Causal Inference on the Marginal Effect of an Exposure: Addressing Biases due to Covariate-Driven Monitoring Times and Confounders

Janie Coulombe

Doctor of Philosophy

Department of Epidemiology, Biostatistics and Occupational Health

McGill University
Montréal, Québec
July 2021

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

Dedication

I dedicate this thesis to my Mom.

Acknowledgements

First and foremost, I would like to express my most sincere gratitude to my PhD co-supervisors, Dr. Erica E. M. Moodie and Dr. Robert W. Platt, for their help and support over these past years. I wish to thank Erica for her continuous encouragements, generosity, optimism, and for the hundreds of discussions we had where she selflessly shared her knowledge and passion with me. I would like to thank Robert for advising me, for his kindness and flexibility, and for his confidence in my work. I am grateful to both of my co-supervisors for the numerous opportunities they have given me, and for making these past few years so enjoyable.

At the Department of Epidemiology, Biostatistics and Occupational Health, I am indebted to the several Professors who generously shared their knowledge and excitement for biostatistics with me; in particular, I wish to thank Dr. Alexandra Schmidt, Dr. Michal Abrahamowicz, Dr. James Hanley, Dr. Andrea Benedetti, and Dr. Shirin Golchi, for their teaching and mentoring. I want to thank my friends and colleagues at McGill University, and in particular, Kevin, Gabrielle, Menglan, Luc, Victoire, Steve, and James, for the great times we had, for the laughters, and for your support all throughout this journey. I am also very grateful to Andre Yves Gagnon, Katherine Hayden, Dolores Coleto, and Nathalie Theoret, for their support and for facilitating my studies at McGill University. I also wish to thank the Professors of the Department of Statistics at McGill University, where I really enjoyed taking classes and truly appreciated their rigor and passion.

I am indebted to a few other mentors whom I would like to acknowledge here, who influenced me in pursuing doctoral studies and taught me a lot along the way. First, I would like to thank Dr. Christel Renoux for mentoring me, for everything she has taught me, and for encouraging me, both to pursue doctoral studies and during my doctoral studies. I wish to thank Dr. Samy Suissa for all the knowledge and wisdom he shared with me as my su-

pervisor at the Lady Davis Research Institute (LDI). I also wish to thank the staff at the LDI, especially Marisa Mancini and Melissa Dahan, for their help and support, and the many colleagues there who learned me a lot, especially Sophie Dell'Aniello and Hui Yin. At l'Université de Montréal, I am thankful to the several Professors in Statistics who triggered my interest for Statistics from the very beginning, and in particular, to Professor David Haziza. I also want to thank my Masters' supervisor, Professor Christian Léger, for all his teachings.

I wish to acknowledge the financial support I received from the Natural Sciences and Engineering Research Council of Canada, as well as from both my PhD co-supervisors, and from the Faculty of Medicine and Health Sciences at McGill University. I would also like to thank Compute Canada for the computing resources they provided, and for their always prompt replies when I asked for help; they greatly facilitated my work.

I would like to thank some good friends who supported me and brought joy to this adventure, in particular, Sébastien, Catherine, Audrey, Justine, Pauline, Pam, Noé, and Charlène; my good *stat* friends, Audrey-Anne, Paule, and Alex; Phil and Serge for the good old days at l'UdeM; and my cousin Simon, for our many interesting discussions about statistics. To all of you: I am forever grateful for your friendship and your support in good and in bad times.

Finally, this thesis belongs to my family and to my partner, Nicolas. To Mom, Dad, Louis-Jean, and my aunt Lucie, I thank you for your unconditional love, support, patience, and never-ending encouragements. And to Nicolas, thank you so much for helping me everyday to cultivate my curiosity, for your kindness, and for your love.

Preface

This manuscript-based thesis contains six chapters: an introduction, an original literature review, three chapters that correspond to three different stand-alone manuscripts, and a conclusion. A complete bibliography is presented after the appendices, at the end of this thesis. Chapters 3, 4 and 5 are linked by the main research topic of this thesis, and each adds novel methodological developments and new insights to the current statistical literature in the area of causal inference with observational data. Each of these three chapters begins with a short preamble that introduces the topic of the chapter and that briefly describes the gap in literature that I seek to fill with the proposed methodology. The methodologies are all illustrated using real-life data analyses.

The introduction and the literature review (Chapters 1 and 2) of this thesis were conceived and written by Janie Coulombe (JC) following enriching discussions with Erica E. M. Moodie (EEMM), and both Chapters 1 and 2 were further edited by EEMM and Robert W. Platt (RWP). The work in Chapter 3 was conceptualized in a series of discussions between JC and EEMM. JC conducted the methodological derivations, designed and conducted the simulation study, performed the data analysis and wrote the manuscript draft. EEMM provided substantial help and guidance with the methodological derivations and troubleshooting with the simulation studies and data analysis. RWP advised and commented the work. EEMM and RWP corrected and edited the chapter. The methodological work in Chapter 4 was conceptualized by JC and EEMM, and the data application to CPRD was conceptualized in a series of discussions between JC, Christel Renoux (CR), and EEMM. JC conducted the methodological derivations, designed and conducted the simulation study, conducted the data management and analysis, and wrote the manuscript draft. The work was advised and edited by EEMM, RWP, and CR. Ideas in Chapter 5 were conceptualized by JC and EEMM. JC conducted the methodological derivations, designed and conducted the simula-

tion studies, performed the data analyses and wrote the manuscript draft. EEMM and RWP advised the work and edited the chapter. The conclusion was conceived and written by JC and edited by EEMM and RWP.

Preface

Chapters 3 and 5 of this thesis use the free and publicly-available data from the *Add Health* study in the United States. *Add Health* was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill. The project was funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development from 1994-2021, with cooperative funding from 23 other federal agencies and foundations. *Add Health* is currently directed by Robert A. Hummer; it was previously directed by Kathleen Mullan Harris (2004-2021) and J. Richard Udry (1994-2004). More information on obtaining *Add Health* data is available on the project website (https://addhealth.cpc.unc.edu). The content of this thesis is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health or the University of North Carolina at Chapel Hill.

The case study in Chapter 4 of this thesis uses data from the Clinical Practice Research Datalink in the United Kingdom. The corresponding study protocol has been approved by the Independent Scientific Advisory Committee of the United Kingdom Clinical Practice Research Datalink (protocol number 19_017R) and by the Research Ethics Committee of the Jewish General Hospital (Montreal, Quebec, Canada).

Abstract

Causal inference focuses on the estimation of effects due to specific, well-defined causes (such as exposures on which we can intervene). With the advent of powerful computers and smart electronic devices, data are now collected more rapidly than ever. That abundance of data provides a rich landscape for research on causal inference. However, the collection of these data does not always rely on a study design made expressly for answering the question of interest. For instance, contrary to some randomized controlled studies where exposure is randomized and observation times are set in advance, longitudinal observational data from medical health records are filled with biasing associations that, should they not be taken into account, could adversely affect the inference. In my doctoral thesis, I focus on two such challenges, the confounding bias, and the bias due to covariate-driven monitoring times, in the inference on the causal marginal effect of an exposure on a longitudinal outcome. While there is a vast statistical literature on how to model covariate-driven monitoring times, it has not been studied in a causal framework, nor considered simultaneously with confounding. This thesis proposes ways to consistently estimate the marginal effect of exposure in settings subject to those biases.

In a first manuscript, I propose two novel estimators for the marginal effect of a binary exposure on a continuous, longitudinal outcome. These estimators allow for the outcome to be observed irregularly across individuals. They consider confounding factors and covariate-driven monitoring times that may affect inference via a monitoring and an exposure models, and the corresponding inverse weights. In extensive simulation studies, they are compared along with other common estimators. The asymptotic properties of the best estimator are developed.

The second manuscript is motivated by the estimation of the marginal effects of two antidepressants, citalopram and fluoxetine, on body mass index, in data from the Clinical Practice Research Datalink (CPRD) in the United Kingdom. It is assumed that the longi-

tudinal characteristics of the patients change with physician visits, and therefore, interact with the monitoring process. Different causal diagrams are used to describe how bias due to covariate-driven monitoring times can arise in different situations, including the complex setting where the endogenous covariate process can be modified by the monitoring process. A new stabilized and cumulated inverse weight is proposed for the latter setting. The weight serves to break the association between the full history of covariates and the monitoring process.

In the third manuscript, I aim to evaluate the marginal (causal) effect of the time spent on video games weekly, on suicide attempts. To investigate that effect, I use longitudinal data from the Add Health Study, in the United States; these data are subject to confounding and monitoring times that may be associated with patients' characteristics. I first extend one of the estimators proposed in the first manuscript of this thesis to allow consideration of a continuous exposure via a generalized inverse probability of treatment weight, along with a categorical ordinal outcome via a proportional odds model. Simulation studies are used to demonstrate the consistency of the approach, which is further used to estimate the marginal odds ratio for a 2-fold or an 8-fold increases in the time spent playing video games on the number of suicide attempts (categorized as 0, 1, or 2 or more).

Using causal diagrams, I provided in this thesis a thorough demonstration of the bias due to covariate-driven monitoring times. I proposed a sound methodology for evaluating causal effects in observational studies subject to confounding and covariate-driven monitoring times. The proposed methods were further used to answer mental health-related research questions.

Abrégé

L'inférence causale est axée sur l'estimation d'effets dûs à des causes spécifiques et bien définies (par exemple, une exposition sur laquelle il est possible d'intervenir). Avec l'avènement des ordinateurs super puissants et outils électroniques intelligents, les données sont collectées plus rapidement que jamais auparavant. Cette abondance de données représente une grande opportunité pour la recherche en inférence causale. Cependant, la collecte de ces données ne dépend que très rarement d'un plan d'étude pensé expressément pour répondre à une question causale d'intérêt. Par exemple, contrairement aux essais cliniques randomisés où l'exposition est randomisée et où les temps de visite sont planifiés d'avance, les données longitudinales provenant des dossiers médicaux électroniques contiennent des associations trompeuses qui peuvent influencer l'inférence si elles ne sont pas considérées. Dans ma thèse de doctorat, je me concentre sur deux défis reliés à ces associations, soient le biais de confusion et le biais dû aux temps de visite qui dépendent des variables, lors de l'inférence causale sur l'effet marginal d'une exposition sur une issue longitudinale. Alors que la littérature statistique sur la modélisation des temps de visite dépendants est riche, cela n'a pas été étudié précisément dans un contexte d'inférence causale, et n'a pas été considéré simultanément avec la confusion. Cette thèse propose des méthodes pour estimer de façon consistente l'effet marginal d'une exposition dans des contextes enclins à ces types de biais.

Dans un premier manuscrit, je propose deux nouveaux estimateurs pour l'effet marginal d'une exposition dichotomique sur une issue longitudinale continue. Ces estimateurs permettent à l'issue longitudinale d'être observée à des temps irréguliers qui varient entre les individus. Ils considèrent les facteurs de confusion et les temps de visite dépendant des covariables qui peuvent affecter l'inférence, grâce à des modèles pour les visites et pour l'exposition et des poids inverses correspondants. Dans des études de simulation exhaustives, ils sont comparés avec des estimateurs plus communs. Les propriétés asymptotiques du meilleur estimateur sont dérivées.

Le deuxième manuscrit est motivé par l'estimation des effets marginaux de deux antidépresseurs, le citalopram et la fluoxétine, sur l'indice de masse corporelle, dans les données de la Clinical Practice Research Datalink (CPRD) au Royaume-Uni. On assume que les caractéristiques longitudinales des patients peuvent changer avec leurs visites chez le médecin, et ainsi, qu'elles interagissent avec le processus de visite. Des diagrammes causaux sont utilisés afin de décrire comment le biais dû aux temps de visite dépendants peut survenir dans différentes situations, incluant le cas complexe où le processus de variable endogène peut être modifié par le processus de visite. Pour ce cas complexe, je propose un nouveau poids stabilisé et cumulé. Ce poids est utilisé pour briser tout lien de dépendance entre l'histoire complète des variables et le processus de visite.

Dans un troisième manuscrit, je souhaite évaluer l'effet marginal causal du temps hebdomadaire passé à jouer aux jeux vidéos sur les tentatives de suicide. Pour évaluer cet effet, j'utilise les données longitudinales de l'étude Add Health aux États-Unis. Ces données sont sujettes à une exposition et des temps de visite qui dépendent des caractéristiques des individus dans l'étude. Je propose d'abord une extension de l'un des deux estimateurs proposés dans le premier manuscrit de cette thèse, afin de permettre la considération d'une exposition continue via un poids inverse à la probabilité de traitement généralisé, ainsi que d'une issue ordinale grâce à un modèle de cotes proportionnelles. Je démontre la consistence de cette approche à partir d'études de simulation, et l'approche est ensuite utilisée afin d'estimer le rapport de cotes marginal pour une augmentation de 2 fois, ou de 8 fois le temps passé à jouer aux jeux vidéos, sur le nombre de tentatives de suicide (catégorisé en 0, 1, ou 2 ou plus).

Dans cette thèse, j'ai offert une démonstration exhaustive du biais dû aux temps de visite dépendant des variables, à partir de diagrammes causaux. J'ai proposé une méthodologie rigoureuse pour évaluer les effets causaux dans les études observationnelles sujettes à la confusion et aux temps de visite qui dépendent de variables. Les méthodes proposées ont

permis de répondre à des questions de recherche en santé mentale.

Table of contents

List of Tables

List of Figures

Abbreviations

AG Andersen and Gill

ATE average treatment effect

BMI body mass index

CI confidence interval

CLT central limit theorem

CPRD Clinical Practice Research Datalink

DAG directed acyclic graph

DGM data generating mechanism

EHR electronic health records

ENG most recent grades in English or language arts

FCA feeling that friends cared about them

FHTR frequency of having trouble relaxing

GAD generalized anxiety disorder

GEE generalized estimating equation

GP general practitioners

GS most recent grades in Science

HES Hospital Episode Statistics

HOF frequency with which they hang out with friends

HSS most recent grades in History or Social Sciences

IIV inverse intensity of visit

IIVW inverse intensity of visit weights

IMD Index of Multiple Deprivation

IPT inverse probability of treatment

IPTMP Inverse Probability of Treatment and Monitoring Proportional Odds Model

IPTW inverse probability of treatment weights

LGR level of grooming of the respondent

MATH most recent grades in mathematics

MI multiple imputation

MLE maximum likelihood estimator

MOR marginal odds ratio

MSE mean squared error

OLS ordinary least squares

ONS Office for National Statistics (United Kingdom)

PS propensity score

PWP Prentice, Williams and Peterson

RSBI the respondent seemed bored or impatient

SSRI selective serotonin reuptake inhibitor

SUTVA Stable Unit Treatment Value Assumption

UK United Kingdom

US United States

Chapter 1

Introduction

This thesis is concerned with the estimation of the causal marginal effect of an exposure on a longitudinal outcome, in studies using observational data such as those from electronic health records (EHR). I focus on two important challenges in those data, which are the bias due to confounding of the relationship between the exposure and the outcome, and that due to the covariate-driven follow-up times.

The causal marginal effect of a binary exposure on an outcome is often referred to as the average treatment effect (ATE) in the literature. For a treatment with levels 0 and 1, it is conceptualized as the average difference in the outcome, had we intervened on everyone to give them treatment 0, as opposed to having given everyone treatment 1. The term *causal* refers to the fact that, in such situation, the effect observed is truly due to the exposure, not to a (spurious) association between the exposure and the outcome. Correct estimation of the ATE, or of other causal estimands related to other exposure effects, is highly important. Ultimately, these quantities provide valuable insights into whether a treatment should be enforced at the population level or not. Therefore, they can help health authorities to take sound decisions regarding global health policies.

Observational data such as those from EHR are not meant for research purposes. When

interest lies in the estimation of the causal effect of an exposure using those data – for instance, that of a treatment prescribed to patients – then that exposure was not randomized to patients. No study design was used in order to collect data, to gain the best insights on the marginal effect of the exposure. In particular, nothing ensures that there is no biasing difference between the different exposure groups. Furthermore, in observational studies (and in some experimental, randomized studies as well), investigators may have no control over the irregularity of the follow-up times, which I also refer to as visit or monitoring times. In this thesis, I assume that the patients' outcomes are assessed (i.e., observed) at those times. These can occur irregularly across individuals and may depend on the patients' characteristics. As a consequence, associations between a patient's covariate, monitoring, and outcome processes are likely to exist in observational data. These (potentially biasing) associations should be considered, as they can affect the causal inference on the marginal effect of an exposure on a longitudinal outcome.

In the statistical literature, several methods have been proposed for making statistical inference on marginal effects while accounting for irregular or covariate-driven monitoring times (see e.g., Lin and Ying [2001], Lipsitz et al. [2002], Lin et al. [2004], Liang et al. [2009], Bůžková and Lumley [2009]). In the causal inference literature, confounding has also been discussed extensively, and methods such as those based on the propensity score (Rosenbaum and Rubin [1983]) or on g-computation (Robins [1986]) have been proposed to remove the association that is due to confounding between an exposure and an outcome, when estimating the marginal effect of the exposure (the ATE). Nevertheless, covariate-driven monitoring times were rarely considered simultaneously with confounding, nor within a causal framework. The causal inference framework often implies that we make several assumptions on the relations between a set of variables, including the monitoring indicators, the exposure and the outcome. Understanding these relations, and how imbalances due to confounding and covariate-driven monitoring times can affect estimators for the ATE, are paramount to its estimation. The causal literature offers tools (such as the directed acyclic graphs) that

can help with describing and understanding these imbalances better.

In this thesis, I therefore build on the previous statistical literature on irregular follow-up times and on causal inference to propose and demonstrate a methodology for the estimation of the causal marginal effect of exposure (or treatment), for observational studies subject to confounding and covariate-driven monitoring times. In Chapter 2, I provide a review of the most important components to consider in the development of that methodology. In particular, I discuss the special features of observational data, causal inference, and the different estimation tools that can be used to estimate marginal effects such as the ATE.

In Chapter 3, two novel estimators for the causal marginal effect of a binary exposure on a longitudinal, continuous outcome are proposed and demonstrated. These estimators extend previous estimators that did not account simultaneously for confounding and covariate-driven monitoring times. The asymptoptic properties of the estimator that performed the best (between the two) are developed. The estimators are demonstrated in extensive simulation studies and they are further used to assess the marginal effect of a depressive mood on weight, in an analysis of data from the *Add Health* study in the United States (US).

In Chapter 4, endogeneity of the covariate process driving monitoring times is considered. Endogeneity can create long-term depencencies between the monitoring and the outcome processes. Causal diagrams are used to describe how long-term dependencies can arise, and a new methodology is proposed that acccounts for endogeneity, and that is used to compare the marginal effects of two commonly used antidepressant drugs on body mass index (BMI) in data from the Clinical Practice Research Datalink (CPRD) in the United Kingdom (UK). In that chapter, the previous literature is also extended to allow for the covariates affecting the visits to be only monitored at times when the outcome is monitored. This relaxes the assumptions commonly postulated by previous authors.

In Chapter 5, one of the estimators proposed in Chapter 3 is extended to study the causal effect of a continuous exposure on a categorical, ordinal longitudinal outcome. Some chal-

lenges encountered in that novel setting are addressed. Finally, the proposed methodology is used to assess the effect of the number of hours spent playing video games on a categorized number of suicide attempts outcome, in individuals from the *Add Health* study in the US.

Chapters 3 to 5 were written as stand-alone manuscripts. Chapter 3 is published in *Biometrics*. Chapters 4 is currently under review in a statistical journal. Chapter 5 has recently been accepted for publication in *Statistics in Medicine*. An effort has been made to keep the notation consistent across all chapters of this thesis, and any important difference in the notation is mentioned in the preamble before a given chapter. This thesis ends with a conclusion in which I review the contributions of this thesis, discuss notable limitations of this work, and mention some ideas for future work.

Chapter 2

Literature review

This chapter reviews the theory that we extended in all three manuscripts that make up this thesis. First, I discuss the challenges associated with the analysis of data that were not collected for research purposes and how their features can affect our causal inference. Then, I overview some concepts commonly used in the causal inference literature, such as the potential outcome framework, the identifiability assumptions that are necessary for inference, and the use of causal diagrams as a complement to causal inference. In a third section, I review different modelling approaches for the effect of interest, including the use of estimating equations. In a fourth section, I discuss the history of the proposed methods for covariate-driven monitoring times. Finally, in a last section, inverse weighting is discussed in the context of sample selection, causal inference, and more specifically, covariate-driven monitoring times.

In this chapter, vectors and matrices are denoted in bold. I use $Y_i(t)$ to denote the longitudinal outcome process of individual i, $i = 1, ..., n$, at time t, and $I_i(t)$ to denote the exposure of individual i at time t; the terms treatment and exposure will be used interchangeably. The outcome process is assumed to be observed at some times $T_{i1}, ..., T_{iV_i}$ for individual i. Observation times are not necessarily common across individuals. I assume that individuals are all

part of a common study cohort, with a maximum follow-up time of τ across the cohort, such that any time t considered is contained in $[0, \tau]$. In some instances, I use the notation $\mathbf{X_i(t)}$ to refer to a vector of predictors for individual i, which may contain the exposure $I_i(t)$. To estimate the marginal effect of exposure, semiparametric models are used. The parameters $\boldsymbol{\beta}$ are those corresponding to the columns of $\mathbf{X(t)}$, when the outcome mean is modelled as a function of $\mathbf{X(t)}'\boldsymbol{\beta}$. The coefficient $\beta_I \subset \boldsymbol{\beta}$ corresponds to the associational parameter between the exposure vector $\mathbf{I(t)}$ and the outcome vector $\mathbf{Y(t)}$ in those models.

2.1 Data not collected for research purposes

This thesis focuses on causal inference methods for data that were not collected for research purposes. These data or studies which use them are often called *observational* (in contrast to experimental), for we *observe* and infer from them (often, restrospectively) rather than conducting a designed experiment aimed specifically at answering a causal research question. For that reason, some of the resulting data features cannot be controlled for by using a well-thought design. In some instances, the use of special methods for causal inference in observational data may enable causal inferences to be drawn. In other instances, it may be impossible to answer a causal study question, as the nature of the data, the availability of covariates, or an ill-defined exposure preclude identifying the causal effect.

Examples of data that were not collected for research purposes include those from EHR. EHR data consist of the computerized notes of general practitioners (GP) or other clinicians, taken during appointments with patients. Characteristics of the patients, such as symptoms, anthropometric measurements (e.g. weight), chronic diseases, previous or current condition, and so on, are noted in a patient's medical file. These notes may be entered directly in an electronic file or be processed with the help of specialized software, and they are made available in a common format (at least within a given healthcare setting; different systems or hospitals may not conform to the same format in general). In most EHR, the drugs

prescribed by the GP are also entered in the computer system automatically, which provides linked information about medication use. Sometimes, data from EHR can be linked with hospital data and mortality databases to supplement information on hospital stay diagnoses and the causes and date of death, or with other specialized databases such as cancer registries. As an example of EHR data, the second project of this thesis uses data from the CPRD in the UK. The CPRD contains EHR data from more than 11 million patients in the UK (Herrett et al. [2015]), including information on the drugs prescribed, weight, height, comorbidities, demographics, and diagnostic codes. Of course, these measurements are not necessarily available for every patient and at every time, and missing data is a common problem in analyses using similar data.

Observational data represent a rich source of information for learning about causal exposure effects. However, they also present some common pitfalls that can affect the causal inference. These pitfalls include the risk of confounding (e.g., by indication), selection bias or covariate-driven monitoring times, measurement error, missing data (e.g., on the covariates), to name only a few. Some of these risks are now briefly discussed in contexts relevant to the work that follows in this thesis.

2.1.1 Confounding bias

Confounding bias is a distortion of the exposure effect of interest, due to a common cause of the exposure and the study outcome (Greenland and Morgenstern [2001]). It is one of the most common sources of bias discussed in epidemiological studies. It can affect the inference on the effect considerably, and can explain large differences between effects due to associations, and effects that are causal.

Suppose that you wish to estimate the causal, marginal effect of an intervention or exposure I (which takes values in $\{0,1\}$) on a continuous outcome Y, where age strongly affects both the exposure indication (such that older patients receive treatment $I = 1$ much more often

than their younger counterparts) as well as the outcome (such that older patients tend to have worse outcomes). When investigating the data, you may find that most patients with treatment $I = 1$ have poorer outcome values. However, these patients are also older, on average. Their outcome is also affected by their age, which tends to worsen the outcome. It is hard to assess whether the exposure really worsens the outcome on its own (causal effect), or whether the observed reduction is due to age, a common cause of exposure indication and outcome. In that example, age is a confounder for the relationship between treatment I and the outcome Y. When confounders are not accounted for, merely evaluating the association between the exposure and the outcome will lead to a measure of association, not a causal effect of the exposure.

Confounding bias is not an issue in well-conducted randomized controlled studies, except under very specific circumstances. Randomization ensures balance in patients' characteristics across the treatment groups or levels (Greenland [1990]). That is, no patient's characteristic predicts whether that patient will, or will not receive treatment $I = 1$, and therefore, there is no common cause of the exposure and the outcome that confounds the exposure effect. In data that were not collected for research purposes, however, confounding bias is likely to occur. For instance, a treatment is prescribed *mostly* according to patients' characteristics, based on their history, medical condition, own preferences, and on the physician's preferences. Often, these same characteristics also affect the study outcome. For an exposure that is not a therapeutic treatment (like, e.g., exposure to pollution), an individual's characteristics may also predict the exposure level, and can affect the outcome too. In this thesis, confounder variables are denoted by the set $\mathbf{K_i(t)}$ at time t if they are time-varying, or else by $\mathbf{K_i}$, in individual i. If they do not vary in time, then we assume that they are defined using only pre-treatment or pre-exposure information.

2.1.2 Selection bias and irregular monitoring times

Selection bias has received less attention than confounding in observational studies. In contrast to confounding bias, which is due to an ancestor common cause of the exposure and the outcome (creating a *backdoor* path from the exposure to the outcome), selection bias generally arises due to common effects of the exposure and the outcome. Often, issues that arise with selection are related to a phenomenon called *collider stratification bias* (Greenland [2003]). A collider is a common effect that blocks a path, for instance, from the exposure to the study outcome (Greenland et al. [1999]). Figure 2.1 shows a causal diagram with a collider S on the path from I to Y. Causal diagrams are reviewed in Section 2.2.3.

$$I \longrightarrow S \longleftarrow Y$$

Figure 2.1: S is a collider on the path from I to Y

Hernán et al. [2004] have provided a structural approach to selection bias (focused predominantly on the cross-sectional setting) and used causal diagrams to explain how selection bias can arise in different situations. According to these authors, the term *selection bias* is used in epidemiology to refer to many types of biases, including informative dropout, volunteer bias, or incorrect selection of controls in case-control studies.

In the cross-sectional case, the most basic example of selection bias is that where both the exposure and the outcome affect the probability of being selected into the study. In such case, the indicator for being in the study acts as a collider on a path from the exposure to the outcome (Figure 2.2); conditioning on the collider opens a biasing path between the exposure and the outcome (Figure 2.3).

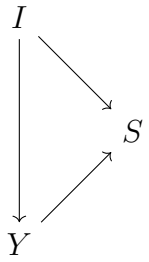


Figure 2.2: I is the exposure, Y the outcome, and S an indicator of being selected in the study. S is a collider on the path $I - S - Y$.

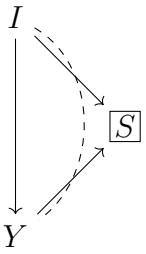


Figure 2.3: A square around S means that we condition upon that variable (for instance, by restricting the analysis to one stratum of S: e.g., those patients selected). This can lead to bias, as it opens a biasing path from I to Y that is not due to the actual causal effect of I.

When data are longitudinal, on the other hand, selection due to informative dropout or to informative monitoring times can occur throughout follow-up time, and the whole visit mechanism may have to be modelled in time. Since selection bias and related biases (such as informative dropout) may have different consequences depending on the target causal estimand, drawing a causal diagram may help greatly in understanding where the monitoring or the selection processes can lead to biases.

In EHR, GP notes are transformed into computerized health data. However, only the patients who visit their physician have their data recorded. Selection bias is therefore likely to occur, as patients' covariates (including the treatment received) are likely to be associated with both their monitoring process and their outcome process, leading to imbalances in the characteristics of the patients visiting and therefore, being represented in the study. In particular, sicker patients, or those with specific chronic diseases, are expected to visit their physician more often than others.

2.1.3 Missing data

In longitudinal studies, covariates are rarely updated at all times, but rather at discrete points in time. When there is no physician visit or hospital stay between two time points, covariates could be extrapolated using the last observation and carrying it forward. Yet, covariates may

have changed in between visits. If these covariates are part of the confounders set or they affect selection bias, not having them measured (or not having their most *up-to-date* value) may lead to a poor adjustment of the exposure, the monitoring, or the outcome models.

In this thesis, the outcome is assumed to be observed only sporadically and at irregular times across patients, which is why we aim to account for the potential for selection bias (discussed in the previous section). A monitoring model is used to account for the outcome monitoring process, and missing data in the outcome are considered via inverse intensity of visit weights (discussed more in detail in Sections 2.4 and 2.5.2). For the monitoring model, one must make assumptions about whether the outcome is observed (or missing) *completely at random*, *missing at random*, or *missing not at random*; this is discussed further in Section 2.4. Other missing data, such as missing values in the covariates from the exposure and the monitoring models, may have to be considered. Multiple imputation (MI) (Rubin [1976]) is a common approach to addressing missing data, if it may be assumed that they are *missing at random*. That is, MI does not provide a solution for data which are missing not at random, a type of missingness which implies that covariates other than those available to us predict the missingness mechanism.

MI replaces the missing values by a "sensible" prediction, allowing standard, complete-case analyses to be performed. The strength of this approach is that it accounts properly for the variance due to imputation. To impute the dataset, sampling from different predictive distributions (one for each covariate to impute) is performed several times (e.g., 5 or 10 times). As an example, the Normal distribution can be used as the predictive distribution for a continuous variable, and missing values can be replaced by a random draw from the Normal distribution, where the Normal mean is a function (e.g., a linear combination) of other covariates associated with the variable which must be predicted. After imputation, the parameter of interest is computed on each (completed) replicate dataset. Estimates are combined across all replicates using what is known as Rubin's rule (Rubin [2004]) -

essentially a simple average for point estimates, and a weighted average for standard errors. In particular, the variance (or standard error) of the final estimator accounts both for the *between* variance (across all replicate datasets) and the average *within* variance (the variance in one dataset). This variance is referred to as the *pooled variance*. Alternatively, a non-parametric bootstrap can be used to obtain an estimate of the pooled variance by resampling individuals or observations and by reproducing the complete multiple imputation procedure within each bootstrap-resampled dataset.

