of Infectious Diseases* 104, pp. 224–231.

Smith, Bonnie, Shu Yang, Andrea J Apter, and Daniel O Scharfstein (2022). "Trials with Irregular and Informative Assessment Times: A Sensitivity Analysis Approach". *arXiv preprint arXiv:2204.11979*.

Song, Xinyuan, Xiaoyun Mu, and Liuquan Sun (2012). "Regression analysis of longitudinal data with time-dependent covariates and informative observation times". *Scandinavian Journal of Statistics* 39.2, pp. 248–258.

Stekhoven, Daniel J. and Peter Buehlmann (2012). "MissForest - non-parametric missing value imputation for mixed-type data". *Bioinformatics* 28.1, pp. 112–118.

Stuart, Elizabeth A. (2010). "Matching methods for causal inference: A review and a look forward". *Stat Sci.* 25.1, pp. 1–21.

Sun, Liu-quan, Xinyuan Song, and Jie Zhou (2011). "Regression analysis of longitudinal data with time-dependent covariates in the presence of informative observation and censoring times". *Journal of Statistical Planning and Inference* 141.8, pp. 2902–2919.

Sun, Liuquan, Xinyuan Song, Jie Zhou, and Lei Liu (2012). "Joint analysis of longitudinal data with informative observation times and a dependent terminal event". *Journal of the American Statistical Association* 107.498, pp. 688–700.

Tsiatis, Anastasios A (2006). *Semiparametric theory and missing data*. Vol. 4. Springer.

van Buuren, Stef and Karin Groothuis-Oudshoorn (2011). "mice: Multivariate Imputation by Chained Equations in R". *Journal of Statistical Software* 45.3, pp. 1–67. DOI: 10.18637/jss.v045.i03.

White, Ian R, Patrick Royston, and Angela M Wood (2011). "Multiple imputation using chained equations: issues and guidance for practice". *Statistics in medicine* 30.4, pp. 377–399.

Yang, Dan, Yang He, Bo Wu, Yan Deng, Menglin Li, Qian Yang, Liting Huang, Yaming Cao, and Yang Liu (2020). "Drinking water and sanitation conditions are associated

with the risk of malaria among children under five years old in sub-Saharan Africa: a logistic regression model analysis of national survey data". *Journal of advanced research* 21, pp. 1–13.

Young, Jessica G, Mats J Stensrud, Eric J Tchetgen Tchetgen, and Miguel A Hernán (2020). "A causal framework for classical statistical estimands in failure-time settings with competing events". *Statistics in medicine* 39.8, pp. 1199–1236.

Zhu, Liang, Hui Zhao, Jianguo Sun, Stanley Pounds, and Hui Zhang (2013). "Joint analysis of longitudinal data and recurrent episodes data with application to medical cost analysis". *Biometrical journal* 55.1, pp. 5–16.

Zhu, Yeying, Donna L Coffman, and Debashis Ghosh (2015a). "A boosting algorithm for estimating generalized propensity scores with continuous treatments". *Journal of causal inference* 3.1, pp. 25–40.

Zhu, Yeying, Maya Schonbach, Donna L. Coffman, and Jennifer S. Willams (2015b). "Variable Selection for Propensity Score Estimation via Balancing Covariates". *Epidemiology* 26.2, e14–e15.

APPENDICES

Appendix A

Additional Results for Chapter 3

The following Chapter contains supplementary material for Chapter 3.

The proof of Theorem 3.3.1 can be found in Section A.1. Section A.2 contains additional results for the simulations performed in the main paper. The additional results for Simulations I (sensitivity to violations of the noninformative censoring assumption), II (variable inclusion), and III (weight trimming) can be found in Sections A.2.3, A.2.4, and A.2.5, respectively. Section A.3 contains sensitivity analyses for the PRISM cohort data analysis presented in the main paper. The sensitivity analysis for violations of the non-informative censoring assumption can be found in Section A.3.1. The sensitivity analysis for violations of the independence assumption (clustering) can be found in Section A.3.2. The sensitivity analysis where individuals who had surrogate censoring times prior to six months were randomly censored between zero and six months can be found in Section A.3.3.

A.1 Proofs

Proof of Theorem 3.3.1. Let $w_i^F(t;\boldsymbol{\alpha},\boldsymbol{\gamma},\pi,h)$ denote the FIPTIW weight for individual i at time t. Without loss of generality, we prove this result using stabilized weights for general function $h(\cdot)$ of covariates $\boldsymbol{X}_i(t)$ for the IIW weights and general function $\pi(\cdot)$ of covariates $\boldsymbol{W}_i(t)$ for the IPTW weights. We begin by taking the expectation and applying the law of iterative expectations as

$$
\begin{aligned}
&E\{U(\boldsymbol{\beta};\boldsymbol{\alpha},\boldsymbol{\gamma},\pi,h)\} \\
&= E\Bigg\{\sum_{i=1}^{n}\int_0^{\tau}\boldsymbol{X}_i(t)\left\{\frac{dg(\mu)}{d\mu}\Big|_{\mu_i(t;\boldsymbol{\beta})}\right\}^{-1} v(\mu_i(t;\boldsymbol{\beta}))^{-1}\{Y_i(t)-\mu_i(t;\boldsymbol{\beta})\}w_i^F(t;\boldsymbol{\alpha},\boldsymbol{\gamma},\pi,h)dN_i(t)\Bigg\} \\
&= E\Bigg\{E\Bigg\{\sum_{i=1}^{n}\int_0^{\tau}\boldsymbol{X}_i(t)\left\{\frac{dg(\mu)}{d\mu}\Big|_{\mu_i(t;\boldsymbol{\beta})}\right\}^{-1} v(\mu_i(t;\boldsymbol{\beta}))^{-1}\{Y_i(t)-\mu_i(t;\boldsymbol{\beta})\} \\
&\qquad \times w_i^F(t;\boldsymbol{\alpha},\boldsymbol{\gamma},\pi,h)dN_i(t)\big|\boldsymbol{X}_i(t)\Bigg\}\Bigg\} \\
&= E\Bigg\{\sum_{i=1}^{n}\int_0^{\tau}\boldsymbol{X}_i(t)\left\{\frac{dg(\mu)}{d\mu}\Big|_{\mu_i(t;\boldsymbol{\beta})}\right\}^{-1} v(\mu_i(t;\boldsymbol{\beta}))^{-1} \\
&\qquad \times E\Big\{\{Y_i(t)-\mu_i(t;\boldsymbol{\beta})\}w_i^F(t;\boldsymbol{\alpha},\boldsymbol{\gamma},\pi,h)dN_i(t)\big|\boldsymbol{X}_i(t)\Big\}\Bigg\}
\end{aligned}
$$

To show this expectation is zero, we can show that the conditional expectation nested inside of the full expression is zero. We show this by again using the law of iterative expectations to re-write the conditional expectation as

$$
\begin{aligned}
&E\Big\{\{Y_i(t)-\mu_i(t;\boldsymbol{\beta})\}w_i^F(t;\boldsymbol{\alpha},\boldsymbol{\gamma},\pi,h)dN_i(t)\big|\boldsymbol{X}_i(t)\Big\} \\
&= E\Big\{E\Big[\{Y_i(t)-\mu_i(t;\boldsymbol{\beta})\}w_i^F(t;\boldsymbol{\alpha},\boldsymbol{\gamma},\pi,h)dN_i(t)\big|\boldsymbol{Z}_i(t),\boldsymbol{X}_i(t),\boldsymbol{W}_i,Y_i(t),C_i\geq t\Big]\big|\boldsymbol{X}_i(t)\Big\} \\
&= E\Big\{E\Big[\{Y_i(t)-\mu_i(t;\boldsymbol{\beta})\}\left(\frac{D_i(t)}{\pi_i(\boldsymbol{W}_i;\alpha)}+\frac{(1-D_i(t))}{(1-\pi_i(\boldsymbol{W}_i;\alpha)}\right)\left(\frac{h(\boldsymbol{X}_i(t))}{\exp(\boldsymbol{Z}_i(t)\boldsymbol{\gamma})}\right) \\
&\qquad \times dN_i(t)\big|\boldsymbol{Z}_i(t),\boldsymbol{X}_i(t),\boldsymbol{W}_i,Y_i(t),C_i\geq t\Big]\big|\boldsymbol{X}_i(t)\Big\} \\
&= E\Big\{\{Y_i(t)-\mu_i(t;\boldsymbol{\beta})\}E\Big[\left(\frac{D_i(t)}{\pi_i(\boldsymbol{W}_i;\alpha)}+\frac{(1-D_i(t))}{(1-\pi_i(\boldsymbol{W}_i;\alpha))}\right)\left(\frac{h(\boldsymbol{X}_i(t))}{\exp(\boldsymbol{Z}_i(t)\boldsymbol{\gamma})}\right) \\
&\qquad \times dN_i(t)\big|\boldsymbol{Z}_i(t),\boldsymbol{X}_i(t),\boldsymbol{W}_i,Y_i(t),C_i\geq t\Big]\big|\boldsymbol{X}_i(t)\Big\}.
\end{aligned}
$$

If we assume the observation and treatment assignment processes are conditionally independent given the outcome, censoring time, and the covariates related to the outcome, probability treatment assignment, and observation intensity, then the conditional expectation can be separated into two multiplicative terms. That is, we can write the expectation above as

$$= E\Big\{\{Y_i(t) - \mu_i(t;\boldsymbol{\beta})\} E\Big[\left(\frac{D_i(t)}{\pi_i(\boldsymbol{W}_i;\alpha)} + \frac{(1-D_i(t))}{(1-\pi_i(\boldsymbol{W}_i;\alpha))}\right) \big| \boldsymbol{Z}_i(t), \boldsymbol{X}_i(t), \boldsymbol{W}_i, Y_i(t), C_i \geq t\Big] \times$$
$$E\Big[\left(\frac{h(\boldsymbol{X}_i(t))}{\exp(\boldsymbol{Z}_i(t)\boldsymbol{\gamma})}\right) dN_i(t) \big| \boldsymbol{Z}_i(t), \boldsymbol{X}_i(t), \boldsymbol{W}_i, Y_i(t), C_i \geq t\Big] \big| \boldsymbol{X}_i(t)\Big\}.$$

Recall that $D_i(t)$ is a covariate contained in the set $\boldsymbol{X}_i(t)$. As such, we can simplify this expression to be

$$= E\Big\{\{Y_i(t) - \mu_i(t;\boldsymbol{\beta})\} \left(\frac{(D_i(t)}{\pi_i(\boldsymbol{W}_i;\alpha)} + \frac{(1-D_i(t))}{(1-\pi_i(\boldsymbol{W}_i;\alpha))}\right) \times$$
$$\left(\frac{h(\boldsymbol{X}_i(t))}{\exp(\boldsymbol{Z}_i(t)\boldsymbol{\gamma})}\right) E\left[dN_i(t) \big| \boldsymbol{Z}_i(t), \boldsymbol{X}_i(t), \boldsymbol{W}_i, Y_i(t), C_i \geq t\right] \big| \boldsymbol{X}_i(t)\Big\}.$$

As $C_i \geq t$, $dN_i(t) = dN_i^*(t)$, and we can write this quantity as

$$= E\Big\{\{Y_i(t) - \mu_i(t;\boldsymbol{\beta})\} \left(\frac{(D_i(t)}{\pi_i(\boldsymbol{W}_i;\alpha)} + \frac{(1-D_i(t))}{(1-\pi_i(\boldsymbol{W}_i;\alpha))}\right) \times$$
$$\left(\frac{h(\boldsymbol{X}_i(t))}{\exp(\boldsymbol{Z}_i(t)\boldsymbol{\gamma})}\right) E\left[dN_i^*(t) \big| \boldsymbol{Z}_i(t), \boldsymbol{X}_i(t), \boldsymbol{W}_i, Y_i(t), C_i \geq t\right] \big| \boldsymbol{X}_i(t)\Big\}.$$

Then under O1, we can simplify this expression to

$$= E\Big\{\{Y_i(t) - \mu_i(t;\boldsymbol{\beta})\} \left(\frac{(D_i(t)}{\pi_i(\boldsymbol{W}_i;\alpha)} + \frac{(1-D_i(t))}{(1-\pi_i(\boldsymbol{W}_i;\alpha))}\right) \times$$
$$\left(\frac{h(\boldsymbol{X}_i(t))}{\exp(\boldsymbol{Z}_i(t)\boldsymbol{\gamma})}\right) E\left[dN_i^*(t) \big| \boldsymbol{Z}_i(t)\right] \big| \boldsymbol{X}_i(t)\Big\}.$$

and under O4, we can further simplify this to

$$= E\Big\{\{Y_i(t) - \mu_i(t;\boldsymbol{\beta})\} \left(\frac{D_i(t)}{\pi_i(\boldsymbol{W}_i;\alpha)} + \frac{(1-D_i(t))}{(1-\pi_i(\boldsymbol{W}_i;\alpha))}\right) h(\boldsymbol{X}_i(t))\lambda_0(t) \big| \boldsymbol{X}_i(t)\Big\}$$
$$= \left(\frac{D_i(t)}{\pi_i(\boldsymbol{W}_i;\alpha)} + \frac{(1-D_i(t))}{(1-\pi_i(\boldsymbol{W}_i;\alpha))}\right) h(\boldsymbol{X}_i(t))\lambda_0(t) E\Big\{\{Y_i(t) - \mu_i(t;\boldsymbol{\beta})\} \big| \boldsymbol{X}_i(t)\Big\}.$$

As $E\Big\{\{Y_i(t) - \mu_i(t;\boldsymbol{\beta})\} \big| \boldsymbol{X}_i(t)\Big\} = 0$, we have shown that the expectation of the estimating equation is zero at $\{\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \pi, h\}$ for any function $h(\cdot)$ of covariates $\boldsymbol{X}_i(t)$. We additionally note that $dN_i(t) = 0$ when the individual is censored ($C_i < t$). □

A.2 Additional Simulation Results

This section contains additional simulation studies not shown in the main paper. Section A.2.1 outlines the data-generating mechanisms for a continuous (normally-distributed) outcome. Section A.2.2 compares the use of cubic splines to estimate the time-varying intercept in the continuous and binary outcome cases. Sections A.2.3 to A.2.5 show the results of Simulations I to III for different sample sizes, and also the continuous case for $n = 100$.