2.2 Causal inference

2.2.1 Potential outcome framework and average treatment effect

The (utopic) gold standard for inferring on a causal marginal effect is an experiment where one has access to the outcomes of all individuals from the population, under all possible levels of a treatment. For such an experiment, inferring whether a treatment causes an outcome would be as easy as contrasting the outcome across the different treatment levels. The fact that this experiment is impossible to conduct relates to what Holland calls the fundamental problem of causal inference (Holland [1986]): we cannot observe the outcomes of an individual under many different levels of a treatment, but only one of them, at a given time. For that reason, causal inference can be seen as a missing data problem where we aim to infer the values of these other, *counterfactual* outcomes that are never observed, i.e. the outcomes corresponding to levels of treatment other than the one actually received, for each individual under study.

The idea of potential outcomes goes back at least as far as 1923. Neyman [1923] used the term *unknown potential yields* in an agricultural science study in which he compared yields of different crop varieties. That study was experimental; the first to propose a similar idea in an observational setting was Rubin [1974]. In his seminar paper, Rubin defined a causal effect as the difference between two outcomes: one, had the patient received treatment

E (experiment), and one, had the patient received the treatment C (control). Following Neyman and Rubin, the ideas of potential, or *counterfactual* outcomes (a term primarily used in philosophy; see e.g. Lewis [1973]) became increasingly used by others (see e.g. Rubin [2005] for an interesting discussion on the use of those terms). It became the most popular framework, or way of reasoning about causal effects in statistics. Effectively, it provided a way to define causal estimands and to translate those estimands to the statistical framework, where models can be used for estimation.

We now give a brief example of the notation used in the potential outcomes framework for the case where the exposure is binary and where the outcome is continuous and longitudinal. Denote by $Y_{i1}(t)$ the potential outcome of individual i at time t, was individual i exposed to treatment 1, and denote by $Y_{i0}(t)$ the potential outcome of individual i at time t, if they were exposed to treatment 0. As mentioned earlier, individual i can only be exposed to one of the two treatment options (0 or 1) at time t, so only one of these potential outcomes will actually be observed. The quantity $\mathbb{E}\left[Y_{i0}(t) - Y_{i1}(t)\right]$ (or, in the cross-sectional study case, $\mathbb{E}\left[Y_{i0} - Y_{i1}\right]$) corresponds to the ATE or the causal marginal effect of treatment. It is the average difference in outcomes, had everyone in the population been treated with treatment $I = 1$, against everyone been treated with treatment $I = 0$ (at time t). The ATE is a common causal estimand, as it can provide guidance on how to improve public health at a global level since this estimand targets policy-type questions (applying to a population rather than an individual or sub-population).

The ATE literature has largely focused on a continuous outcome setting, however marginal causal estimands can be defined for other outcome types as well. For example, in studies with binary outcomes, $\mathbb{E}\left[Y_{i0}(t)\right]$ is the probability of $Y_{i0}(t)$ taking the value 1. The marginal risk difference, defined as $\mathbb{E}\left[Y_{i1}(t)\right] - \mathbb{E}\left[Y_{i0}(t)\right]$, or other causal contrasts, such as the marginal

odds ratio (MOR):

$$MOR = \left\{ \frac{\mathbb{E}\left[Y_{i1}(t)\right]}{1 - \mathbb{E}\left[Y_{i1}(t)\right]} \right\} \Big/ \left\{ \frac{\mathbb{E}\left[Y_{i0}(t)\right]}{1 - \mathbb{E}\left[Y_{i0}(t)\right]} \right\},$$

are contrasts of interest. The MOR is also a common causal contrast of interest in studies where the outcome is categorical.

The potential outcome framework allows causal contrast of interest to be defined clearly at a conceptual level, however such contrasts cannot be estimated without some identifiability assumptions that link together the causal and the statistical frameworks and, importantly, the potential and actually observed outcomes.

2.2.2 Identifiability assumptions

Some common identifiability assumptions are needed in order for the causal estimand to identified or estimated via statistical models (Hernán and Robins [2016], Chapter 3). These assumptions are:

1. (A1) The *Stable Unit Treatment Value Assumption* (SUTVA) generally encompasses two assumptions, the *no interference* assumption, and the *consistency* assumption. *No interference* means that the exposure of an individual does not affect the outcome of another individual in the study. Under that assumption, we do not have to consider other individuals' exposures in the estimation of the exposure effect for one individual and, importantly, the potential outcome for an individual depends only on their exposure and not that of others. The *consistency* assumption, or the consistency of the potential outcome, states that the outcome we observe in an individual who received treatment/exposure 1 is equal to their potential outcome under exposure 1, and reciprocally for $I = 0$. That is, the exposure or intervention is well-defined (there are not several versions of the exposure encompassed in a single level and, therefore, there are not several versions of the outcome under that exposure), and the outcome we observe

is indeed equal to the potential outcome. The assumption thus states that $Y_{i1} = Y_i$ if $I_i = 1$, and $Y_{i0} = Y_i$ if $I_i = 0$.

2. (A2) The assumption of *conditional exchangeability* (of exposure groups). Conditioning on a set of covariates (e.g., a set of potential confounders $\mathbf{K}$, generally the pre-treatment variables), we assume that the exposure assigned does not actually affect the potential outcomes. That assumption is expressed by

$$Y_{i0}, Y_{i1} \perp I_i | \mathbf{K_i}.$$

That is, each individual "possesses" two potential outcomes (for an exposure that is binary), and the fact they will be assigned exposure 1 or 0 will not affect the "fixed" values (deterministic, although this is controversial: see e.g. Hernán and Robins [2016], technical point 1.2) of their potential outcomes, conditional on the set of covariates $\mathbf{K}$. Alternatively, this may be interpreted as the assignment of treatment is not done with knowledge of how the individual would respond to the two (or multiple) treatment alternatives. That is, within levels or strata defined by the confounders, those individuals who would respond best to one treatment option do not preferentially receive it. Exchangeability, in general, means that the potential outcome averages would be the same in two different treatment groups randomly selected from the population of interest, no matter which treatment they are assigned. It relies on the assumption that the groups are equivalent in their characteristics to random samples from the population of interest, and therefore, it does not matter to which group we assign exposure 1 or 0; both can be used to estimate the effect of all exposure levels. That assumption also relates to the confounding factors we discussed earlier. The set $\mathbf{K}$ must contain all potential confounders (discussed in Section 2.1.1) such that conditioning on these covariates will adjust for any *backdoor path* between the exposure and outcome (these paths are not due to the actual causal effect of exposure but that creates an association

between the exposure and the outcome which precludes exchangeability of the groups unless accounting for them covariates).

3. (A3) *Positivity of the exposure* (conditional on the set of variables that led to conditional exchangeability, e.g. $\mathbf{K}$) requires $0 < P(I_i = 0|\mathbf{K_i}), P(I_i = 1|\mathbf{K_i}) < 1$. If patients have no chance of receiving the intervention conditional on their characteristics, or if they have no chance of not receiving the intervention conditional on their characteristics, then estimating the exposure effect in the corresponding regions of the covariates space relies on an extrapolation (which is possible via a parametric, unsaturated model, but could provide an unreliable estimate for the ATE). Empirical positivity is sometimes necessary to compute an estimate of the ATE, such as when using inverse probability of treatment weights. In that case, if positivity is not met, the estimator is undefined.

Suppose now that we aim to estimate the causal ATE in a setting where the exposure is binary (0 or 1), the outcome process is longitudinal and measured repeatedly, and where the set of potential cofounders is denoted by $\mathbf{K}(\mathbf{t})$ such that we have $Y_{i0}(t), Y_{i1}(t) \perp I_i(t)|\mathbf{K_i}(\mathbf{t})$. Then, when putting (A1), (A2) and (A3) together, we obtain:

(no interference)

$$
\begin{aligned}
\mathbb{E}[Y_{i1}(t) - Y_{i0}(t)] =& \mathbb{E}[Y_{i1}(t)] - \mathbb{E}[Y_{i0}(t)] \text{ by the linearity of the expectation} \\
=& \mathbb{E}_K[\mathbb{E}[Y_{i1}(t)|\mathbf{K_i}(\mathbf{t})]] - \mathbb{E}_K[\mathbb{E}[Y_{i0}(t)|\mathbf{K_i}(\mathbf{t})]] \text{ by the iterated expectation} \\
=& \mathbb{E}_K[\mathbb{E}[Y_{i1}(t)|I_i(t) = 1, \mathbf{K_i}(\mathbf{t})]] - \mathbb{E}_K[\mathbb{E}[Y_{i0}(t)|I_i(t) = 0, \mathbf{K_i}(\mathbf{t})]] \text{ by} \\
& \text{conditional exchangeability and positivity} \\
=& \mathbb{E}_K[\mathbb{E}[Y_i(t)|I_i(t) = 1, \mathbf{K_i}(\mathbf{t})]] - \mathbb{E}_K[\mathbb{E}[Y_i(t)|I_i(t) = 0, \mathbf{K_i}(\mathbf{t})]] \text{ by consistency.}
\end{aligned}
$$

That is, we can estimate the causal contrast of interest by using the last row above, using for instance a parametric model for the expectation of $Y_i(t)$ given the exposure and covariates

$\mathbf{K_i(t)}$ and averaging across the $\mathbf{Ks}$. That last row is sometimes referred to as the *g-formula* (Robins [1986]), a generalization of the more common standardization method. Another decomposition of the last row above will prove useful in Section 2.5.1 of this chapter, when we discuss inverse probability of treatment weights.

In practice, there exist methods to assess the different identifiability assumptions discussed in this section. For instance, the component of the SUTVA assumption corresponding to the "well-defined intervention" is typically subjective, but methods are currently developed to assess the hypothesis of no interference across individuals, or to explicitly model the spillover when it exists (see e.g., Hudgens and Halloran [2008], Aronow [2012], Saveski et al. [2017], Arpino and Mattei [2016], Forastiere et al. [forthcoming]). To assess the positivity assumption, one may model the probability of exposure conditional on covariates, and can compare the overlap of that probability in the treated and the untreated (Crump et al. [2009], Garrido et al. [2014]). Different sensitivity analyses have been developed, too, to assess the effect of unmeasured confounders (intimately related to the exchangeability condition), see e.g., Robins et al. [2000b], McCandless et al. [2007], Groenwold et al. [2016].

The use of causal diagrams is often suggested in the literature to complement the potential outcome framework and the related causal estimation and inference. In particular, causal diagrams can be used to assess the assumption of conditional exchangeability (given, of course, the additional assumptions on data structure that are made and encoded in the diagram). I now formally introduce directed acyclic graphs, a type of causal diagram commonly used in the causal inference literature.

2.2.3 Directed acyclic graphs

Directed acyclic graphs (DAGs) are a nonparametric tool based on structural equation models (Pearl [1998]) which are used to depict the assumed causal links between a network of covariates (examples were presented in Figures 2.1, 2.2 and 2.3). DAGS are made up of

nodes (vertices) that represent each variable, and edges (the arrows), which link the vertices together. The word *directed* refers to the edges between variables, which are supplemented with an arrowhead on one side, while the term *acyclic* means that, starting from a given node (e.g. node A), one cannot return to A from a causal path (i.e., one with all arrows pointing in the same direction). In a setting where DAGs are used to assess the causal effect of an exposure on an outcome, both the exposure and outcome should be included, as well as any variable that is a common parent to any pair of variables in the DAG (Pearl [1995]).

DAGs can be used to assess which set of covariates must be conditioned upon to reach conditional eachangeability. Effectively, they can be used to complement inference under the potential outcome framework, as they help in understanding whether an association between two covariates in the DAG is truly causal, or if (part of) it is not. The *backdoor* criteria (Pearl [1995]) is easily verified by using a causal diagram. That criteria ensures that no backdoor path exists, from the exposure to the outcome, that is unblocked and that is not due to the actual causal effect of the exposure on the outcome (see, for instance, Shrier and Platt [2008] for a six-step approach to ensuring no biasing backdoor path after adjusting for a set of covariates). Once these paths are blocked by a set of covariates, then we have conditional exchangeability (as discussed in Section 2.2.2) conditional upon that set. Potential bias due to selection on colliders can also be verified by depicting relations in a causal diagram. The DAG should contain any covariate for which there is a selection, such that we can determine whether there is potential for collider stratification bias. As noted above, the DAG itself represents an encoding of assumptions about the causal dependencies between variables and so any conclusions on, for example, the variables that must be accounted for to break or block backdoor paths implicitly rely on the DAG correctly representing true variables dependencies.

2.3 Outcome modelling and estimation

The estimators we propose in all three manuscripts of this thesis rely on estimating equations (also called M-estimators). Generalized estimating equations, or M-estimators, which are slightly different from the former, are an effective way of modelling the marginal effect of an exposure, as they were proposed to model population average quantities without explicit concern for the correlation structure between measurements. Furthermore, these equations can easily incorporate weights. In the problem that interests us, confounding and covariate-driven monitoring times can be accounted for by using "double" weights. The choice of using estimating equations is therefore natural, and existing theory on two-step estimators (discussed further in this section) allows for the consideration of variance components due to both the parameters in the weight models and the parameter(s) of interest.

2.3.1 Generalized estimating equations

The study of correlated data can be divided in two main categories, where one either wishes to study the trend of correlated measurements over time (in which case the correlation is paramount), or to study the effect of covariates on e.g. the mean outcome (in which case the parameters of interest are the covariates' effects, and the correlation in measurements are nuisance parameters). Generalized estimating equations (GEEs) were proposed by Liang and Zeger [1986] as a way to solve the latter issue. The authors discussed their methodology in a series of papers, including two seminal papers; see Zeger and Liang [1986] and Liang and Zeger [1986].

GEEs are extensions of generalized linear models for correlated data, and utilize and extend the quasi-likelihood theory (Wedderburn [1974], McCullagh [1983]) to the multivariate case. Assuming that the variance is a function of the mean of the measurements (further referred to as outcomes), they rely on a correct specification of the mean outcome and variance to provide unbiased estimates for covariates' effects on a function of the mean outcome (thus,

they work under less parametric assumptions for the measurements than usual likelihood-based method, and only weak assumptions on the correlation across outcomes of the same individual are required). In particular, a misspecification of the correlation matrix in the GEE yet leads to consistent estimators, while correctly specifying the correlation matrix generally improves the efficiency of the estimation. In cases where covariates vary in time, however, as discussed in the cautionary note from Sullivan Pepe and Anderson [1994], either an independence working covariance matrix should be used, or a specific assumption on the dependence of the outcome marginal mean on the time-varying covariates should be verified to ensure consistency of GEEs.

Denote by $\boldsymbol{\mu_i} = h(\mathbf{X_i}\boldsymbol{\beta})$ the mean of the outcome vector $\mathbf{Y_i}$ (for simplicity, assume in this section that the outcomes are observed at some pre-specified times, common across individuals, such that $Y_i(T_{i1}), ..., Y_i(T_{iV_i})$ can be denoted by a vector of same length for each individual, $\mathbf{Y_i}$). Let $\boldsymbol{\nu_i} = g(\boldsymbol{\mu_i})/\phi$ be the outcome variance, with ϕ a scale parameter. Under correct specifications of the mean and variance, a standard score equation can be used to estimate the parameters $\boldsymbol{\beta}$ in settings with *i.i.d.* outcomes. That equation is provided by

$$S_k(\boldsymbol{\beta}) = \sum_{i=1}^{n} \mathbf{D_i}\boldsymbol{\nu_i}^{-1}(\mathbf{Y_i} - \boldsymbol{\mu_i}) = \mathbf{0} \tag{2.1}$$

for an outcome distribution in the exponential family (note, $\mathbf{D_i} = \partial\boldsymbol{\mu_i}/\partial\beta_k$ is the matrix of partial derivatives). Liang and Zeger [1986] innovated by incorporating information from a working correlation matrix in those equations, which allowed for several measurements from the same individual (and thus naturally dependent) to be considered. They demonstrated that efficiency can be gained when the working correlation matrix is correctly specified.

For $\mathbf{R_i}$ the correlation matrix between measurements of the same individual (the authors actually used a parametric working matrix), and for $\mathbf{A_i}$ a diagonal matrix with the variance elements $g(\mu_{ij}) = \nu_{ij} \times \phi$ on its diagonal, they defined a new covariance matrix $\mathbf{V_i} =$

$\mathbf{A_i^{1/2}R}(\boldsymbol{\alpha})\mathbf{A_i^{1/2}}\phi$ and plugged it in the equation, to obtain

$$S'_k(\boldsymbol{\beta}) = \sum_{i=1}^{n} \mathbf{D_i}\mathbf{V_i}^{-1}(\mathbf{Y_i} - \boldsymbol{\mu_i}) = \mathbf{0}. \tag{2.2}$$

The authors showed that estimators based on these GEEs are consistent and asymptotically normal (under some additional assumptions), with a consistent variance estimator provided by the so called *sandwich* estimator.

To solve GEEs, iterative methods (that fluctuate between the regression coefficients and the scale and correlation parameters to provide the best fit) were proposed, in which nuisance parameters are estimated using the method of moments (or alternatively, another set of estimating equations), and the parameters of interest, $\boldsymbol{\beta}$, using a modified Fisher scoring algorithm. These approaches are implemented in standard statistical software; see e.g. the R packages *gee* and *geepack* (ported to R: Carey [2019], Halekoh et al. [2006]).

2.3.2 M-estimators and estimating equations

M-estimators, equivalent to the Generalized Method of Moments in econometrics (Huber [1964], Huber [1967], Chamberlain [1987], Newey [1988], Newey and McFadden [1994]), were "rediscovered" in the GEE seminal papers of Liang and Zeger in the eighties. They, too, are based on the work of Huber. A great discussion on the generality of the M-estimation approach and a review can be found in Stefanski and Boos [2002].

A large variety of asymptotically normal estimators can be thought of as M-estimators. Assuming that a solution exists, an M-estimator $\widehat{\boldsymbol{\beta}}_M$ solves the following vector equation for $\boldsymbol{\beta}$:

$$S_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{\psi}(\mathbf{Y_i}, \boldsymbol{\beta}) = \mathbf{0}. \tag{2.3}$$

That is, if one can transform an existing estimator into the solution for the equation in (2.3),

then the respective estimator is also an M-estimator. The maximum likelihood estimator (MLE), under regularity conditions, is an example of an M-estimator, as the derivative of the log-likelihood is zero at the value of the MLE. When an estimator does not clearly correspond to an M-estimator, it could still be part of a multidimensional system of equations leading to (2.3), such that the estimator may be called a *partial* M-estimator (see e.g. the example in Stefanski and Boos [2002], for the average absolute mean difference). In the statistical literature, M-estimators were developed using semiparametric theory of influence functions.

2.3.3 Asymptotic properties of M-estimators

The general proof of asymptotic normality of M-estimators calls upon the weak law of large numbers, the central limit theorem (CLT) and Slutsky's theorem. Let $\xrightarrow{l}$ refer to convergence in law, and $\xrightarrow{p}$ refer to convergence in probability. Consider the definition of an M-estimator given in (2.3). If the function $\psi(\cdot)$ is smooth, and if there is a unique solution $\boldsymbol{\beta}_0$ to (2.3), there exists a sequence of M-estimators that converges in probability to the true parameter (Huber [1967]). The function S_n in (2.3) can be decomposed into a Taylor series expansion around the true value, as follows:

$$S_n(\widehat{\boldsymbol{\beta}}_M) = S_n(\boldsymbol{\beta}_0) + \left.\frac{\partial S_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} (\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta}_0) + \mathcal{O}(n), \tag{2.4}$$

for $\mathcal{O}(n)$ a remainder term of order n.

Assuming that the matrix $\left.\frac{\partial S_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$ is non singular and, as such, that we can multiply by the inverse of that matrix on each side, and further rearranging the terms and multiplying by the square root of n, we obtain

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta}_0) = \left[\left.\frac{\partial S_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right]^{-1} \sqrt{n} S_n(\boldsymbol{\beta}_0) + \sqrt{n}\mathcal{O}^*(n) \tag{2.5}$$

for $\mathcal{O}^*(n)$ a new remainder term of order n. We then call upon all three theorems mentioned earlier; we have

$$-\frac{\partial S_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \xrightarrow{p} \mathbb{E}\left[-\frac{\partial \boldsymbol{\psi}(\mathbf{Y_1},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right] \text{ by the weak law of large numbers, and} \quad (2.6)$$

$$\sqrt{n}S_n(\boldsymbol{\beta}_0) \xrightarrow{l} MVN(0, \mathbb{E}[\boldsymbol{\psi}(\mathbf{Y_1},\boldsymbol{\beta_0})\boldsymbol{\psi}(\mathbf{Y_1},\boldsymbol{\beta_0})^{\mathbf{T}}]) \text{ by the central limit theorem,} \quad (2.7)$$

assuming that $\mathbb{E}\left[-\frac{\partial \boldsymbol{\psi}(\mathbf{Y_1},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right]$ and $\mathbb{E}[\boldsymbol{\psi}(\mathbf{Y_1},\boldsymbol{\beta_0})\boldsymbol{\psi}(\mathbf{Y_1},\boldsymbol{\beta_0})^{\mathbf{T}}]$ exist and are finite.

Under further regularity conditions detailed in Huber [1967], and provided that (2.3), (2.6) and (2.7) hold, using Slutsky's theorem, we obtain that

$$\widehat{\boldsymbol{\beta}}_M \xrightarrow{l} MVN\left(\boldsymbol{\beta}_0, \frac{\mathbf{V}(\boldsymbol{\beta}_0)}{n}\right) \text{ as } n \to \infty, \quad (2.8)$$

where $\mathbf{V}(\boldsymbol{\beta}_0) = \mathbb{E}\left[-\frac{\partial \boldsymbol{\psi}(\mathbf{Y_1},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right]^{-1} \mathbb{E}[\boldsymbol{\psi}(\mathbf{Y_1},\boldsymbol{\beta_0})\boldsymbol{\psi}(\mathbf{Y_1},\boldsymbol{\beta_0})^{\mathbf{T}}] \left(\mathbb{E}\left[-\frac{\partial \boldsymbol{\psi}(\mathbf{Y_1},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right]^{-1}\right)^T.$

2.3.4 Extension to two-step estimators

The estimators developed in this thesis belong to the class of two-step estimators, a class of M-estimators that rely upon a first step (the estimation of nuisance parameters) and a second step (the estimation of the parameter of interest, relying on substitution or plug-in values of the nuisance parameters). In our case of interest, nuisance parameters are those from the exposure and the monitoring models and the marginal effect of exposure is the parameter of interest. Newey and McFadden [1994] have shown how to develop the variance of these estimators. Their approach provides a consistent estimator for the variance that accounts for the variability due to both steps in the two-step procedure described above. Two-step estimators are discussed more in depth in the supplementary material of the first manuscript of this thesis.

2.4 Methods for covariate-driven monitoring times

Several methods have been proposed to account for covariate-driven monitoring times in contexts where the aim is to estimate marginal or conditional effects, with most methods being extensions of GEEs. In this thesis, I assume throughout that the monitoring times coincide perfectly with the observation of the outcome process, and that a (sufficient) set of covariates that predicts the monitoring times is available for analysis, thereby the assumption of covariate-driven monitoring times is equivalent to that of the outcome being missing *at random* (as opposed to *completely at random*, or *not at random*).

Robins et al. [1995] proposed the first extension of GEEs which allowed for the longitudinal outcome, the vector $\mathbf{Y_i}$, to contain missing values. The authors primarily considered the monotone missingness patterns for $\mathbf{Y_i}$ (which gave the name to their method: inverse probability of *censoring* weights). Individuals were allowed to have repeated outcome measurements in time, but these had to correspond to a set of common observation times across patients. The missingness was modelled as a function of patients' characteristics, typically by using a logistic regression and including information on previous outcomes or covariate history in the set of predictors; the resulting fitted probabilities were used to construct inverse weights. However, the method does not apply to observational studies with highly irregular visit patterns across the patients.

Lin and Ying [2001] proposed a new methodology, based on a slightly different set of estimating equations than the previous authors. These authors also focussed on the estimation of marginal effects of covariates, but they aimed to avoid estimating the intercept function $\alpha_0(t)$ in the mean outcome model given by $\mathbb{E}[Y_i(t)|\mathbf{X_i(t)}] = \alpha_0(t) + \boldsymbol{\beta}_0'\mathbf{X_i(t)}$, all while allowing irregular observation times. For that, they re-centered the more common GEEs by a mean function of the design matrix that considered the monitoring process, avoiding the estimation of the intercept function and therefore providing a less parametric estimator for $\boldsymbol{\beta}_0$. For the monitoring model, they innovated on Robins et al. [1995] by using a propor-

tional rate model, following previous work of Lawless and Nadeau [1995], Lin et al. [2000], and Pepe and Cai [1993]. This model considered observation times that occurred at any (continuously measured) time, and varied across patients. The proportional rate model assumed that $\mathbb{E}[dN_i(t)|\mathbf{Z_i(t)}] = \exp(\boldsymbol{\gamma}'\mathbf{Z_i(t)})d\Delta_0(t)$ for $dN_i(t)$ an indicator of visit/observation for individual i at time t, $N_i(t)$ a counting process for observation times in individual i that counts the number of previous observation times by time t, and $\Delta_0(t)$ any non-decreasing function in time. Due to the structure of their set of estimating equations, the method of Lin and Ying yet only allowed for covariates $\mathbf{Z(t)} \subset \mathbf{X(t)}$ (included in the design matrix) to affect the monitoring rate.

Several extensions of the Lin and Ying equations followed. Sun et al. [2005] included an additional term in the outcome mean model considered by Lin and Ying which depended directly on the monitoring process $N_i(t)$. Thus, they considered conditional effects of covariates $\mathbf{X(t)}$, rather than marginal effects. The authors argued that this type of model is more useful for situations in which prediction is more of interest than inference, or where independence tests for the outcome and monitoring processes would be necessary. As the Lin and Ying equations, Sun et al.'s method requires for the outcome to be missing at random. In cases where the outcome is missing *not a random*, however, the covariates available in the analysis alone cannot be used to break the dependence between the monitoring and the outcome processes. To solve this issue, several authors proposed different extensions of the Lin and Ying equations in which random effects are incorporated to consider dependencies due to unmeasured (latent) variables (see e.g. Sun et al. [2007], Liang et al. [2009], Sun et al. [2011]). Among them, Sun et al. [2011] proposed to jointly model the monitoring, the outcome, and the censoring processes all together via a latent effect.

The Lin and Ying equations considered the monitoring process directly within their estimating equations. A similar model for the monitoring as considered in Lin and Ying was also used by Lin et al. [2004], in the form of inverse intensity of visit (IIV) weighting. That method

is discussed more in depth in Section 2.5.2 where I discuss inverse probability weights. The term *proportional rate* model (as used by Lin and Ying [2001]) is used in settings where the monitoring indicator is modelled marginally, as a function of current covariates, while the term *proportional intensity* model (as used by Lin et al. [2004]) corresponds to cases where the monitoring indicator at time t is modelled as a function of the past, including previous covariates, outcomes or monitoring times (something that has been noted in Lindsey [2004] and others, although there seems to be confusion about the distinctions between the rate and the intensity models in the literature; I come back to this distinction in Section 2.5.2). An innovation of Lin et al. [2004] over the Lin and Ying equations is that the inverse intensity of visit weight can account for more covariates in the monitoring model, since these weights can incorporate any covariate, even those not contained in the design matrix $\mathbf{X}(\mathbf{t})$.

Combinations of the previous methods followed. First, note that the methodology proposed by Lin and Ying [2001] is especially useful when one prefers to avoid estimating the mean baseline function, a function of time, in the outcome model. However, the weights proposed in Lin et al. [2004] allow the Lin and Ying equations some more flexibility to account for other covariates (such as mediators of the design matrix covariates) to affect monitoring times. Bůžková and Lumley [2009] proposed an extension where they combined both methods to estimate marginal effects of time-dependent covariates (as opposed to the marginal effects of time-fixed covariates considered in Lin et al. [2004]); this extension is discussed more in depth in Chapter 3 of this thesis. Similarly, Tan et al. [2014] proposed a few possible extensions or combinations of the previous methods based on the Lin and Ying equations; in particular, they considered time-dependent covariates in the method of Liang et al. [2009], and added flexibility to that method as well as that of Sun et al. [2007] by incorporating IIV weights.

In parallel to methods based on estimating equations, some authors proposed fully parametric methods. Assuming that the outcome measurements are multivariate normally distributed,

Lipsitz et al. [2002] computed the full likelihood for the outcome and decomposed it in a monitoring and an outcome components. Under certain assumptions about ignorability of the monitoring process, this resulted in the possibility of ignoring completely the monitoring process to make inference on the marginal effects via the conditional outcome process likelihood only. However, that method relied on the strong assumption of outcome normality. Another fully parametric method was proposed by Ryu et al. [2007], who extended the work of Lipsitz et al. They first extended the monitoring process using a nonhomogeneous mixed Poisson process for the intensity (for that, they incorporated a random effect multiplier to the monitoring intensity function). These authors, unlike Lipsitz et al., also worked under the Bayesian paradigm to estimate the regression coefficients in the outcome model, which allowed them to fit different correlation matrices for the measurements of each individual in time.

A doubly robust approach was proposed by Pullenayegum and Feldman [2013] – one of the only methods proposed for covariate-driven monitoring times that relies on a strategy where only one of either the outcome or the monitoring process needs to be correctly specified to yield a consistent estimator. For the outcome model, the authors used a model for the time-dependent increment in the outcome as a function of covariates and a martingale residual, such that the distribution of residuals could be left unspecified. Doubly robust equations were then built by adding a new term in the standard IIV-weighted GEEs. That additional term is a product of residuals from the outcome increment model and the monitoring model, such that a zero-mean estimating equation would ensure orthogonality between the outcome and monitoring processes simultaneously with the double robustness property.

A thorough summary of methods for covariate-driven monitoring times was put forward by Pullenayegum and Lim [2016], who discussed the different assumptions of several of those methods mentioned above. Other methods for tackling covariate-driven monitoring times followed; see e.g. McCulloch et al. [2016], Dai and Pan [2018], Shahn et al. [2019]

for similar methods or extensions of those discussed above, which do not apply directly in our setting of interest (either because they use random effects and hence are not marginal, they focus only on survival types of outcomes, or they do not allow for the same extent in visit irregularity as we wish to tackle). See Lokku et al. [2020] and Pullenayegum [2020] for guidance on visit irregularity (e.g., for suggestions on observational study designs or on how to summarize the irregularity). Further, see Pullenayegum [2016] for an extension of a strategy called *outputation* to irregular visits. That methodology thins the visit process to discard observations in such a way as to induce independence between the covariates and the monitoring process in the data remaining. Outputation leads to results that are similar to those found using inverse intensity of visit weights, and can prove useful in settings where inverse weights cannot be implemented (e.g., in joint models).

In the causal inference framework, few authors have tackled the issue of covariate imbalances related to the monitoring times. For instance, Zhu et al. [2017] proposed weighted estimating equations for the parameters in a parametric survivor function $S(\cdot)$, for settings with covariate-driven monitoring times (accounted for via weights). They considered a function $S(\cdot)$ of time-fixed covariates, in which case the model parameters could correspond to marginal effects. In the case where an exposure was incorporated in the design matrix, they proposed to use inverse probability of treatment weights to adjust for potential confounding. Their method applies to survival-type of outcomes, and for time-fixed variables (e.g., exposure) in the design matrix. Shahn et al. [2019] introduced a method set within the causal inference framework. They proposed to use g-computation (Robins [1986]) rather than inverse probability of treatment weighting, to account for confounding bias in the estimation of the effect of a sequence of treatments. They defined the causal contrast of interest as a function of a longitudinal potential outcome, and estimated it by using g-computation and conditioning on a fixed visit scheme in the outcome model used for prediction of the counterfactual outcomes. Although the approach considered sequences of treatments as measured at different monitoring times, only slight deviations from a common visit scheme across

individuals can be accommodated.

In this thesis, I focus on approaches similar to that of Zhu et al. [2017] but for other types of outcomes, and extensions of the method of Bůžková and Lumley [2009] so as to address confounding. I assume that, due to the visit process, the outcome is missing at random (but not completely at random) and treat the parameters in the exposure and the monitoring models as nuisance parameters. I further assume that we can create a DAG faithful to the generating mechanisms in nature. I aim to provide a methodology for situations where the aim is to estimate a population average (marginal) quantity, and thus focus on estimating equation approaches rather than random effect models.

2.5 Inverse probability weighting

Inverse probability weights have been used in many different settings, but one of their first uses can be traced back to Horvitz and Thompson [1952] who aimed to estimate a summary measure from a population, using a sample where units were sampled with unequal probabilities. In their problem, the probability of selection of a unit u_i (associated with measurement Y_i) was unequal across individuals as it depended on the order of selection into the sample. Using the sample, the authors wished to estimate the sum of an outcome Y in the population. The authors focused on estimators of the form

$$\hat{T} = \sum_{i=1}^{n} \Omega_i Y_i,$$

for Ω_i, $i = 1, ..., n$ a term that would be used as a weight when unit u_i is selected into the sample. To obtain $\mathbb{E}[\hat{T}] = T$, which corresponds to

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^{n} \Omega_i Y_i\right] &= \sum_{i=1}^{n} p(u_i)\Omega_i Y_i \\ &= \sum_{i=1}^{n} Y_i,\end{aligned}$$

they showed that the weights Ω_i should be equal to the quantity $1/p(u_i)$, which corresponds to the inverse probability of selection into the sample for each unit.

2.5.1 Weighting in causal inference

In the causal inference literature, inverse probability weights are frequently used to create balance across exposure groups, in settings where a comparison of their outcome is of interest. These inverse probability weights, called the inverse probability of treatment weights (IPTW), are a function of the propensity score (Rosenbaum and Rubin [1983]), which is the coarsest possible *balancing score*, i.e. a function of the covariates $\mathbf{K}$ that, when conditioned upon, makes the covariates and treatment assignment $\mathbf{I}$ independent. In particular, the propensity score can be used in place of the confounding variables $\mathbf{K}$ to perform adjustment (e.g. weighting, stratification or matching) when it is difficult to stratify or match on a large set of covariates.

Rosenbaum and Rubin [1983] proved in their seminal paper that the propensity score (PS), defined as the conditional probability of being exposed, given pretreatment variables (typically, the confounders), was the coarsest balancing score. Denoting the propensity score by $e(\mathbf{K})$, *balance* means that

$$\mathbf{I} \perp \mathbf{K} | e(\mathbf{K}),$$

such that $e(\mathbf{K})$ represents a summary of covariates $\mathbf{K}$ that can be used alone to break the dependence between the exposure and covariates $\mathbf{K}$, rather than the full set of $\mathbf{K}$. In

particular, if we have that $\{\mathbf{Y(1)}, \mathbf{Y(0)}\} \perp \mathbf{I}|\mathbf{K}$, then we also have that $\{\mathbf{Y(1)}, \mathbf{Y(0)}\} \perp \mathbf{I}|e(\mathbf{K})$ where $e(\mathbf{K})$ is the true propensity score (associated with the true exposure generating mechanism).

Adjustment via the propensity score can be convenient in studies where there are imbalances across treatment groups, particularly when it is easier to correctly specify the treatment allocation than the outcome process. IPTW, as opposed to a direct adjustment for the covariates or some function of them such as the PS in the outcome model, separates the study design from the analysis, and it has been shown to provide good performance in terms of bias reduction (Austin [2011]). Thus, to make our methodology generalizable, we focus on the IPTW as the key method of adjusting for confounding. The IPTW are defined as the inverse of $\mathbb{I}(I_i = 1)P(I_i = 1|\mathbf{K_i}) + \mathbb{I}(I_i = 0)P(I_i = 0|\mathbf{K_i})$ for person i; they create a pseudo-population in which exposure is not associated with the potential confounders.

In Section 2.2.2, we presented the *g-formula*. To explain why inverse probability of treatment weighting is an effective approach to adjusting for confounding, I now continue the development from that section:

$$
\begin{aligned}
&\mathbb{E}_K[\mathbb{E}\left[Y_i(t)|I_i(t) = 1, \mathbf{K_i(t)}\right]] \\
&\quad = \sum_K \mathbb{E}[Y_i(t)|I_i(t) = 1, \mathbf{K_i(t)}]P(\boldsymbol{K} = \mathbf{K_i(t)}) \\
&\quad = \sum_K \frac{P(I_i(t) = 1|\mathbf{K_i(t)})}{P(I_i(t) = 1|\mathbf{K_i(t)})}\mathbb{E}[Y_i(t)|I_i(t) = 1, \mathbf{K_i(t)}]P(\boldsymbol{K} = \mathbf{K_i(t)}) \\
&\quad = \sum_K \frac{1}{P(I_i(t) = 1|\mathbf{K_i(t)})}\mathbb{E}[Y_i(t)|I_i(t) = 1, \mathbf{K_i(t)}]P(I_i(t) = 1|\mathbf{K_i(t)})P(\boldsymbol{K} = \mathbf{K_i(t)}) \\
&\quad = \mathbb{E}\left[\frac{\mathbb{I}(I_i(t) = 1)Y_i(t)}{P(I_i(t) = 1|\mathbf{K_i(t)})}\right].
\end{aligned}
$$

Similarly, it can be shown that $\mathbb{E}_K[\mathbb{E}\left[Y_i(t)|I_i(t) = 0, \mathbf{K_i(t)}\right]] = \mathbb{E}\left[\frac{\mathbb{I}(I_i(t)=0)Y_i(t)}{1-P(I_i(t)=1|\mathbf{K_i(t)})}\right]$. That is, IPTW can be used to obtain an estimator of the causal contrast of interest, provided conditioning in the PS is on the set of confounders necessary to eliminate all backdoor paths

between exposure and outcome.

The seminal paper of Rosenbaum [1987] is one of the firsts to discuss such a weighting (without relating it to what are now referred to as *g-methods*, simply calling the IPTW method a model-based direct adjustment). He compared the inverse probability of treatment weighting adjustment to direct adjustment for the PS in the regression model, and presented arguments in favour of inverse probability of treatment weighting, including the fact that this approach does not require a correct specification of the outcome model and a somewhat weaker assumption on positivity.

2.5.2 Inverse intensity of visit and recurrent events models

As mentioned in Section 2.4, inverse probability weights were used by Robins et al. [1995] to account for missingness in a longitudinal outcome. However, Lin et al. were the first to propose a more flexible type of weight that applied in settings where monitoring times for the outcome can occur at any time (i.e., in continuous time). Their weights are called the inverse intensity of visit (IIV) weights (Lin et al. [2004]) and can be used to create a pseudo-population (Hernán and Robins [2006]) in which monitoring times are independent of a given set of covariates, that we denote by $\mathbf{Z}$. In particular, we could have that $\mathbf{Z} = \mathbf{X}$ where $\mathbf{X}$ is the design matrix in the outcome mean model considered, but covariates $\mathbf{Z}$ could also include other covariates (such as those on the causal paths from covariates $\mathbf{X}$ to the outcome). The authors incorporated their IIV weights in GEE-type estimating equations, and showed the unbiasedness of their equations for the marginal effects of (time-fixed) covariates $\mathbf{X}$ using iterated expectation. They derived the asymptotic properties of the proposed estimators; effectively, their estimator corresponds to an M-estimator, such that the corresponding influence function can be used to derive asymptotic variance and to prove asymptotic normality. The estimating equation they used to intoduce the IIV weights

is given by

$$\sum_{i=1^n} \int_0^\tau \frac{\{Y_i(t) - \mu(t, \mathbf{X}_i; \boldsymbol{\beta}_0)\}\, \boldsymbol{c}(t, \mathbf{X}_i; \boldsymbol{\beta}_0)}{\widehat{\lambda_0(t)} \exp[\hat{\gamma}_0^T H\left\{t, \mathcal{F}_i^{obs}(t-)\right\}]} = \mathbf{0}, \tag{2.9}$$

where $\mu(t, \mathbf{X}_i; \boldsymbol{\beta}_0)$ is a marginal mean model for the outcome specified as a function of covariates $\mathbf{X}$ and the parameters of interest, and where the estimated monitoring intensity is found in the denominator. The function $\mathbf{c}(\cdot)$ can be used to improve efficiency of the resulting estimator. The authors used a proportional intensity model where they estimate the instantaneous probability of visit at time t as $\widehat{\lambda_0(t)} \exp[\hat{\gamma}_0^T H\left\{t, \mathcal{F}_i^{obs}(t-)\right\}]$, with a function H at time t of previous covariates $\mathcal{F}_i^{obs}(t-)$. This function could be multidimensional, with $\hat{\gamma}_0 = \hat{\boldsymbol{\gamma}}_0$ a vector. Methods from recurrent events modelling can be used to model that intensity function. The Andersen and Gill (AG) model (Andersen and Gill [1982]), an extension of the Cox proportional hazards model (Cox [1972]) to recurrent events has been used to estimate the parameters $\boldsymbol{\gamma}_0$ (see e.g. Lin et al. [2004], Bůžková and Lumley [2009]).

Recurrent events modelling

Monitoring times are recurrent events for a given individual, and they can be represented by a corresponding counting process $N_i(t) = \sum_{j=1}^t dN_i(j)$ for individual i. The modelling of covariates' effects on those recurrent events can be accomplished using standard modelling tools for recurrent events. Some reviews of methods for recurrent events in biostatistics can be found in Amorim and Cai [2015] and Ozga et al. [2018]. A wide-ranging reference book on recurrent events is that of Cook and Lawless [2007]. The different models proposed in the literature effectively differ by the assumptions that must be met to obtain valid inference. The most popular methods include the AG model; the Prentice, Williams and Peterson (PWP) models (we discuss two of them); and the Wei-Lin-Weissfeld model. Some other approaches, such as the marginal mean or rate modelling, the frailty models, or the multi-

state models, are also briefly considered in this section.

The AG model (Andersen and Gill [1982]) is a straightforward extension of the Cox proportional hazards model (Cox [1972]) to recurrent events. It uses the same partial likelihood theory to fit the main effects in the instantaneous probability function (called the *intensity* rather than the *hazard*, which emphasizes that it is not conditional on having survived up to a given time point), but it differs from the Cox proportional hazards model by the risk-sets being considered. With the AG model, patients with a previous event remain in the study until they are censored or until end of study, such that other future events are still considered. Lin et al. [2000] discussed the differences between the AG model and the more marginal models for the visit rate, such as the marginal rate or the marginal mean models (discussed further below); the latter of these models require weaker assumptions on the structure of dependence across the recurrent events of an individual. In contrast, the Andersen and Gill model relies on the assumption that the correlation between measurements of the same individual in time can be explained by the (time-varying) covariates in the model, such that conditioning on these variables leads to conditional independence. The AG model, as opposed to the other models discussed in the next paragraph, does not allow for the main effects of covariates in the intensity function to vary after a new event occurs (but functions of the previous events can be added as covariates in the intensity model).

The next two methods rely on stratified versions of the same intensity function considered by Andersen and Gill. Either the first event is treated differently, or the intensity is allowed to vary for each new event stratum. PWP proposed two useful models for the intensity (Prentice et al. [1981]). The first, also an extension of the Cox proportional hazards model, assumes a different baseline intensity function, and different covariate effects for each recurrent event:

$$\lambda_{ij}(t) = \lambda_{0j}(t) \exp(\boldsymbol{\gamma}_j' \mathbf{Z}_j)$$

for j the order of the event for an individual (where covariates $\mathbf{Z}_j$ can be re-assessed, for

instance, after each new event). The second model they proposed allows for the baseline function to depend on the time between events (i.e. the *gap time*) rather than the time since cohort entry:

$$\lambda_{ij}(t) = \lambda_{0j}(t - t_{j-1}) \exp(\boldsymbol{\gamma}_j' \mathbf{Z}_j).$$

The Wei-Lin-Weissfeld model (Wei et al. [1989]) is similar to those proposed by PWP, except that individuals are considered at risk for an higher-order event (e.g., a fourth one) even if they have not experienced the previous events yet (and similarly, for other orders). That is, individuals are considered to be at risk of a j^{th} event at all times.

Other methods can be used to add flexibility into the modelling of recurrent events. For instance, frailty models can be used to account for missing outcomes that are missing not at random, in which case covariates alone cannot necessarily break the dependence between the recurrent events and the covariates (in the case of covariate-driven monitoring times, this has been discussed in Section 2.4). Marginal models for the events rates or means can also be used (Lin et al. [2000]). These models are generally similar to the AG model, and consider the entirety of the recurrent events process as a counting process. However, the marginal models make weaker assumptions on the dependence between the recurrent events of the same individual by removing the assumption of the AG model that the influence of previous events on a recurrent event at time t must be mediated via the time-varying covariates in the model (e.g., $\mathbf{Z}(\mathbf{t})$). For marginal models, the structure of dependence across the events of an individual is allowed to be quite complex, and could go beyond the dependence due to the covariates $\mathbf{Z}(\mathbf{t})$.

Note that the AG model can be used to estimate the main effects both in the proportional intensity and in marginal rate models (under their respective assumptions, and under the AG assumption that subsequent events from an individual are independent conditional on the time-varying covariate process) (Lin et al. [2000]). In that thesis, I therefore fit intensity models rather than more marginal models, even in settings where assumptions on the

dependence across recurrent monitoring times are weak, because intensity models provide consistent estimates for covariates main effects (on the rate or visit intensity) for all the situations I consider, and because they are easier to implement. A Cox proportional hazards model is fitted using the *coxph* function in R (Therneau [2020]), where the data must first be transformed into a counting process format (e.g., one day per data row), and this method provides the estimates from the AG model. Finally, multi-state models can be used for estimating the transition intensities in settings with several (possibly, competing) recurrent events. These models and the corresponding theory can be used to compute other quantities such as the stage occupancy probability, which is the probability to be in one of the possible states (i.e. visit or not) at a given time. See Cook and Lawless [2018] for more details on multi-state models.

2.6 Summary

In this literature review, I discussed some of the key components for developing a methodology for estimation of the causal ATE using observational data that are subject to confounding, missing data, and covariate-driven monitoring times. These included a discussion on the challenges associated with the analysis of observational data, an introduction to some important statistical concepts in causal inference, a review of previously proposed methods that tackle some of the special features of observational data, and an introduction to the different modelling tools that are used in this thesis to estimate the ATE on a longitudinal outcome, along with the rationale for using them.

Chapter 3

Weighted regression analysis to correct for informative monitoring times and confounders in longitudinal studies

Preamble to Manuscript 1.

The authors Lin and Ying [2001] and Bůžková and Lumley [2009] have proposed semi-parametric estimators for the marginal effects of covariates on a longitudinal outcome, for settings with covariate-driven monitoring times, with their theory being mostly discussed outside of the causal framework. That, and other previous work on covariate-driven monitoring times, did not focus on exposure effects nor considered confounding simultaneously with the covariate-driven monitoring times. This gap motivated the proposal and the evaluation of different methodologies for tackling these features simultaneously.

In this manuscript, the methods of Lin and Ying [2001] and Bůžková and Lumley [2009] are extended, and two novel estimators are proposed and compared, with both accounting for covariate-driven follow-up times and confounding. The two estimators also account for mediators on the causal path from the exposure to the outcome that may affect monitoring

times. The two estimators are based on different sets of estimating equations and present with different theoretic properties. The first proposed estimator does not require the estimation of a time-varying intercept function in the outcome mean model. The second estimator incorporates splines to make the modelling of the intercept function more flexible, and its variance can easily be derived using semiparametric theory for two-step estimators.

The original contributions of this chapter are i) proposing the first consistent estimators for the ATE in settings with covariate-driven monitoring times and confounders, when the exposure is binary and the outcome is continuous, ii) proposing the first thorough design for simulation studies with confounding and covariate-driven monitoring times in the specific context where the exposure is binary and the outcome is continuous, and iii) deriving the asymptotic properties of the estimator that performed the best between the two proposed estimators.

The corresponding manuscript was published in *Biometrics* (Coulombe et al. [2021a]).

Weighted regression analysis to correct for informative monitoring times and confounders in longitudinal studies

Janie Coulombe[1], Erica E. M. Moodie[1], Robert Platt[1,2].

[1]*Department of Epidemiology, Biostatistics, and Occupational Health, McGill University*

[2]*Lady Davis Institute for Medical Research, Montréal*

This thesis contains the accepted version of the corresponding paper published in *Biometrics* (Coulombe et al. [2021a]).

Abstract

We address estimation of the marginal effect of a time-varying binary treatment on a continuous longitudinal outcome in the context of observational studies using electronic health records, when the relationship of interest is confounded, mediated and further distorted by an informative visit process. We allow the longitudinal outcome to be recorded only sporadically and assume that its monitoring timing is informed by patients' characteristics. We propose two novel estimators based on linear models for the mean outcome that incorporate an adjustment for confounding and informative monitoring process through generalized inverse probability of treatment weights and a proportional intensity model respectively. We allow for a flexible modelling of the intercept function as a function of time. Our estimators have closed-form solutions, and their asymptotic distributions can be derived. Extensive simulation studies show that both estimators outperform standard methods such as the ordinary least squares estimator or estimators that only account for informative monitoring or confounders. We illustrate our methods using data from the *Add Health* study, assessing the effect of depressive mood on weight in adolescents.

3.1 Introduction

Consider a setting where we are interested in understanding the cross-sectional impact of an exposure on an outcome, for example a physician is interested in the impact of a current clinical measurement on the current presence of an illness, and their patients are repeatedly assessed over time. To learn about such an association, we may turn to electronic health records (EHRs). Longitudinal outcomes, particularly those drawn from EHR data, may be measured irregularly across patients. The time points at which they are recorded may depend on patients' health condition, which may on its turn be linked with the value of the outcome measured at those visit times, leading to imbalances in the data similar to those observed in selection bias. Moreover, confounders and mediating variables (Greenland and Robins [1986]) occur simultaneously with informative monitoring times in most observational studies, and thus must also be accounted for.

In this work, we focus on making inference on the marginal effect of a binary, time-varying treatment on a continuous outcome, measured repeatedly over time. To model longitudinal outcomes in contexts where monitoring times are irregular or informative, several methods have been proposed, but none of them simultaneously considered confounding in a setting with continuous outcomes. When monitoring times are informative, Robins et al. [1995] proposed a weighted extension of the generalized estimating equations of Zeger and Liang [1986] to estimate the marginal effect of intervention on a longitudinal outcome. In their method, inverse probability of response weights were incorporated into estimating equations to adjust for nonrandom missingness, which addressed the problem of informative monitoring times but was restricted to the case where there is a common set of monitoring times for all individuals, which is often not the case in observational studies. Lin and Ying [2001] developed a class of closed-form estimators for the marginal effect of variables on the mean outcome that accounted for informative monitoring times and allowed for those times to vary across individuals. Several innovations followed, which we detail further in the following

section, many of which are covered in the review of Pullenayegum and Lim [2016].

We extend the existing literature further and propose two new and flexible estimators for the marginal effect of a (potentially time-varying) binary treatment on a longitudinal continuous outcome for settings in which the exposure is not randomized. In our methods, we allow the mediators, the exposure and other covariates to affect the timing of the outcome monitoring, and both the confounders and the mediators to vary in time. The first estimator is a semiparametric extension that builds on the work of Bůžková and Lumley [2009]. The second estimator is a weighted least squares type estimator that incorporates two time-varying weights. This latter flexible estimator provides a simpler and more intuitive alternative to the first, with comparable performance. Its asymptotic variance is derived. In simulation studies, we compare both estimators and more standard ones in different contexts of dependency between covariates and monitoring times.

The remainder of this article is organized as follows: Section 3.2 introduces the notation, assumptions and inference procedure. Section 3.3 presents the details of the simulation studies and the results. Section 3.4 applies the methodology to the analysis of the data from the *Add Health* study (Harris [2009]). Finally, we provide some concluding remarks in Section 3.5.

3.2 Methods

3.2.1 Background

Lin and Ying [2001] considered the following marginal model:

$$E\left[Y_i(t)|\mathbf{X_i(t)}\right] = \alpha(t) + \boldsymbol{\beta}'\mathbf{X_i(t)}, \tag{3.1}$$

with $\alpha(t)$ an arbitrary function of time t, $\mathbf{X(t)}$ a design matrix and $Y_i(t)$ a continuous longitudinal outcome. They assumed a proportional intensity model for the monitoring

times of the outcome, which monitoring times were only allowed to depend on covariates in the outcome model, $\mathbf{X(t)}$. They proposed a semiparametric estimator for $\boldsymbol{\beta}$ in (3.1), which does not require estimation of the intercept $\alpha(t)$. In 2009, Bůžková and Lumley proposed to incorporate a weight in Lin and Ying's estimator that accounts for the dependency between monitoring times and any covariates that are *not* in the design matrix $\mathbf{X(t)}$. In particular, their approach allows for any mediators of the treatment-outcome relationship to affect monitoring times.

Tan et al. [2014] presented a summary of some of the extensions of Lin and Ying's estimator and proposed a few developments of existing methods. Other authors have proposed fully parametric methods to jointly model the visit and outcome processes (Lipsitz et al. [2002]; Ryu et al. [2007]), or introduced shared latent effects to link the outcome and the visit processes (e.g., Sun et al. [2012]; Cai et al. [2012]). Most recently, Zhu et al. [2017] proposed an estimator for interval-censored outcomes when confounding and irregular visit times may be present. They were among the first to consider these two features, however the method is focused on a very particular outcome type.

The problem of accounting for mediators and confounding variables in observational studies has been addressed via several methods. It is now well-known that mediating variables should not be included in the design matrix of the outcome model if the estimand is the total effect of exposure on the outcome (Rosenbaum [1984]). Propensity score methods such as inverse probability of treatment (IPT) weights are commonly used to adjust for imbalances across treatment groups due to confounders (Rosenbaum and Rubin [1983]; Robins et al. [2000a]). The standard IPT weight for a binary and time-fixed treatment I_i, baseline confounders $\mathbf{K_i}$ and parameters $\boldsymbol{\omega}$ is given by

$$\mathrm{e}_i(\boldsymbol{\omega}) = \frac{1}{\mathbb{I}_{(I_i=1)}P(I_i = 1|\mathbf{K_i};\boldsymbol{\omega}) + \mathbb{I}_{(I_i=0)}(1 - P(I_i = 1|\mathbf{K_i};\boldsymbol{\omega}))}, \tag{3.2}$$

where $\mathbb{I}_{(I_i=1)}$ is an indicator function for treatment $I_i = 1$. We typically estimate the pa-

rameters $\boldsymbol{\omega}$ by fitting a logistic regression model with dependent variable I_i and predictors $\mathbf{K_i}$.

3.2.2 Assumptions

Suppose that we have a random sample of n individuals, indexed by $i = 1, ..., n$. We are interested in the marginal effect of a binary intervention $I_i(t)$ on the longitudinal, continuous outcome $Y_i(t)$, where t represents the time. We use the Neyman-Rubin potential outcome framework (Neyman [1923]; Rubin [1974]) to express the estimand of interest, which is the causal contrast $E\left[Y_{i1}(t) - Y_{i0}(t)\right]$ where $Y_{i1}(t)$ corresponds to the outcome that would have been observed at time t, had individual i been treated by $I_i(t) = 1$, and $Y_{i0}(t)$, had the individual been treated by $I_i(t) = 0$. More specifically, we focus on a time-invariant marginal effect of the exposure on the outcome recorded at the same monitoring time.

The terms *treatment*, *intervention* and *exposure* are used interchangeably to refer to $I_i(t)$, and *monitoring times* and *visit times* refer to the times when the outcome $Y_i(t)$ is observed. We use bold notation to refer to vectors and matrices. We now detail the assumptions required about the outcome model (O1-O2), the visit process (V1-V2), the treatment model (P1-P3), and the total follow-up time (C1); these assumptions are required for consistency of our proposed estimators.

To estimate $E\left[Y_{i1}(t) - Y_{i0}(t)\right]$, one can also estimate the contrast $E\left[Y_i(t)|I_i(t) = 1\right] - E\left[Y_i(t)|I_i(t) = 0\right]$ in a pseudo-population where there is no imbalance in covariates between treatment groups due to confounding and the monitoring process. In a setting with no such imbalance, we assume that treatment groups are similar in their characteristics and that patient groups are interchangeable prior to exposure. The following marginal linear model for the mean can be used for estimation:

$$E[Y_i(t)|I_i(t)] = \alpha(t) + \beta I_i(t). \tag{O1}$$

The parameter β in (O1) is exactly equal to $E\left[Y_i(t)|I_i(t)=1\right]-E\left[Y_i(t)|I_i(t)=0\right]$ so it might represent a valid estimate for the causal contrast of interest. However, we are aware in our setting of an underlying confounding process, and the following conditional model for the mean is a sensible model to use to estimate the conditional effect of treatment β_I:

$$E[Y_i(t)|I_i(t),\mathbf{K_i(t)}]=\alpha(t)+\beta_I I_i(t)+\boldsymbol{\beta}_{\boldsymbol{K}}'\mathbf{K_i(t)} \tag{O2}$$

for $\mathbf{K_i(t)}$ the confounders of the relationship $(I_i(t),Y_i(t))$. Depending on the distribution of confounders $\mathbf{K_i(t)}$ in the sample under study, an estimator of β in the marginal model in (O1) might be biased for $E\left[Y_{i1}(t)-Y_{i0}(t)\right]$, due to imbalances in the confounders between treatment groups. Moreover, the model in (O2) is marginalized over other covariates not included in $I_i(t)$ or $\mathbf{K_i(t)}$ that may affect both the outcome and the monitoring times. For now, we do not consider explicit modeling of the covariates affecting monitoring times which could bias an estimate for β_I in (O2), and focus on the conditional model (O2). We see later how we can obtain an estimate for the average treatment effect in a pseudo-population where there are no imbalances between treatment groups with respect to $\mathbf{K_i(t)}$, and no imbalances in observed/unobserved outcomes due to an informative monitoring process.

Let the intercept $\alpha(t)$ remain unspecified in (O2). In addition to confounding, assume that the relationship between $I_i(t)$ and $Y_i(t)$ may be mediated by a vector of (potentially time-varying) covariates $\mathbf{Z_i(t)}$ which are in the causal path from the exposure $I_i(t)$ to the outcome $Y_i(t)$; see Bůžková and Lumley [2005] for an asthma-related example.

Assume that the longitudinal outcome $Y_i(t)$ is only observed at times $T_{i1},...,T_{iK_i}$, with $N_i(t)=\sum_{k=1}^{K_i} I(T_{ik}\leq t)$. Note that other patient features might be recorded and available in between times when the outcome is recorded. $N_i(t)$ is used to denote the number of monitoring times by time t, for individual i. The quantity $dN_i(t)$ is equal to 1 if $Y_i(t)$ is measured at time t and 0 otherwise, and τ represents the maximum follow-up time in the cohort under study.

We suppose that the relationship between $I_i(t)$ and $Y_i(t)$ may be distorted by an informative monitoring process, and that monitoring at time t depends on the set of covariates $\mathbf{V_i(t)} = \{\mathbf{Z_i(t)}, I_i(t)\}$ through a proportional intensity model for the monitoring times:

$$E[dN_i(t)|\mathbf{V_i(t)}] = \xi_i(t) \exp\left(\boldsymbol{\gamma}_V' \mathbf{V_i(t)}\right) d\Lambda_0(t), \tag{V1}$$

where the function $\Lambda_0(\cdot)$ is arbitrary and nondecreasing, and $\xi_i(t)$ is the indicator of individual i still being in the study at time t. We assume that for each time $0 < t < C_i$, for a certain time granularity (e.g. daily), and for each individual i, we have $0 < P[dN_i(t)|\mathbf{V_i(t)}] < 1$. We restrict the assumption to a particular granularity, as positivity is unlikely to hold when time is continuous.

Suppose that $\mathbf{V_i(t)}$ contains all common predictors of the monitoring times and the outcome,

$$N_i(t) \perp Y_i(t)|\mathbf{V_i(t)}. \tag{V2}$$

In fact, note that $\mathbf{Z(t)} \subset \mathbf{V(t)}$ may contain any mediator of the relationship between $I_i(t)$ and $Y_i(t)$, but also any other variable that is not the intervention but that affects monitoring times.

Note that the modelling of monitoring times through equation (V1) requires all covariates affecting monitoring times to be available at any time t, $0 \leq t < C_i$, $\forall i$ during follow-up (again, under a particular time granularity, e.g. daily). We note that administrative databases or EHRs often contain the information on drugs prescribed or previous diagnostics at any time (even in between times when the outcome is recorded) and these values can be carried forward in between monitoring times so as to use as much information as possible. In clinical practice, in the absence of new measurements, this information may be relied on to make decisions (Cao et al. [2016]).

For the exposure, we assume conditional exchangeability, stable-unit treatment value and positivity of treatment, which respectively correspond to:

$$I_i(t) \perp \{Y_{i0}(t), Y_{i1}(t)\} \,|\mathbf{K_i(t)} \tag{P1}$$

$$\{Y_{i0}(t), Y_{i1}(t)\} \,|I_i(t) = \{Y_{i0}(t), Y_{i1}(t)\} \,|I_i'(t) \text{ if } I_i(t) = I_i'(t) \tag{P2}$$

$$0 < P(I_i(t) = 1|\mathbf{K_i(t)}), P(I_i(t) = 0|\mathbf{K_i(t)}) < 1. \tag{P3}$$

These conditions are necessary to use propensity score methods to adjust for confounding, along with correct model specifications.

While the maximum follow-up time is τ, it may be that some individuals are not followed after a certain point. Let C_i denote the end of follow-up ("censoring" time, though we are not working in a time-to-event context) for individual i; we consider that the end of follow-up is administrative and non-informative, that is

$$E\left[Y_i(t)|I_i(t), \mathbf{K_i(t)}, C_i \geq t\right] = E\left[Y_i(t)|I_i(t), \mathbf{K_i(t)}\right]. \tag{C1}$$

We note that this assumption could be circumvented by using inverse probability of censoring weights to adjust for informative dropout. See, for instance, Robins et al. [2000a].