A.2.1 Data Generation for Continuous Outcomes

For simulations where a continuous outcome is simulated, we generate the outcome as

$$Y_i(t) = \mu(t) + \beta_1 D_i + \beta_2(G_i(t) - E\{G_i(t)|D_i\}) + \beta_3(Z_i - E\{Z_i|D_i\}) + \phi_i + \epsilon_i(t), \quad \text{(A.1)}$$

where $\mu(t) = (2 - t)$, $\epsilon_i(t) \sim N(0, 1)$ is the random error term (for each individual and time) and $\phi_i \sim N(0, 1.25)$ is an individual random effect. We let $\beta_1 = 0.5$, $\beta_2 = 2$, and $\beta_3 = 1$. As the outcome $Y(t)$ is continuous, modeling via a marginal model should have very little effect on the bias of the estimated model parameters $\boldsymbol{\beta}$, even though the data was generated using a subject-specific (mixed effects) model. By centering the auxiliary covariates, we allow these covariates to be related to the longitudinal outcome but omitted from the marginal model of interest

$$E\{Y_i(t)|D_i\} = (2 - t) + \beta_1 D_i. \quad \text{(A.2)}$$

The estimation of β_1, the average treatment effect, is of primary interest for each simulation study. To estimate the time-varying intercept in the simulation studies, we turn to cubic splines along with a constant intercept as in Coulombe et al.Coulombe et al., 2021b. The knots of the spline were chosen by finding the tertiles of the observation times to estimate $\mu(t)$. Although the simulations assume the outcome is normally distributed, we note that we can also simulate data with non-identity link functions, as in Bůžková and Lumley.Buzkova and Lumley, 2007

A.2.2 Simulation Study Examining the Use of Cubic Splines

We perform a simulation study examining the use of cubic splines for continuous and binary outcomes.

We simulate data as in Section 3.4.1 for the binary case and Section A.2.1 for the continuous case. We let $\gamma_2 = \{0, 0.3\}$, $\gamma_3 = \{0, 0.6\}$, and $\alpha_1 = \{0, 1\}$. When γ_2 or γ_3 are non-zero, the observation process is conditionally ignorable and must be accounted for in the analysis. When $\alpha_1 \neq 0$, the treatment assignment process is conditionally ignorable and must be accounted for in the analysis.

We simulate 1000 data sets for each combination of $\{\gamma_2, \gamma_3, \alpha_1\}$, and fit an unweighted model (Naive), IIW-GEE, inverse probability of treatment weighted generalized estimating equation (IPTW-GEE), and FIPTIW-GEE using two different methods for specifying or estimating the time-varying intercept $\mu(t)$ in the outcome model. The first is by specifying the parametric form of $\mu(t)$ and treating it as an offset term in the outcome model. This assumes that the form of the time-varying intercept is known, which is often not the case in practice. Thus, we also estimate $\mu(t)$ by employing cubic splines where the knots are chosen as the tertiles of the potential observation times. We aggregate the results and calculate the bias, mean squared error (MSE), average standard error (Avg SE), and coverage of the 95% confidence interval over the 1000 data sets.

The results are shown in Table A.1 for the binary case and Table A.2 for the continuous case. In all cases, the FIPTIW-GEE is unbiased, even when it is not necessary to adjust for the observation and/or treatment assignment processes. When cubic splines are used to estimate the time-varying intercept, we tend to see an increase in the MSE, however the results remain unbiased. As the time-varying intercept will often be unknown in practice, we see cubic splines as a useful way to estimate them in the outcome model.

In most cases, we tend to see a slight under-coverage for the FIPTIW method.

				Parametric Offset Used				Cubic Splines Used			
γ_2	γ_3	α_1		Naive	IIW	IPTW	FIPTIW	Naive	IIW	IPTW	FIPTIW
0	0	0	Bias:	-0.011	-0.009	-0.015	-0.013	-0.022	-0.023	-0.016	-0.017
			MSE:	0.136	0.136	0.136	0.135	0.241	0.241	0.234	0.234
			Avg SE:	0.359	0.360	0.362	0.363	0.447	0.477	0.451	0.451
			Coverage:	0.914	0.916	0.924	0.920	0.924	0.740	0.930	0.936
		1.5	Bias:	0.097	0.097	-0.022	-0.021	0.211	0.213	0.024	0.025
			MSE:	0.093	0.093	0.073	0.073	0.223	0.224	0.175	0.175
			Avg SE:	0.292	0.292	0.287	0.287	0.405	0.713	0.418	0.418
			Coverage:	0.940	0.950	0.958	0.952	0.912	0.910	0.954	0.956
	0.6	0	Bias:	0.269	-0.004	0.268	-0.005	-0.216	0.016	-0.219	0.013
			MSE:	0.238	0.152	0.234	0.151	0.289	0.227	0.292	0.228
			Avg SE:	0.374	0.366	0.379	0.371	0.448	0.516	0.451	0.426
			Coverage:	0.880	0.930	0.878	0.936	0.874	0.766	0.882	0.926
		1.5	Bias:	0.409	0.127	0.286	0.012	-0.014	0.222	-0.223	0.024
			MSE:	0.274	0.110	0.179	0.088	0.209	0.208	0.240	0.145
			Avg SE:	0.303	0.293	0.299	0.289	0.414	0.722	0.417	0.386
			Coverage:	0.722	0.918	0.814	0.944	0.924	0.914	0.896	0.948
0.3	0	0	Bias:	0.076	0.019	0.072	0.013	-0.013	-0.017	-0.024	-0.028
			MSE:	0.148	0.132	0.147	0.131	0.202	0.194	0.185	0.179
			Avg SE:	0.366	0.356	0.369	0.359	0.439	0.483	0.441	0.436
			Coverage:	0.938	0.940	0.940	0.946	0.942	0.778	0.946	0.950
		1.5	Bias:	0.196	0.127	0.068	0.002	0.189	0.175	0.001	-0.007
			MSE:	0.131	0.099	0.084	0.074	0.200	0.193	0.174	0.171
			Avg SE:	0.291	0.285	0.284	0.282	0.396	0.675	0.409	0.405
			Coverage:	0.900	0.936	0.954	0.950	0.920	0.892	0.950	0.948
	0.6	0	Bias:	0.332	-0.003	0.338	0.004	-0.255	-0.008	-0.258	-0.008
			MSE:	0.260	0.131	0.261	0.125	0.313	0.179	0.303	0.170
			Avg SE:	0.383	0.359	0.389	0.365	0.451	0.492	0.454	0.415
			Coverage:	0.856	0.948	0.860	0.960	0.856	0.774	0.878	0.952
		1.5	Bias:	0.479	0.121	0.354	0.008	0.012	0.202	-0.201	0.006
			MSE:	0.337	0.101	0.217	0.077	0.199	0.178	0.232	0.132
			Avg SE:	0.306	0.284	0.298	0.279	0.416	0.702	0.418	0.379
			Coverage:	0.636	0.912	0.768	0.950	0.908	0.934	0.896	0.962

Table A.1: Results of the simulation study investigating the impact of using cubic splines for the **binary** outcome case and $n = 100$. Average treatment effect (ATE, β_1) is calculated by fitting an independent GEE with various weights for each simulation scheme over the 1000 generated data sets. The weighting methods include no weighting (unweighted), inverse intensity weighting (IIW), inverse probability of treatment weighting (IPTW), and flexible inverse probability of treatment and intensity weighting (FIPTIW). Standard errors (SEs) and 95% confidence intervals are calculated using the Robins et al., 2000 type standard error estimates.

				Parametric Offset Used				Cubic Splines Used			
γ_2	γ_3	α_1		Naive	IIW	IPTW	FIPTIW	Naive	IIW	IPTW	FIPTIW
0	0	0	Bias:	-0.007	-0.005	0.008	0.010	-0.010	-0.011	-0.008	-0.008
			MSE:	0.120	0.118	0.116	0.114	0.205	0.202	0.180	0.177
			Avg SE:	0.342	0.342	0.346	0.346	0.422	0.489	0.424	0.425
			Coverage:	0.942	0.934	0.946	0.940	0.916	0.774	0.946	0.940
		1.5	Bias:	0.162	0.158	0.017	0.013	0.250	0.250	0.004	0.004
			MSE:	0.103	0.099	0.078	0.074	0.230	0.230	0.155	0.155
			Avg SE:	0.273	0.273	0.285	0.285	0.388	0.750	0.399	0.399
			Coverage:	0.890	0.898	0.938	0.946	0.888	0.918	0.950	0.948
	0.6	0	Bias:	0.295	0.012	0.286	0.004	-0.259	0.039	-0.255	0.044
			MSE:	0.241	0.132	0.228	0.124	0.294	0.198	0.264	0.172
			Avg SE:	0.356	0.343	0.359	0.348	0.432	0.539	0.433	0.406
			Coverage:	0.840	0.932	0.866	0.934	0.888	0.796	0.894	0.928
		1.5	Bias:	0.443	0.147	0.300	0.004	-0.047	0.240	-0.289	-0.002
			MSE:	0.299	0.114	0.193	0.097	0.182	0.199	0.264	0.132
			Avg SE:	0.286	0.276	0.295	0.286	0.409	0.740	0.420	0.383
			Coverage:	0.636	0.870	0.776	0.906	0.944	0.942	0.902	0.942
0.3	0	0	Bias:	0.097	0.018	0.092	0.012	0.023	0.024	0.029	0.030
			MSE:	0.130	0.115	0.120	0.107	0.206	0.199	0.187	0.181
			Avg SE:	0.341	0.337	0.345	0.340	0.419	0.524	0.423	0.416
			Coverage:	0.918	0.928	0.934	0.944	0.910	0.810	0.940	0.938
		1.5	Bias:	0.209	0.119	0.064	-0.021	0.276	0.266	0.022	0.022
			MSE:	0.119	0.083	0.075	0.069	0.238	0.224	0.147	0.140
			Avg SE:	0.269	0.266	0.276	0.277	0.387	0.766	0.397	0.391
			Coverage:	0.872	0.914	0.944	0.940	0.876	0.940	0.952	0.960
	0.6	0	Bias:	0.358	-0.006	0.362	0.000	-0.289	-0.010	-0.291	-0.012
			MSE:	0.287	0.133	0.275	0.124	0.314	0.184	0.299	0.166
			Avg SE:	0.371	0.349	0.374	0.351	0.445	0.490	0.448	0.412
			Coverage:	0.830	0.920	0.846	0.936	0.862	0.780	0.886	0.948
		1.5	Bias:	0.518	0.139	0.383	0.005	-0.028	0.240	-0.275	0.001
			MSE:	0.357	0.091	0.234	0.076	0.186	0.207	0.256	0.144
			Avg SE:	0.283	0.267	0.288	0.280	0.413	0.740	0.423	0.379
			Coverage:	0.552	0.912	0.712	0.950	0.936	0.928	0.908	0.948

Table A.2: Results of the simulation study investigating the impact of using cubic splines for the **continuous** outcome case and $n = 100$. Average treatment effect (ATE, β_1) is calculated by fitting an independent GEE with various weights for each simulation scheme over the 1000 generated data sets. The weighting methods include no weighting (unweighted), inverse intensity weighting (IIW), inverse probability of treatment weighting (IPTW), and flexible inverse probability of treatment and intensity weighting (FIPTIW). Standard errors (SE) and 95% confidence intervals are calculated using the Robins et al., 2000 type standard error estimates.

A.2.3 Additional Results for Simulation I

Simulation I was performed using sample sizes of $n = 50$ and $n = 500$ for the binary outcome case. The results are shown in Figures A.1 and A.2 and Tables A.3 and A.4, respectively. The results of the simulation study are similar to the case of $n = 100$, aside from the variances. The variance of the estimated ATEs is much larger for the smaller sample size of $n = 50$, and there were extreme outliers of the estimates ATEs present (omitted from Figure A.1). The variance of the estimated ATEs is much smaller for the larger sample size of $n = 500$.

Simulation I was also performed using a continuous outcome, as described in Section A.2.1 for the sample size of n = 100. In this simulation, we let $\beta_2 = 2$. These results are shown in Figure A.3 and table A.5.