The causal diagram in Figure 3.1, panel A depicts the structure of the data generating mechanism at time t. Panel B shows the presumed underlying data mechanism for our analysis of the *Add Health* study, presented in Section 3.4. Note in Figure 3.1 that we assume that the confounders and the mediators vary in time, which is allowed but is not necessary. Even when these variables vary in time, their effects on the monitoring times are assumed constant over time (i.e. we estimate $\boldsymbol{\gamma}_V$ rather than $\boldsymbol{\gamma}_V(t)$). Finally, we note that knowledge about the problem under study should inform the best choice for the set $\mathbf{K_i(t)}$, which may incorporate covariates measured at time t, as well as at previous time s, for $s < t$. Settings with time-dependent confounding are allowed, as long as the set of

confounders include all potential confounders of the marginal relationship under study at a given time and that mediators are not conditioned upon in the outcome model.

Dotted arrows in Figure 3.1 refer to potential relationships we may want to consider.

3.2.3 Existing methods

Lin and Ying [2001] proposed a semiparametric estimator for β in the marginal model (O1) without reference to a particular covariate or intervention of interest. Their method did not account for the variables that affected the monitoring times whenever those variables were not contained in the design matrix for the outcome model. Bůžková and Lumley [2009] extended their work to account for those other variables. They built an estimator for the marginal effect of treatment based on the stochastic process $M_i(t;\beta,\boldsymbol{\gamma_V},\mathscr{A})$ which, in our case, is defined by

$$M_i(t;\beta,\boldsymbol{\gamma_V},\mathscr{A}) = \int_0^t \left(Y_i(s) - \beta I_i(s)\right) dN_i(s) - \xi_i(s)\exp\left(\boldsymbol{\gamma_V'}\mathbf{V_i(s)}\right) d\mathscr{A}(s), \tag{3.3}$$

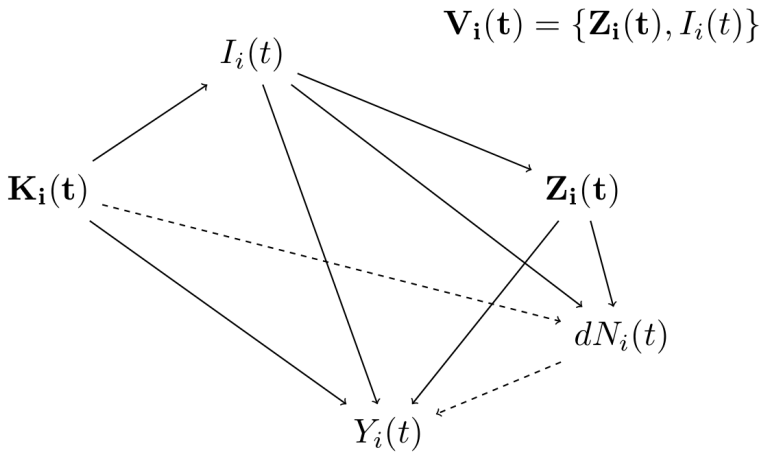
where $\mathscr{A}(t) = \int_0^t \alpha(s)d\Lambda(s)$. They defined a weighted version $G_i(t;\beta,\boldsymbol{\gamma},\mathscr{A})$ of that process, with

$$G_i(t;\beta,\boldsymbol{\gamma},\mathscr{A}) = \int_0^t \frac{1}{\rho_i(s;\boldsymbol{\gamma})} dM_i(s;\beta,\boldsymbol{\gamma_V},\mathscr{A}) \tag{3.4}$$

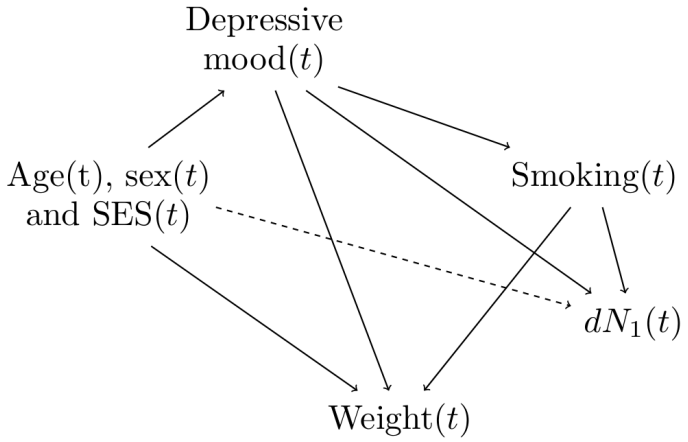
with the stabilized rate ratio weight $\rho_i(s;\boldsymbol{\gamma})$, given in our setting by

$$\rho_i(s;\boldsymbol{\gamma}) = \frac{\exp\left(\boldsymbol{\gamma_1'}\mathbf{Z_i(s)} + \gamma_2 I_i(s)\right)}{\exp\left(\gamma_I I_i(s)\right)}. \tag{3.5}$$

Note that $\boldsymbol{\gamma_1'}\mathbf{Z_i(s)} + \gamma_2 I_i(s) = \boldsymbol{\gamma_V'}\mathbf{V_i(s)}$. The weight (3.5) allows their estimator to consider the dependency between $\mathbf{Z(t)}$ and the monitoring times while not adding $\mathbf{Z(t)}$ directly into the design matrix. It also accounts for the dependency between the covariates in the design matrix of the outcome model (here, $I_i(t)$) and the monitoring times. The parameters $\boldsymbol{\gamma_1}$ and



(a) Causal diagram overlaying the monitoring process onto data generating process at time t



(b) Causal diagram for the *Add Health* study data at time t in individual $i = 1$

Figure 3.1: Structure of the data generating and monitoring process for (a) a general setting and (b) the analysis of *Add Health* data

γ_2 in (3.5) can be estimated by fitting a proportional intensity model for monitoring times with $\mathbf{Z(t)}$ and $\mathbf{I(t)}$ as covariates, while γ_I is estimated using the same type of model with only $\mathbf{I(t)}$ as a covariate.

Bůžková and Lumley [2009] show that $E\left[dG_i(t;\beta,\boldsymbol{\gamma},\mathscr{A})|I_i(t)\right]=0$ under assumptions (O1), (C1), (V1) and (V2). They further build estimating equations for β in (O1). In our setting where the design matrix is $\mathbf{I(t)}$, their procedure yields the following estimator:

$$\begin{aligned}\widehat{\beta}_{BL} = &\left[\sum_{i=1}^{n}\int_0^\tau \frac{W(t)}{\rho_i(t;\boldsymbol{\gamma})}\left(I_i(t)-\overline{I}(t;\gamma_I)\right)^2 dN_i(t)\right]^{-1}\\ &\times\sum_{i=1}^{n}\int_0^\tau \frac{W(t)}{\rho_i(t;\boldsymbol{\gamma})}\left(I_i(t)-\overline{I}(t;\gamma_I)\right)\left(Y_i(t)-\overline{Y}^*(t;\gamma_I)\right)dN_i(t),\end{aligned} \tag{3.6}$$

which is a least squares type estimator where the design matrix is the vector $\left(\mathbf{I(t)}-\overline{\mathbf{I}}(\mathbf{t};\boldsymbol{\gamma_I})\right)$, the outcome vector is given by $\left(\mathbf{Y(t)}-\overline{\mathbf{Y}}^*(\mathbf{t};\boldsymbol{\gamma_I})\right)$, $W(t)$ is an arbitrary time-dependent weight that may be used to reduce the variance, and $\overline{Y}^*(t;\gamma_I)$ a weighted average of the nearest-neighbor approximation to Y at time t (which is also used to reduce the variance of the estimator). The re-centering of $I_i(t)$ by its adjusted mean in (3.6) eliminates from the estimation the intercept $\alpha(t)$ in (O1) and avoids having to model the relationship between the mean outcome and time t, hence the semiparametric and more flexible nature of the estimator. The estimating equations that Bůžková and Lumley [2009] used are sums of independent zero-mean random vectors, and the variance of their estimator can be derived using standard asymptotic theory along with Taylor expansions. In what follows, we use $W(t)=1\ \forall t$.

In order to estimate the adjusted means, the proportional intensity model (V1) is fitted with only the predictor $I_i(t)$. The coefficient $\widehat{\gamma}_I$ for $I_i(t)$ is used to compute the weighted means. For any vector $\mathbf{R(t)}$ in general, we have:

$$\overline{R}(t;\widehat{\gamma}_I)=\sum_{i=1}^{n}R_i(t)\frac{\xi_i(t)\exp\left(\widehat{\gamma}_I I_i(t)\right)}{\sum_{j=1}^{n}\xi_j(t)\exp\left(\widehat{\gamma}_I I_j(t)\right)}. \tag{3.7}$$

The estimator $\widehat{\beta}_{BL}$ is, however, biased for the marginal effect of the intervention $I_i(t)$ in our setting of interest, because it is limited to randomized controlled settings and does not consider imbalances between treatment groups which are due to confounders $\mathbf{K_i(t)}$. With a conditional expectation such as in (O2), the process used to build the estimator (which was based on assumption (O1)) is no longer zero-mean and the estimator may thus not converge to the true parameter.

3.2.4 The Inverse Probability of Centered Treatment and Monitoring Estimator $\widehat{\beta}_{IPCTM}$

Under similar assumptions to Bůžková and Lumley [2009], but now further including covariates as in (O2), we first develop an estimator for the conditional effect of $I_i(t)$ on $Y_i(t)$, as in the setting depicted in Figure 3.1(A). Note that this estimator is marginalized over the predictors $\mathbf{V(t)}$ of the monitoring times and, as in Bůžková and Lumley, we use a monitoring weight to account for any imbalance in those predictors that could bias the effect of $\mathbf{I(t)}$ conditional on $\mathbf{K(t)}$. We define a new process $P_i(t) = P_i(t; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathscr{A})$ as

$$P_i(t) = \int_0^t \frac{1}{\rho_i(s;\boldsymbol{\gamma})} \left[\left(Y_i(s) - \beta_I I_i(s) - \boldsymbol{\beta}_{\boldsymbol{K}}' \mathbf{K_i(s)}\right) dN_i(s) - \xi_i(s) \exp\left(\boldsymbol{\gamma}_{\boldsymbol{V}}' \mathbf{V_i(s)}\right) d\mathscr{A}(s) \right],$$

with $\mathscr{A}(t) = \int_0^t \alpha(s) d\Lambda(s)$. In Appendix A.1, we show that $E\left[dP_i(t) | I_i(t), \mathbf{K_i(t)}\right] = 0$, and the derivation of the estimating equations and estimators for the conditional effects. We

obtain the following estimators for the conditional effects of $\left[I_i(t)\ \mathbf{K_i(t)}\right]'$ in (O2):

$$[\widehat{\beta}_I\ \widehat{\boldsymbol{\beta}}_{\boldsymbol{k}}]' = \left[\sum_{i=1}^{n}\int_0^\tau \frac{W(t)}{\rho_i(t;\widehat{\boldsymbol{\gamma}})}\begin{pmatrix} I_i(t)-\overline{I}(t;\widehat{\gamma}_I) \\ \mathbf{K_i(t)}-\overline{\mathbf{K}}(\mathbf{t};\widehat{\gamma}_\mathbf{I})\end{pmatrix}^{\otimes 2} dN_i(t)\right]^{-1} \times \sum_{i=1}^{n}\int_0^\tau \frac{W(t)}{\rho_i(t;\widehat{\boldsymbol{\gamma}})}\begin{pmatrix} I_i(t)-\overline{I}(t;\widehat{\gamma}_I) \\ \mathbf{K_i(t)}-\overline{\mathbf{K}}(\mathbf{t};\widehat{\gamma}_\mathbf{I})\end{pmatrix}' \left(Y_i(t)-\overline{Y}^*(t;\widehat{\gamma}_I)\right) dN_i(t). \tag{3.8}$$

Using the estimating equation for conditional effects to estimate the parameters β_I and $\boldsymbol{\beta_K}$ in (O2) corresponds to using a weighted least squares regression with predictors $(I_i(t)-\overline{I}(t;\widehat{\gamma}_I))$ and $(\mathbf{K_i(t)}-\overline{\mathbf{K}}(\mathbf{t};\widehat{\gamma}_\mathbf{I}))$, a dependent variable $(Y_i(t)-\overline{Y}^*(t;\widehat{\gamma}_I))$ and weights $W(t)/\rho_i(t;\widehat{\boldsymbol{\gamma}})$. To rather estimate the marginal effect of $I_i(t)$ on the mean outcome, we propose to use weights to create a pseudo-population in which there is no imbalance due to confounders, and so we change focus to the corresponding estimating equation for the marginal model (O1), and its corresponding estimator given in (3.6), when there is no imbalance due to confounders.

The re-weighting procedure we use is reminiscent of standard inverse probability of treatment weighting. Our goal is to break any dependency between the columns of the design matrix in (3.8), given by $I_i(t)-\overline{I}(t;\widehat{\gamma}_I)$ and $\mathbf{K_i(t)}-\overline{\mathbf{K}}(\mathbf{t};\widehat{\gamma}_\mathbf{I})$. Note that the quantity $\left(I_i(t)-\overline{I}(t;\widehat{\gamma}_I)\right)$ is typically *not* binary so we cannot use a logistic regression to model $E\left[I_i(t)-\overline{I}(t;\widehat{\gamma}_I)|\mathbf{K_i(t)}-\overline{\mathbf{K}}(\mathbf{t};\widehat{\gamma}_\mathbf{I})\right]$. We model the conditional mean using a linear model. Suppose

$$E\left[I_i(t)-\overline{I}(t;\widehat{\gamma}_I)|\mathbf{K_i(t)}-\overline{\mathbf{K}}(\mathbf{t};\widehat{\gamma}_\mathbf{I})\right] = \psi_0 + \boldsymbol{\psi}_\mathbf{1}'(\mathbf{K_i(t)}-\overline{\mathbf{K}}(\mathbf{t};\widehat{\gamma}_\mathbf{I})). \tag{3.9}$$

Estimates for $E\left[I_i(t)-\overline{I}(t;\widehat{\gamma}_I)|\mathbf{K_i(t)}-\overline{\mathbf{K}}(\mathbf{t};\widehat{\gamma}_\mathbf{I})\right]$ are obtained via the predictions from the

linear regression model (3.9) with estimated coefficients. To transform these values into pseudo probabilities that lie between 0 and 1 so as to further re-weight the marginal estimating equation corresponding to the estimator in (3.6), we use an approach suggested by Robins et al. [2000a]. We then stabilize these pseudo probabilities, using a marginal model for the mean of $I_i(t) - \overline{I}(t;\widehat{\gamma}_I)$ that is equal to ψ_m so as to compute a final *stabilized generalized weight* given by

$$\text{sgw}_i(t;\widehat{\boldsymbol{\psi}}) = \text{sgw}_i(t;\widehat{\psi}_0,\widehat{\boldsymbol{\psi}}_{\mathbf{1}},\widehat{\psi}_m) = \frac{g^{-1}\left(\widehat{\psi}_0 + \widehat{\boldsymbol{\psi}}'_{\mathbf{1}}(\mathbf{K_i(t)} - \overline{\mathbf{K}}(\mathbf{t};\widehat{\gamma}_{\mathbf{I}}))\right)}{g^{-1}\left(\widehat{\psi}_m\right)} \tag{3.10}$$

for $g^{-1}(\widehat{a}_i(t)) = 1/\sqrt{2\pi\widehat{\sigma}_a^2}\exp\left(-\widehat{\epsilon}_{a,i}(t)^2/(2\widehat{\sigma}_a^2)\right)$ the Normal density function evaluated at the linear regression residuals $\widehat{\epsilon}_{a,i}(t) = \left(I_i(t) - \overline{I}(t;\widehat{\gamma}_I) - \widehat{a}_i(t)\right)$, with $\widehat{\sigma}_a^2$ the empirical variance of $\widehat{\epsilon}_{a,i}(t)$. Another way of modelling the variable $I_i(t) - \overline{I}(t;\widehat{\gamma}_I)$ would be to categorize it into quantiles (Naimi et al. [2014]). That procedure could work particularly well if the distribution of $I_i(t) - \overline{I}(t;\widehat{\gamma}_I)$ is not unimodal and is asymmetric. This latter approach was evaluated in sensitivity analyses.

The weight (3.10) is incorporated into the estimating equations corresponding to the estimator of Bůžková and Lumley in (3.6), and we obtain the new estimating equation

$$\begin{aligned} U^{mar}(\beta,\alpha,\widehat{\boldsymbol{\gamma}},\widehat{\boldsymbol{\psi}}) = \sum_{i=1}^{n}\int_0^\tau \frac{W(t)}{\rho_i(t;\widehat{\boldsymbol{\gamma}})}\frac{1}{\text{sgw}_i(t;\widehat{\boldsymbol{\psi}})}\left(I_i(t) - \overline{I}(t;\widehat{\gamma}_I)\right) \\ \times\left[Y_i(t) - \overline{Y}^*(t;\widehat{\gamma}_I) - \beta\left(I_i(t) - \overline{I}(t;\widehat{\gamma}_I)\right)\right]dN_i(t). \end{aligned} \tag{E2}$$

Solving equation (E2) for $U^{mar}(\beta,\alpha,\widehat{\boldsymbol{\gamma}},\widehat{\boldsymbol{\psi}}) = 0$ leads to the closed-form solution of our proposed *Inverse Probability of Centered Treatment and Monitoring* (IPCTM) estimator,

that is given by:

$$\widehat{\beta}_{IPCTM} = \left[\sum_{i=1}^{n}\int_0^\tau \frac{W(t)}{\rho_i(t;\widehat{\boldsymbol{\gamma}})}\frac{\left(I_i(t)-\overline{I}(t;\widehat{\gamma}_I)\right)^2}{\text{sgw}_i(t;\widehat{\boldsymbol{\psi}})}\,dN_i(t)\right]^{-1} \times \sum_{i=1}^{n}\int_0^\tau \frac{W(t)}{\rho_i(t;\widehat{\boldsymbol{\gamma}})}\frac{\left(I_i(t)-\overline{I}(t;\widehat{\gamma}_I)\right)}{\text{sgw}_i(t;\widehat{\boldsymbol{\psi}})}\left(Y_i(t)-\overline{Y}^*(t;\widehat{\gamma}_I)\right)dN_i(t) \quad (3.11)$$

for the estimand of interest, the marginal effect of $I_i(t)$ on $Y_i(t)$.

Note that the intercept function $\alpha(t)$ is left unspecified in (O1) so that one need not assume any particular form for the dependence of the outcome $Y(t)$ on time t. More details on the unbiasedness of the IPCTM estimator are presented in Appendix A.2. Similarly to Bůžková and Lumley [2009], the asymptotic variance of the IPCTM estimator can be developed using standard asymptotic theory. It is also possible to directly account for the components of variance due to the weights using theory on two-step estimators (Newey and McFadden [1994]) along with the variance formula provided by Bůžková and Lumley [2009].

3.2.5 The Flexible Inverse Probability of Treatment and Monitoring Estimator $\widehat{\beta}_{FIPTM}$

A second estimator, which is also a weighted least squares type estimator, is proposed to estimate the marginal effect of treatment on a longitudinal and continuous outcome. It requires slightly stronger parametric specifications for the intercept $\alpha(t)$ in (O1), which is modelled through cubic splines. However, it is easier to implement in practice, and as we will demonstrate in Section 3.2, it often provides equivalent performance as the IPCTM estimator in simulation studies. Given its more parametric nature, we also expect it to exhibit smaller variance than the IPCTM estimator.

Let us assume again the conditional mean model (O2) along with assumptions (P1), (P2) and (P3) and that monitoring times can be modelled through a proportional intensity model as in (V1). We use a weighted least squares method, and aim to create a pseudo-population

in which imbalances due to confounders and covariate-dependent monitoring times are eliminated through re-weighting. We first readjust the observations for the monitoring process using an inverse probability of monitoring weight defined by the inverse of $\varphi_i(t;\boldsymbol{\gamma}_V)$, with

$$\varphi_i(t;\boldsymbol{\gamma}_V) = \exp\left(\boldsymbol{\gamma_1'}\boldsymbol{Z_i(t)} + \gamma_2 I_i(t)\right). \tag{3.12}$$

Again, assuming a proportional intensity model for the monitoring times, one does not need to estimate the function $\Lambda_0(t)$ in (V1) since this term at time t will cancel out across individuals. The parameters $\boldsymbol{\gamma_1}$ and γ_2 can be estimated by fitting a proportional intensity model.

We use a standard approach to adjust for imbalances due to confounders, and add an inverse probability of treatment weight into the weighted least squares regression. That weight is given by:

$$\mathrm{e}_i(t;\boldsymbol{\omega}) = \frac{1}{\mathbb{I}_{(I_i(t)=1)}P(I_i(t)=1|\mathbf{K_i(t)};\boldsymbol{\omega}) + \mathbb{I}_{(I_i(t)=0)}(1-P(I_i(t)=1|\mathbf{K_i(t)};\boldsymbol{\omega}))}. \tag{3.13}$$

The quantities $P(I_i(t)=1|\mathbf{K_i(t)};\boldsymbol{\omega})$ and $P(I_i(t)=0|\mathbf{K_i(t)};\boldsymbol{\omega})$ in (3.13) can be estimated via logistic regression with $\mathbf{K_i(t)}$ as covariates and $I_i(t)$ as the dependent variable. Once again, knowledge about the problem under study should inform selection of $\mathbf{K_i(t)}$ for inclusion in the treatment model used to estimate the IPT weights in (3.13).

The intercept $\alpha(t)$ in (O2) is modelled using cubic splines along with a constant intercept. We use splines with two knots and choose the tertiles of the distribution of t for the knots. The final estimator has a closed-form solution given by

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{FIPTM}} = \left[\sum_{i=1}^{n}\int_0^\tau \frac{\mathrm{e}_i(t;\boldsymbol{\omega})}{\varphi_i(t;\widehat{\boldsymbol{\gamma}}_{\boldsymbol{V}})}\mathbf{S_i(t)}^{\otimes 2}dN_i(t)\right]^{-1}\sum_{i=1}^{n}\int_0^\tau \frac{\mathrm{e}_i(t;\boldsymbol{\omega})}{\varphi_i(t;\widehat{\boldsymbol{\gamma}}_{\boldsymbol{V}})}\mathbf{S_i(t)}'Y_i(t)dN_i(t) \tag{3.14}$$

with $\mathbf{S(t)}$ a matrix with $s+2$ columns, for s the number of columns in the basis of the cubic spline. The leading column of $\mathbf{S(t)}$ is a vector of 1 for the constant intercept, and the last column corresponds to the intervention $\mathbf{I(t)}$. We are interested in the last coefficient of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{FIPTM}}$, which corresponds to the estimator for the marginal effect of treatment, that we further refer to as $\widehat{\beta}_{FIPTM}$.

The asymptotic variance of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{FIPTM}}$ is computed using standard theory on weighted least squares estimator, with the components of variance due to the weights incorporated into the sandwich estimator using theory on two-step estimators (Newey and McFadden [1994]). For derivations, see Appendix A.3. A comparison of the empirical, the bootstrapped and the asymptotic variances in simulation studies is presented in Table A.3 of Appendix A.4, along with the coverage of the FIPTM estimator.

3.3 Simulation study

Simulation studies were conducted to assess the performance of both estimators $\widehat{\beta}_{IPCTM}$ and $\widehat{\beta}_{FIPTM}$ for the marginal effect of $I_i(t)$ on the mean of $Y_i(t)$, for different levels of dependency ($\boldsymbol{\gamma_V}$) between covariates and monitoring times and for different forms of intercept $\alpha(t)$. The data generating mechanism was similar to the one presented in Figure 3.1(A) and inspired by Bůžková and Lumley [2009], but incorporates (possibly time-varying) confounders. In a first study described below, the intervention and the confounders were kept time fixed. In a second study, they could vary in time (details are presented in Appendix A.5).

Simulation study 1: Time-fixed confounders and treatment

For all patients i, three baseline confounders $\{K_{1i}, K_{2i}, K_{3i}\}$ were generated with $K_{1i} \sim \mathrm{N}(1,1), K_{2i} \sim$ Bernoulli(0.55), and $K_{3i} \sim \mathrm{N}(0,1)$. The intervention $I_i(t)$ was binary and time-fixed, and was simulated as $I_i \sim$ Bernoulli(p_{Ii}) with $p_{Ii} = \text{expit}\left(0.5 + 0.8K_{1i} + 0.05K_{2i} - 1K_{3i}\right)$. One time-varying mediator $Z_i(t)$ was generated, conditional on I_i, as $Z_i(t)|I_i = 1 \sim \mathrm{N}(2,1)$

and $Z_i(t)|I_i = 0 \sim \mathrm{N}(4,4)$. The outcome $Y_i(t)$ was simulated as $Y_i(t) = \alpha(t) + 1\,I_i + 3\,[Z_i(t) - E\,[Z_i(t)|I_i]] + 0.4\,K_{1i} + 0.05\,K_{2i} - 0.6\,K_{3i} + \epsilon_i(t)$ with $\epsilon_i(t) \sim \mathrm{N}(\phi_i, 0.01)$ and $\phi_i \sim \mathrm{N}(0, 0.04)$.

The quantities above were first simulated in continuous time, with time discretized over a grid of 0.01 units, from 0 to τ. Then, monitoring times (i.e. when the outcome was observed) were simulated according to a nonhomogeneous Poisson process, with intensity at time t equal to $\lambda_i(t|I_i, Z_i(t)) = \eta_i \exp(\gamma_1 I_i + \gamma_2 Z_i(t))$, with η_i a gamma distributed random variable with mean 1 and variance 0.01. Bernoulli draws with probabilities proportional to these intensities could be used at each time point to assign monitoring times. Monitoring times could be drawn up until the maximum follow-up time τ; we fixed $\tau = 2$ and obtained different mean numbers of visits which depended on parameters (γ_1, γ_2). We tested three combinations: $(\gamma_1, \gamma_2) = (0, 0)$, which corresponded to no dependency on covariates; $(\gamma_1, \gamma_2) = (-0.3, 0.2)$; and $(\gamma_1, \gamma_2) = (0.6, 0.3)$. The follow-up time was further censored at time C_i for each individual, with $C_i \sim \mathrm{Uniform}(\tau/2, \tau)$. For $\alpha(t)$, five different functions of time were tested: $\alpha(t) = 3$; $\alpha(t) = 2.5t$; $\alpha(t) = \sin(t)$; $\alpha(t) = \exp(t)$; and $\alpha(t) = \exp(2|\sin(3t)|)$. Two sample sizes, respectively $n = 250$ and $n = 500$, were tested. We used a total of 1000 simulations in each study.

The proposed estimators were compared to more standard ones, i.e. an OLS estimator, a visit-weighted estimator and an IPT-weighted estimator. The OLS estimator $\widehat{\beta}_{OLS}$ was obtained by fitting a linear regression model with outcome $Y_i(t)$, a constant intercept and the independent variable I_i. The estimator $\widehat{\beta}_{VW}$ was a weighted least squares estimator in which a time-dependent monitoring weight was incorporated. The monitoring time model was correctly specified and included I_i and $Z_i(t)$ as explanatory variables. The IPT-weighted estimator was a weighted linear regression estimator in which an inverse probability of treatment weight was incorporated. For the estimators $\widehat{\beta}_{IPCTM}$ and $\widehat{\beta}_{FIPTM}$, the treatment and the monitoring models were correctly specified.

In Appendix A.4, we present the results for 9 additional simulation scenarios in which treatment and confounding variables were also time fixed. Scenarios i) and ii) respectively correspond to the cases where confounder variables $\{K_{1i}, K_{2i}, K_{3i}\}$ were correlated, or where confounder variables $\{K_{1i}, K_{2i}, K_{3i}\}$ affected the monitoring intensity. Scenarios iii) and iv) correspond to the cases where generalized IPT weights in the IPCTM estimator were computed from a cumulative logistic regression, with the variable $I_i(t) - \overline{I}(t; \widehat{\gamma}_I)$ binned into 10 quantiles, or with 20 quantiles, respectively. Sensitivity analyses v), vi), vii) and viii) aim to assess sensitivity to model misspecification via studies where we: v) changed the error distribution for a Log-Normal distribution centered in 0, in the mean outcome model, rather than the Normal errors we previously simulated; vi) incorporated non-linear functions of the confounder covariates in the generative outcome model; vii) incorporated non-linear terms of the covariates in the generative proportional intensity model for the visits; and viii) drew, for each individual, a different intercept function $d\Delta_0(t)$ from 3 possible functions: $d\Delta_0(t) \in \{1; 1.5t; \sin(t)\}$, with respective probability $1/2, 1/4, 1/4$. Finally, the simulation scenario ix) explored the effect of conditioning on confounders in the outcome mean model, for all the estimators that were being compared.

Results

Summary statistics (including empirical bias) for each estimator are found in Appendix A.4. Figure 3.2 shows absolute biases and empirical mean squared errors (MSEs) for each of the five estimators we compared; each boxplot summarizes the distribution of bias or MSE, over all 15 scenarios of dependency and intercept functions that we considered. We also present results for one of the scenarios where $\alpha_0(t) = 3$, in Table 3.1 in this manuscript. The results in Table 3.1 were based on a simulation study where exposure and confounders were kept as time-fixed.

Table 3.1: Simulation study with confounding and covariate-dependent monitoring times ($\tau = 2$, $n = 250$, $\alpha(t) = 3$, time-fixed exposure and confounders)

(γ_1, γ_2)	Median no. visits (IQR)	Absolute bias (Empirical variance)				
		$\hat{\beta}_{OLS}{}^{\dagger}$	$\hat{\beta}_{VW}{}^{\ddagger}$	$\hat{\beta}_{IPT}{}^{\star}$	$\hat{\beta}_{FIPTM}$	$\hat{\beta}_{IPCTM}$
$(0, 0)$	1 (1-2)	0.72 (0.41)	0.71 (0.30)	0.06 (1.06)	0.09 (0.77)	0.08 (0.99)
$(-0.3, 0.2)$	2 (1-3)	1.05 (0.19)	0.72 (0.18)	1.77 (0.40)	0.04 (0.39)	0.01 (0.44)
$(0.6, 0.3)$	5 (4-7)	1.98 (0.12)	0.76 (0.19)	2.65 (0.30)	0.00 (0.38)	0.02 (0.47)

† Ordinary least squares regression with outcome $Y_i(t)$ and exposure $I_i(t)$ with a constant intercept
‡ Weighted least squares regression with outcome $Y_i(t)$ and exposure $I_i(t)$ with a constant intercept and an inverse probability of monitoring weight computed from a proportional intensity model with $I_i(t)$ and $Z_i(t)$ as predictors
⋆ Weighted least squares regression with outcome $Y_i(t)$ and exposure $I_i(t)$ with a constant intercept and one an inverse probability of treatment weight computed from a logistic regression model with $\mathbf{K_i(t)}$ as predictors

As we notice in Figure 3.2, the OLS estimator, which we can see is biased, generally provides variable MSEs due to the different sets of $\boldsymbol{\gamma_V}$ parameters. When adjusting for the monitoring process only, we observe that $\widehat{\beta}_{VW}$ varies much less. However, it remains biased due to confounding. The IPT estimator, on the other hand, is only unbiased when there is no informative visit process. Most importantly, $\widehat{\beta}_{IPCTM}$ and $\widehat{\beta}_{FIPTM}$ exhibit almost zero bias and a quite narrow distribution for their absolute bias. As expected, different parameters (γ_1, γ_2) lead to different mean numbers of visits. Typically, the greater the mean number of visits, the smaller the bias for the two latter estimators (see Tables A.1 and A.2 in Appendix A.4). In Table 3.1 of this manuscript, we find simular results which are representative of the results from across scenarios. In particular, we find that the absolute bias of the two proposed estimators $\widehat{\beta}_{IPCTM}$ and $\widehat{\beta}_{FIPTM}$ is near 0, but that their variance tends to be greater than that of their comparators, as the $\boldsymbol{\gamma_V}$ coefficients increase. The two proposed estimators dramatically outperform their comparators in terms of bias as those coefficients increase.