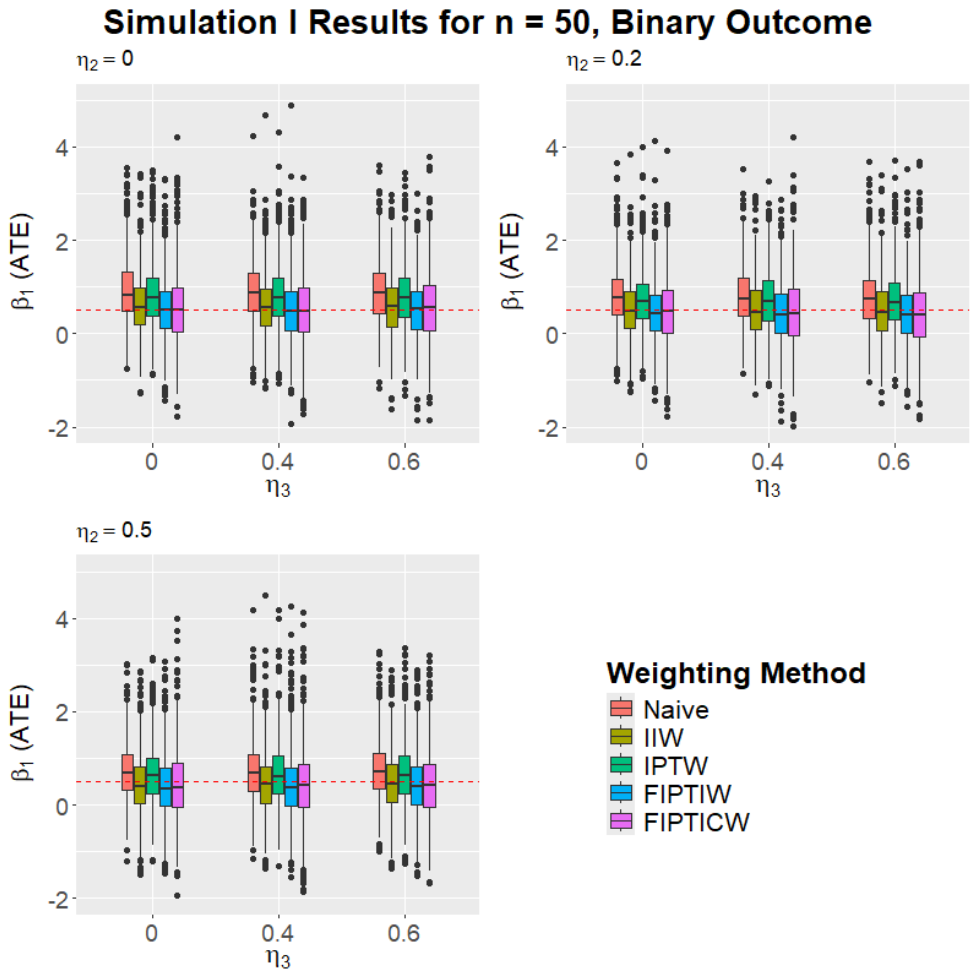


Figure A.1: Results of Simulation I for $n = 50$ with a **binary outcome**. **Extreme outliers (ATE> 5) were present and removed from the data visualization**. Average treatment effect (ATE, β_1) is calculated by fitting an independent GEE with various weights for each simulation scheme over the 1000 generated data sets. The weighting methods include no weighting (unweighted), inverse intensity weighting (IIW), inverse probability of treatment weighting (IPTW), flexible inverse probability of treatment and intensity weighting (FIPTIW) and flexible inverse probability of treatment weighting with inverse probability of censoring weights included (FIPTICW). The true value of the ATE is 0.5, and is denoted by the red horizontal line.

		Unweighted			IIW			IPTW		
η_2	η_3	Bias	AvgSE	Coverage	Bias	AvgSE	Coverage	Bias	AvgSE	Coverage
0.0	0.0	0.423	0.559	0.844	0.114	0.563	0.907	0.327	0.557	0.862
	0.4	0.404	0.554	0.854	0.092	0.556	0.905	0.304	0.550	0.865
	0.6	0.431	0.549	0.855	0.130	0.552	0.909	0.347	0.545	0.871
0.2	0.0	0.295	0.534	0.899	-0.002	0.541	0.927	0.229	0.529	0.899
	0.4	0.309	0.539	0.870	0.013	0.550	0.912	0.234	0.539	0.878
	0.6	0.303	0.534	0.882	0.021	0.539	0.892	0.239	0.533	0.894
0.5	0.0	0.296	0.527	0.907	0.028	0.542	0.910	0.229	0.526	0.918
	0.4	0.212	0.523	0.902	-0.045	0.537	0.902	0.165	0.522	0.901
	0.6	0.236	0.527	0.880	-0.021	0.545	0.909	0.171	0.528	0.895

		FIPTIW			FIPTICW		
η_2	η_3	Bias	AvgSE	Coverage	Bias	AvgSE	Coverage
0.0	0.0	0.036	0.560	0.902	0.043	0.595	0.859
	0.4	0.013	0.552	0.894	0.033	0.585	0.852
	0.6	0.065	0.548	0.909	0.095	0.580	0.867
0.2	0.0	-0.056	0.538	0.914	-0.010	0.571	0.884
	0.4	-0.052	0.550	0.903	-0.044	0.583	0.877
	0.6	-0.032	0.539	0.879	-0.034	0.569	0.856
0.5	0.0	-0.032	0.540	0.899	-0.002	0.566	0.866
	0.4	-0.089	0.536	0.891	-0.089	0.570	0.857
	0.6	-0.079	0.544	0.887	-0.072	0.566	0.873

Table A.3: Empirical bias, 95% confidence interval coverage, and average standard error estimates for Simulation I for $n = 50$ with a **binary outcome**. Average treatment effect (ATE, β_1) is calculated by fitting an independent GEE with various weights for each simulation scheme over the 1000 generated data sets. The weighting methods include no weighting (unweighted), inverse intensity weighting (IIW), inverse probability of treatment weighting (IPTW), flexible inverse probability of treatment and intensity weighting (FIPTIW) and flexible inverse probability of treatment weighting with inverse probability of censoring weights included (FIPTICW). Standard errors are calculated using the Robins et al., 2000 type standard error estimates.

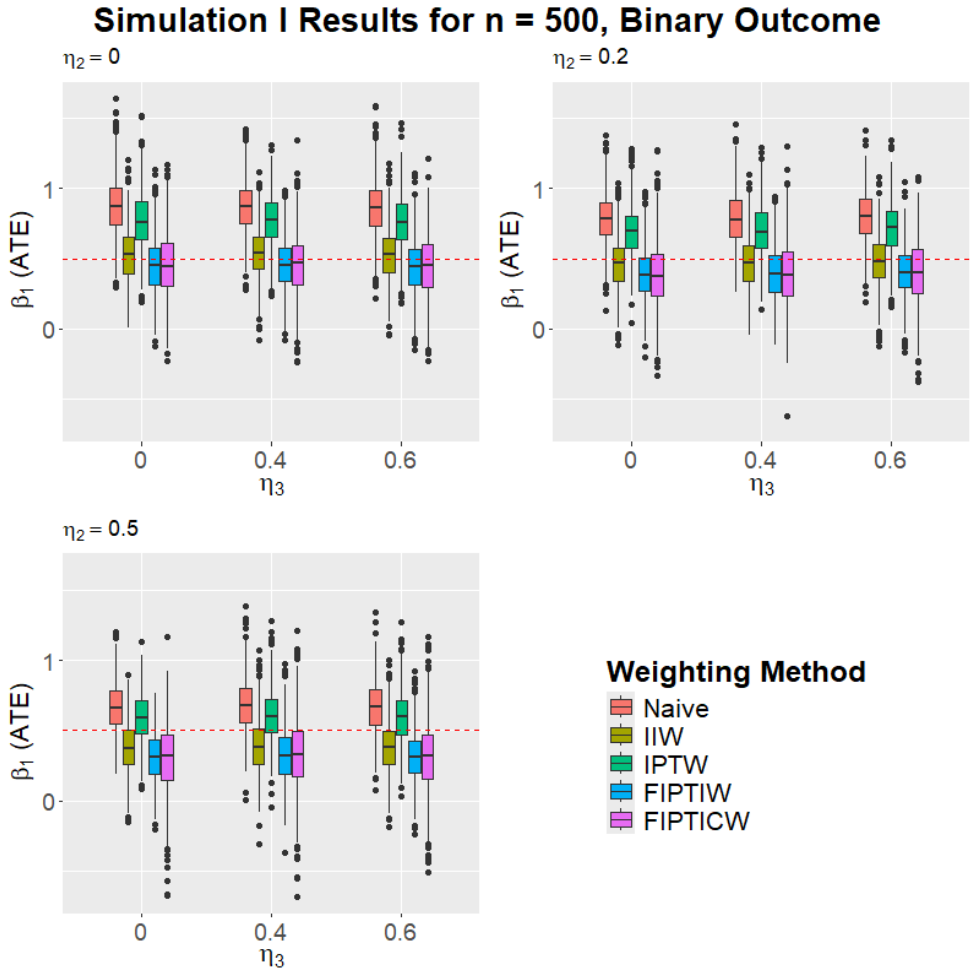


Figure A.2: Results of Simulation I for $n = 500$ with a **binary outcome**. Average treatment effect (ATE, β_1) is calculated by fitting an independent GEE with various weights for each simulation scheme over the 1000 generated data sets. The weighting methods include no weighting (unweighted), inverse intensity weighting (IIW), inverse probability of treatment weighting (IPTW), flexible inverse probability of treatment and intensity weighting (FIPTIW) and flexible inverse probability of treatment weighting with inverse probability of censoring weights included (FIPTICW). The true value of the ATE is 0.5, and is denoted by the red horizontal line.

		Unweighted			IIW			IPTW		
η_2	η_3	Bias	AvgSE	Coverage	Bias	AvgSE	Coverage	Bias	AvgSE	Coverage
0.0	0.0	0.371	0.191	0.516	0.033	0.183	0.929	0.270	0.187	0.691
	0.4	0.376	0.191	0.498	0.039	0.184	0.943	0.275	0.187	0.687
	0.6	0.364	0.192	0.517	0.028	0.184	0.942	0.262	0.188	0.729
0.2	0.0	0.285	0.184	0.676	-0.037	0.180	0.928	0.196	0.180	0.827
	0.4	0.289	0.185	0.674	-0.031	0.182	0.936	0.199	0.181	0.802
	0.6	0.303	0.184	0.625	-0.019	0.182	0.951	0.216	0.181	0.789
0.5	0.0	0.163	0.176	0.861	-0.126	0.179	0.888	0.092	0.174	0.927
	0.4	0.179	0.177	0.839	-0.113	0.180	0.881	0.108	0.174	0.899
	0.6	0.168	0.175	0.855	-0.119	0.178	0.889	0.095	0.173	0.917

		FIPTIW			FIPTICW		
η_2	η_3	Bias	AvgSE	Coverage	Bias	AvgSE	Coverage
0.0	0.0	-0.051	0.182	0.927	-0.044	0.212	0.935
	0.4	-0.043	0.183	0.946	-0.040	0.215	0.932
	0.6	-0.056	0.183	0.933	-0.055	0.213	0.934
0.2	0.0	-0.113	0.179	0.888	-0.116	0.218	0.898
	0.4	-0.106	0.181	0.892	-0.107	0.222	0.909
	0.6	-0.092	0.181	0.928	-0.097	0.223	0.925
0.5	0.0	-0.190	0.178	0.810	-0.192	0.232	0.856
	0.4	-0.176	0.178	0.811	-0.172	0.231	0.867
	0.6	-0.184	0.177	0.811	-0.183	0.227	0.853

Table A.4: Empirical bias, 95% confidence interval coverage, and average standard error estimates for Simulation I for $n = 500$ with a **binary outcome**. Average treatment effect (ATE, β_1) is calculated by fitting an independent GEE with various weights for each simulation scheme over the 1000 generated data sets. The weighting methods include no weighting (unweighted), inverse intensity weighting (IIW), inverse probability of treatment weighting (IPTW), flexible inverse probability of treatment and intensity weighting (FIPTIW) and flexible inverse probability of treatment weighting with inverse probability of censoring weights included (FIPTICW). Standard errors are calculated using the Robins et al., 2000 type standard error estimates.

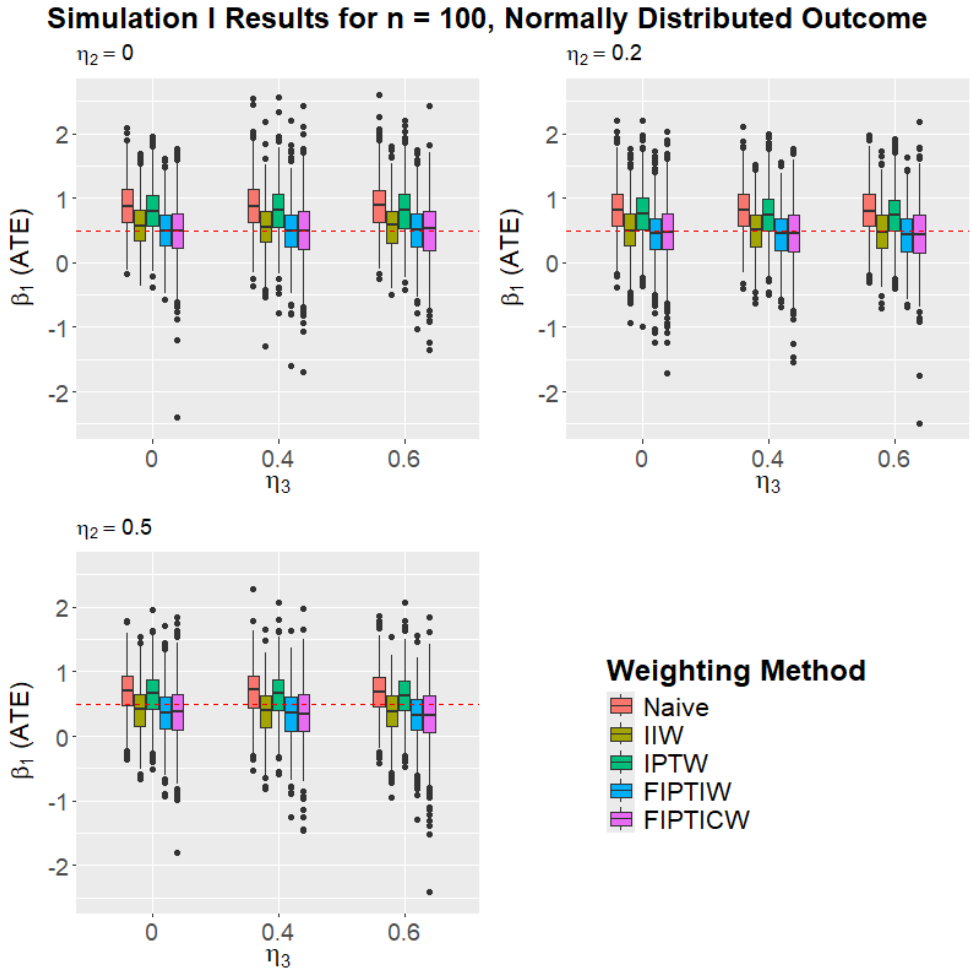


Figure A.3: Results of Simulation I for $n = 100$ with a **continuous** outcome. Average treatment effect (ATE, β_1) is calculated by fitting an independent GEE with various weights for each simulation scheme over the 1000 generated data sets. The weighting methods include no weighting (unweighted), inverse intensity weighting (IIW), inverse probability of treatment weighting (IPTW), flexible inverse probability of treatment and intensity weighting (FIPTIW) and flexible inverse probability of treatment weighting with inverse probability of censoring weights included (FIPTICW). The true value of the ATE is 0.5, and is denoted by the red horizontal line.

		Unweighted			IIW			IPTW		
η_2	η_3	Bias	AvgSE	Coverage	Bias	AvgSE	Coverage	Bias	AvgSE	Coverage
0.0	0.0	0.387	0.341	0.771	0.068	0.334	0.916	0.309	0.350	0.818
	0.4	0.386	0.344	0.781	0.056	0.333	0.916	0.305	0.352	0.830
	0.6	0.382	0.341	0.768	0.056	0.332	0.910	0.309	0.350	0.819
0.2	0.0	0.321	0.337	0.806	0.008	0.334	0.914	0.251	0.344	0.857
	0.4	0.311	0.335	0.820	-0.007	0.335	0.920	0.246	0.345	0.855
	0.6	0.303	0.340	0.828	-0.016	0.336	0.918	0.237	0.348	0.848
0.5	0.0	0.201	0.326	0.870	-0.099	0.336	0.914	0.151	0.334	0.905
	0.4	0.191	0.329	0.888	-0.111	0.337	0.911	0.140	0.339	0.910
	0.6	0.180	0.331	0.891	-0.122	0.341	0.925	0.122	0.339	0.899

		FIPTIW			FIPTICW		
η_2	η_3	Bias	AvgSE	Coverage	Bias	AvgSE	Coverage
0.0	0.0	0.002	0.345	0.918	-0.010	0.380	0.892
	0.4	-0.011	0.346	0.913	-0.012	0.372	0.898
	0.6	-0.007	0.344	0.913	-0.007	0.374	0.884
0.2	0.0	-0.051	0.345	0.928	-0.037	0.378	0.896
	0.4	-0.064	0.347	0.919	-0.054	0.380	0.888
	0.6	-0.074	0.348	0.916	-0.063	0.372	0.884
0.5	0.0	-0.146	0.346	0.909	-0.135	0.373	0.891
	0.4	-0.160	0.348	0.907	-0.158	0.380	0.870
	0.6	-0.177	0.351	0.906	-0.182	0.375	0.871

Table A.5: Empirical bias, 95% confidence interval coverage, and average standard error estimates for Simulation I for $n = 100$ with a **continuous** outcome. Average treatment effect (ATE, β_1) is calculated by fitting an independent GEE with various weights for each simulation scheme over the 1000 generated data sets. The weighting methods include no weighting (unweighted), inverse intensity weighting (IIW), inverse probability of treatment weighting (IPTW), flexible inverse probability of treatment and intensity weighting (FIPTIW) and flexible inverse probability of treatment weighting with inverse probability of censoring weights included (FIPTICW). Standard errors are calculated using the Robins et al., 2000 type standard error estimates.

A.2.4 Additional Results for Simulation II

Simulation II was performed using sample sizes of $n = 50$ and $n = 500$. The results are shown in Tables A.6 and A.7, respectively.

We also performed Simulation II with a continuous outcome, simulated as described in Appendix A.2. The results of this simulation are shown in Table A.8.

The same trends are observed as described in Section 3.4.3. In general, we have larger variances and better coverage as the sample sizes increase. The same trends are observed for the continuous and binary outcome simulations for $n = 100$.

			Variables used to estimate intensity							
γ_2	β_2		**Naive**	D	$G(t)$	Z	$D, G(t)$	D, Z	$G(t), Z$	$D, G(t), Z$
0	0									
		Bias:	-0.184	-0.184	-0.168	-0.008	-0.176	0.038	-0.008	0.038
		Variance:	0.384	0.384	0.385	0.297	0.376	0.298	0.295	0.295
		MSE:	0.417	0.417	0.413	0.297	0.407	0.299	0.295	0.297
		Avg SE:	0.547	0.547	0.553	0.515	0.547	0.518	0.515	0.518
		Coverage:	0.907	0.907	0.907	0.940	0.909	0.934	0.939	0.935
	2									
		Bias:	-0.197	-0.197	-0.171	-0.027	-0.184	0.018	-0.028	0.018
		Variance:	0.362	0.362	0.371	0.285	0.356	0.287	0.284	0.287
		MSE:	0.401	0.401	0.399	0.285	0.389	0.287	0.284	0.287
		Avg SE:	0.552	0.552	0.559	0.520	0.552	0.523	0.520	0.523
		Coverage:	0.900	0.900	0.906	0.942	0.907	0.944	0.942	0.942
0.3	0									
		Bias:	-0.223	-0.223	-0.206	-0.050	-0.213	-0.001	-0.050	-0.002
		Variance:	0.360	0.360	0.355	0.289	0.348	0.292	0.286	0.289
		MSE:	0.409	0.409	0.398	0.292	0.393	0.291	0.288	0.289
		Avg SE:	0.543	0.543	0.544	0.509	0.539	0.510	0.507	0.509
		Coverage:	0.893	0.893	0.893	0.935	0.899	0.939	0.934	0.939
	2									
		Bias:	-0.163	-0.163	-0.153	-0.004	-0.165	0.041	-0.016	0.030
		Variance:	0.417	0.417	0.413	0.332	0.396	0.334	0.322	0.324
		MSE:	0.444	0.444	0.436	0.332	0.423	0.335	0.321	0.325
		Avg SE:	0.551	0.551	0.550	0.515	0.543	0.517	0.509	0.511
		Coverage:	0.889	0.889	0.894	0.916	0.898	0.916	0.919	0.918
0.6	0									
		Bias:	-0.197	-0.197	-0.192	-0.024	-0.197	0.023	-0.032	0.013
		Variance:	0.378	0.378	0.360	0.277	0.350	0.277	0.268	0.270
		MSE:	0.416	0.416	0.397	0.278	0.388	0.277	0.269	0.270
		Avg SE:	0.542	0.542	0.537	0.506	0.532	0.506	0.499	0.501
		Coverage:	0.904	0.904	0.904	0.944	0.906	0.946	0.942	0.946
	2									
		Bias:	-0.190	-0.190	-0.197	-0.023	-0.207	0.023	-0.045	0.000
		Variance:	0.395	0.395	0.363	0.311	0.352	0.315	0.292	0.297
		MSE:	0.430	0.430	0.402	0.311	0.394	0.315	0.293	0.296
		Avg SE:	0.556	0.556	0.545	0.519	0.538	0.519	0.504	0.506
		Coverage:	0.891	0.891	0.894	0.923	0.897	0.926	0.924	0.927

Table A.6: Simulation results for Simulation II for $n = 50$ for the **binary outcome** case. Bias, mean squared error (MSE), average standard error (Avg SE), and coverage of the 95% confidence interval of the average treatment effect (ATE) is calculated by weighting the outcome model in Equation (A.2) by inverse intensity weighting (IIW) for each simulation scheme over the 1000 generated data sets. Variables included in the IIW model are listed in the table. The true value of the ATE is 0.5.

			Variables used to estimate intensity							
γ_2	β_2		**Naive**	D	$G(t)$	Z	$D, G(t)$	D, Z	$G(t), Z$	$D, G(t), Z$
0	0									
		Bias:	-0.252	-0.252	-0.249	-0.055	-0.251	-0.003	-0.055	-0.004
		Variance:	0.039	0.039	0.040	0.030	0.039	0.031	0.030	0.031
		MSE:	0.103	0.103	0.102	0.033	0.102	0.031	0.033	0.031
		Avg SE:	0.184	0.184	0.185	0.163	0.184	0.164	0.163	0.164
		Coverage:	0.709	0.709	0.713	0.922	0.712	0.932	0.920	0.931
	2									
		Bias:	-0.232	-0.232	-0.229	-0.048	-0.231	0.001	-0.048	0.001
		Variance:	0.034	0.034	0.035	0.027	0.034	0.028	0.027	0.028
		MSE:	0.088	0.088	0.087	0.030	0.087	0.028	0.029	0.028
		Avg SE:	0.186	0.186	0.187	0.166	0.186	0.166	0.166	0.166
		Coverage:	0.756	0.756	0.759	0.936	0.757	0.955	0.936	0.954
0.3	0									
		Bias:	-0.249	-0.249	-0.250	-0.056	-0.251	-0.004	-0.059	-0.007
		Variance:	0.033	0.033	0.032	0.026	0.031	0.026	0.025	0.026
		MSE:	0.095	0.095	0.094	0.029	0.094	0.026	0.029	0.026
		Avg SE:	0.184	0.184	0.182	0.162	0.180	0.162	0.161	0.161
		Coverage:	0.726	0.726	0.721	0.942	0.722	0.950	0.936	0.955
	2									
		Bias:	-0.227	-0.227	-0.230	-0.045	-0.231	0.004	-0.051	-0.002
		Variance:	0.035	0.035	0.033	0.026	0.032	0.026	0.025	0.025
		MSE:	0.086	0.086	0.086	0.028	0.086	0.026	0.028	0.025
		Avg SE:	0.189	0.189	0.184	0.166	0.182	0.166	0.162	0.163
		Coverage:	0.777	0.777	0.754	0.952	0.751	0.954	0.946	0.949
0.6	0									
		Bias:	-0.251	-0.251	-0.253	-0.054	-0.255	-0.002	-0.059	-0.007
		Variance:	0.033	0.033	0.030	0.024	0.030	0.024	0.023	0.023
		MSE:	0.096	0.096	0.095	0.027	0.095	0.024	0.027	0.023
		Avg SE:	0.188	0.188	0.181	0.163	0.179	0.162	0.159	0.159
		Coverage:	0.739	0.739	0.709	0.960	0.705	0.962	0.956	0.964
	2									
		Bias:	-0.209	-0.209	-0.221	-0.028	-0.223	0.021	-0.043	0.005
		Variance:	0.039	0.039	0.033	0.027	0.032	0.027	0.025	0.025
		MSE:	0.082	0.082	0.082	0.028	0.082	0.028	0.027	0.025
		Avg SE:	0.194	0.194	0.181	0.168	0.180	0.167	0.160	0.160
		Coverage:	0.812	0.812	0.756	0.950	0.756	0.955	0.948	0.958

Table A.7: Simulation results for Simulation II for $n = 500$ for the **binary outcome** base. Bias, mean squared error (MSE), average standard error (Avg SE), and coverage of the 95% confidence interval of the average treatment effect (ATE) is calculated by weighting the outcome model in Equation (3.7) by inverse intensity weighting (IIW) for each simulation scheme over the 1000 generated data sets. Variables included in the IIW model are listed in the table. The true value of the ATE is 0.5.

			Variables used to estimate intensity							
γ_2	β_2		**Naive**	D	$G(t)$	Z	$D, G(t)$	D, Z	$G(t), Z$	$D, G(t), Z$
0	0									
		Bias:	-0.269	-0.269	-0.262	-0.041	-0.263	0.016	-0.041	0.015
		Variance:	0.157	0.157	0.157	0.120	0.156	0.121	0.120	0.121
		MSE:	0.229	0.229	0.225	0.121	0.225	0.121	0.121	0.121
		Avg SE:	0.376	0.376	0.375	0.342	0.375	0.343	0.342	0.343
		Coverage:	0.878	0.878	0.877	0.938	0.878	0.941	0.938	0.943
0	2									
		Bias:	-0.269	-0.269	-0.256	-0.047	-0.260	0.008	-0.048	0.008
		Variance:	0.206	0.206	0.204	0.156	0.203	0.156	0.155	0.156
		MSE:	0.278	0.278	0.269	0.158	0.270	0.156	0.157	0.156
		Avg SE:	0.407	0.407	0.406	0.371	0.406	0.372	0.371	0.372
		Coverage:	0.881	0.881	0.885	0.936	0.886	0.937	0.933	0.935
0.3	0									
		Bias:	-0.268	-0.268	-0.261	-0.045	-0.262	0.012	-0.045	0.012
		Variance:	0.158	0.158	0.153	0.120	0.153	0.121	0.116	0.118
		MSE:	0.230	0.230	0.221	0.122	0.221	0.121	0.118	0.118
		Avg SE:	0.375	0.375	0.371	0.339	0.371	0.339	0.336	0.336
		Coverage:	0.881	0.881	0.887	0.939	0.887	0.943	0.941	0.944
0.3	2									
		Bias:	-0.253	-0.253	-0.243	-0.037	-0.247	0.019	-0.037	0.020
		Variance:	0.191	0.191	0.177	0.156	0.178	0.160	0.150	0.154
		MSE:	0.255	0.255	0.236	0.158	0.239	0.160	0.151	0.154
		Avg SE:	0.409	0.409	0.400	0.372	0.400	0.372	0.366	0.367
		Coverage:	0.885	0.885	0.893	0.929	0.895	0.929	0.936	0.944
0.6	0									
		Bias:	-0.246	-0.246	-0.238	-0.031	-0.239	0.026	-0.030	0.026
		Variance:	0.158	0.158	0.153	0.121	0.152	0.122	0.117	0.117
		MSE:	0.219	0.219	0.210	0.122	0.209	0.123	0.117	0.118
		Avg SE:	0.378	0.378	0.368	0.344	0.368	0.344	0.335	0.335
		Coverage:	0.875	0.875	0.883	0.941	0.885	0.946	0.936	0.944
0.6	2									
		Bias:	-0.257	-0.257	-0.250	-0.038	-0.254	0.018	-0.045	0.012
		Variance:	0.193	0.193	0.172	0.141	0.173	0.143	0.132	0.133
		MSE:	0.259	0.259	0.235	0.143	0.237	0.143	0.134	0.133
		Avg SE:	0.417	0.417	0.399	0.377	0.399	0.377	0.363	0.364
		Coverage:	0.907	0.907	0.894	0.955	0.898	0.955	0.948	0.952

Table A.8: Simulation results for Simulation II for $n = 100$ for the **normally distributed outcome** case. Bias, mean squared error (MSE), average standard error (Avg SE), and coverage of the 95% confidence interval of the average treatment effect (ATE) is calculated by weighting the outcome model in Equation (A.2) by inverse intensity weighting (IIW) for each simulation scheme over the 1000 generated data sets. Variables included in the IIW model are listed in the table. The true value of the ATE is 0.5.