In Figure 3.2, we also observe that the IPCTM estimator exhibits a greater MSE than the flexible estimator ($\widehat{\beta}_{FIPTM}$) in studies with time-fixed treatment and confounder variables, while it exhibits a smaller mean squared error than the FIPTM estimator in studies with time-varying treatment and confounder variables. As expected, the range of MSE narrows as the sample size increases. Given that both $\widehat{\beta}_{IPCTM}$ and $\widehat{\beta}_{FIPTM}$ exhibit a bias that tends

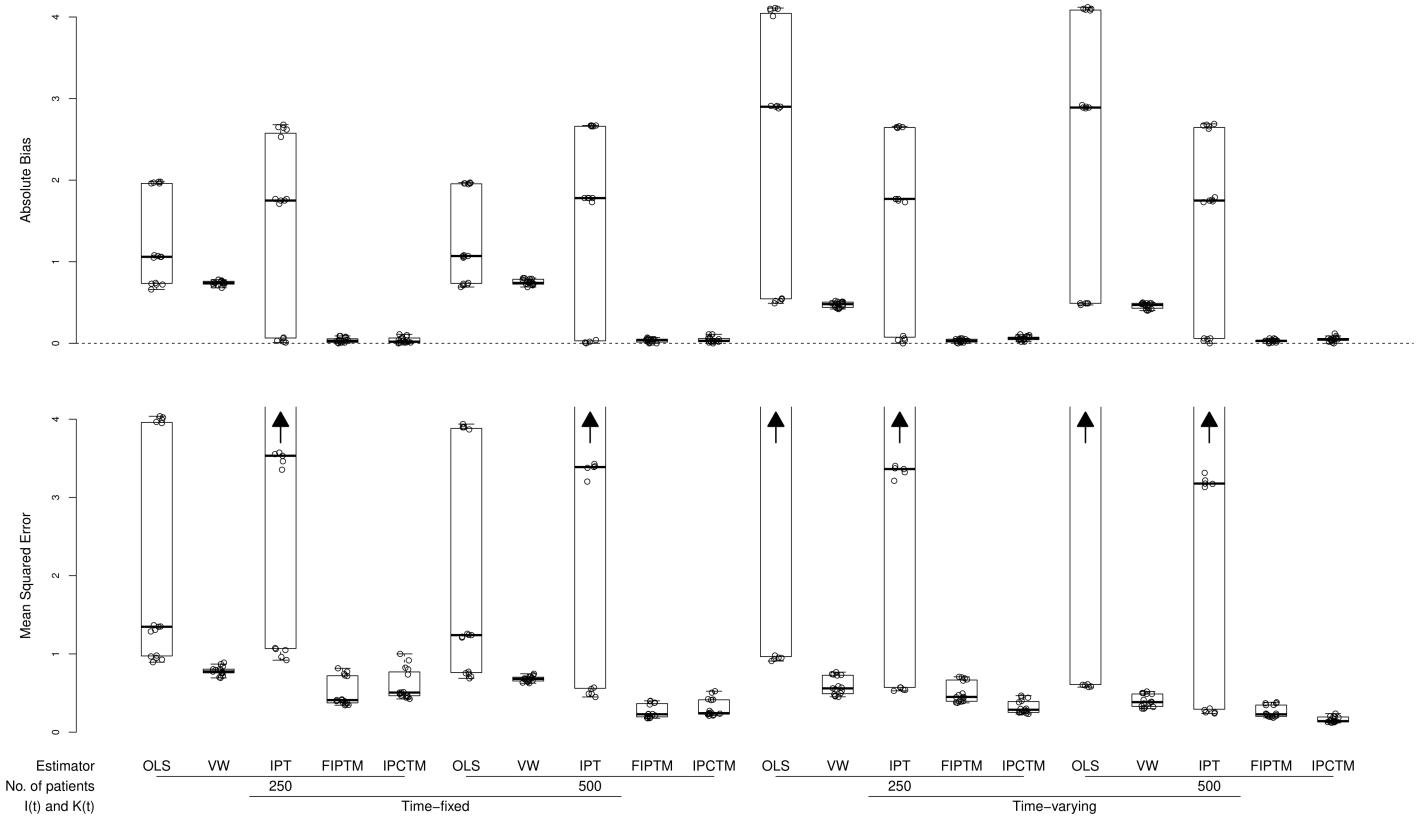


Figure 3.2: Boxplots of the distribution of absolute bias (top panel) and of MSE (bottom panel) from all 15 simulation scenarios, for the five estimators: Ordinary least squares, visit weighted only, inverse probability of treatment weighted estimator, FIPTM and IPCTM estimator, for time-fixed (left) or time-varying variables (right) and different sample sizes ($\tau = 2$, 1000 simulations).

towards 0, and that $\widehat{\beta}_{FIPTM}$ is easier to implement in practice, we contend that it should be preferred. We present in Table A.3 of Appendix A.4 a comparison of its bootstrapped, empirical and asymptotic variances, which were generally very similar. In studies with time-varying treatment and confounding variables, the IPCTM estimator may be more efficient. Further investigation of whether the centered estimator may be more competitive in a wider range of scenarios will be an important avenue of future work.

Sensitivity analyses for the first simulation study with time-fixed treatment and confounders

The results (distributions of biases and MSEs) for all 9 sensitivity analyses can be found

in Tables A.4 (i), A.5 (ii), A.6 (iii and iv), A.7 (v, vi, vii, viii), and A.8 (ix) in Appendix A.4. A brief summary of these results can also be found in Appendix A.6. Overall, our proposed methods were not too sensitive to misspecification of the different models, except for the sensitivity analysis where we incorporated non-linear functions of the covariates in the proportional intensity model for monitoring times. In that latter case, the FIPTM estimator has shown great bias, while the IPCTM estimator was not as affected by the misspecification of the monitoring model.

3.4 Application to the *Add Health* Study

The proposed estimators were applied to data from the National Longitudinal Study of Adolescents to Adult Health (*Add Health*) (Harris [2009]) to assess the marginal effect of depressive mood on weight in pounds, in adolescents. Our estimators were also compared to more standard estimators that do not account for informative monitoring process and/or confounding.

Add Health is a four-wave longitudinal study on adolescents who, over the course of the study, age to become young adults. A pool of participants who were well representative of adolescents in United States were enrolled during the years 1994-5 while they were in grades 7 to 12, and followed until 2008 (Wave IV). For each of the four waves, an in-home questionnaire was completed by the participants. A parent questionnaire was completed by one of the participants' parents at baseline only (Wave I). Data collected from in-home interviews are publicly available online for all four waves (Harris [2013]). For the purpose of this analysis, we assumed that longitudinal data are made up of a maximum of four time points where the outcome is potentially recorded. Hence, $time = 1, 2, 3, 4$ respectively correspond to all four waves. For simplicity, none of our analyses considered the sampling weights used in *Add Health* study.

We first defined the time-varying exposure that consisted of a binary depression score, using

a question from the in-home interview that was related to the current depressive mood of the participant. For the question *How often was the following true during the past week? You felt depressed.*, a participant's score was set to 0 if they answered *Never/rarely* or *Sometimes* and to 1 if their answer was contained in *A lot of the time* or *Most/All of the time*. The longitudinal outcome consisted in the weight in pounds, which was recorded at every in-home interview. We assumed that the relationship between depression status and weight was mediated by smoking, since depressive mood exacerbates smoking (e.g., Stepankova et al. [2017]), which in turn affects weight (e.g., Grunberg [1985]). We used as a proxy for smoking the number of cigarettes smoked during the past 30 days, also recorded at each of the four in-home interviews. A participant who had smoked at least one cigarette in the previous 30 days was considered to be a smoker. For confounders of the relationship between depression and weight, we included age, sex and socioeconomic status (SES). SES was defined using the two following in-home questions asked to one of the participants' parents: *About how much total income, before taxes did your family receive in 1994?* and *How far did you go in school?*. The answers were transformed into quintiles and summed up to give a total score contained between 0 and 10, with 10 corresponding to the highest SES.

A total of 6504 participants were enrolled at Wave I. Data presented missing values due to patients' dropout or their refusal to answer questions during the course of the study. We assumed that monitoring times (i.e. times when weight was recorded) depended on the depression status, the smoking status, age, sex and SES, which variables were included in a proportional intensity model for the monitoring times. In the exposure model, we adjusted for the potential confounders age, sex and SES. If patients had a value at their first interview, this value was used to impute values at other waves (it remained fixed in time). Recall that variables predicting the visit process are required to be available at all time. Thus, we employed multiple imputation with M=5 imputations, using predictive mean matching to impute any remaining missing values in age, sex and SES, as well as for missing values in exposure and mediator. Following imputation and analysis, the coefficient for exposure

of interest was combined across the imputations (Rubin [1976]). One thousand stratified bootstrap samples were drawn, with strata taken to be the individual, and they were used to assess the variance of each of the 5 estimates we compared. Table 3.2 presents a summary of the characteristics of the cohort at baseline, stratified by their depressive mood. Table 3.3 presents the average rate ratios for the 5 variables that were incorporated into the proportional intensity model for the visit times, along with confidence intervals computed using Rubin's rule for multiply imputed datasets (Rubin [2004]). Table 3.4 shows all estimated effects of depressive mood on weight with corresponding 95% Wald-type confidence intervals (CIs) using bootstrap standard errors.

Table 3.2: Characteristics at baseline of children enrolled in the *Add Health* study, stratified by depressive mood

	Depressive mood	
Variable	No	Yes
Smoking (N, %)	1367 (23.3)	280 (44.0)
Age (median, IQR)	15 (14-16)	16 (14-17)
Sex=female (N, %)	2914 (49.8)	433 (68.0)
SES (median, IQR)	6 (4-8)	5 (4-7)

The two exposure groups (depressed/not depressed) presented differences at baseline, with more smokers, older participants, more females and lower SES on average in the participants with depressive mood than in those without. Smoking and sex (female) were associated with a higher probability of the outcome being reported, and age with a lower probability (Table 3.3).

Table 3.3: Average rate ratios and 95% confidence intervals for variables in the proportional intensity model for monitoring times

Variable	Rate ratio	95 % CI
Depressive mood	0.93	0.84; 1.02
Smoking	1.08	1.03; 1.13
Age	0.94	0.93; 0.94
Sex=female	1.04	1.01; 1.07
SES	1.00	0.99; 1.01

Table 3.4: Comparison of the estimates of the marginal effect of depression status on average weight in pounds

	Estimate	95% CI
$\widehat{\beta}_{OLS}$	-3.83	-5.55; -2.11
$\widehat{\beta}_{VW}$	-3.69	-5.44; -1.94
$\widehat{\beta}_{IPT}$	-1.56	-3.45; 0.33
$\widehat{\beta}_{FIPTM}$	1.43	-0.35; 3.21
$\widehat{\beta}_{IPCTM}$	1.12	-0.59; 2.83

An important difference was found between the estimates for the marginal effect of depressive mood computed using $\widehat{\beta}_{OLS}$, $\widehat{\beta}_{VW}$, or $\widehat{\beta}_{IPT}$, and those obtained with our proposed estimators. The change in estimate seemed to be due to both confounding and informative monitoring times, with an important difference between $\widehat{\beta}_{OLS}$ and $\widehat{\beta}_{IPT}$, and an important remaining difference between $\widehat{\beta}_{IPT}$ and $\widehat{\beta}_{FIPTM}$ or $\widehat{\beta}_{IPCTM}$. The methods that did not account for confounding and informative monitoring times suggest that depressive mood leads to decreases in weight of nearly four pounds.

After adjusting for confounding and informative monitoring times, the estimates were consistent with those found in the literature. We found a small increase in weight due to depressive mood, with the lower limit of the confidence interval that corresponded to a weight loss of about half a pound, and an upper limit that consisted of a weight gain of just over three pounds. Wurtman [1993] explained the complex relationships leading to weight increase in patients with depression and the link with smoking. Studies such as van Strien et al. [2016] found no significant direct effect of depression on weight gain but only a positive effect through emotional eating as a mediator.

The differences observed and the sign reversal of the estimates after accounting for important features that may bias the estimates echo the results of Hernán et al. [2000]. The fact that we observed a reversal between the IPT-weighted estimator and the FIPTM and IPCTM estimators supports the message that informative visit process-induced bias should be accounted for.

3.5 Discussion

Electronic health records are increasingly available and a common source of data to study the effect of treatments on longitudinal outcomes in pharmacoepidemiological studies (Hennessy [2006]). Given their real-world nature, monitoring times in EHRs are often covariate-dependent and the outcome recorded may be associated with the same covariates, which introduces selection bias in the analysis. Most often, that feature is ignored. However, when it is considered, confounding bias is rarely accounted for, as – until now – no simple method has been described to account for the two sources of bias simultaneously. In this article, we proposed two novel estimators for the marginal effect of a treatment on a longitudinal outcome which account for imbalances due to covariate-dependent monitoring times, confounding and mediation. Neither estimator requires the longitudinal outcome to be measured at all times in continuous but rather only sporadically. The asymptotic properties of both estimators can be derived. These estimators are relevant to EHRs and to studies where irregular monitoring times were planned.

The proposed estimators were compared to more standard ones in simulation studies and both outperformed the OLS estimator, the weighted least squares estimator with an inverse monitoring weight and the inverse probability of treatment weighted estimator. Their empirical absolute bias tended towards 0, and the FIPTM estimator has shown good coverage. Moreover, we provided a practical framework for analysts, with both estimators being flexible with regards to the modelling of the intercept function. We recommend the use of the FIPTM estimator, which is easy to implement in practice and for which we have derived the asymptotic variance. For situations where the intercept function $\alpha(t)$ is expected to vary extensively in time, or for time-dependent treatment and confounders settings, the IPCTM estimator could be preferred and has shown to be well-behaved.

The estimators we propose rely on important assumptions. One challenge related to this work is the need for the treatment model to be correctly specified, and the risk for un-

measured confounding. Unmeasured confounding has been widely discussed, and sensitivity analyses are available to evaluate the degree at which it could impact the estimate of interest (Robins et al. [2000b]; Lash and Fink [2003]; Schneeweiss [2006]). In the situation where the treatment model is misspecified, the IPT weights may not provide adequate adjustment for confounding. Knowledge about the research problem should inform the set of potential confounders to incorporate into the treatment model. The use of directed acyclic graphs may help in determining which predictors should be included in the treatment model (Pearl [1995]; Greenland et al. [1999]), however these encode the analyst's beliefs and may themselves overlook important variables.

Another challenge is the need for the predictors of the monitoring process to be recorded at all times. In administrative databases and EHRs, information on drugs prescribed or dispensed, diagnostics and interventions are often recorded even in between physician visits when the outcome is monitored. For instance, in a study where the question is whether a particular drug impacts the outcome of blood pressure, blood pressure might be measured only when a patient's physician suspects changes in blood pressure and yet the patient potentially visited the physician at several other points, with data such as the exposure and comorbidities being recorded. In some observational studies, however, it will not be possible to assess covariate values in between the times when the outcome is measured. In that case, our methods could be extended to incorporate only the covariates measured at monitoring times, and to use them to predict the future monitoring times.

Acknowledgements

We would like to thank Professor Thomas Lumley for his generous assistance with the application of methods proposed in *Semiparametric modeling of repeated measurements under outcome-dependent follow-up* (Bůžková and Lumley [2009]). We are also very grateful to the Editor, the Associate Editor and two referees who provided helpful comments that greatly

improved the manuscript.

This work is supported by a doctoral scholarship from the Natural Sciences and Engineering Research Council (NSERC) of Canada (Ref. 401223940) to author JC. EEMM acknowledges support from a Discovery Grant from NSERC and a chercheur-boursier career award from the Fonds de recherche du Québec – Santé.

This research uses data from *Add Health*, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. No direct support was received from grant P01-HD31921 for this analysis.

Data availability statement

The data that support the findings in this paper are openly available in the Data Sharing for Demographic Research repository, at https://doi.org/10.3886/ICPSR21600.v21. The analysis was restricted to the Add Health public-use data and did not include restricted-use data (Harris and Udry [2018]).

Chapter 4

Estimation of the Marginal Effect of Antidepressants on Body Mass Index under Confounding and Endogenous Covariate-Driven Monitoring Times

Preamble to Manuscript 2. The endogeneity of a covariate process can be defined as the process interacting and possibly being modified by another process of interest (for instance, the monitoring process). Motivation for this chapter comes from the realization that the endogeneity of the covariates affecting the monitoring process may, in certain instances, affect the causal inference on the ATE. For instance, in the analysis of the CPRD data which motivates this work, where we aim to compare the causal marginal effects of two antidepressants on BMI, patients' characteristics affect visit timing, and their characteristics can be modified following physicians' visits. Yet, this idea of endogeneity has not been tackled in any previous work related to covariate-driven monitoring times. Some authors explicitly mentioned that their proposed method applies to situations where the covariate

process is exogenous (see e.g., Pullenayegum and Feldman [2013]), however most did not mention that explicitly. It would thus be useful to understand how the bias due to covariate-driven visits arises, the link between that bias and the endogeneity of the covariate process, and to study the consequences of endogeneity in a setting similar to that of Chapter 3.

This manuscript provides a thorough description of the bias due to covariate-dependent monitoring times, via the use of heuristic demonstrations with causal diagrams. A new weight is proposed to tackle endogeneity when that endogeneity may lead to long-term biasing dependencies between the monitoring and the outcome processes. Simulation studies are used to demonstrate that endogeneity in the covariate process may have to be considered. The proposed method is used to answer the research question, in an analysis of the CPRD data.

The methodological contributions of this manuscript are i) to provide the first thorough description of the bias due to covariate-driven monitoring times in longitudinal settings (i.e., longitudinal collider stratification bias) via causal diagrams, ii) to propose and demonstrate the first cumulated and stabilized weight for endogenous covariate-driven monitoring times (and to propose the first stabilizer for that type of weight), and iiii) relaxing the assumption on the availability of the covariate process affecting the monitoring process, whereby the method proposed in this manuscript only requires the covariate process to be observed at the same times as the outcome is observed. The substantive contribution of this work is iv) providing the first estimate for the marginal effect of citalopram and fluoxetine on body mass index when accounting for an extensive list of potential confounders and the covariate-driven monitoring times which may depend on an endogenous covariate process.

That manuscript is currently under review in *Annals of Applied Statistics*.

Estimation of the Marginal Effect of Antidepressants on Body Mass Index under Confounding and Endogenous Covariate-Driven Monitoring Times

Janie Coulombe[1], Erica EM. Moodie[1], Robert W. Platt[1], Christel Renoux[1,2,3].

[1]*Department of Epidemiology, Biostatistics, and Occupational Health, McGill University*

[2]*Department of Neurology and Neurosurgery, Faculty of medicine, McGill University*

[3]*Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal*

Abstract

In studying the marginal effect of antidepressants on body mass index using electronic health records data, we face several challenges. Patients' characteristics can affect the exposure (confounding) as well as the timing of routine visits (measurement process), and those characteristics may be altered following a visit which can create dependencies between the monitoring and body mass index when viewed as a stochastic or random processes in time. This may result in a form of selection bias that distorts the estimation of the marginal effect of the antidepressant. Inverse intensity of visit weights have been proposed to adjust for these imbalances, however no approaches have addressed complex settings where the covariate and the monitoring processes affect each other in time so as to induce endogeneity, a situation likely to occur in electronic health records. We review how selection bias due to outcome-dependent follow-up times may arise and propose a new cumulated weight that models a complete monitoring path so as to address the above-mentioned challenges and produce a reliable estimate of the impact of antidepressants on body mass index. More specifically, we do so using data from the Clinical Practice Research Datalink in the United Kingdom, comparing the marginal effect of two commonly used antidepressants, citalopram and fluoxetine, on body mass index. The results are compared to those obtained with simpler methods that do not account for the extent of the dependence due to an endogenous covariate process.

4.1 Introduction

Citalopram and fluoxetine are two selective serotonin reuptake inhibitor (SSRI) commonly prescribed as a first-line treatment in patients with depression. Weight gain is a side effect associated with the use of antidepressants (including SSRIs) which may lead to treatment discontinuation or non-adherence (see e.g. De las Cuevas et al. [2014]). Literature is undecided on whether there is a differential effect of these two drugs on weight (Sussman and Ginsberg [1998]; Serretti and Mandelli [2010]). No previous studies using electronic health records (EHR) data have compared their effect on weight or body mass index (BMI) after adjusting for potential confounders and selection bias due to outcome-dependent follow-up time. However, EHR and administrative databases do not collect data for research purposes, and patient BMI is not recorded at every follow-up visit. Its recording may be associated with patient characteristics.

When making inference on the causal effect of an exposure on a longitudinal outcome that is measured repeatedly over time using observational data, the analyst only observes data according to a measurement frequency that is driven by the characteristics of the individuals under observation. For instance, the smoking status of a patient, their comorbidities and general health state, their prescription drugs, or other characteristics can affect the frequency with which they visit their physician, and these same characteristics can change following a routine visit, as the physician may recommend lifestyle changes or the mere fact of having a visit may increase a patient's awareness of their habits. The imbalances related to the visit process can lead to biased estimates of the exposure effect in situations where the monitoring indicator is a collider and that conditioning on it opens a path from the exposure to the longitudinal outcome. This phenomenon is sometimes referred to as outcome-dependent follow-up times (Lipsitz et al. [2002]).

Several methods have been proposed to handle imbalances due to outcome-dependent follow-up, with most discussed outside of a causal framework. Some authors proposed joint mod-

elling of the monitoring and the outcomes processes via random effects or shared parameters (see e.g. Liang et al. [2009]). Pullenayegum [2016] discussed outputation, an approach that thins the visit process by repeatedly discarding selected observations, to make the monitoring process independent of covariates. Pullenayegum and Lim [2016] also presented an insightful review on methods for irregular visits, including inverse intensity of visit weights (IIVW). The proposed IIVW of Lin et al. [2004] apply to settings in which the monitoring rate is modelled as a function of exogenous covariates. It is particularly convenient to use inverse weighting in the context of causal inference on the marginal effect of exposure, as the weights can easily be incorporated into estimating equations to obtain an estimate of treatment effect, and they can be function of mediators of the exposure-outcome effect that affect monitoring.

Monitoring times and the outcome process are both stochastic processes which occur in continuous time. In the estimation of the marginal effect of antidepressants on BMI, a common set of characteristics affects the monitoring times and the BMI values, and the dependence between the monitoring and the BMI processes may be complex. Yet, to date, no weighting strategy has been proposed to address the difficult problem of an *endogenous* covariate process interacting with the monitoring process and the outcome process in continuous time in a setting where monitoring times are informative (note, a covariate process is said to be endogenous if the covariate is affected by the fact of there being a visit, which can create long-term dependencies between the monitoring and the outcome processes). Beyond the research question that motivates this work, several other examples of such processes exist, including the setting in which monitoring times allow a physician to take a treatment decision and prescribe a new treatment which may affect a patient's future outcome and monitoring time, or that where the advice of the physician may affect a patient's future habits (such as smoking). In all each of these examples, the gap time between visits may also act as an endogenous process that is modified by a new visit (and returns to zero when there is such visit). If the gap time is affected by variables that make the monitoring and the outcome

processes dependent in time, this too may induce biasing dependencies.

In this work, we focus on the estimation of the (causal) marginal effect of two antidepressants, citalopram and fluoxetine, on BMI which is measured repeatedly over time. We wish to address the possibility that the relationship of interest is distorted by imbalances due to the monitoring process and its relation with the BMI process, as well as confounding, which two features were never considered simultaneously in the assessment of the marginal effect of antidepressants on weight or BMI. Our methodology generalizes easily to the estimation of the causal marginal effect of a binary intervention, on a longitudinal continuous outcome, in observational studies. We provide a thorough description of the bias due to covariate-driven monitoring times in longitudinal settings, with a demonstration that relies on causal diagrams. In addition, we propose the first weighting method for addressing situations with longitudinal collider stratification bias (Greenland [2003]) that is due to an endogenous covariate process affecting the monitoring times. Another key contribution of our method, as opposed to those previously proposed, is that it loosens the restrictions on when the covariate process (affecting and being affected by monitoring process) has to be observed; standard inverse intensity of visit weights rely on the assumption that the monitoring model's covariates are assessed in continuous time while our methodology uses modelling of the gap times as functions of the last covariates observed, throughout patients' follow-up. This represents an important relaxation of the assumptions needed for consistent estimation.

This paper is structured as follows: in Section 4.2, we present the notation and assumptions, some background and the proposed extension. In Section 4.3, we present details on the simulation studies that were conducted to assess the performance of the proposed methods along with their results. In Section 4.4, we present the analysis of the data from the Clinical Practice Research Datalink (CPRD) (Herrett et al. [2015]), where we compare different estimators for the marginal effects of citalopram and fluoxetine on BMI using the novel weighting method. A discussion follows in Section 4.5.

4.2 Methods

4.2.1 Notation and Causal Assumptions

Let i denote the index of an individual, $i = 1, ..., n$, and t denote the time, which is continuous, with $t \in [0, \tau]$ and τ the maximum follow-up time in the cohort under study. We are interested in estimating the causal marginal effect of a binary point intervention $I_i(0)$ (i.e., a choice between two antidepressants) on the continuous, longitudinal outcome $Y_i(t)$ (BMI). Our interest is in a time-fixed ("point") intervention for two antidepressant drugs, however extensions to the weighting approach for the time-varying intervention case are straightforward. Bold is used to refer to vectors and matrices.

We use the Neyman-Rubin potential outcome framework (Neyman [1923]; Rubin [1974]) to express the estimand of interest, which is the causal contrast $\mathbb{E}\left[Y_{i1}(t) - Y_{i0}(t)\right]$ where $Y_{i1}(t)$ corresponds to the outcome that would have been observed at time t, had individual i received intervention $I_i(0) = 1$, and $Y_{i0}(t)$, had the individual received intervention $I_i(0) = 0$. In the analysis of the CPRD data, this contrast corresponds to the average difference in BMI had everyone been treated with citalopram, versus had everyone been treated with fluoxetine. Our interest is in the time-invariant effect of citalopram vs fluoxetine (i.e. we assume that the impact is not modified by time: $\mathbb{E}\left[Y_{i1}(t) - Y_{i0}(t)\right] = \theta$).

We are interested in addressing two important sources of bias that may distort estimators of our parameter of interest. First, we consider confounding bias, where confounders, denoted by $\mathbf{K(t)} = \mathbf{K(0)}$, are covariates that affect both the intervention and the outcome $Y_i(t)$ under study. As the intervention is time-fixed, confounders are measured at baseline, prior to receiving the intervention. Second, we consider selection bias caused by covariate-informed monitoring times (also referred to as *visit times* throughout). In this work, individuals are allowed to have completely different sets of monitoring times (and thus, a unique pattern of visits). In practice, we choose a certain coarsening of time (e.g., daily) over which we observe

monitoring times. To distinguish these two characterizations of time, we will assume, when needed, that the coarsening is daily and will denote discrete times by $t = 0$ (baseline), $1, 2, 3, ..., \tau$. For continuous time ($t \in [0, \tau]$), we will use the notation $t-$ to refer to the moment immediately before time t.

We suppose that monitoring times coincide with the observation of the longitudinal continuous outcome $Y_i(t)$. This means that we consider a monitoring time to be any time, and only those times, when BMI is assessed. We denote by $N_i(t)$ a counting process for monitoring times in individual i, which counts the number of previous visits they had by time t. A monitoring indicator at time t is denoted by $dN_i(t)$ in individual i. We also introduce the notation $l_i(t) = t - B_i(t)$, with $l_i(t)$ the previous (most recent) visit time of individual i at time t, where $B_i(t)$ is a time-dependent gap time which gives, at time t, the delay since the last visit in individual i. For each individual, let C_i denote the censoring time, that is, the time until which we can potentially observe an individual, their covariate and their outcome processes. Let $\xi_i(t) = \mathbb{I}(C_i \geq t)$ be the indicator of individual i still being in the study at time t. We assume throughout that censoring is uninformative except through the monitoring process. That is, we assume that we capture in modelling the monitoring process any (possibly biasing) imbalances in censoring times across the treatment groups. Denote by $\mathcal{H}_{\mathbf{i}}^{\mathbf{o}}(\mathbf{t}-)$ the *observed* history of covariates (personal characteristics, monitoring indicators, outcome values; these will be discussed in depth in Section 4.4, for the analysis of CPRD data) by time $t-$ in individual i. Since the intervention is given at baseline, $\mathcal{H}_{\mathbf{i}}^{\mathbf{o}}(\mathbf{t}-)$ may contain mediators of the relationship between the intervention and the longitudinal outcome $Y_i(t)$. We acknowledge that monitoring indicators $dN_i(t), t \in [0, \tau]$ can be colliders and block the path between the intervention and the outcome of interest throughout follow-up time. Conditioning on these colliders by using only the observed data can unblock the path between intervention and outcome and risks biasing the estimator of the marginal effect of intervention of interest.