A.2.5 Additional Results for Simulation III

Simulation III was performed again using a continuous outcome, simulated as described in Appendix A.2.1.

Degree of Informativeness		**Mean Proportion of IPTW Weights**			**Mean Proportion of IIW Weights**			**Mean Proportion of FIPTIW Weights**		
Treatment Assignment Process	Observation Process	>5	>10	>20	>5	>10	>20	>5	>10	>20
Low	Low	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Moderate	Low	0.61	0.14	0.03	0.00	0.00	0.00	0.02	0.00	0.00
High	Low	0.49	0.16	0.05	0.00	0.00	0.00	0.03	0.01	0.00
Low	Moderate	0.00	0.00	0.00	0.34	0.00	0.00	0.57	0.01	0.00
Low	High	0.00	0.00	0.00	4.44	0.16	0.00	5.18	0.25	0.01
Moderate	Moderate	0.08	0.01	0.00	0.44	0.00	0.00	1.41	0.09	0.01

Table A.9: Distribution of weights under each simulation scheme for Simulation III when a **continuous, normally-distributed outcome** is simulated.

When the time-varying intercept is known and is treated as an offset term in the outcome model, weight trimming does slightly reduce the bias in some cases. However, the difference in bias, MSE, and coverage is unsubstantial when compared to the case where no trimming is performed.

The results again show sensitivity to extreme weights when a cubic spline is used, which is not alleviated by weight trimming. In many settings, large biases are present both with and without weight trimming. These results are similar to what was seen in Section 3.4.4 with a binary outcome simulated.

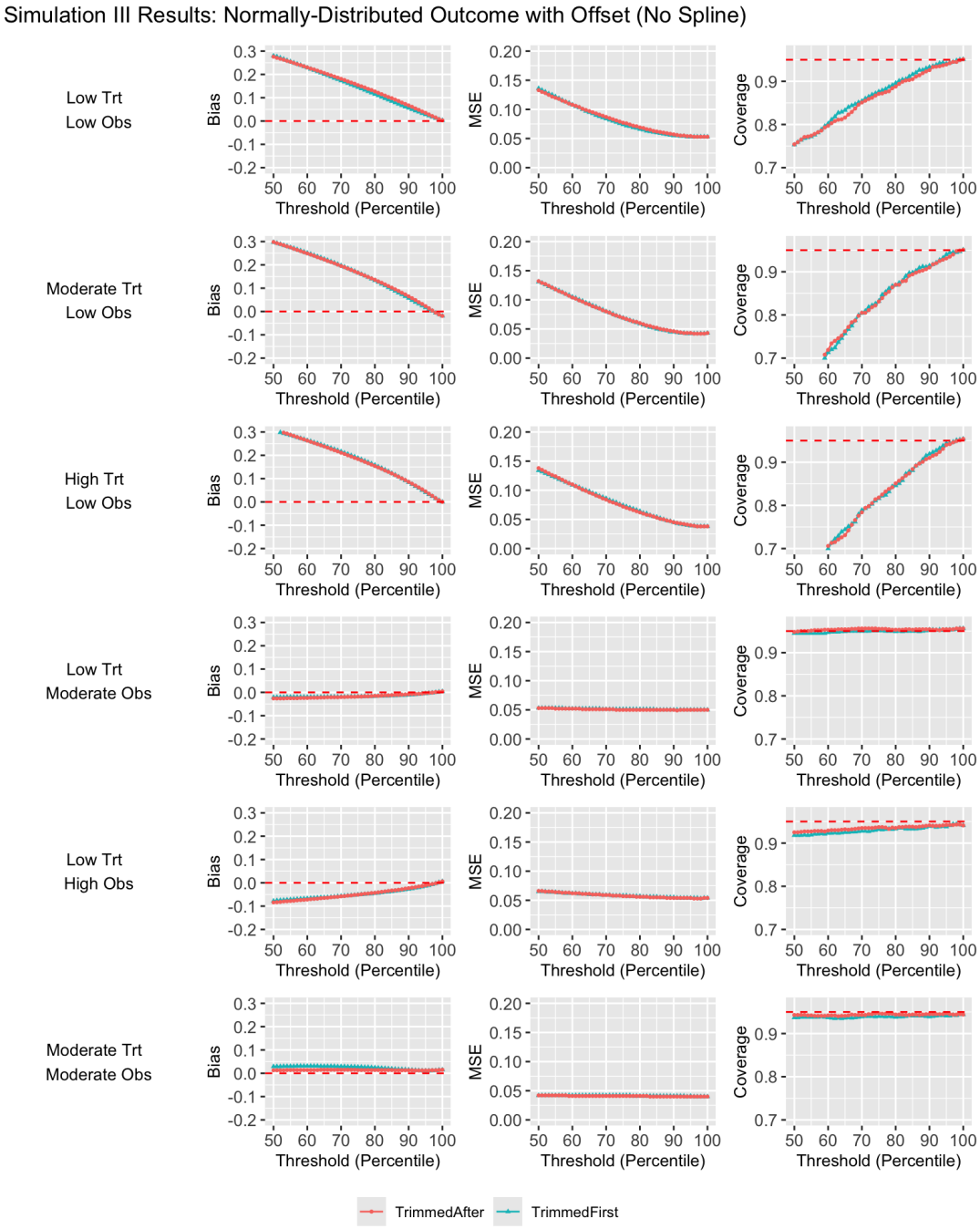


Figure A.4: Results of the simulation under various degrees of informativeness (low, moderate, and high) in the treatment assignment ("Trt") and observation ("Obs") processes when the time-varying intercept is known. "Trimmed before" refers to trimming the individual IPTW and IIW weights prior to multiplying and calculating the FIPTIW weights. "Trimmed after" refers to trimming the FIPTIW weights after multiplying the IPTW and IIW weights together. The true average treatment effect β_1 is simulated to be 0.5. The parametric form of the time-varying intercept is known, and included in the model as an offset.

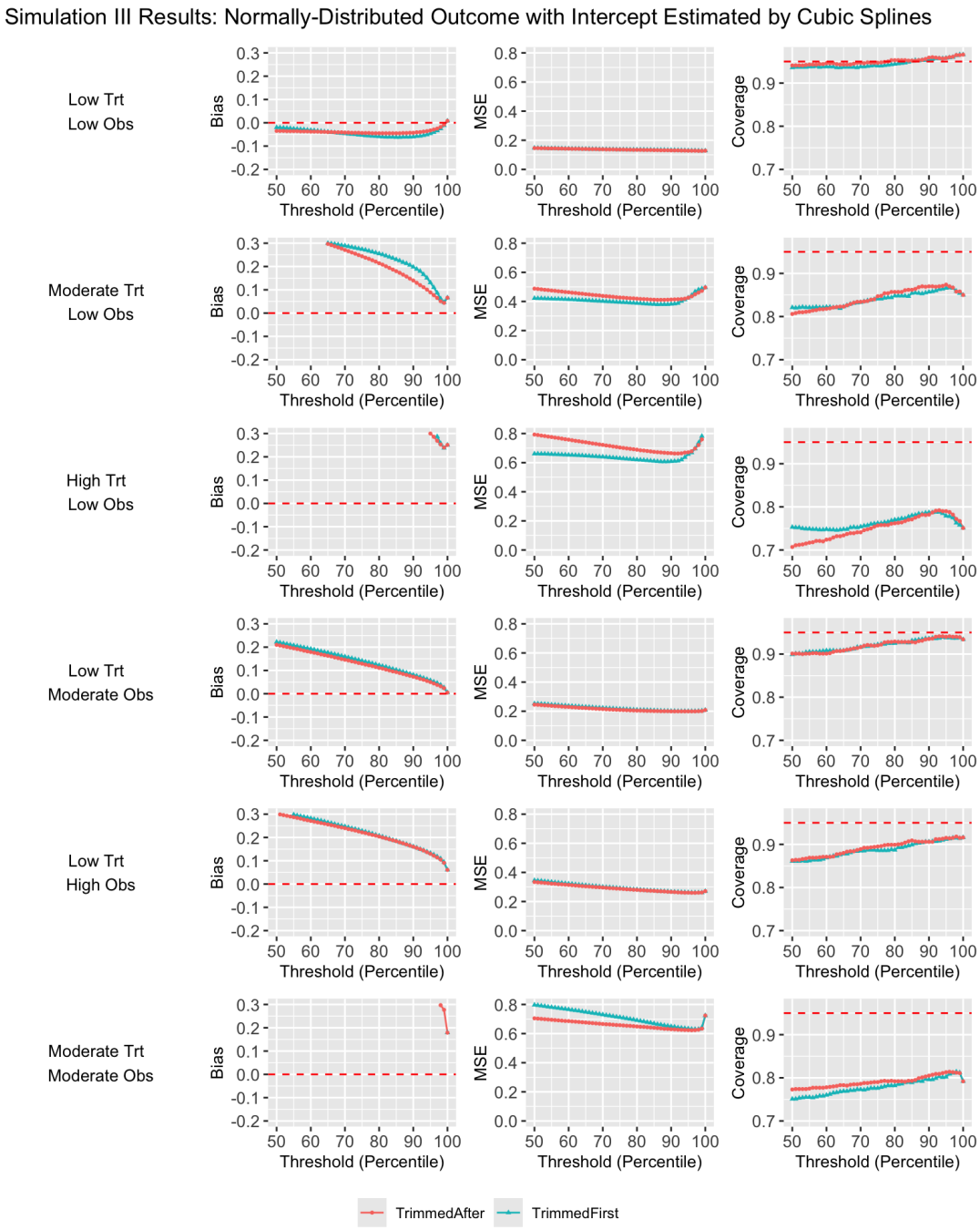


Figure A.5: Results of the simulation under various degrees of informativeness (low, moderate, and high) in the treatment assignment ("Trt") and observation ("Obs") processes when the time-varying intercept is known. "Trimmed before" refers to trimming the individual IPTW and IIW weights prior to multiplying and calculating the FIPTIW weights. "Trimmed after" refers to trimming the FIPTIW weights after multiplying the IPTW and IIW weights together. The true average treatment effect β_1 is simulated to be 0.5. The parametric form of the time-varying intercept is unknown, and estimated using cubic splines.

A.3 Sensitivity Analyses for the Malaria Data Set

A.3.1 Ignoring Censoring

The analysis in Section 3.5 was repeated without the use of artificial censoring. The same 287 individuals were included in the study, but we considered all observations in the data set. In this data set, 69.34% of individuals were censored for nonignorable reasons (i.e., reasons potentially related to the longitudinal outcome). The number of observations per individual ranged from one to 124, with a mean number of observations of 33.19.

The variables included in the propensity score model were the same as in Section 3.5. The maximum IPTW weight was 3.23. The same variables were included in the intensity model as in Section 3.5, and the maximum IIW weight was 11.16. 2.78% of weights were above 5, and 0.00% of weights were above 10. After multiplication, the maximum FIPTIW weight was 13.20. 6.58% of FIPTIW weights were above 5, and 0.34% were above 10. When employing weight trimming, any weight above 5.37 was trimmed.

Weighting Method	β_1	**SE**(β_1)	**95% CI for** β_1	**Odds Ratio (OR)**	**95% CI for OR**
None	0.730	0.061	(0.611, 0.848)	2.075	(1.843, 2.336)
IPTW	0.619	0.139	(0.346, 0.892)	1.857	(1.413, 2.439)
IIW	0.911	0.137	(0.644, 1.179)	2.487	(1.903, 3.252)
FIPTIW	0.781	0.140	(0.507, 1.054)	2.184	(1.660, 2.869)
FIPTIW (Trimmed)	0.792	0.139	(0.520, 1.065)	2.208	(1.682, 2.900)

Table A.10: Results of the estimation of the odds ratio (OR) of malaria diagnoses for those residing in households with unprotected water sources versus those residing in households with protected water sources (β_1) by different weighting methods for an independent GEE in the PRISM cohort. Sample includes those who were censored, and likely violates the noninformative censoring violation.

The results are given in Table A.10. The untrimmed FIPTIW model estimates a 2.18 times increase in the odds of malaria diagnosis for those residing in households with unprotected versus protected water sources. The result is similar, but slightly larger, when weight trimming is employed. When comparing to the analysis in Section 3.5, the estimates are much larger. We again see sensitivity to violations of the noninformative censoring assumption, as in Simulation II in Section A.2.4. As such, analysts must be cautious when individuals who were censored for reasons that may be related to the longitudinal outcome are included in the analysis.

A.3.2 Ignoring Clustering

The analysis performed in Section 3.5 was repeated with the inclusion of all individuals in each family. This resulted in a sample size of 725 children in 287 unique households. Only 1.5% of individuals were censored for reasons potentially related to the longitudinal outcome when artificial censoring at six months was used. The baseline demographics for this sample are given in Table A.11.

The variables included in the propensity score model were the same as those described in Section 3.5. The maximum IPTW weight was 3.86. The same variables were included in the intensity model as in Section 3.5, and the maximum IIW weight was 4.15. After multiplying the weights together, the maximum FIPTIW weight was 7.07. 0.33% of FIPTIW weights were above 5. When employing weight trimming, any FIPTIW weight above 2.622 was trimmed.

Covariate	Mean (SD) or n (%)	Covariate	Mean (SD) or n (%)
Age at Enrollment	5.56 (2.13)	**Food Problems per Week**	
Sex		Sometimes	262 (36.14%)
Female	358 (49.38%)	Never	115 (15.86%)
Male	367 (50.62%)	Often	100 (13.79%)
Sub-County		Seldom	114 (15.72%)
Walukuba	198 (27.31%)	Always	134 (18.48%)
Kihihi	272 (37.52%)	**Waste Facilities**	
Nagongera	255 (35.17%)	Covered pit latrine, no slab	244 (33.66%)
Household Wealth Index		Covered pit latrine with slab	44 (6.07%)
Least poor	235 (32.41%)	Composting toilet	9 (1.24%)
Poorest	257 (35.45%)	Uncovered pit latrine, no slab	298 (41.1%)
Middle	233 (32.14%)	Flush toilet	21 (2.9%)
Drinking Water Source		Uncovered pit latrine with slab	15 (2.07%)
Public tap	249 (34.34%)	Vip latrine	8 (1.1%)
Protected public well	63 (8.69%)	No facility	86 (11.86%)
River/stream	68 (9.38%)	**Number of Persons in House**	6.75 (2.93)
Protected spring	61 (8.41%)	**Reason for Withdrawal**	
Borehole	169 (23.31%)	Moved out of area	120 (16.55%)
Open public well	61 (8.41%)	Reached 11 years of age	227 (31.31%)
Pond/lake	15 (2.07%)	Withdrawal of informed consent	9 (1.24%)
Unprotected spring	39 (5.38%)	Inability to comply	17 (2.34%)
Unprotected Water Source		Unable to be located for 120 days	37 (5.1%)
No	542 (74.76%)	Completed study	90 (12.41%)
Yes	183 (25.24%)	Unknown	225 (31.03%)
Dwelling Type			
Modern	206 (28.41%)		
Traditional	519 (71.59%)		

Table A.11: Baseline demographics for the 725 children when we allow multiple children from the same household to be included in the analysis.

Weighting Method	β_1	SE(β_1)	95% CI for β_1	Odds Ratio (OR)	95% CI for OR
None	0.456	0.110	(0.240, 0.673)	1.578	(1.271, 1.959)
IPTW	0.333	0.127	(0.084, 0.583)	1.395	(1.087, 1.791)
IIW	0.498	0.137	(0.231, 0.766)	1.645	(1.259, 2.151)
FIPTIW	0.371	0.161	(0.056, 0.686)	1.449	(1.058, 1.986)
FIPTIW (Trimmed)	0.400	0.134	(0.137, 0.663)	1.492	(1.147, 1.94)

Table A.12: Results of the estimation of the odds ratio (OR) of malaria diagnoses for those residing in households with unprotected water sources versus those residing in households with protected water sources (β_1) by different weighting methods for an independent GEE in the PRISM cohort. Sample includes those who resided in the same household, and likely violates the independence assumption.

The results are given in Table A.12. The untrimmed FIPTIW model estimates the odds of malaria diagnosis for those residing in households with unprotected versus protected water sources is 1.45 times greater. The trimmed FIPTIW OR estimate is similar in magnitude (OR = 1.45). The estimated effect sizes are slightly smaller than in the analysis presented in Section 3.5. There appears to be some sensitivity to violations of the independence assumption and as such, we identify extending existing methodology to account for the correlations between individuals as an area of future work.

A.3.3 Employing Random Censoring for Individuals Censored Prior to Six Months

The analysis performed in Section 3.5 was repeated where individuals who were censored prior to six months were randomly censored between zero and six months, instead of at six months exactly.

The same 287 children were included as in the original analysis. However, the distribution of observation times varied slightly due to the introduction of random sampling for a small sample of individuals who were likely censored prior to six months. The number of observations per individual ranged from one to 11, with an average of 4.78 clinic visits. Of the 1373 observations, 287 were for enrollment, 516 were scheduled, and 570 were unscheduled. Malaria was diagnosed in 13.03% of observations. 38.68% of patients had a Malaria diagnosis at some point.

The same covariates in the propensity score (IPTW) and observation intensity (IIW) models were used in this analysis as described in Section 3.5. No IPTW, IIW, or FIPTIW weights were above 5. When employing trimming, any FIPTIW weight above 2.17 was trimmed. Table A.13 show the results of the analysis.

Weighting Method	β_1	**SE**(β_1)	**95% CI for** β_1	**Odds Ratio (OR)**	**95% CI for OR**
None	0.423	0.186	(0.059, 0.786)	1.527	(1.060, 2.196)
IPTW	0.323	0.230	(-0.128, 0.774)	1.381	(0.880, 2.168)
IIW	0.558	0.222	(0.123, 0.993)	1.747	(1.131, 2.699)
FIPTIW	0.434	0.223	(-0.004, 0.872)	1.543	(0.996, 2.391)
FIPTIW (Trimmed)	0.444	0.219	(0.014, 0.874)	1.559	(1.014, 2.397)

Table A.13: Results of the estimation of the odds ratio (OR) of malaria diagnoses for those residing in households with unprotected water sources versus those residing in households with protected water sources (β_1) by different weighting methods for an independent GEE in the PRISM cohort. In this analysis, those who were assumed to have been censored prior to six months were randomly censored between zero and 182.5 days.

The results of this analysis show are similar to those shown in Table 3.6, showing very little sensitivity to including a small proportion of individuals who dropped out of the study (potentially for reasons related to the longitudinal outcome) before the study end date.

Appendix B

Additional Results for Chapter 4

B.1 Additional Results for Simulation I

B.1.1 Additional Results for Simulation I, Binary Outcome

In this section, we present the results for Simulation I where a binary outcome is simulated and extreme estimates are not removed.

When IIW is not employed and an independent GEE is fitted, the estimated ATE is biased. When the observation process covariates $\boldsymbol{V}(t)$ are known at all times, the we see unbiased results, and close to 95% coverage.

When the covariates $\boldsymbol{V}(t)$ are only observed at the times at which Y is observed, none of the existing methods are unbiased. High MSE, and extreme undercoverage of the 95% confidence interval are present.

When the missingness mechanism is MCAR and the missingness is 25%, CCA, SI-PPM, and MICE-PPM are unbiased. At 50% missingness, SI-PPM produces some extreme estimates which cause the bias and MSE to be close to infinity. However, CCA and MICE-PPM remain unbiased. CCA and MICE-PPM produce similar MSE and coverage. At 90% missingness, CCA, SI-PPM, and missForest produce extreme estimates. MICE-PPM remains unbiased. LOCF and missForest are biased in all cases, with missForest producing extreme estimates of the ATE at 90% missingness.

When the data are MAR, again LOCF and missForest are biased in all settings. CCA, SI-PPM, and MICE-PPM are unbiased at all degrees of missingness. The three methods

Simulation I (Binary Outcome): Bias

Missingness Mechanism	p_{miss}	CCA	LOCF	SI-PMM	MICE-PMM	missForest
No IIW/Naive	-	0.341	-	-	-	-
No Missingness	0	-0.010	-	-	-	-
Observation Times Only	-	0.355	0.343	0.342	0.340	0.331
MCAR	0.25	-0.004	0.072	-0.010	-0.013	-0.085
	0.50	-0.001	0.156	-inf	-0.011	-0.192
	0.90	inf	0.303	inf	-0.008	-inf
MAR	0.25	-0.009	0.040	-0.011	-0.012	-0.053
	0.50	-0.006	0.110	-0.015	-0.015	-0.125
	0.90	0.007	0.281	-0.008	-0.011	-inf
MNAR	0.25	0.008	0.082	-0.010	-0.012	-0.087
	0.50	inf	0.173	inf	-0.013	-inf
	0.90	inf	0.309	inf	-inf	-inf

Table B.1: Empirical bias of β_1 calculated across the 1000 simulation runs for each missing data method and missingness scheme when a binary outcome is simulated in Simulation I. $\widehat{\beta}_1$ was estimated by Equation (4.1) using an independent GEE weighted by IIW (unless otherwise stated) where the missing observation model covariates were imputed using the indicated method. The true value of β_1 is 0.5.

have similar MSE in all cases. MICE-PPM tends to have the best coverage, aside form when the missingness is 90% where CCA has the best coverage. At 90% missingness, missForest also produces extreme estimates.

When the data are MNAR, CCA, SI-PPM, and MICE-PPM are unbiased at 25% missingness. At 50% missingness, CCA, SI-PPM, and missForest produce extreme estimates of the ATEs. MICE-PPM is the only unbiased method. At 90% missingness, only LOCF produced stable estimates of the ATE, however the results were biased.

Overall, CCA and MICE-PPM tend to provide unbiased estimates of β_1, aside from when $V(t)$ is only known when $Y(t)$ is. In this setting, none of the existing methods provided unbiased estimation of β_1.

Simulation I (Binary Outcome): MSE

Missingness Mechanism	p_{miss}	CCA	LOCF	SI-PMM	MICE-PMM	missForest
No IIW/Naive	-	0.202	-	-	-	-
No Missingness	0	0.103	-	-	-	-
Observation Times Only	-	0.230	0.208	0.208	0.205	0.285
MCAR	0.25	0.103	0.100	0.106	0.104	0.121
	0.50	0.105	0.114	inf	0.104	0.177
	0.90	inf	0.179	inf	0.123	inf
MAR	0.25	0.103	0.099	0.103	0.102	0.112
	0.50	0.103	0.105	0.106	0.105	0.140
	0.90	0.118	0.166	0.129	0.119	inf
MNAR	0.25	0.116	0.101	0.106	0.103	0.133
	0.50	inf	0.120	inf	0.112	inf
	0.90	inf	0.182	inf	inf	inf

Table B.2: MSE of β_1 calculated across the 1000 simulation runs for each missing data method and missingness scheme when a binary outcome is simulated in Simulation I. $\widehat{\beta_1}$ was estimated by Equation (4.1) using an independent GEE weighted by IIW (unless otherwise stated) where the missing observation model covariates were imputed using the indicated method. The true value of β_1 is 0.5.

B.1.2 Additional Results for Simulation I: Continuous Outcome

We also consider the setting where a continuous outcome is generated. In this setting, the marginal outcome model of interest is

$$E\{Y_i(t)|D_i\} = \mu(t) + \beta_1 D_i \tag{B.1}$$

where β_1 is the ATE. We let $\mu(t) = (2-t)$ and let $D_i = 1$ indicate that individual i is in the treatment group and $D_i = 0$ indicate that individual i is in the control for the duration of the study. The probability of being assigned to the treatment or control group is $\pi_D = 0.5$.

$V_1(t)$ to $V_4(t)$, and $R(t)$ are generated as described in Section 4.3.1.

We generate the continuous outcome as

$$\begin{aligned} Y_i(t) = \mu(t) + \beta_1 D_i + \beta_2(V_{i1}(t) - E\{V_{i1}|D_i\}) + \beta_3(V_{i2}(t) - E\{V_{i2}|D_i\}) \\ + \beta_4(V_{i3}(t) - E\{V_{i3}|D_i\}) + \beta_5(V_{i4}(t) - E\{V_{i4}|D_i\}) \end{aligned} \tag{B.2}$$

Simulation I (Binary Outcome): Coverage

Missingness Mechanism	p_{miss}	CCA	LOCF	SI-PMM	MICE-PMM	missForest
No IIW/Naive	-	0.755	-	-	-	-
No Missingness	0	0.952	-	-	-	-
Observation Times Only	-	0.764	0.773	0.755	0.752	0.772
MCAR	0.25	0.949	0.947	0.958	0.952	0.939
	0.5	0.951	0.910	0.940	0.950	0.916
	0.9	0.943	0.788	0.926	0.940	0.787
MAR	0.25	0.952	0.954	0.948	0.953	0.942
	0.50	0.954	0.936	0.944	0.949	0.933
	0.90	0.946	0.816	0.927	0.929	0.802
MNAR	0.25	0.947	0.943	0.946	0.951	0.942
	0.50	0.948	0.894	0.945	0.939	0.898
	0.90	0.913	0.788	0.912	0.925	0.800

Table B.3: 95% Coverage of β_1 calculated across the 1000 simulation runs for each missing data method and missingness scheme when a binary outcome is simulated in Simulation I. $\widehat{\beta}_1$ was estimated by Equation (4.1) using an independent GEE weighted by IIW (unless otherwise stated) where the missing observation model covariates were imputed using the indicated method. Standard errors for the 95% confidence intervals are calculated using the Robins et al., 2000 type standard error estimates. The true value of β_1 is 0.5.

where $\mu(t) = (2-t)$, $\epsilon_i(t) \sim N(0,1)$ is the random error term (for each individual and time) and $\phi_i \sim N(0, 1.25)$ is an individual random effect. As in the binary case, we let $\beta_1 = 0.5$, $\beta_2 = 1$, $\beta_3 = 0.7$, $\beta_4 = 0.5$, and $\beta_5 = 0.5$. As the outcome $Y(t)$ is continuous, modeling via a marginal model should have very little effect on the bias of the estimated model parameters $\boldsymbol{\beta}$, even though the data was generated using a subject-specific (mixed effects) model. By centering the auxiliary covariates, we allow these covariates to be related to the longitudinal outcome but omitted from the marginal model of interest

$$E\{Y_i(t)|D_i\} = (2-t) + \beta_1 D_i. \tag{B.3}$$

We consider the same model for the observation process as described in Section 4.3.1.