We assume:

$$I_i(0) \perp \{Y_{i0}(t), Y_{i1}(t)\} \,|\mathbf{K_i}(\mathbf{0}) \tag{I1}$$

$$I_i(0) \not\perp \{Y_{i0}(t), Y_{i1}(t)\} \,|dN_i(t), \mathbf{K_i}(\mathbf{0}) \tag{I2}$$

$$I_i(0) \perp \{Y_{i0}(t), Y_{i1}(t)\} \,|\mathcal{H}^{\mathbf{o}}_{\mathbf{i}}(\mathbf{t}-), dN_i(t), \mathbf{K_i}(\mathbf{0}) \tag{I3}$$

$$0 < P(dN_i(t) = 1|\mathbf{K_i}(\mathbf{0}),\mathcal{H}^{\mathbf{o}}_{\mathbf{i}}(\mathbf{t}-)), P(dN_i(t) = 0|\mathbf{K_i}(\mathbf{0}), \mathcal{H}^{\mathbf{o}}_{\mathbf{i}}(\mathbf{t}-)) < 1 \tag{P1}$$

$$0 < P(I_i(0) = 1|\mathbf{K_i}(\mathbf{0})),P(I_i(0) = 0|\mathbf{K_i}(\mathbf{0})) < 1, \tag{P2}$$

such that $\mathcal{H}^{\mathbf{o}}_{\mathbf{i}}(\mathbf{t}-)$ and $\mathbf{K_i}(\mathbf{0})$ are sufficient sets to break the dependency between the potential outcome and the intervention, even when conditioning on the potential collider $dN_i(t)$. We further assume that we have positivity for intervention and monitoring (assumptions P1 and P2), where positivity for monitoring essentially means that there is at least a non-zero probability for any patient to have a visit on any given day (or week, depending on time granularity) while there is no day when it is 100% certain that a visit will occur. We also assume the stable unit treatment value assumption (SUTVA), a condition that encompasses a well-defined exposure or intervention, as well as no interference between individuals' effects.

Suppose that both the mean outcome process and the monitoring process at time t may depend on the confounders $\mathbf{K}(\mathbf{0})$, the intervention $\mathbf{I}(\mathbf{0})$, as well as a longitudinal, possibly vector-valued covariate process $\mathbf{Z_i}(\mathbf{l_i}(\mathbf{t})) \subset \mathcal{H}^{\mathbf{o}}_{\mathbf{i}}(\mathbf{t}-)$ that may be affected by monitoring times in the past. Like the outcome, we assume that this additional covariate process may only be assessed at monitoring times (such that $Z_i(t) = Z_i(l_i(t)) \,\forall t$), a not unrealistic scenario. We further assume the following conditional outcome mean model

$$\mathbb{E}\left[Y_i(t)|\mathbf{K_i}(\mathbf{0}), \mathbf{Z_i}(\mathbf{l_i}(\mathbf{t})), I_i(0)\right] = \alpha_0(t) + \boldsymbol{\beta_K}\mathbf{K_i}(\mathbf{0}) + \boldsymbol{\beta_Z}\mathbf{Z_i}(\mathbf{l_i}(\mathbf{t})) + \beta_I I_i(0), \tag{CO}$$

with $\alpha_0(t)$ a flexible intercept function. The variables $\mathbf{Z}(\mathbf{l_i}(\mathbf{t}))$ may include mediators of

the relationship between $I_i(0)$ and $Y_i(t)$ (in the context of our analysis, for example, a variable like having a diagnosis of diabetes is affected by antidepressant drugs and may itself affect individuals' weight). We aim to estimate the total effect of the intervention, without distortion by confounding or blocking paths which act through mediators. We henceforth use weights to create a pseudo-population in which there is no imbalance between confounder variables across intervention groups, and in which the monitoring and the outcome processes are independent. Using data from that pseudo-population, one can further use the marginal outcome mean model that follows

$$\mathbb{E}\left[Y_i(t)|I_i(0)\right] = \alpha(t) + \beta I_i(0); \tag{MO}$$

β in (MO) represents the average intervention effect of $\mathbf{I(0)}$ for which we seek an estimate. Denote by $\alpha(t)$ the intercept in the model, which may vary with time. We will also consider scenarios in which the intercept varies by individual ($\alpha_i(t)$).

In settings with outcome-dependent follow-up, Lin et al. (2004) have proposed the following continuous-time estimating equation to estimate the effect $\boldsymbol{\beta_0}$ of some time-fixed covariates $\mathbf{X}$ on an outcome $\mathbf{Y(t)}$ (which is not necessarily continuous):

$$\mathbb{E}\left[\int_0^\tau \left\{\mathbf{Y(t)} - \boldsymbol{\mu}(\boldsymbol{t}, \mathbf{X}; \boldsymbol{\beta_0})\right\} \frac{\mathbf{c}(\boldsymbol{t}, \mathbf{X}; \boldsymbol{\beta_0})}{\boldsymbol{\lambda}(\boldsymbol{t}|\mathcal{H}^{\mathbf{o}}(\mathbf{t}-))} \mathbf{dN(t)}\right] = \mathbf{0},$$

in which $\mathbf{c}(\boldsymbol{t}, \mathbf{X}; \boldsymbol{\beta_0})$ is an arbitrary weight, $\boldsymbol{\mu}(\boldsymbol{t}, \mathbf{X}; \boldsymbol{\beta_0})$ a mean function, and $\boldsymbol{\lambda}(\boldsymbol{t}|\mathcal{H}^{\mathbf{o}}(\mathbf{t}-))$, an IIVW which is a function of the history of observed variables.

Similarly, Coulombe et al. [2021a] proposed two flexible estimators for the marginal effect of the intervention; these accounted for confounding along with informative monitoring times and were tested in settings where the intervention can vary in time. In the context that interests us, where we aim to estimate β in the marginal outcome mean model (MO), the corresponding design matrix ($\mathbf{X(t)}$) is composed of the intervention $\mathbf{I(0)}$ and may also con-

tain functions of time such as a cubic spline basis to model the intercept function $\alpha(t)$. However, the authors did not consider individualized intercepts that can vary in time, a setting which raises additional challenges, and no previous author has considered endogeneity of the covariate process, which can create long-term dependence in the monitoring path that goes beyond the current covariates. Further, in both Bůžková and Lumley [2009] and Coulombe et al. [2021a], it is assumed that variables that affect monitoring time are observed at all times, a frequently unrealistic assumption. In the estimation of the marginal effect of antidepressants on BMI, in particular, considering these features will allow more flexibility. It will further allow us to postulate weaker assumptions on the monitoring process. In the CPRD data, patients' characteristics that lead them to visiting (or not) their physician are unlikely to be updated in continuous time; rather, their measurement mostly coincides with that of the outcome, BMI.

For the inverse weight (and intensity function) $\boldsymbol{\lambda}(\boldsymbol{t}|\mathcal{H}^{\mathbf{o}}(\mathbf{t}-))$, different modelling strategies have been considered by previous authors, such as different time scales or conditioning on different sets of covariates (Bůžková and Lumley [2009]; Zhu et al. [2017]; Coulombe et al. [2021a]). The intensity for a counting process is defined by the instantaneous rate, which is given by $\lambda_i(t|\mathcal{H}^{\mathbf{o}}_{\mathbf{i}}(\mathbf{t}-)) = \lim_{dt\to 0} P(N_i(t+dt) - N_i(t) = 1|\mathcal{H}^{\mathbf{o}}_{\mathbf{i}}(\mathbf{t}-))/dt$. The rate is preferred to a discrete probability model, as the monitoring "events" can occur at any time on a continuous time scale. In practice, and assuming that as dt gets closer to 0, the time units are so small that only one jump can occur per time unit dt, one can view this as a Bernoulli experiment over each small time unit dt with a certain probability of visit.

By definition, a conditional intensity model uniquely defines the counting process and its dependency on the past (Lindsey [2004]; Cook and Lawless [2007]), including previous monitoring times. The model may be affected by whether or not the covariates affecting the intensity function are endogenous. If they are exogenous, and if monitoring at time t does not depend on previous monitoring times, a marginal approach where the marginal effect of

covariates is estimated will suffice (e.g., as proposed in Lin et al. [2000]). If the covariate process is endogenous and if visit at time t depends on both the covariate process interacting with it and previous monitoring times, the conditional intensity function may have to account for complex functions of the past. In particular, links between the covariate, the monitoring and the outcome processes can exist due to the endogeneity.

4.2.2 Visit Process Scenarios and their Data Generating Mechanisms

We now propose different general scenarios for the monitoring times process, and describe the associated data generating mechanisms (DGMs). The first two DGMs refer to scenarios often encountered (and postulated) in the literature. The third and fourth DGMs correspond to situations where monitoring times affect the endogenous covariate process, such as we postulate in our analysis of the CPRD data. They are used to demonstrate the potential selection bias and the proposed methodology.

For each DGM, we review how the selection bias due outcome-dependent follow-up times arises, and which IIVW can be incorporated in the estimating equations for the marginal effect of intervention to make correct inferences. In each diagram, we depict all time points over which a visit can occur. We also assume that bias due to confounder variables is appropriately accounted for via classic adjustment methods such as the inverse probability of treatment (intervention) weight (Rosenbaum and Rubin [1983]).

The first DGM we consider is depicted in Figure 4.1 and is reminiscent of the causal diagram considered in Coulombe et al. [2021a] in which the marginal intensity of a visit at time t can be modelled using exogenous covariates measured at time t. Suppose that we aim to estimate the causal effect of the baseline intervention $I(0)$ on the longitudinal outcome $Y(t)$. In that DGM, at each time t, $Z(t)$ acts as a mediator of the relationship of interest. The selection bias due to outcome-dependent follow-up times comes from those mediators, as

conditioning on $dN(t)$, $t \geq 0$ unblocks the path going through colliders at each node $dN(t)$ $(t \in 0, 1)$, which opens the path $I(0) - dN(t) - Z(t) \rightarrow Y(t)\ \forall t$.

In that setting, to address the selection bias due to conditioning on the observed data, one can fit a proportional rate model for the rate of visit while conditioning on both $I(0)$ and $Z(t)$ at each time t as we have that $dN(t) \perp Y(t)|I(0), Z(t), \mathbf{K}(\mathbf{0})$. Further, there is no dependency structure across monitoring times that must be considered. We assume

$$\mathbb{E}\left[dN_i(t)|I_i(0), Z_i(t)\right] = \xi_i(t) \exp\left(\gamma_I I_i(0) + \gamma_Z Z(t)\right) \lambda_0(t)dt, \tag{4.1}$$

a proportional rate model where the effect of covariates is captured via the $\boldsymbol{\gamma} = \{\gamma_I, \gamma_Z\}$ parameters, and the effect of time via $\lambda_0(t)$. For the estimation of the $\boldsymbol{\gamma}$ parameters, one can use the Andersen and Gill model (Andersen and Gill [1982]), an extension of the Cox proportional hazards model for recurrent events. The baseline rate $\lambda_0(t)$ need not be estimated if the time scale is time since cohort entry, as the term will cancel out across individuals (Bůžková and Lumley [2009]). If the time scale is otherwise (e.g., if it is the gap time, $B_i(t)$), then the weight should incorporate a baseline function of that alternate time scale (Zhu et al. [2017]); Breslow-type estimators can be used to estimate the baseline rate (Cox [1972]). For fitting the model in (4.1), covariates $\{I_i(0), Z_i(t)\}$ must be measured (recorded) at all times during each patient's follow-up, which is not necessarily straightforward in observational longitudinal studies.

The second DGM we consider is presented in Figure 4.2 and is similar to the DGM considered in Zhu et al. [2017]. As compared to the causal diagram in Figures 4.1, covariates measured *before time* t are now assumed to affect both the monitoring and the outcome at time t, and to mediate the effect of interest. In this second DGM scenario, conditioning on colliders $dN(\cdot)$ by analyzing available data unblocks a path from the intervention and the outcome, via e.g. $I(0) - dN(1) - Z(0) \rightarrow Y(1)$, or, similarly, via $I(0) - dN(1) - Z(0) \rightarrow Y(2)$. A proportional rate model can be used for blocking these biasing paths from the intervention

to the outcome. We must now condition on covariates measured or updated in the past to adjust for selection bias, and must assume that no covariate measured after that point in the past affects both the next outcome and monitoring indicator. Suppose that we denote by $Z(t-)$ the last covariate value of $Z(\cdot)$ that affects, at time t, the monitoring indicator and the outcome; then the following model for monitoring will appropriately address the covariate-dependent monitoring times:

$$\mathbb{E}\left[dN_i(t)|I_i(0), Z_i(t-)\right] = \xi_i(t) \exp\left(\gamma_I I_i(0) + \gamma_Z Z_i(t-)\right) \lambda_0(t)dt. \tag{4.2}$$

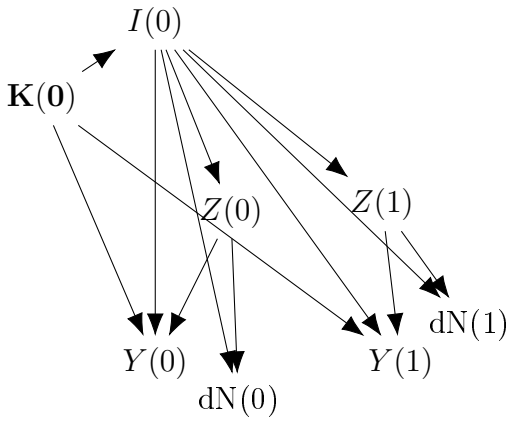


Figure 4.1: Causal diagram for the first DGM (patient index i removed) $I(0)$ is an intervention of interest whose marginal effect on a longitudinal outcome, $Y(t)$ – assumed to be time-invariant – is of interest. $\mathbf{K(0)}$ represent confounding variables, $Z(t)$ are mediators, and $dN(t)$ indicates the monitoring process through which the outcome is observed.

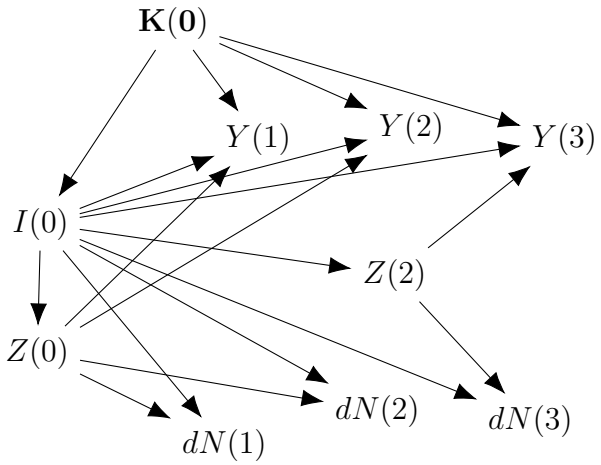


Figure 4.2: Causal diagram for the second DGM (patient index i removed) $I(0)$ is an intervention of interest whose marginal effect on a longitudinal outcome, $Y(t)$ – assumed to be time-invariant – is of interest. $\mathbf{K(0)}$ represent confounding variables, $Z(t)$ are mediators, and $dN(t)$ indicates the monitoring process through which the outcome is observed. Covariates $Z(t)$ are only "updated" at times 0 and 2 and affect next outcomes and monitoring indicators.

In the DGM in Figures 4.1 and 4.2, the visit rate depends on the covariate process but the fact of being monitored has no impact on the outcome or covariate processes. However, we postulate that our analysis of the marginal effect of antidepressants on BMI is such that previous monitoring affect the future monitoring path. For instance, the gap time (i.e. the time since a last visit) is an endogenous covariate process which is modified by each subsequent monitoring indicator. The smoking status of each patient, or other health

indicators, as well as drug prescriptions, could also be modified by a visit having taken place. When it is realistic to assume such an endogenous covariate process, and thus a "joint" process for the covariate, the monitoring and the outcome, dependencies between these processes may arise throughout the follow-up of a patient. This may mean that conditional on only the covariates measured most recently, the monitoring and the outcome processes are not independent. In particular, this may include situations where the monitoring path depends not only on the current covariates, but where it is also modified by what happened in the past (e.g., having had a physician visit yesterday makes my probability of visit today much lower). Effectively, previous monitoring indicators have interacted with covariates, such as the gap time, and the probability of visit on a given day may depend on the whole path the monitoring process went on (and all previous transitions). It may then be necessary to look at the monitoring path as a whole, and in particular to weight for the entire monitoring process – and not simply the most recent monitoring event – to ensure no unblocked paths between the intervention and outcome.

As an example of a scenario where these long-term dependencies may arise, we present a third DGM where the covariates are affected by the monitoring indicator through follow-up; practically, we implement this by including an interaction between the covariate and the visit indicator (Figure 4.3). In Figures 4.3 and 4.4, we use a notation as in Moodie and Stephens [2020] and use the symbol * to refer to an interaction between the covariates whose arrows point into the *. The interaction terms are necessarily deterministic, as an interaction term is solely determined by the respective variables that interact together. However, the monitoring indicators themselves are random, and an individual can transition to a visit or not on any given day.

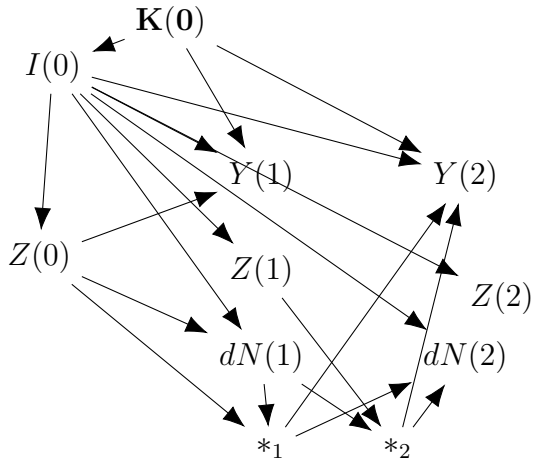


Figure 4.3: Causal diagram for the third DGM (patient index i removed) $I(0)$ is an intervention of interest whose marginal effect on a longitudinal outcome, $Y(t)$ – assumed to be time-invariant – is of interest. $\mathbf{K(0)}$ represent confounding variables, $Z(t)$ are mediators, and $dN(t)$ indicates the monitoring process through which the outcome is observed. Asterisks represent interactions between the covariates whose arrows point into it.

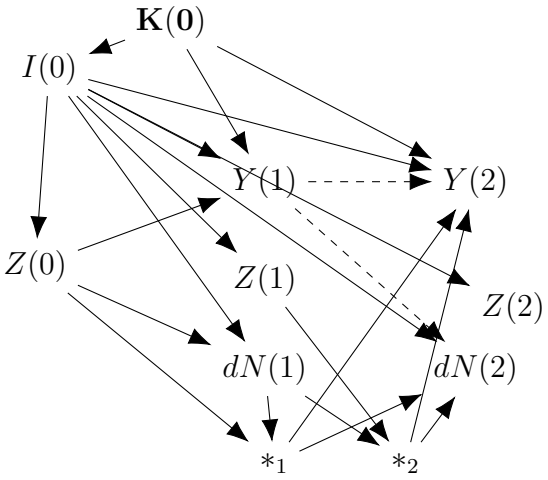


Figure 4.4: Causal diagram for the fourth DGM (patient index i removed) $I(0)$ is an intervention of interest whose marginal effect on a longitudinal outcome, $Y(t)$ – assumed to be time-invariant – is of interest. $\mathbf{K(0)}$ represent confounding variables, $Z(t)$ are mediators, and $dN(t)$ indicates the monitoring process through which the outcome is observed. Asterisks represent interactions between the covariates whose arrows point into it. Dashed lines represent the new links added from the causal diagram presented in Figure 4.3.

In Figure 4.3, the covariate $Z(0)$ mediates the effect of $I(0)$ on $Y(1)$. The covariate process $Z(\cdot)$ interacts with monitoring and whenever there is a new monitoring time s (where $dN(s) = 1$ for some $s > 0$), the covariate process $Z(\cdot)$ is updated while still depending on the intervention at baseline. In that case, the selection bias due to outcome-dependent follow-up times cannot be adjusted for by using only the standard IIVW put forward by previous authors, as presented in equations (4.1) or (4.2) because a biasing path remains via the interactions between covariates $Z(\cdot)$ and the monitoring indicators $dN(\cdot)$. Using only the observed data and therefore conditioning on $dN(\cdot)$ opens colliders at the $dN(t)$, $\forall t > 0$. After conditioning, one example biasing path from the intervention to the outcome is given by $I(0) \rightarrow Z(0) \rightarrow *_1 - dN(2) - *_2 \rightarrow Y(2)$. This path remains open even if we adjust for the last covariates observed $\mathbf{Z_i}(\mathbf{l_i}(\mathbf{t}))$ and for the intervention. Further, adjusting for the previous interaction term (between the last monitoring indicator and the most recent co-

variates observed) will not suffice, as this last interaction depends on the whole monitoring path (for instance, the interaction term for the previous monitoring indicator will be 0 if there was no visit yesterday, providing no adjustment for the previous non-null interaction term that occurred before yesterday). We graphically demonstrate some examples of biasing paths that arise after conditioning on the collider $dN(t)$ in the causal diagrams shown in Figures 4.1 to 4.4 along with a heuristic demonstration of why simpler weights do not always account properly for outcome-dependent monitoring times in Appendix B.1.

In Figure 4.4, the outcome process $Y(\cdot)$ also affects the monitoring rate for any given time in between the current monitoring time and the next monitoring time, as well as affecting the next outcome value. One consequence of this is that conditioning on $dN(2)$ opens a path between the intervention and the outcome through e.g. the path $I(0) \rightarrow Z(0) \rightarrow *_1 - dN(2) - Y(1)$. Other biasing paths due to colliders $dN(t), t > 0$, similar to those discussed for the third DGM, can also be found.

In the following section, we present our proposed inverse weighting method that can – unlike previous approaches – appropriately account for endogenous covariate-dependent monitoring processes as in the third and fourth described DGM above.

4.2.3 A new weighting approach: Extension using the joint monitoring path

To ensure that we break the dependence between the outcome and the monitoring processes in our estimation of the marginal effect of antidepressants on weight (or, similarly, in settings such as in the third and fourth DGMs depicted in Figures 4.3 and 4.4), we propose an approach inspired by transition intensities and occupation probabilities used in the multistate models literature (Cook and Lawless [2018]). In this approach, we effectively account for the full (observed) covariates history and the joint monitoring process.

The first step in our proposed approach is to model what we term a *partial* conditional visit

intensity at each point in time, which will represent the "transition intensity" to the state of being monitored (or a *visit*). In this context, partial refers to the fact that we only condition on a subgroup of covariates measured in the past, and thus only require for these covariates to make subsequent monitoring indicators independent. We make the following assumption on monitoring indicators:

$$dN_i(t) \perp Y_i(t) | \mathcal{H}_{\mathbf{i}}^{\mathbf{o}}(\mathbf{t}-) \tag{I4}$$

and further assume that only the subset $\{\mathbf{Z_i}(\mathbf{l_i}(\mathbf{t})), I_i(0), B_i(t-)\}$ of $\mathcal{H}_{\mathbf{i}}^{\mathbf{o}}(\mathbf{t}-)$ affecting the partial intensity at time t are sufficient to break the dependence between subsequent monitoring indicators at each time t:

$$dN_i(t) \perp dN_i(t-) | \mathbf{Z_i}(\mathbf{l_i}(\mathbf{t})), I_i(0), B_i(t-), \tag{I5}$$

where $dN_i(t-)$ is the last visit indicator observed prior to time t, and $B_i(t-)$, the last gap time. (In discrete time these would correspond to e.g. $dN_i(t-1)$ and $B_i(t-1)$; note that these do not encompass the history of gap times and visits, only the values attached to the previous time unit.) Assumption (I5) implies that given the previous gap time and the covariates observed at the previous visit time, the two subsequent monitoring indicators are independent of one another. This is a conditional Markov assumption for the monitoring process that allows us to decompose the process into a series of monitoring indicators. Note that $\mathbf{Z_i}(\mathbf{l_i}(\mathbf{t}))$ could contain many different kinds of predictors of the visit intensity, including mediators of the relationship between $I_i(0)$ and $Y_i(t)$ and functions of gap times or of time since cohort entry.

We quote Theorem 1 in Pearl [2009] (pp. 110), which we rely on to compute the joint intensity of a given monitoring path:

Theorem 1 (The Causal Markov Condition). *Any distribution generated by a Markovian*

model M can be factorized as:

$$P(v_1, v_2, ..., v_n) = \prod_i P(v_i|pa_i)$$

where $V_1, V_2, ...V_n$ are the endogenous variables in M, and pa_i are (values of) the endogenous "parents" of V_i in the causal diagram associated with M.

To model the partial conditional (visit transition) intensity at time t, based on assumption (I5), we propose:

$$\lambda_i(t|\mathbf{Z_i}(\mathbf{l}(\mathbf{t_i})), I_i(0), B_i(t-)) = \lambda_0(B_i(t)) \exp(\gamma_I I_i(0) + \boldsymbol{\gamma_Z}\mathbf{Z_i}(\mathbf{l_i}(\mathbf{t}))), \tag{4.3}$$

where $B_i(t)$ is a function of $B_i(t-)$ (which justifies the condition on that covariate in I5).

To model a *personalized* baseline intensity $\lambda_0(B_i(t))$, inspired by Zhu et al. [2017], we propose a Breslow-type estimator (Cox [1972]), modified to be a function of the gap time since a last visit. For a gap time $B(t)$, it is given by

$$\hat{\lambda}_{0,1}(B(t)) = \frac{\sum_{i=1}^{n} \int_{s=0}^{\tau} \mathbb{I}(dN_i(s) = 1 \cap B_i(s) = B(t))}{\sum_{i=1}^{n} \int_{s=0}^{\tau} \exp\left(\hat{\gamma}_I I_i(0) + \hat{\boldsymbol{\gamma}}_{\boldsymbol{Z}}\mathbf{Z_i}(\mathbf{l_i}(\mathbf{s}))\right) \mathbb{I}(dN_i(s) = 1 \cap B_i(s) = B(t))}, \tag{S1}$$

for $\mathbb{I}(\cdot)$ an indicator function. The logical statement $dN_i(s) = 1 \cap B_i(s) = B(t)$ means that patient i both has a visit at time s and that their gap time is $B_i(s) = B(t)$.

The intensities in (4.3) are fitted using the Andersen and Gill model's main effects. As time is continuous, we compute the product integral of the transition intensities in (4.3) to compute the probability of having a given monitoring path up to time t. The product integral consists in the extension of the sum integral to the product (Gill and Johansen [1990]). A well-known estimator utilizing the product integral is the Kaplan-Meier estimator (Kaplan and Meier [1958]). Here, transitions refer to those from the *non-visit* to the *visit* state, and vice-versa.

For each time t, we assume the following *simplified* transition matrix for individual i:

$$\begin{bmatrix} 1-\xi_i(t)\exp\left(\gamma_I I_i(0)+\boldsymbol{\gamma_Z}\mathbf{Z_i}(\mathbf{l_i}(\mathbf{t}))\right)\lambda_0(B_i(t))dt & \xi_i(t)\exp\left(\gamma_I I_i(0)+\boldsymbol{\gamma_Z}\mathbf{Z_i}(\mathbf{l_i}(\mathbf{t}))\right)\lambda_0(B_i(t))dt \\ 1-\xi_i(t)\exp\left(\gamma_I I_i(0)+\boldsymbol{\gamma_Z}\mathbf{Z_i}(\mathbf{l_i}(\mathbf{t}))\right)\lambda_0(B_i(t))dt & \xi_i(t)\exp\left(\gamma_I I_i(0)+\boldsymbol{\gamma_Z}\mathbf{Z_i}(\mathbf{l_i}(\mathbf{t}))\right)\lambda_0(B_i(t))dt \end{bmatrix},$$

and, depending on which state was occupied at the very previous time unit, only some of these transitions will be non-null for each individual, at each time (this is why we call it *simplified*; in reality, each element from the matrix above should be augmented with an indicator for the previous monitoring indicator, $\mathbb{I}(dN_i(t-))$, and the gap time and the covariates $\{I_i(0), \mathbf{Z_i}(\mathbf{l_i}(\mathbf{t}))\}$ will potentially be different on each row of the matrix, depending on whether or not there was a visit at time $t-$). We rely on assumptions (I4), (I5), and **Theorem 1**, and take the product integral of the intensities for a given patient i; this leads to the probability of a given monitoring path conditional on the observed history of covariates, which is shown in (4.4). The product symbol in (4.4) refers to the product integral (as opposed to the standard product term).

$$\begin{aligned} usw_i(t|\mathcal{H}_{\mathbf{i}}^{\mathbf{o}}(\mathbf{t}-)) = \prod_{s=0}^{t} &\left\{\xi_i(s)\exp\left(\gamma_I I_i(0)+\boldsymbol{\gamma_Z}\mathbf{Z_i}(\mathbf{l_i}(\mathbf{s}))\right)\lambda_0(B_i(s))ds\right\}^{\mathbb{I}(dN_i(s)=1)}\times \\ &\left\{1-\xi_i(s)\exp\left(\gamma_I I_i(0)+\boldsymbol{\gamma_Z}\mathbf{Z_i}(\mathbf{l_i}(\mathbf{s}))\right)\lambda_0(B_i(s))ds\right\}^{\mathbb{I}(dN_i(s)=0)}. \end{aligned} \quad (4.4)$$

The weight in (4.4) risks being highly variable. Furthermore, the inverse weight risks leading to extreme values, so that any estimator relying on it would consequently also have high variability. To address this, we propose two alternative stabilizers to incorporate in (4.4) and to be cumulated over each dt time units. The first stabilizer is given by $\hat{\lambda}_{0,1}(B(t))$ as shown in equation (S1), such that the baseline rate over each dt time unit in (4.4) cancels out. We compare this to a second proposed stabilizer (S2) that conserves the true effect of gap time from the fitted baseline rate. The second stabilizer uses a different estimator for

the baseline rate that only depends on baseline covariates ($\mathbf{I(0)}$) and is given by

$$\hat{\lambda}_{0,2}(B(t)) = \frac{\sum_{i=1}^{n} \int_{s=0}^{\tau} \mathbb{I}(dN_i(s) = 1 \cap B_i(s) = B(t))}{\sum_{i=1}^{n} \int_{s=0}^{\tau} \exp\left(\hat{\delta}_I I_i(0)\right) \mathbb{I}(dN_i(s) = 1 \cap B_i(s) = B(t))}, \tag{S2}$$

for δ the parameter in a proportional intensity model for monitoring times with the covariate $\mathbf{I(0)}$ as the only predictor. The second stabilizer (S2) does not depend on the endogenous covariate process $\mathbf{Z_i}(\mathbf{l_i}(\cdot))$, but rather depends only on covariates measured at cohort entry. Unlike the stabilizer (S1) which results in a weight that does not adjust for the impact of the gap time on monitoring, the stabilizer (S2) may more completely account for structures where gap time itself affects the monitoring process.

Including the stabilizers $\lambda_{0,1}(B(t))$ or $\lambda_{0,2}(B(t))$ in the denominator in equation (4.4) leads to the second proposed weight ($j \in 1, 2$):

$$sw_{i,j}(t|\mathcal{H}^{\mathbf{o}}(\mathbf{t}-)) = \prod_{s=0}^{t} \left(\frac{\xi_i(s) \exp\left(\gamma_I I_i(0) + \boldsymbol{\gamma_Z} \mathbf{Z_i}(\mathbf{l_i}(\mathbf{s}))\right) \lambda_0(B_i(s)) ds}{\lambda_{0,j}(B_i(s)) ds} \right)^{\mathbb{I}(dN_i(s)=1)} \times$$
$$\left(\frac{1 - \xi_i(s) \exp\left(\gamma_I I_i(0) + \boldsymbol{\gamma_Z} \mathbf{Z_i}(\mathbf{l_i}(\mathbf{s}))\right) \lambda_0(B_i(s)) ds}{1 - \lambda_{0,j}(B_i(s)) ds} \right)^{\mathbb{I}(dN_i(s)=0)}. \tag{4.5}$$

By cumulating the intensity through time and by using its product as an inverse weight, we control (under stated assumptions) for the entire monitoring process conditional on the covariates' history. This weighting results in independence between the covariates and the monitoring process, so that their effect on the longitudinal outcome process can be estimated without bias.

Once the weights are defined, similarly to Coulombe et al. [2021a], we use the following

estimating equation for the marginal effect of intervention

$$\mathbb{E}\left(\int_0^\tau \frac{\mathbf{Y(t)} - [\boldsymbol{\beta}_s' \mathbf{S(t)}]}{\boldsymbol{w}(\mathbf{t}|\mathbf{K(0)})\boldsymbol{sw_j}(\mathbf{t}|\mathcal{H}^{\mathbf{o}}(\mathbf{t}-))} \mathbf{dN(t)}\right) = \mathbf{0}, \tag{4.6}$$

where $\mathbf{S(t)}$ is a matrix containing a column of ones and a cubic spline basis to flexibly model time t (together representing $\boldsymbol{\alpha}(\mathbf{t})$), and the intervention variable $\mathbf{I(0)}$, and where $\beta \subset \boldsymbol{\beta_s}$ is the parameter of interest. That estimating equation is unbiased for the parameter of interest (proof given in Appendix B.2). The weight $\mathbf{sw_j}(\mathbf{t}|\mathcal{H}^{\mathbf{o}}(\mathbf{t}-))$ in (5.8) can be replaced by another with equivalent properties (balancing properties, positivity) such as $\mathbf{usw}(\mathbf{t}|\mathcal{H}^{\mathbf{o}}(\mathbf{t}-))$, without modifying the rest of the estimating equation. The function $w_i(t|\mathbf{K(0)})$ can be estimated by using a correctly specified function of the covariates $\mathbf{K(0)}$ that breaks the dependence between the intervention and the confounders $\mathbf{K(0)}$. For instance, if the intervention and the confounders are time-fixed and assessed at time 0, an inverse probability of treatment weight can be used, and defined as the inverse of $w(t|\mathbf{K_i(0)};\boldsymbol{\omega}) = P(I_i(0) = 1|\mathbf{K_i(0)};\boldsymbol{\omega})\mathbb{I}(I_i(0) = 1) + (1 - P(I_i(0) = 1|\mathbf{K_i(0)};\boldsymbol{\omega}))\mathbb{I}(I_i(0) = 0)$. That weight will account for imbalances due to confounders under the assumptions (I1) and (P2) if there is no conditioning on monitoring indicators, or under assumptions (I3) and (P2) if there is, and that the monitoring intensity is also taken into account. If the intervention is not time-fixed and if time-varying confounding exists that simultaneously acts as intermediate variables, methods such as presented in the seminal paper by Robins et al. [2000a] can be used to recover balance.

For details on asymptoptic properties and how to compute a conservative asymptotic variance of the estimator for the marginal effect of intervention resulting from equation (5.8), we refer the reader to Appendix B.3. In simulation studies, we use a nonparametric bootstrap to estimate the variance, re-sampling *individuals* rather than observations to ensure within-person correlation of measures is preserved.

4.3 Simulation studies

We performed several simulation studies to assess the performance of the proposed weights to adjust for imbalances due to the informative monitoring process in settings similar to that of our research question which we answer using data from the CPRD. Our aim was to estimate the marginal effect of a binary (point) intervention on a continuous longitudinal outcome in contexts with confounding and informative monitoring times, where monitoring times were simulated in a sequential manner, dependent on previous information.

In the main study, we simulated for each patient i three baseline confounders as $K_{1i} \sim \text{N}(1,1), K_{2i} \sim \text{Bernoulli}(0.55)$, and $K_{3i} \sim \text{N}(0,1)$. The intervention $I_i(t)$ was binary and time-fixed: $I_i \sim \text{Bernoulli}(p_{Ii})$ with $p_{Ii} = \text{expit}\left(0.5 + 0.8K_{1i} + 0.05K_{2i} - 1K_{3i}\right)$. One time-varying mediator $Z_i(\cdot)$ was generated, conditional on I_i. It was only updated whenever there was a new visit ($dN_i(\cdot) = 1$), and was simulated as $Z_i(t)|I_i = 1 \sim \text{N}(2,1)$ and $Z_i(t)|I_i = 0 \sim \text{N}(4, 2^2)$ on those visit days. On other (non-visit) days, we denote the process by $Z_i(l_i(t))$, simply carrying forward the last observed value. Time was discretized over a grid of 0.01 units, from 0 to τ. The intensity of monitoring at each time point over that grid was simulated as $\lambda_i(t|I_i, Z_i(l_i(t))) = 0.02B_i(t)\exp\left(\gamma_1 I_i + \gamma_2 Z_i(l_i(t))\right)$. The outcome $Y_i(t)$ was generated according to $Y_i(t) = 0.2B_i(t) + 1\,I_i - 0.8\left(Z_i(l_i(t)) - E\left[Z_i(l_i(t))|I_i\right]\right) + 0.4\,K_{1i} + 0.05\,K_{2i} - 0.6\,K_{3i} + \epsilon_i(t)$ with $\epsilon_i(t) \sim \text{N}(0, 0.5^2)$. Monitoring times were drawn up until the maximum follow-up time τ, which we fixed to $\tau = 5$. Data were simulated to correspond to a study cohort of 500 patients. For each patient, the follow-up time was "censored" (stopped) at time C_i, with $C_i \sim \text{Uniform}(\tau/2, \tau)$; the censoring was non-informative. A total of 1000 replicate datasets were simulated for each simulation study scenario. More details on the simulation study can be found in Appendix B.4.

We compared a naive ordinary least squares estimator that did not account for the confounding or the informative monitoring process ($\hat{\beta}_{LS}$) and an inverse probability of treatment weighted least squares estimator in which the treatment model was correctly specified

but that did not account for the monitoring process ($\hat{\beta}_{IPT}$) with four "doubly weighted" approaches that incorporated a correctly specified IPT weight and alternate versions of IIVW to account for the monitoring process. Specifically, these four estimators relied on intensity of visit weights that:

- did not account for the full history of covariates but only for the last covariates observed, $I_i(0)$ and $Z_i(l_i(t))$ and for the correct baseline intensity through the following inverse weight $\hat{\lambda}_0(B_i(t)) \exp\left(\hat{\gamma}_1 I_i + \hat{\gamma}_2 Z_i(l_i(t))\right)$ (namely $\hat{\beta}_{IH}$),
- used the novel inverse cumulated unstabilized weight ($\hat{\beta}_{USW}$),
- used the cumulated stabilized weights $SW1$ ($\hat{\beta}_{SW1}$), or
- used the cumulated stabilized weights $SW2$ ($\hat{\beta}_{SW2}$)

where the latter two fully account for the full covariate process and its interaction with visit times but used different stabilizing strategies.

The cumulated weights were censored at the respective 2.5 and 97.5th percentiles of their distribution for all three corresponding estimators.

In four sensitivity analyses, we 1) fitted a constant intercept in the outcome model rather than a cubic spline basis for the effect of time; 2) varied the maximum follow-up time ($\tau = 10$ rather than $\tau = 5$); 3) changed the intercept function in the outcome from $0.2B_i(t)$ to simply 0.02; and 4) varied the endogenous process $Z(\cdot)$ such that its mean depended on the cumulative number of previous visits, as: $Z_i(t)|I_i = 1 \sim \mathrm{N}(2 + 0.2 \int_0^{t-} dN_i(s), 1)$ and $Z_i(t)|I_i = 0 \sim \mathrm{N}(4 + 0.2 \int_0^{t-} dN_i(s), 2^2)$ updated on visit days, with the rest unchanged. The results for the sensitivity analyses are presented in Appendix B.6.

Table 4.1 shows the results of the main simulation study in terms of empirical absolute bias and bootstrap variance. The Andersen and Gill model consistently provided unbiased estimates of the parameters $\boldsymbol{\gamma}$ in the monitoring model in all these scenarios, and average numbers of visits varied between 1.9 and 7.1 (Appendix B.5). The Breslow-type estimator

used to estimate $\lambda_0(B_i(t))$ provided consistent estimates (results not shown).

In the main analysis, the least squares estimator was biased (bias ranging between 0.35-0.73). Accounting for the confounding improved the performance of the estimator, as seen with the absolute bias of $\hat{\beta}_{IPT}$ (range between 0.01 and 0.37). However, both the IPW estimator and the doubly weighted estimator $\hat{\beta}_{IH}$ remained biased, with the range of bias depending on the strength of the effect of covariates in the visit process (as controlled by the $\boldsymbol{\gamma}$s). This supports our claim that existing approaches do not fully adjust for a visit process in settings with endogeneity and long-term dependencies.

With respect to the cumulated weights, only $\hat{\beta}_{SW2}$ consistently provided unbiased estimates. Its variance was also relatively small, compared with the other simpler estimators. The unstabilized weighted estimator $\hat{\beta}_{USW}$ was highly variable, and censoring its cumulated intensity weights at the 2.5th and 97.5 percentiles did not reduce its variance to a satisfactory degree (Table 4.1). This could be due to the sample size, or the number of simulations conducted. We present the absolute bias; the empirical bias varied between negative and positive, so the direction of the bias was not systematic. For the first of the stabilized weighted estimators, $\hat{\beta}_{SW1}$, we hypothesize that the weights might not adequately account for the effect of time since the baseline rate canceled out after stabilization. In fact, we observed that the estimator $\hat{\beta}_{SW1}$ performed slightly better than $\hat{\beta}_{IH}$ in general. This may be due to its adjustment for part of the endogeneity/dependence due to covariates since the data were generated such that $Z(\cdot)$ was simulated according to a random Normal variable with a mean that did not vary across follow-up. Although on average, the process remained centered around the same value, adjusting for the process $Z(\cdot)$ in $\hat{\beta}_{SW1}$ nonetheless accounted for the variation of $Z(\cdot)$ around its mean.

In sensitivity analyses (Appendix B.6), we found very similar results, with $\hat{\beta}_{SW2}$ performing better across the board. All results remained similar when using a constant intercept as compared to a more flexible intercept with a cubic spline basis, as well as when increasing

the maximum follow-up time, τ, to 10. When changing the endogenous covariate distribution that affected the monitoring and the outcome processes to make it dependent on the cumulative number of previous visit times, bias was not, in general, much greater. In the fourth sensitivity analysis where we changed the intercept function to a constant in the definition of the outcome, similar results were observed.

Table 4.1: Main analysis: Mean absolute bias and bootstrap variance for the estimators compared (1000 simulations, $n = 500$ patients).

γ	Mean absolute bias of the estimator						Bootstrap variance of the estimator					
	$\hat{\beta}_{LS}$	$\hat{\beta}_{IPT}$	$\hat{\beta}_{IH}$	$\hat{\beta}_{USW}$	$\hat{\beta}_{SW1}$	$\hat{\beta}_{SW2}$	$\hat{\beta}_{LS}$	$\hat{\beta}_{IPT}$	$\hat{\beta}_{IH}$	$\hat{\beta}_{USW}$	$\hat{\beta}_{SW1}$	$\hat{\beta}_{SW2}$
-0.3; 0.1	0.35	0.37	0.12	0.04	0.14	0.03	0.03	0.07	0.16	1.56	0.13	0.08
-0.2; 0.2	0.49	0.24	0.11	0.00	0.09	0.01	0.03	0.06	0.11	1.23	0.08	0.06
-0.1; 0.2	0.64	0.08	0.09	0.14	0.00	0.02	0.02	0.05	0.09	1.11	0.06	0.05
-0.1; -0.3	0.69	0.01	0.11	0.07	0.01	0.03	0.02	0.05	0.08	0.59	0.06	0.05
0; 0	0.73	0.01	0.03	0.16	0.01	0.01	0.02	0.05	0.09	0.56	0.05	0.05
0.1; -0.3	0.69	0.03	0.11	0.19	0.04	0.02	0.02	0.04	0.06	0.40	0.05	0.05
0.2; -0.2	0.64	0.12	0.26	0.18	0.19	0.02	0.02	0.03	0.06	0.36	0.05	0.05
0.3; 0.2	0.67	0.08	0.34	0.24	0.30	0.05	0.02	0.03	0.06	0.26	0.07	0.05

The results for the variance were as expected: the use of unstabilized cumulated weight in $\hat{\beta}_{USW}$ led to a large variance, while the variance of other estimators remained low. Stabilizing the cumulated weights in $\hat{\beta}_{SW1}$ and $\hat{\beta}_{SW2}$ led to smaller variance, in general, than the more classical (and, in these settings, biased) estimator that accounted for the visit process, $\hat{\beta}_{IH}$. A comparison of the empirical and the bootstrap variances of all estimators compared in the main analysis can be found in Appendix B.7; the bootstrap variance was slightly greater than the empirical variance, in general.

4.4 Comparison of the effect of citalopram and fluoxetine on BMI using the CPRD data

We use the proposed methodology to estimate the marginal effect of citalopram and fluoxetine on BMI, relying on data from the Clinical Practice Research Datalink in the United Kingdom

(UK). The CPRD is one of the largest primary care databases of de-identified data. It contains data from more than 13 million patients treated in general practices from across the UK, including demographics, anthropometric measurements such as BMI, lifestyle factors, all prescriptions issued by general practitioners (recorded according to the British National Formulary), and medical diagnoses (coded using the Read Classification System). The CPRD data were linked with the Hospital Episode Statistics (HES) repository and the Office for National Statistics (United Kingdom) (ONS) mortality database. These provided access to further patients information on related diagnoses for each hospital stay (coded using the International Classification of Diseases version 10), and dates of death.

Our study protocol was approved by the Independent Scientific Advisory Committee of the UK Clinical Practice Research Datalink (protocol number 19_017R) and the Research Ethics Committee of the Jewish General Hospital (Montreal, Quebec, Canada).

We defined a cohort of adult new users of citalopram or fluoxetine who had a confirmed diagnosis of depression in the year prior to initiation. To be included, patients had to initiate their treatment for one of the two study drugs between April 1st, 1998, and December 31st, 2017. The final cohort comprised 246,503 patients (56% citalopram new users); see Coulombe et al. [2021b] for details of the cohort construction. Patients were followed until a first code for pregnancy, treatment discontinuation for citalopram or fluoxetine, switch to any other antidepressant drug, end of CPRD coverage, administrative end of study (December 31st, 2017), death, or when reaching a maximum follow-up time set to 18 months, whichever happened first. Patients were considered continuously exposed if a subsequent prescription for the initiating drug was issued less than 30 days after the end of the duration of the last prescription for the corresponding drug.

Over the follow-up period, we collected any data on BMI and considered as a monitoring time any day when BMI was recorded by the general practitioner. Other days when either 1) no physician or hospital visit occurred; or 2) such visit occurred but no BMI was measured

and recorded, were considered to be the same (i.e. *no monitoring*); we did not model them any differently.

The intervention (citalopram or fluoxetine), age, sex, and Index of Multiple Deprivation (IMD) were defined at baseline and were included as predictors in the BMI monitoring model. The IMD is a measure of relative deprivation for small areas in England (Deas et al. [2003]); it was used as a proxy for socioeconomic status. We assumed that certain time-varying covariates were potentially modified by a visit and could affect the next BMI outcome and the timing of the next visit. These covariates were defined for each day during follow-up, and included the smoking status, diabetes or antidiabetic drug use, alcohol abuse, a diagnosis for anxiety or generalized anxiety disorder (GAD), other psychiatric diseases (including autism spectrum disorder, obsessive compulsive disorder, and schizophrenia), the number of hospitalisations in the previous month, and benzodiazepine drug, lipid lowering drug, or antipsychotic drug use (all three considered separately). We fitted a proportional intensity model as a function of all covariates mentioned above for the visit model. The visit intensity was modelled as a function of the gap time $B_i(t)$, accounted for via the baseline intensity $\lambda_0(B_i(t))$, and all predictors in the model. It was used to compute the different weights and corresponding estimators described in Section 4.2 and compared in Section 4.3. For every time unit (one day), we obtained an estimate of the visit intensity.

We assumed that the relationship between the intervention (citalopram or fluoxetine) and BMI was potentially confounded by a set of covariates measured at baseline. These included age, sex, the IMD, smoking status, diabetes, alcohol abuse, anxiety or GAD, other psychiatric diseases, the number of previous hospitalisations, as well as the use of lipid lowering therapy, benzodiazepine drugs, or antipsychotic drugs. We fitted a logistic regression model to predict the intervention at baseline (citalopram or fluoxetine) and included the confounder variables mentioned above as predictors in the model. The fit led to a propensity score that was used in an inverse probability of treatment weight, to address confounding.

For the monitoring and the exposure models mentioned above, covariates defined once at baseline were defined in the same way as in Coulombe et al. [2021b] (any medication use was measured in the year prior to cohort entry, and comorbidities using any data recorded prior to cohort entry). For time-varying covariates, we used different definitions; we considered a patient exposed to a given medication for the duration of a prescription and unexposed otherwise. After any first diagnosis for a chronic disease during follow-up (diabetes, alcohol abuse, anxiety or GAD and other psychiatric diseases), a patient was considered as having the disease for the remainder of the follow-up. At any given time during follow-up, the time-varying smoking status was defined using the very last code recorded for smoking. Contrary to the simulation studies, visits to general practices could occur even in between times when BMI was monitored, such that patients' characteristics other than BMI were being updated in between BMI monitoring times, regardless of whether BMI was monitored or not on those days.

The (different versions of the) inverse weight for the monitoring, along with the inverse probability of treatment weight were used to perform a weighted regression for the outcome, which included a cubic spline basis to model the effect of time since a last visit ("gap time") on the mean outcome, and the intervention of interest. For the proposed estimators ($\hat{\beta}_{USW}$, $\hat{\beta}_{SW1}$, and $\hat{\beta}_{SW2}$), the cumulated weights were truncated using the 2.5th and 97.5th percentiles as in simulation studies. There were a few differences across the two treatment groups at baseline (Appendix B.8). The most important difference was in the prevalence of anxiety or GAD (30.5% in citalopram users vs 22.1%). The average BMI at cohort entry was 26.8 in both treatment groups. During follow-up, the average BMI in citalopram users shifted to 28.3 (median 27.2), as compared to 28.9 (median 27.6) in fluoxetine users (results not shown), indicating a slightly different shift between the two groups as compared to baseline.

In the citalopram group, we found an average of 0.28 visit per patient over the entire follow-up, with an average follow-up time of 0.45 years, yielding a crude visit (or BMI monitoring)

rate of 0.62 visit per year. In the fluoxetine users, the average number of visits per year was 0.25, with an average follow-up time of 0.40 years, yielding the same crude monitoring rate of 0.62 visit/year.

Table 4.2: Rate ratios from the visit intensity model, Clinical Practice Research Datalink, UK, 1998-2017

Variable	Rate Ratio	95% CI
Citalopram (Ref.: Fluoxetine)	0.95	0.93, 0.96*
Age at baseline	1.00	0.99, 1.00
Sex (Ref.: Female)	0.77	0.76, 0.78*
IMD at baseline	1.06	1.05, 1.06*
Smoking (Ref.: Never)		
Ever	0.92	0.91, 0.94*
Missing	0.23	0.23, 0.23*
Diabetes	2.07	2.02, 2.12*
Alcohol abuse	1.14	1.07, 1.22*
Anxiety or GAD	1.00	0.98, 1.03
Psychiatric diagnosis	1.03	0.93, 1.13
Number of hospitalisations in prior month	0.98	0.96, 1.01
Antipsychotic drugs	1.12	1.07, 1.18*
Benzodiazepine drugs	1.20	1.16, 1.23*
Lipid lowering drugs	1.21	1.18, 1.25*

Abbreviations: IMD, Index of multiple deprivation; GAD, Generalized anxiety disorder.
* Confidence interval does not contain 1.

Table 4.2 shows the adjusted rate ratios for monitoring estimated from the Andersen and Gill model, for all covariates in the multivariate monitoring intensity model, along with corresponding 95% confidence intervals (CIs) obtained from the model.

We found that males, citalopram users (as opposed to fluoxetine), and those with a previous record of smoking or no smoking information at all (as opposed to non-smokers) were less likely to have their weight recorded. A greater IMD, alcohol abuse, diabetes as well as the use of antipsychotic, benzodiazepine, or lipid lowering drugs were all associated with an increase in the rate of BMI monitoring. It is unclear whether these time-dependent covariates could lead to long-term biasing dependencies between the monitoring and the BMI processes, as both of these processes vary in time as a function of these covariates.

Estimating the marginal effect of antidepressants on BMI after re-weighting only for a point inverse intensity of visit weight, as opposed to a cumulated weight, could provide different estimates if the cumulated weight indeed provides further adjustment (e.g., if the probability of visit at time t in individual i is only proportional to that of another individual j on their set of covariates, when accounting for the full history of intensities).

When we incorporated the same set of covariates in an outcome model for the continuous outcome BMI, we found that being older, male, using citalopram (as opposed to fluoxetine), alcohol abuse, smoking, a greater number of hospitalisations in the previous month, and the use of benzodiazepine drugs were statistically significantly associated with a lower BMI (Appendix B.9). On the other hand, a greater IMD, diabetes, and the use of lipid lowering therapy were significantly associated with a greater BMI. Of note, several covariates were associated with both monitoring and BMI value (Tables 4.2 and Appendix B.9). These covariates may have induced selection bias due to outcome-dependent monitoring times. In particular, diabetes was strongly associated with both the monitoring rate, and the outcome value. Previous literature suggested that diabetes is a mediator of antidepressant drugs' effect on weight, as antidepressant therapy is associated with poor glycemic control (see e.g. Gagnon et al. [2018]).

In Table 4.3 we present the estimates for the marginal effect of citalopram (as compared to fluoxetine) on BMI, for each of the six estimators we compare, along with the 95% robust CIs. Using the two cumulated weights and accounting for the possibility for long-term dependencies brought the estimates further away from the null (coefficients around -0.61 to -0.73) as compared to the more standard inverse intensity of visit weighted estimator ($\hat{\beta}_{IH}$, coefficient -0.40, 95% CI: -0.58, -0.22).

Table 4.3: Comparison of six estimators for the marginal effect of citalopram (as opposed to fluoxetine) on BMI, Clinical Practice Research Datalink, UK, 1998-2017

Estimator	Estimate (Robust 95% CI)
$\hat{\beta}_{LS}$	-0.58 (-0.70, -0.46)
$\hat{\beta}_{IPT}$	-0.65 (-0.78, -0.53)
$\hat{\beta}_{IH}$	-0.40 (-0.58, -0.22)
$\hat{\beta}_{USW}$	-0.73 (-1.03, -0.44)
$\hat{\beta}_{SW1}$	-0.62 (-0.78, -0.46)
$\hat{\beta}_{SW2}$	-0.61 (-0.76, -0.46)

Using our proposed cumulated intensity weight ($\hat{\beta}_{SW2}$), as opposed to a simpler weight ($\hat{\beta}_{IH}$), resulted in a change of approximately 50% in the point estimate of the marginal effect of citalopram, in this study. The difference could indicate that long-term dependencies between the covariate, the monitoring and the outcome processes indeed exist. However, all estimators and the associated CIs suggest the same conclusion: that citalopram leads to less weight gain than fluoxetine. The effect of the two study drugs on BMI, as well as the difference in effects, remains modest, although we remind the reader that the follow-up time was relatively short.

4.5 Discussion

In studies using electronic health records or administrative data, *when* patients' information is recorded often depends on patients' characteristics. The informative nature of monitoring times may be associated with biasing paths between the intervention and the outcome under study, as monitoring can be a source of selection bias. In our analysis of the CPRD data, several patient characteristics were associated both with the monitoring rate and the BMI values, potentially inducing selection bias in the estimation. No previous studies have estimated the marginal effect of citalopram and fluoxetine on BMI while accounting for this type of bias (along with confounding).

It can be unclear whether the dependence between the monitoring process and the BMI

process extends beyond the last covariates observed. This work proposes some first insights into this. We proposed and demonstrated a new methodology to address that dependence by accounting for the potential for longitudinal collider-stratification bias due to an endogenous covariate process. In the CPRD data, we estimated the marginal effect of prescribing citalopram versus fluoxetine on BMI. The proposed weights did provide different estimates for that effect, as compared to the more standard IIVW that does not fully account for the covariate-dependent monitoring path. However, the differences were clinically modest. In general, a comparison of our proposed estimators and other simpler estimators that are not cumulated over time could provide indications of whether long-term dependencies are present if estimates differ substantially across approaches. In simulation studies, the proposed stabilized cumulated weighted estimator ($\hat{\beta}_{SW2}$) was the only estimator to be consistently unbiased for the marginal effect of intervention across all scenarios with endogeneity. The stabilization yielded more efficient estimation.

The proposed weights that account for the covariate-driven monitoring times are similar to that used in marginal structural models to address confounding in longitudinal treatment sequences (Robins et al. [2000a]) and to the calculation of stage occupation probability in the multistate models literature for settings with continuous time (Cook and Lawless [2018]) but they tackle imbalances due to the monitoring process rather than that due to confounding factors. We combined them with inverse probability of treatment weights to account simultaneously for confounding. Together, these weights create a pseudo-population in which the monitoring and the BMI processes are independent, and in which the two antidepressant groups are exchangeable, so as to permit inference about the marginal effect of both antidepressants on BMI. This study is the first to assess the marginal effect of citalopram and fluoxetine on BMI while considering that monitoring times are driven by an endogenous covariate process, and the first to propose a methodology for it. Another key strength of this work is our simplifying assumption on the covariate process; it allows for the covariates to be assessed occasionally (at monitoring times) for the proposed weights to break

the dependence between the monitoring and BMI processes. Further, our weights allow for the analyst to account for the effect of mediators of the antidepressant-BMI relationship that affect monitoring times without blocking the total effect of interest.

The proposed estimators rely on important assumptions. In particular, the model for the visit intensity function must be correctly specified. If covariates "occurring" or being updated in between monitoring times induce dependency between the monitoring and the BMI processes, and that they are not accounted for in the intensity model, then the proposed estimator could be biased. We also made the strong assumption of positivity of the monitoring process. When it is unrealistic, this assumption could be circumvented by smoothing the intensity function, e.g. by coarsening the monitoring indicators. Moreover, the proposed estimators rely on the standard identifiability assumptions in causal inference. For instance, the presupposition of conditional exchangeability assumes that we measured all potential confounders for the relationship between the antidepressants and BMI, an assumption that cannot be verified in practice. Sensitivity analyses were proposed to assess the extent to which unmeasured confounding can affect the estimator for the marginal effect of exposure (see e.g. Streeter et al. [2017] for a review of methods, for longitudinal settings). The assumption about the positivity of treatment, on the other hand, could be implausible in other settings, but in this work, it is unlikely to be violated as citalopram and fluoxetine are often prescribed interchangeably in patients with depression. In other situations where this assumption is not plausible, patients who have no chance of receiving some treatment options could be removed (or part of their person-time in the study) at the cost of reduced generalizability.

Although citalopram and fluoxetine are front-line treatments for depression and hence very commonly prescribed, side effects remain a significant challenge for users. In particular, weight gain may be substantial and so it is of considerable interest to use data from a general population to determine the impact of these drugs and to see whether one might lead to a lower burden of this particular side effect. In the first analysis of electronic health

records data from a large, population-based sample, we have found that citalopram leads to (modestly) less weight gain than fluoxetine, after adjusting for biases due to confounding and the covariate-induced visit process. These findings must be interpreted with caution as, of course, clinical decisions must balance a number of additional factors. Nevertheless, this analysis serves as an important model for considerations that are required when working with EHR data.

Appendices

The following material is available in Appendix B:

B.1 Causal diagrams and biasing paths due to the monitoring process

B.2 Estimating equation for the marginal effect of treatment on a continuous longitudinal outcome

B.3: Asymptotic properties of the proposed estimator

B.4: Details of the simulation studies

B.5: Results of the main simulation study, including the average number of visits and estimated parameters in the visit model

B.6: Results of all sensitivity analyses

B.7: Comparison of the bootstrap and the empirical variance of the estimators

B.8: Table of baseline characteristics stratified by intervention group in the CPRD

B.9: Multivariate outcome model in the analysis of the CPRD

Acknowledgments

This work is supported by a doctoral scholarship from the Natural Sciences and Engineering Research Council (NSERC) of Canada (Ref. 401223940) to author JC. EEMM acknowledges

support from a Discovery Grant from NSERC and a chercheur-boursier career award from the Fonds de recherche du Québec–Santé. RWP acknowledges support from a Discovery Grant from NSERC and a Foundation Scheme Grant from CIHR.

Chapter 5

Estimating the Marginal Effect of a Continuous Exposure on an Ordinal Outcome using Data Subject to Covariate-Driven Treatment and Visit Processes

Preamble to Manuscript 3. In the previous two chapters, a general framework was proposed for causal inference on the average treatment effect in settings with covariate-driven monitoring times and confounding. The methodology was extended to settings where there is endogeneity, meaning that the covariate, the monitoring and the outcome processes are allowed to interact together in time.

The motivation for Chapter 5 comes from the desire to generalize the previously proposed methodology for cases with a continuous exposure and an ordinal outcome. This new methodology is to be used to estimate the marginal effect of the number of hours spent

playing video games weekly on a categorized outcome related to the number of suicide attempts in the *Add Health* study.

In this chapter, the exposure is denoted by $D_i(t)$ rather than the previous notation $I_i(t)$, to emphasize that the exposure is continuous (i.e. a dose, and to avoid confusion with the individual index i when the notation and assumptions are discussed).

The original contributions of this work are i) to propose the first complete methodology for assessing the marginal effect of a continuous exposure on an ordinal outcome when data are subject to confounding and covariate-driven monitoring times, and ii) to discuss and assess different weighting strategies previously proposed in the literature in a real-life application, to ultimately estimate the marginal impact of the time spent playing video games on suicide attempts in young adults.

This manuscript has been accepted for publication in *Statistics in Medicine* and is currently in press.

Estimating the Marginal Effect of a Continuous Exposure on an Ordinal Outcome using Data Subject to Covariate-Driven Treatment and Visit Processes

Janie Coulombe[1], Erica EM Moodie[1], Robert W. Platt[1].

[1]Department of Epidemiology, Biostatistics, and Occupational Health, McGill University

This thesis contains the accepted version of the corresponding paper accepted for publication in *Statistics in Medicine* (Coulombe et al. [forthcoming]).