As observation times are simulated from an intensity model, some individuals may not have any observation times, which can reduce the overall sample size. Across the 1000

data sets that are generated, the average sample size is 92.5. Among those who had at least one observation in each data set, and the average number of events per individual is 7.97. The average minimum and maximum number of events per individual across the 1000 data sets are 1.00 and 21.99, respectively. The bias, MSE, and coverage are shown in Tables B.4, B.5, and B.6, respectively.

Simulation I (Continuous Outcome): Bias

Missingness Mechanism	p_{miss}	**CCA**	**LOCF**	**SI-PMM**	**MICE-PMM**	**missForest**
No IIW/Naive	-	0.399	-	-	-	-
No Missingness	0	-0.012	-	-	-	-
Observation Times Only	-	0.422	0.395	0.401	0.402	0.389
MCAR	0.25	-0.006	0.092	-0.015	-0.014	-0.112
	0.50	0.000	0.194	-0.015	-0.013	-0.261
	0.90	0.037	0.359	0.002	-0.003	-0.653
MAR	0.25	-0.010	0.054	-0.017	-0.015	-0.067
	0.50	-0.006	0.140	-0.017	-0.019	-0.166
	0.90	0.010	0.332	-0.022	-0.022	-0.592
MNAR	0.25	0.007	0.102	-0.010	-0.014	-0.119
	0.50	0.029	0.214	-0.016	-0.014	-0.267
	0.90	0.078	0.366	-0.007	-0.015	-0.563

Table B.4: Empirical bias of β_1 calculated across the 1000 simulation runs for each missing data method and missingness scheme. $\widehat{\beta}_1$ was estimated by Equation (B.1) using an independent GEE weighted by IIW (unless otherwise stated) where the missing observation model covariates were imputed using the indicated method. The true value of β_1 is 0.5.

When IIW is not employed and an independent GEE is fitted, the estimated ATE is biased. When the observation process covariates $\boldsymbol{V}(t)$ are known at all times, the results are unbiased, and close to 95% coverage.

When the covariates $\boldsymbol{V}(t)$ are only observed at the times at which Y is observed, none of the existing methods are unbiased. We have high MSE, and extreme undercoverage of the 95% confidence interval.

When the missingness mechanism is MCAR, CCA, SI-PPM, and MICE-PPM are unbiased while LOCF and missForest are biased up to 50% missingness. At 90% missingness, CCA is slightly biased, and SI-PPM and MICE-PPM perform similarly in terms of bias.

Simulation I (Continuous Outcome): MSE

Missingness Mechanism	p_{miss}	CCA	LOCF	SI-PMM	MICE-PMM	missForest
No IIW/Naive	-	0.188	-	-	-	-
No Missingness	0	0.042	-	-	-	-
Observation Times Only	-	0.213	0.186	0.193	0.193	0.240
MCAR	0.25	0.043	0.045	0.045	0.044	0.065
	0.50	0.044	0.071	0.051	0.047	0.144
	0.90	0.067	0.159	0.075	0.062	0.750
MAR	0.25	0.042	0.040	0.042	0.041	0.051
	0.50	0.042	0.054	0.045	0.044	0.087
	0.90	0.052	0.139	0.079	0.066	0.583
MNAR	0.25	0.050	0.047	0.046	0.044	0.072
	0.50	0.063	0.079	0.054	0.048	0.172
	0.90	0.211	0.164	0.141	0.091	1.547

Table B.5: MSE of β_1 calculated across the 1000 simulation runs for each missing data method and missingness scheme. $\widehat{\beta_1}$ was estimated by Equation (B.1) using an independent GEE weighted by IIW (unless otherwise stated) where the missing observation model covariates were imputed using the indicated method.

However, MICE-PPM has lower MSE, and better coverage than SI in all cases.

When the data are MAR, again LOCF and missForest are biased in all settings. In all settings, CCA, SI-PPM, and MICE-PPM have less than 5% relative bias. CCA sees the smallest bias in all cases, and tends to have better MSE and coverage than SI-PPM and MICE-PPM, even at 90% missingness.

When the data are MNAR, CCA, SI-PPM, and MICE-PPM are unbiased at 25% missingness. As the proportion of missing observation increases, the empirical bias of CCA becomes more substantial. However, even at 90% missingness, SI-PPM and MICE-PPM remain unbiased. SI-PPM sees smaller bias in most cases, but tends to have higher MSE and worse coverage than MICE-PPM . In all settings, undercoverage was present.

Overall, MICE-PPM tends to provide unbiased estimates of β_1, aside from when $V(t)$ is only known when $Y(t)$ is. In this setting, none of the existing methods provided unbiased estimation of β_1.

Simulation I (Continuous Outcome): Coverage

Missingness Mechanism	p_{miss}	CCA	LOCF	SI-PMM	MICE-PMM	missForest
No IIW/Naive	-	0.345	-	-	-	-
No Missingness	0	0.946	-	-	-	-
Observation Times Only	-	0.360	0.390	0.347	0.342	0.496
MCAR	0.25	0.944	0.923	0.935	0.941	0.891
	0.50	0.945	0.803	0.919	0.927	0.769
	0.90	0.927	0.458	0.878	0.903	0.521
MAR	0.25	0.948	0.941	0.943	0.943	0.922
	0.50	0.945	0.877	0.933	0.933	0.852
	0.90	0.930	0.516	0.864	0.887	0.515
MNAR	0.25	0.940	0.918	0.935	0.943	0.892
	0.50	0.933	0.764	0.912	0.934	0.774
	0.90	0.890	0.437	0.831	0.865	0.586

Table B.6: 95% confidence interval coverage for β_1 calculated across the 1000 simulation runs for each missing data method and missingness scheme. $\widehat{\beta}_1$ was estimated by Equation (B.1) using an independent GEE weighted by IIW (unless otherwise stated) where the missing observation model covariates were imputed using the indicated method. Standard errors for the 95% confidence intervals are calculated using the Robins et al., 2000 type standard error estimates.

B.2 Additional Results for Simulation II

B.2.1 Additional Results for Simulation II: Binary Outcome

We present the results of Simulation II when a binary outcome is simulated without omitting the extreme estimates. The bias, MSE, and coverage are shown in Tables B.7, B.8, and B.9, respectively.

When IIW is not employed, the outcome model produces biased estimates of the ATE. When IIW is employed and there is no missingness, the results are unbiased.

When $\boldsymbol{Z}(t)$ is known only at the observation times (a special case of MNAR), LOCF provides the best estimation in terms of bias, MSE, and coverage, with a relative empirical bias of 14.2%. CCA, SI-PPM, MICE-PPM are biased, and missForest produces extreme

estimates, as the bias is near infinity.

When the data are MCAR and the missingness is 25% CCA, LOCF, and missForest are unbiased while SI-PPM and MICE-PPM are slightly biased (<6% relative bias). At 50% bias, missForest becomes biased. At 90% missingness, LOCF is unbiased and provides the best estimation in terms of bias, MSE, and coverage. CCA is slightly more biased, while SI-PPM and MICE-PPM are biased. Again, missForest produces extreme estimates of the ATE.

When the data are MAR, LOCF provides unbiased estimation at all degrees of missingness. CCA is unbiased at 25% and 50% missingness. SI-PPM and MICE-PPM have small biases at 25% missingness, but are biased above 50% missigingness. missForest is unbiased at 25%, has a small bias at 50% missingness, and produces extreme estimates again at 90% missingness. In all cases, LOCF produces the smallest MSE and best coverage.

When the data are MNAR and the missingness is 25% or 50%, LOCF is unbiased with small MSE and coverage close to 95%. All other methods have small biases at 25% missingness, and more substantial biases at 50% missingness. At 90% missingness, LOCF has a small relative bias (4.6%) while the other methods are largely biased or produce extreme estimates.

B.2.2 Additional Results for Simulation II: Continuous Outcome

We also consider the setting where a continuous outcome is generated. In this setting, the marginal outcome model of interest is the same as Equation (B.1) where β_1 is the ATE. We again let $\mu(t) = (2 - t)$ and let $D_i = 1$ indicate that individual i is in the treatment group and $D_i = 0$ indicate that individual i is in the control for the duration of the study. The probability of being assigned to the treatment or control group is $\pi_D = 0.5$.

$V_1(t)$ to $V_4(t)$, and $R(t)$ are generated as described in Section 4.3.2.

We generate the continuous outcome as in Equation (B.2), and consider the same model for the observation process as described in Section 4.3.2. We let $\boldsymbol{\beta} = (0.5, 1, 0.4, 0.5, 0.6)$ and $\boldsymbol{\gamma} = (0.6, 0.6, 0.5, 0.7, 0.6)$.

As observation times are simulated from an intensity model, some individuals may not have any observation times, which can reduce the overall sample size. Across the 1000 data sets that are generated, the average sample size is 89.27. Among those who had at least one observation in each data set, and the average number of events per individual is 5.59. The average minimum and maximum number of events per individual across the

Simulation II (Binary Outcome): Bias

Missingness Mechanism	p_{miss}	CCA	LOCF	SI-PMM	MICE-PMM	missForest
No IIW/Naive	-	0.187	-	-	-	-
No Missingness	0	0.008	-	-	-	-
Observation Times Only	-	0.284	0.071	0.186	0.183	**inf**
MCAR	0.25	0.004	0.007	0.026	0.027	-0.009
	0.50	-0.002	0.007	0.050	0.044	-0.031
	0.90	0.024	-0.004	0.080	0.062	**-inf**
MAR	0.25	0.003	0.007	0.026	0.027	-0.011
	0.50	-0.003	0.006	0.048	0.042	-0.028
	0.90	0.022	-0.005	0.086	0.080	**-inf**
MNAR	0.25	0.026	0.006	0.027	0.026	-0.026
	0.50	0.052	0.004	0.048	0.044	-0.053
	0.90	**inf**	-0.023	0.102	0.089	**inf**

Table B.7: Empirical bias of β_1 calculated across the 1000 simulation runs for each missing data method and missingness scheme when a binary outcome is simulated under Simulation II. $\widehat{\beta}_1$ was estimated by Equation (4.1) using an independent GEE weighted by IIW (unless otherwise stated) where the missing observation model covariates were imputed using the indicated method. The true value of β_1 is 0.5.

1000 data sets are 1.00 and 18.74, respectively. The bias, MSE, and coverage are shown in Tables B.10, B.11, and B.12, respectively.

When IIW is not employed and an independent GEE is fitted, the estimated ATE is biased. When the observation process covariates $\boldsymbol{V}(t)$ are known at all times, the we have unbiased results, and close to 95% coverage.

When the covariates $\boldsymbol{V}(t)$ are only observed at the times at which Y is observed, only LOCF is unbiased, with low MSE and coverage close to 95%.

When the missingness mechanism is MCAR, SI-PPM and MI-PMM have the smallest biases up to 50% missingness. However, LOCF again tends to have the lowest MSE and best coverage. At 90% missingness, CCA, SI-PPM, and MICE-PPM outperform the other methods in terms of bias, but LOCF again has the lowest MSE and best coverage. MissForest is biased in all cases. Similar results are seen when the data is MAR.

When the data are MNAR, CCA has the lowest bias and MSE at 25% missingness. At

Simulation II (Binary Outcome): MSE

Missingness Mechanism	p_{miss}	CCA	LOCF	SI-PMM	MICE-PMM	missForest
No IIW/Naive	-	0.067	-	-	-	-
No Missingness	0	0.038	-	-	-	-
Observation Times Only	-	0.237	0.053	0.069	0.066	**inf**
MCAR	0.25	0.038	0.038	0.044	0.042	0.040
	0.50	0.039	0.038	0.055	0.050	0.047
	0.90	0.106	0.037	0.117	0.090	**inf**
MAR	0.25	0.038	0.038	0.043	0.042	0.040
	0.50	0.039	0.038	0.055	0.049	0.045
	0.90	0.103	0.036	0.130	0.102	**inf**
MNAR	0.25	0.038	0.038	0.043	0.042	0.042
	0.50	0.047	0.038	0.057	0.051	0.057
	0.90	**inf**	0.037	0.158	0.111	**inf**

Table B.8: MSE of β_1 calculated across the 1000 simulation runs for each missing data method and missingness scheme when a binary outcome is simulated under Simulation II. $\widehat{\beta}_1$ was estimated by Equation (4.1) using an independent GEE weighted by IIW (unless otherwise stated) where the missing observation model covariates were imputed using the indicated method. The true value of β_1 is 0.5.

50% missingness, SI-PPM and MICE-PPM perform similarly and are unbiased. At 90% missingness, all methods aside from SI-PPM and MICE-PPM are biased. MICE-PPM has better coverage than SI-PPM in this setting.

Overall, MICE-PPM tends to provide the most unbiased estimates of β_1, aside from when $V(t)$ is only known when $Y(t)$ is. In this setting, LOCF is the only method to provide unbiased estimation of β_1.

Simulation II (Binary Outcome): Coverage

Missingness Mechanism	p_{miss}	CCA	LOCF	SI-PMM	MICE-PMM	missForest
No IIW/Naive	-	0.819	-	-	-	-
No Missingness	0	0.929	-	-	-	-
Observation Times Only	-	0.836	0.910	0.817	0.827	0.740
MCAR	0.25	0.927	0.930	0.928	0.936	0.926
	0.50	0.919	0.928	0.908	0.912	0.915
	0.90	0.901	0.930	0.869	0.881	0.850
MAR	0.25	0.924	0.930	0.919	0.928	0.922
	0.50	0.924	0.933	0.916	0.921	0.927
	0.90	0.892	0.932	0.861	0.862	0.856
MNAR	0.25	0.938	0.933	0.922	0.933	0.920
	0.50	0.921	0.928	0.909	0.921	0.911
	0.90	0.837	0.920	0.821	0.839	0.780

Table B.9: 95% confidence interval coverage of β_1 calculated across the 1000 simulation runs for each missing data method and missingness scheme when a binary outcome is simulated under Simulation II. $\widehat{\beta}_1$ was estimated by Equation (4.1) using an independent GEE weighted by IIW (unless otherwise stated) where the missing observation model covariates were imputed using the indicated method. The true value of β_1 is 0.5.

Simulation II (Continuous Outcome): Bias

Missingness Mechanism	p_{miss}	CCA	LOCF	SI-PMM	MICE-PMM	missForest
No IIW/Naive	-	0.102	-	-	-	-
No Missingness	0	-0.018	-	-	-	-
Observation Times Only	-	0.136	-0.002	0.097	0.096	0.093
MCAR	0.25	-0.020	-0.019	-0.010	-0.011	-0.029
	0.50	-0.023	-0.019	-0.004	-0.005	-0.042
	0.90	-0.008	-0.026	-0.004	-0.001	-0.059
MAR	0.25	-0.020	-0.019	-0.010	-0.010	-0.030
	0.50	-0.022	-0.019	-0.002	-0.005	-0.040
	0.90	-0.010	-0.026	-0.001	0.003	-0.054
MNAR	0.25	-0.001	-0.019	-0.013	-0.013	-0.039
	0.50	0.020	-0.022	-0.009	-0.011	-0.058
	0.90	0.056	-0.043	-0.012	-0.013	-0.071

Table B.10: Empirical bias of β_1 calculated across the 1000 simulation runs for each missing data method and missingness scheme when a continuous outcome is simulated under Simulation II. $\widehat{\beta}_1$ was estimated by Equation (B.1) using an independent GEE weighted by IIW (unless otherwise stated) where the missing observation model covariates were imputed using the indicated method. The true value of β_1 is 0.5.

Simulation II (Continuous Outcome): MSE

Missingness Mechanism	p_{miss}	CCA	LOCF	SI-PMM	MICE-PMM	missForest
No IIW/Naive	-	0.018	-	-	-	-
No Missingness	0	0.009	-	-	-	-
Observation Times Only	-	0.038	0.009	0.017	0.017	0.027
MCAR	0.25	0.009	0.009	0.010	0.010	0.010
	0.50	0.009	0.009	0.014	0.012	0.013
	0.90	0.019	0.009	0.023	0.020	0.045
MAR	0.25	0.009	0.009	0.011	0.010	0.010
	0.50	0.009	0.009	0.013	0.012	0.013
	0.90	0.018	0.008	0.028	0.022	0.055
MNAR	0.25	0.008	0.009	0.010	0.010	0.012
	0.50	0.009	0.009	0.013	0.012	0.019
	0.90	0.042	0.010	0.029	0.024	0.112

Table B.11: MSE of β_1 calculated across the 1000 simulation runs for each missing data method and missingness scheme when a continuous outcome is simulated under Simulation II. $\widehat{\beta}_1$ was estimated by Equation (B.1) using an independent GEE weighted by IIW (unless otherwise stated) where the missing observation model covariates were imputed using the indicated method.

Simulation II (Continuous Outcome): Coverage

Missingness Mechanism	p_{miss}	CCA	LOCF	SI-PMM	MICE-PMM	missForest
No IIW/Naive	-	0.771	-	-	-	-
No Missingness	0	0.949	-	-	-	-
Observation Times Only	-	0.841	0.948	0.781	0.783	0.745
MCAR	0.25	0.947	0.950	0.946	0.945	0.937
	0.50	0.944	0.949	0.928	0.934	0.924
	0.90	0.907	0.944	0.900	0.903	0.839
MAR	0.25	0.946	0.949	0.937	0.945	0.943
	0.50	0.948	0.947	0.931	0.934	0.926
	0.90	0.923	0.945	0.865	0.893	0.805
MNAR	0.25	0.948	0.948	0.937	0.938	0.93
	0.50	0.932	0.943	0.928	0.931	0.898
	0.90	0.834	0.932	0.849	0.873	0.713

Table B.12: 95% confidence interval coverage for β_1 calculated across the 1000 simulation runs for each missing data method and missingness scheme when a continuous outcome is simulated under Simulation II. $\widehat{\beta}_1$ was estimated by Equation (B.1) using an independent GEE weighted by IIW (unless otherwise stated) where the missing observation model covariates were imputed using the indicated method. Standard errors for the 95% confidence intervals are calculated using the Robins et al., 2000 type standard error estimates.

Glossary

average treatment effect (ATE)

a statistical measurement that compares the average outcomes of a treatment group to a control group. xiii, xv, xvii, xviii, xix, xx, xxi, xxii, 1, 2, 11, 13, 14, 18, 19, 22, 24, 25, 26, 27, 28, 30, 31, 35, 41, 42, 49, 51, 54, 55, 59, 60, 61, 62, 63, 64, 65, 66, 67, 102, 103, 104, 105, 106, 107, 108, 109, 110, 112, 113, 114, 129, 130

censoring

any event occurring in the study by time t that ensures the future values of all covariates of interest are unknown. By definition, loss to follow-up and administrative censoring at the study end date are a form of censoring. xiv, 5, 17, 21, 22, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 84

classification and regression trees (CART)

decision trees that are appropriate for predicting both categorical and continuous data 18

complete case analysis (CCA)

a missing data method where any observations with missingness are omitted 46, 47, 49, 53, 54, 60, 66, 67, 123, 124, 127, 128, 129, 130, 131

conditionally ignorable

the assumption that a process is ignorable (see definition of ignorability) conditional on the observed covariates 8, 14, 15, 84

consistency

the assumption that the observed outcome under a specific treatment is equal to that treatment's potential outcome, and we can only observe one potential outcome at a time 14

directed acyclic graph (DAG)

a graphical representation containing edges and nodes that is used to represent the assumptions made in a model. xiv, xv, 20, 22, 27, 29, 69, 71, 72, 73, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 88

flexible inverse probability of treatment and intensity weighted generalized estimating equation (FIPTIW-GEE)

a generalized estimating equation weighted by the inverse probability of being observed and treated 16, 17, 21, 30, 43, 101

flexible inverse probability of treatment, intensity, and censoring weighting (FIPTICW)

a weighting method used to account for conditionally ignorable observation processes and conditionally ignorable treatment assignment processes that incorporates inverse probability of censoring weights in an attempt to account for informative censoring. xiii, xv, xvii, xx, xxi, 21, 22, 24, 25, 26, 105, 106, 107, 108, 109, 110

flexible inverse probability of treatment and intensity weighting (FIPTIW)

a weighting method used to account for conditionally ignorable observation processes and conditionally ignorable treatment assignment processes vi, xiii, xiv, xv, xvi, xvii, xx, xxi, 12, 15, 16, 17, 18, 20, 21, 22, 24, 25, 26, 29, 30, 31, 32, 33, 37, 39, 40, 41, 42, 43, 44, 49, 86, 87, 98, 101, 102, 103, 105, 106, 107, 108, 109, 110, 116, 117, 118, 119, 121

g-estimation

a method used in causal inference to obtain causal estimates of the effect of an exposure on an outcome, typically the mean difference in an outcome after adjusting for confounders 85

generalized boosted model (GBM)

a tree-based machine learning method that aggregates the results of multiple weak decision trees 15

homogeneous Poisson process (HPP)

a counting process such that the number of events occurring in an interval $(0, t)$ follows a Poisson distribution with constant rate λt 20

informative censoring

the violation of the noninformative censoring assumption, where the probability of being censored is related to the longitudinal outcome 26, 39, 43, 44, 86

informative observation process

an observation process that is related to the longitudinal outcome 2, 11, 12, 42, 81, 84

inverse probability of treatment weighted generalized estimating equation (IPTW-GEE)

a generalized estimating equation weighted by the inverse probability of being treated 101

inverse probability of censoring weighting (IPCW)

a weighting method used to account for conditionally ignorable censoring processes 12, 17, 18, 21, 22, 26, 43, 86

inverse intensity weighted generalized estimating equation (IIW-GEE)

a generalized estimating equation weighted by the inverse probability of being observed 8, 10, 53, 101

inverse probability of treatment weighting (IPTW)

a weighting method used to account for conditionally ignorable treatment assignment processes xiii, xiv, xv, xvi, xvii, xx, xxi, 11, 12, 13, 14, 15, 18, 22, 24, 25, 26, 27, 29, 30, 31, 32, 37, 40, 41, 42, 75, 77, 80, 83, 86, 98, 102, 103, 105, 106, 107, 108, 109, 110, 116, 117, 118, 119, 121

inverse intensity weighting (IIW)

a weighting method used to account for conditionally ignorable observation processes xiii, xiv, xv, xvi, xvii, xviii, xix, xx, xxi, xxii, xxiii, xxiv, 2, 4, 8, 10, 11, 12, 15, 17, 18, 22, 24, 25, 26, 27, 28, 29, 30, 31, 32, 35, 37, 40, 41, 42, 43, 44, 45, 46, 47, 49, 53, 54, 55, 56, 57, 58, 60, 61, 62, 63, 64, 65, 66, 69, 73, 75, 76, 80, 81, 84, 85, 86, 87, 88, 98, 102, 103, 105, 106, 107, 108, 109, 110, 112, 113, 114, 116, 117, 118, 119, 121, 123, 124, 125, 126, 127, 128, 129, 131, 132, 133, 134, 135, 136

irregular longitudinal data

data involving repeated observations on individuals where the observation times may vary across individuals 1, 5, 86, 88

last observation carried forward (LOCF)

a missing data method that imputes missing longitudinal data using the last known value. 7, 47, 49, 53, 54, 60, 61, 65, 66, 67, 82, 87, 88, 123, 124, 127, 128, 129, 130, 131, 132

malaria

a disease caused by parasites, which can be spread through mosquitoes xviii, 12, 31, 33, 39, 40, 42, 65, 66, 86

missForest

a nonparametric impuation method used for mixed-type data that uses random forests trained on the observed values to predict missing values 48, 49, 53, 54, 60, 66, 67, 88, 123, 124, 127, 128, 129, 130

missing not at random (MNAR)

the type of missing data mechanism where the probability of being observed/missing can depend on the missing data as well as the observed data 47, 51, 53, 54, 60, 87, 124, 128, 129, 130, 131

missing at random (MAR)

the type of missing data mechanism where the probability of being observed/missing does not depend on the missing data 47, 49, 50, 53, 54, 59, 60, 123, 128, 130, 131

missing completely at random (MCAR)

the type of missing data mechanism where the probability of being observed/missing does not depend on the observed or missing data 47, 53, 54, 60, 123, 127, 130, 131

multiple imputation by chained equations using predictive mean matching (MICE-PPM)

a method where missing values are iteratively replaced by multiple estimates using predictive mean matching, which are then used in aggregate to obtain model estimates 48, 49, 53, 54, 55, 60, 66, 67, 87, 123, 124, 127, 128, 129, 130, 131, 132

multiple imputation by chained equations (MICE)

a method where missing values are iteratively replaced by multiple estimates which are then used in aggregate to obtain model estimates 47, 48, 82

non-homogeneous Poisson process (NHPP)

a counting process such that the number of events occurring in an interval $(0, t)$ follows a Poisson distribution with time-varying rate $\lambda(t)$ 20

non-stabilized weights

weights which are estimated by the inverse probability, with one in the numerator 9

noninformative censoring

the assumption that the probability of being censored is unrelated to the longitudinal outcome xiv, xxii, 2, 7, 12, 17, 18, 21, 22, 26, 39, 43, 73, 74, 76, 79, 80, 81, 83, 84, 86, 87, 88, 97, 118

observation process

the underlying process driving the times at which an individual is observed in the data set. Also referred to as a visit process or monitoring process in the literature 2, 5, 6, 8, 9, 10, 11, 13, 14, 15, 16, 17, 22, 27, 29, 30, 31, 40, 42, 43, 45, 69, 70, 75, 76, 79, 81, 82, 84, 85, 86

odds ratio (OR)

a quantity representing the multiplicative increase or decrease in odds between two groups xviii, 42, 65, 66

positivity

the assumption that the probability of receiving a given treatment is greater than zero for every individual 14

potential outcome

the hypothetical/counterfactual outcomes an individual would have experienced had they been assigned to any treatment group 13, 14

potential outcomes framework

a popular framework for casual inference problems where we consider the hypothetical outcomes that an individual would have experienced had they been in any treatment group 13

predictive mean matching (PMM)

a nonparametric imputation method that uses observed values from observations with similar predictive means 48

propensity score

the probability of being assigned to the treatment group 14, 15, 18, 29, 30, 40, 43

proportional hazards (PH)

a statistical method often used in time-to-event or recurrent event analysis to relate event times to a set of covariates 7, 9, 17, 22, 28, 30, 40, 43, 45, 46, 47, 48, 82, 87

separability

Two model parameters are separable (i.e., the models do not share parameters) 7

single imputation with predictive mean matching (SI-PPM)

a method where missing values are replaced by an estimate in the data using predictive mean matching 48, 49, 53, 54, 60, 66, 67, 123, 124, 127, 128, 129, 130, 131, 132

single imputation (SI)

a method where missing values are replaced by an estimate in the data 47, 48

stabilized weights

weights which are estimated by replacing the numerator with a marginal probability 9, 10, 15

stable unit treatment value assumption (SUTVA)

an assumption that states the treatment assignment of a given individual does not affect the outcome of another individual 14

strongly ignorable treatment assignment (SITA)

an assumption that states that the treatment assignment is independent of the potential outcomes given the measured covariates 13

surrogate censoring times

the censoring times used to estimate the true censoring time, which is unknown 39, 97

treatment-confounder feedback

a phenomenon that occurs when a time-varying treatment or exposure affects a confounder, and the confounder affects the treatment 70, 85

treatment assignment process

the underlying process that determines whether or not an individual is treated. 13, 14, 22

visiting at random (VAR)

an assumption made on the observation process where the observation times are conditionally independent of the outcome at time t, given data up to that time xiv, xv, 77, 80, 81, 82, 83, 84, 85, 88