Abstract

In the statistical literature, a number of methods have been proposed to ensure valid inference about marginal effects of variables on a longitudinal outcome in settings with irregular monitoring times. However, the potential biases due to covariate-driven monitoring times and confounding have rarely been considered simultaneously, and never in a setting with an ordinal outcome and a continuous exposure. In this work, we propose and demonstrate a methodology for causal inference in such a setting, relying on a proportional odds model to study the effect of the exposure on the outcome. Irregular observation times are considered via a proportional rate model, and a generalization of inverse probability of treatment weights is used to account for the continuous exposure. We motivate our methodology by the estimation of the marginal (causal) effect of the time spent on video or computer games on suicide attempts in the *Add Health* study, a longitudinal study in the United States. Although in the *Add Health* data, observation times are pre-specified, our proposed approach is applicable even in more general settings such as when analyzing data from electronic health records where observations are highly irregular. In simulation studies, we let observation times vary across individuals and demonstrate that not accounting for biasing imbalances due to the monitoring and the exposure schemes can bias the estimate for the marginal odds ratio of exposure.

5.1 Introduction

Suppose a setting where longitudinal data are assessed irregularly across patients. In particular, we focus on non-experimental settings where, contrary to randomized studies, the mere fact of being exposed to a certain level of an exposure can be related to individuals' characteristics. Our interest lies in the estimation of the causal effect of a continuous exposure on a longitudinal, ordinal (categorical) outcome which is measured irregularly. Assuming that the observation of the outcome process coincides with monitoring times, which depend on patients' characteristics, we can then think of the outcome process as being missing *at random* over the course of follow-up time. In observational data, imbalances such as those due to covariate-driven monitoring times or confounding can bias the estimators for the marginal effect of variables (e.g. that of an exposure). The bias due to covariate-driven monitoring times relates to *selection bias*, often due to a phenomenon called *collider-stratification bias* (Greenland [2003]) where restricting a study cohort on a covariate, or conditioning on a selection indicator (which both could be colliders on the path from the exposure to the outcome) may induce dependence between the exposure and outcome. For instance, consider the causal diagram in the right panel of Figure 5.1. Under that data generating mechanism, the monitoring indicator acts as a collider on one of the paths from *Video games* to *Depressive mood*. Conditioning on that indicator (i.e., keeping only in an analysis the observations such that the outcome, *Suicide attempts*, is monitored) can induce spurious association between the variables *Video games* and *Suicide attempts*. Other examples of collider-stratification bias have been reported in studies on birthweight (Whitcomb et al. [2009]) and obesity (Banack and Kaufman [2015]), among others. Collider-stratification bias was also demonstrated in simulation studies, in settings with covariate-driven monitoring times (Bůžková and Lumley [2009], Coulombe et al. [2021a]). In such longitudinal settings, the imbalances related to selection are repeated all throughout follow-up time. For causal inference on an exposure effect, monitoring times may have to be modelled continuously in time if they are related to covariates that are simultaneously associated with the study

outcome and the study exposure.

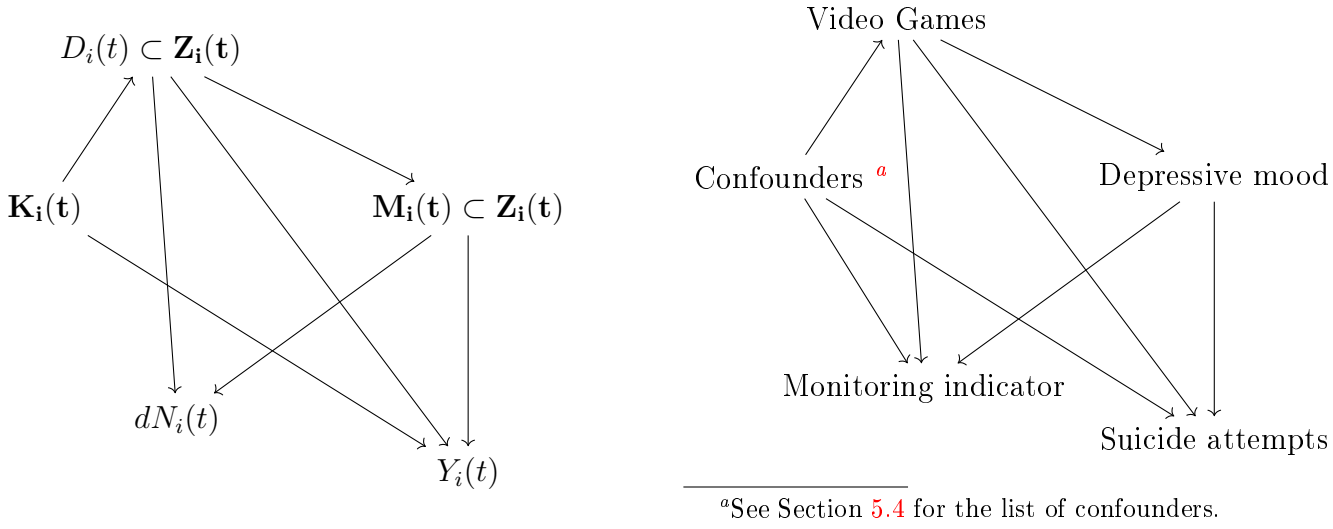


Figure 5.1: The assumed causal diagram in the simulation studies (left panel) and the corresponding causal diagram for the application on the effect of video games on suicide attempts (right panel).

Covariate-driven monitoring times can arise both in a study with presspecified observation times or when using data from electronic health records. We propose and demonstrate a methodology for the estimation of the causal effect of a continuous exposure on a categorical outcome that applies in both settings where observation times are pre-specified (such as the *Add Health* study) or are irregular across individuals (such as when using data from electronic health records). The methodology allows for observation times to depend on individuals' characteristics, and also accounts for potential confounders of the relationship under study. We model the monitoring indicators (which indicate whether there is a visit, or not, during which the outcome is assessed) using a proportional intensity model and the exposure model using a generalized propensity score which is then used to adjust for confounding via inverse probability of treatment weights. The outcome is modelled as a function of the exposure, using a proportional odds model.

Our aim, to estimate the marginal effect of the time spent on video games (weekly) on suicide attempts, motivates the methodological developments. We conduct our analysis

using longitudinal data from the first four waves of the *Add Health* study (Harris [2009]) in the United States. In that study, individuals were followed from their adolescence and until adulthood, and were assessed at different points in time. Individuals' personal data were collected via several questionnaires, and the collected information included demographics, social, biological and behavioral factors as well as parental factors.

This manuscript is divided as follows: in Section 5.2, we discuss the background, our assumptions and notation, as well as the proposed methodology for estimating the causal effect of a continuous exposure on a categorical, ordinal outcome. In Section 5.3, we present the details of the simulation studies that were conducted and corresponding results. The application of the proposed methodology to data from the *Add Health* study to estimate the effect of the number of hours spent on video games on suicide attempts is presented in Section 5.4. A discussion follows in Section 5.5.

5.2 Methods

5.2.1 Background

This work extends previous research (Coulombe et al. [2021a]) in which two estimators were proposed for the marginal effect of a binary (time-varying) exposure $\mathbf{I}(\mathbf{t})$ on a longitudinal continuous outcome $\mathbf{Y}(\mathbf{t})$. These estimators were also proposed for settings where data are subject to being recorded (or monitored) at covariate-dependent visit times, allowed to be irregular across patients, and where there is potential confounding. Inverse weights were used to adjust both for confounding and bias due to covariate-driven monitoring times. In addition, the monitoring weights were allowed to be functions of mediators of the relationship of interest.

In that previous work, two estimators were proposed and compared, namely:

- a first estimator which extended the estimating equations proposed by Lin and Ying

[2001] to account for confounding via a generalized inverse probability of treatment weight, and incorporated the stabilized weight proposed by Bůžková and Lumley [2009] for the visit intensity (denoted by $\hat{\beta}_{IPCTM}$ where IPCTM denotes the inverse probability of centered treatment and monitoring estimator); and

- a second estimator which used the standard weighted least squares estimating equations, weighted with both an inverse probability of treatment weight and an inverse intensity of visit weight (Lin et al. [2004]). That estimator also included a cubic spline basis as a function of time since cohort entry to allow for more flexibility in modelling the effect of time since study baseline on the mean outcome. It was denoted by $\hat{\beta}_{FIPTM}$, where FIPTM refers to the flexible inverse probability of treatment and monitoring estimator (Coulombe et al. [2021a]).

Both estimators were demonstrated via theoretical derivations to be unbiased for large samples, and extensive simulation studies showed that both were unbiased in finite samples and their variances were quite comparable. However, $\hat{\beta}_{FIPTM}$ was easier to implement. To answer the current research question, we aim to extend this estimator to the setting where exposure is continuous, and where the outcome is categorical and ordinal. For that, we use the proportional odds model.

Proportional Odds Model

Let i be a patient index, and t denote time. Suppose that one is interested in the association between a vector of covariates $\mathbf{X_i(t)}$ (of size $1 \times p$) and the categorical outcome $Y_i(t)$. For instance, the vector $\mathbf{X_i(t)}$ may contain a continuous exposure of interest, which we denote by $D_i(t)$ for individual i at time t. The proportional odds model (POM) proposed by McCullagh [1980] models the outcome cumulative probabilities as

$$P(Y_i(t) \leq j|\mathbf{X_i(t)}) = \text{expit}(\alpha_j - \boldsymbol{\beta}'\mathbf{X_i(t)}) \tag{5.1}$$

$\forall j = 1, ..., J$, where the expit function is the inverse logit function and the coefficients $\boldsymbol{\alpha}$ are category-specific intercepts. The effects of covariates $\mathbf{X(t)}$ are assumed to be constant for all j (that is, $\boldsymbol{\beta}_j = \boldsymbol{\beta}$). Effectively, the POM is an extension of the logistic regression model to the multinomial case where, rather than merely comparing the probability of an event occurring (as compared to no event), the outcome domain is stratified at each possible "split value" between two subsequent ordered categories j and $j+1$ in $\{1, ..., J\}$. The probability that is modelled is thus that for the outcome to be smaller or equal to some value in the outcome domain, as opposed to larger.

When studying the marginal effect of a variable, for instance the exposure $D_i(t) \subset \mathbf{X_i(t)}$, on the ordinal outcome $Y_i(t)$, different quantites may be of interest. If the outcome is made of only two categories, the POM is equivalent to the logistic regression model, and the odds ratio (OR) or the relative risk of exposure can be of interest. In the current work, we focus on ordinal outcomes that are made of more than two categories, and use the marginal OR (or log-OR) as our *population-average* quantity of interest. We next review the notation and statistical methodology that we will employ for a general setting.

5.2.2 Assumptions and notation

Vectors and matrices are denoted in bold. Let $Y_i(t)$ be a longitudinal outcome (that is either observed or not) at time t, in individual i ($i = 1, ..., n$), and assume it is categorical and it takes values in $\{1, 2, ..., J\}$. Let the exposure be denoted by $D_i(t)$ at time t, which is continuous and takes values in $\mathbb{R}$ (or, possibly, in $\mathbb{R}^+$). We are interested in the estimation of the marginal effect of an increment in exposure $D_i(t)$ on the outcome $Y_i(t)$. We suppose that the outcome is ordinal – that is, it can be ordered such that categories 1 and J respectively represent the lowest, and the highest levels of that variable. Further, the subsequent categories of the outcome are not considered to be equidistant, and going from e.g. category j to $j+1$ may imply a more important change than going from category $j-1$ to j, etc.

Further suppose that the outcome $Y_i(\cdot)$ of individual i is only observed sporadically, at times denoted by $T_{i1}, ..., T_{iQ_i}$ (called *monitoring* or *visit* times). The quantity Q_i denotes the number of observation times of individual i in the time lapse $[0, \tau]$, with τ the maximum follow-up time in the whole study. As the indices indicate, the monitoring times and number of visits are allowed to vary across individuals. The monitoring indicator $dN_i(t)$ indicates whether there is $(dN_i(t) = 1)$ or is not $(dN_i(t) = 0)$ a visit of individual i at time t. That indicator is expected to depend on certain patients' characteristics that we denote by the set $\mathbf{Z_i(t)}$. The set may contain the exposure of interest, $D_i(t)$, as well as mediators of the relationship of interest between $D_i(t)$ and $Y_i(t)$ or confounding factors. We denote the mediators of the relationship between $D_i(t)$ and $Y_i(t)$ by $\mathbf{M_i(t)}$; they are on the causal path from $D_i(t)$ do $Y_i(t)$. Confounders are denoted by $\mathbf{K_i(t)}$ and are assumed to affect both the outcome $Y_i(t)$ and the exposure $D_i(t)$ at time t such that merely assessing the effect of the exposure on the outcome will lead to a distorted estimate.

Figure 5.1 (left panel) shows the causal diagram that we assume for the data generating mechanism discussed here, at each time t. Effectively, that diagram represents one *slice* or *snapshot* in time, but the relations depicted in that Figure are assumed to be true for each point in time as well as for each individual. This means that some covariates ($\mathbf{Z_i(t)}$ here) affect the monitoring propensity at each point in time, such that a prediction for $dN_i(t)$ would require the information on covariates $\mathbf{Z_i(t)}$ even on times $t \subset [0, \tau]$ when $dN_i(t) = 0$ and there is no visit. In that same causal diagram, the set $\mathbf{Z_i(t)}$ contains all covariates associated with $D_i(t)$ which themselves affect $dN_i(t)$; these are all the covariates affecting monitoring. In the right panel of Figure 5.1, we show the causal diagram that is posited to describe the data collected in the *Add Health* study. We recall that, in the application, we are interested in the total causal effect of video games on the categorical suicide attempt outcome. This effect comprises the direct effect of exposure, but also that mediated through depressive mood.

Our aim is to build a pseudopopulation (Robins et al. [2000a]) in which patients under different exposures are comparable with respect to confounding factors, and in which we adjust for any biasing path between the exposure $D_i(t)$ and the outcome $Y_i(t)$ that would be caused by conditioning on observed data ("visits"). To connect the causal and the statistical frameworks, we first need to reformulate the problem as a causal inference problem by using standard identifiability assumptions. To express the causal contrast of interest, we use the Neyman-Rubin potential outcome framework (Neyman [1923], Rubin [1974]) and denote by $Y_{id}(t)$ the potential outcome of individual i, at time t, and under exposure d. Necessarily, individual i will only receive one level of the continuous exposure D at time t, and therefore, only one of an infinite quantity of possible outcomes will be observed. The estimand we seek is the causal marginal OR for 1-unit increase in the continuous exposure $D_i(t)$. It is given by

$$OR = \left(\frac{\mathbb{P}\left[Y_{id}(t) \leq j\right]}{1 - \mathbb{P}\left[Y_{id}(t) \leq j\right]} \right) \Big/ \left(\frac{\mathbb{P}\left[Y_{i(d+1)}(t) \leq j\right]}{1 - \mathbb{P}\left[Y_{i(d+1)}(t) \leq j\right]} \right) \tag{5.2}$$

and is constant across j. Note that the numerator in (5.2) shows the potential outcome under exposure d, and the denominator, under exposure $d+1$, as we are presenting in (5.2) the OR as a function of the probability for the outcome to be smaller or equal to a certain category j, rather than larger. Depending on the parameterization of the POM model, the estimand could vary. Furthermore, the estimator of the OR for a 1-unit increment in the exposure given in equation (5.2) can be modified to account for any incremental value (e.g., 2, 3, or 10-unit increments) of the exposure, as, under the POM assumptions, we have that

$$OR_w = \left(\frac{\mathbb{P}\left[Y_{id}(t) \leq j\right]}{1 - \mathbb{P}\left[Y_{id}(t) \leq j\right]} \right) \Big/ \left(\frac{\mathbb{P}\left[Y_{i(d+w)}(t) \leq j\right]}{1 - \mathbb{P}\left[Y_{i(d+w)}(t) \leq j\right]} \right) = \exp\left(\log(OR) \times w\right),$$

where OR is given in equation (5.2) and OR_w is the odds ratio for a w-unit increment in the exposure.

The estimation of the OR in (5.2) requires assumptions regarding the data generating mechanism and its associated causal diagram. The causal diagram that we assume (presented in the left panel of Figure 5.1) is subject to both biasing paths due to backdoor paths (confounders), and to selection due to conditioning on monitoring indicators. These *two* biases are due to the lack of exchangeability of the individuals across all exposure levels (Hernán and Robins [2016], Chapters 7-8). For identifiability of the causal exposure effect, we must assume conditional exchangeability, consistency, as well as positivity of the exposure and the monitoring indicators, which are denoted by:

$$D_i(t) \perp Y_{id}(t) | dN_i(t), \mathbf{K_i(t)}, \mathbf{Z_i(t)} \quad \forall d \in \mathbb{R}^+ \tag{P1}$$

$$\text{If } D_i(t) = d \text{ then } Y_{id}(t) = Y_i(t) \tag{P2}$$

$$0 < P(D_i(t) = d | \mathbf{K_i(t)}) < 1 \ \ \forall d \in \mathbb{R}^+ \tag{P3}$$

$$0 < P(dN_i(t) = 1 | \mathbf{Z_i(t)}) < 1 \ \ \forall t \in [0, \tau]. \tag{M1}$$

The assumption (P1) means that the two sets (confounders $\mathbf{K_i(t)}$, and those affecting the monitoring times $\mathbf{Z_i(t)}$) are sufficient to block any biasing path from the exposure to the outcome of interest *even* after conditioning on monitoring indicator $dN_i(t)$. Effectively, what we mean by this assumption is that even when using only the observed outcomes, adjusting for the set $\mathbf{K_i(t)}$ is sufficient to block the biasing paths due to potential confounders, and the set $\mathbf{Z_i(t)}$ contains the common predictors of the monitoring and the outcome processes that can create other biasing paths from the exposure to the outcome after conditioning on $dN_i(t)$. (Note, the common predictors in $\mathbf{Z_i(t)}$ will typically be related to the exposure in some way, such that conditioning on $dN_i(t)$ opens a path from the exposure to the outcome that is due to one of the covariates in $\mathbf{Z_i(t)}$. Otherwise, there may be no need to adjust for these covariates.). Positivity for monitoring means that there is no time point (or time period, if time is considered to be continuous) when the probability that a visit will occur is null, or when it is 1 such that the visit is sure to occur. If such time point existed, this could

lead to computational issues with the inverse weighting methodology that will follow, as well as concerns regarding the interpretation of findings. In addition to the assumptions above, we assume that censoring times (or the individual times when patients' follow-up stops) are uninformative, in the sense that we capture through assumption P1 the only possible differences in follow-up that could ultimately bias our estimator for the effect of exposure on the outcome.

5.2.3 Methodology

Following the assumptions we made in Section 5.2.2, we can estimate the causal marginal OR using the POM. We now explain how we model the monitoring rate, the exposure, and how these models are combined to estimate the estimand.

For the monitoring model, we use an inverse intensity of visit (IIV) weight (Lin et al. [2004]) and model the intensity by using a proportional intensity model as a function of the covariates $\mathbf{Z_i(t)}$, which covariates (if considered together with the confounders) are assumed to create conditional exchangeability of the potential outcomes. The model is as follows:

$$\mathbb{E}\left[dN_i(t)|\mathbf{Z_i(t)}\right] = \xi_i(t)\exp(\boldsymbol{\gamma}'\mathbf{Z_i(t)})\lambda_0(t)dt, \tag{5.3}$$

where $\xi_i(t)$ is an indicator for still being at risk, for individual i at time t, and $\lambda_0(t)dt = d\Delta_0(t)$ with $\Delta_0(t)$ any non-decreasing function (Lawless and Nadeau [1995]). We use the Andersen and Gill model (Andersen and Gill [1982]), an extension of the Cox proportional hazards model (Cox [1972]) to recurrent events, to estimate the parameters $\boldsymbol{\gamma}$ in (5.3). In this case, the "recurrent events" consist of the monitoring times. The baseline rate $\lambda_0(t)$ need not be estimated if it does not vary across individuals and if it is a function of the time since cohort entry (or time into study), rather than e.g. time since the last visit. That is, because any function of time since cohort entry would be the same at time t into the study across individuals, the function cancels out across individuals (Bůžková and Lumley [2009]).

After adjusting for the monitoring intensity, under the mentioned assumptions, we can attain conditional exchangeability by fitting a correctly specified exposure model that is a function of covariates $\mathbf{K_i(t)}$, that will be used to break the links between the exposure and those covariates. However, we are interested in the case where exposure is continuous such that the standard inverse probability of treatment (IPT) weight cannot be implemented. To account for confounding under a continuous exposure, methods such as parametric g-computation were proposed (see eg. Ahern et al. [2009]), however that method requires a correctly specified parametric outcome model, and it combines less naturally with our IIV weight. If the exposure distribution is (approximately) normally distributed, we propose to use a generalization of the standard IPT weight that uses the conditional normal density, rather than the more common propensity score for a binary exposure. This strategy has been discussed by Robins et al. [2000a] and Imai and Van Dyk [2004] and requires fitting two linear models: one where potential confounders are included as predictors (which will lead to a conditional normal density), and another stabilizing model which only includes a constant intercept, as follows:

$$\mathbb{E}[D_i(t)|\mathbf{K_i(t)}] = \psi_0 + \boldsymbol{\psi_1'}\mathbf{K_i(t)} \tag{L1}$$

$$\mathbb{E}[D_i(t)] = \psi_m. \tag{L2}$$

Once the parameters of these models are estimated from the data, we obtain estimates for the parameters $\left\{\hat{\psi}_0, \hat{\boldsymbol{\psi}}_1, \hat{\psi}_m\right\}$ and may compute the residuals from both models. The continuous exposure is assumed to follow a normal distribution that is fully defined by its mean and standard deviation, which can be estimated from the residuals from the fitted model corresponding to (L1).The generalized IPT weight, to be added to our weighted estimating equations, is then computed as

$$\mathbf{e_i(t;\boldsymbol{\psi})} = \frac{h^{-1}(\hat{\psi}_0 + \hat{\boldsymbol{\psi}}_1'\mathbf{K_i(t)})}{h^{-1}(\hat{\psi}_m)} \tag{5.4}$$

for $h^{-1}(\widehat{D}_{l,i}(t)) = 1/\sqrt{2\pi\widehat{\sigma}_l^2}\exp\left(-\widehat{\epsilon}_{l,i}(t)^2/(2\widehat{\sigma}_l^2)\right)$ the Normal density function evaluated at the corresponding linear regression residuals $\widehat{\epsilon}_{l,i}(t) = \left(D_i(t) - \widehat{D}_{l,i}(t)\right)$, with $\widehat{\sigma}_l^2$ the empirical variance of $\widehat{\epsilon}_{l,i}(t)$, $\widehat{D}_{l,i}(t)$ the model predictions, and $l = 1, 2$ the index corresponding to models (L1) and (L2) respectively (Robins et al. [2000a]). Importantly, a misspecified conditional distribution for the exposure may lead to biased estimators, as the generalized IPT weight in (5.4) may then not be proportional to the actual density of the exposure conditional on the confounders. If the exposure is not approximately normally distributed (e.g., if its distribution is highly skewed), a log transformation of the exposure may lead to a more normally distributed (transformed) exposure. Other alternatives to the generalized IPT weight above include that where the exposure is binned in quantiles (e.g., deciles) and where a categorical model is fitted to estimate the probability for the exposure to belong to a respective category (as a function of covariates). Naimi et al. [2014] compared that option to the other weight mentioned above and they have found that the quantile binning approach performs better when the exposure is not normally distributed. Schulz and Moodie [2021] also discussed inverse weights in the context of a continuous dose for an exposure, when developing optimal adaptive dosing strategies. They considered binning the exposure in quantiles to compute the IPT weight, while leaving the exposure in its continuous form in the outcome model. They also used a similar quantiles approach as in Naimi et al. [2014] where they estimated the probability for the exposure to belong to a respective category using the POM model (rather than a conditional logistic regression model as used in Naimi et al. [2014]). They found that the binning approach can reduce the volatility in the weights. In simulation studies below, we demonstrate the proposed methodology with the generalized inverse weight in (5.4) and in the case where the exposure is normally distributed. In the application to the *Add Health* study, we assess the marginal effect of a 1-unit increase in the exposure (the number of hours spent playing video games) in its current form, and after it was log transformed. We assess the sensitivity of the results after using different inverse weighting strategies.

In this work, we focus on the estimation of the marginal OR for a 1-unit increase in the exposure $D_i(t)$ (or the log transformed exposure, in the application), with the odds being those of the outcome $Y_i(t)$. The OR was presented in equation (5.2). Recall that if desired, the OR derived from our methodology could also be used to compute the marginal OR for an arbitrary w-unit increase in the exposure $D_i(t)$, as mentioned in Section 5.2.2. Under the identifiability assumptions in Section 5.2.2, we use a parametric POM model which models the mean outcome as follows:

$$P(Y_i(t) \leq j | D_i(t)) = \text{expit}(\alpha_j - \beta_D D_i(t)) \tag{5.5}$$

$\forall j = 1, ..., J$. Trivial algebra leads to the marginal OR:

$$\begin{aligned}
\frac{\left(\frac{\mathbb{P}[Y_i(t)\leq j|D_i(t)+1]}{1-\mathbb{P}[Y_i(t)\leq j|D_i(t)+1]}\right)}{\left(\frac{\mathbb{P}[Y_i(t)\leq j|D_i(t)]}{1-\mathbb{P}[Y_i(t)\leq j|D_i(t)]}\right)} &= \frac{\frac{\exp(\alpha_j-\beta_D(D_i(t)+1))}{1+\exp(\alpha_j-\beta_D(D_i(t)+1)}}{\frac{1}{1+\exp(\alpha_j-\beta_D(D_i(t)+1)}} \times \frac{\frac{1}{1+\exp(\alpha_j-\beta_D D_i(t))}}{\frac{\exp(\alpha_j-\beta_D D_i(t))}{1+\exp(\alpha_j-\beta_D D_i(t))}} \\
&= \frac{\exp(\alpha_j - \beta_D(D_i(t)+1))}{\exp(\alpha_j - \beta_D D_i(t))} \\
&= \exp(-\beta_D).
\end{aligned} \tag{5.6}$$

Therefore, for making our causal inference, we are left with the estimation of the parameter β_D in (5.6). Using our models for exposure and monitoring, and re-weighting estimating equations by the IPT and the IIV weights, we create a pseudopopulation in which we have conditional exchangeability, so that the parameter β_D can be estimated using directly the POM on the re-weighted data. That is, we extend the *Flexible Inverse Probability of Treatment and Monitoring* weighted estimator ($\hat{\beta}_{FIPTM}$) (Coulombe et al. [2021a]) to the case where the exposure is continuous, and where the outcome mean is not assumed to be a linear function of covariates. The new, proposed, doubly-weighted estimator is further referred to as $\hat{\beta}_{IPTMP}$ for the *Inverse Probability of Treatment and Monitoring POM* model.

Denote by $\zeta_{i,j}(t) = P(Y_i(t) \leq j | D_i(t))$ and by $F_{i,j}(t) = \mathbb{I}(Y_i(t) \leq j)$ with $\mathbb{I}(\cdot)$ the indicator

function, which also correspond to the vectors

$$\boldsymbol{\zeta}_i(\mathbf{t}) = \begin{bmatrix} P(Y_i(t) \leq 1 | D_i(t)) \\ P(Y_i(t) \leq 2 | D_i(t)) \\ \cdots \\ P(Y_i(t) \leq J | D_i(t)) \end{bmatrix}, \text{ and } \mathbf{F_i}(\mathbf{t}) = \begin{bmatrix} \mathbb{I}(Y_i(t) \leq 1) \\ \mathbb{I}(Y_i(t) \leq 2) \\ \cdots \\ \mathbb{I}(Y_i(t) \leq J) \end{bmatrix}, \tag{5.7}$$

where $\mathbf{F_i}(\mathbf{t})$ is our modified outcome which accounts for the fact that the study outcome $Y_i(t)$ is made of several categories. Then, effectively, our methodology is equivalent to using the following estimating equation (an exension of Lin et al. [2004] and Coulombe et al. [2021a]) to estimate the marginal effect of exposure via the log-OR for 1-unit increase in exposure, denoted by β_D:

$$\mathbb{E}\left[\int_0^\tau \frac{\mathbf{e}(\mathbf{t};\boldsymbol{\psi})\left(\mathbf{F}(\mathbf{t}) - \boldsymbol{\zeta}(\mathbf{t})\right)}{\boldsymbol{\varphi}(\mathbf{t};\boldsymbol{\gamma})} \mathbf{dN}(\mathbf{t})\right] = \mathbf{0}, \tag{5.8}$$

where $\mathbf{e}(t;\boldsymbol{\psi})$ is our generalized IPT weight for a continuous exposure, $\boldsymbol{\varphi}(\mathbf{t};\boldsymbol{\gamma})$ an IIV weight for the monitoring, and where the parameterization of $\boldsymbol{\zeta}(\mathbf{t})$ is

$$\zeta_{i,j}(t) = \text{expit}(\alpha_j - \beta_D D_i(t)). \tag{5.9}$$

For the inverse intensity of visit function, we simply plug in our estimated model from (5.3); that is

$$\boldsymbol{\varphi_i}(\mathbf{t};\widehat{\boldsymbol{\gamma}}) = \exp\left(\widehat{\boldsymbol{\gamma}}' \mathbf{Z_i}(\mathbf{t})\right),$$

where the parameters $\boldsymbol{\gamma}$ can be estimated using the *coxph* function in R, from the *survival* package (Therneau [2020]). The baseline rate in (5.3) need not to be estimated, as underlined

earlier. For the IPT weight, we use an estimate of (5.4) that is computed by fitting both linear models and computing the respective residuals as discussed earlier (and assess a few other weighting options in the application to the *Add Health* study). The weighted POM can be fitted using, for instance, the *polr* function from the *MASS* package in R (Venables and Ripley [2002]). Diagnostic checks, such as the surrogate residuals (Greenwell et al. [2018]) implemented in the *Sure* package (Greenwell et al. [2017]) can be used to assess the POM fit. To derive the asymptotic variance of the estimator for the coefficient (or OR) of interest, one can use the theory of two-step estimators (Newey and McFadden [1994]) and compute a sandwich estimator that accounts for the variance components due to the parameter of interest as well as those from both nuisance models. The derivation of the asymptotic variance for the case where the outcome is continuous and the exposure is binary has been derived previously, and shown to be well-approximated via a non-parametric bootstrap approach (Coulombe et al. [2021a]). In current work, we used nonparametric bootstrap to compute an estimate of the variance of the OR in the application to the *Add Health* study, which method has shown to perform well in previous, similar work (Coulombe et al. [2021a]).

5.3 Simulation study

We conducted several simulation studies to assess the proposed methodology in a setting where, in contrast to the *Add Health* study, monitoring times can occur at any time during follow-up, for every individual. Our aim was to estimate the causal marginal OR for a 1-unit increase in the exposure $D_i(t)$ on a categorical and ordinal outcome $Y_i(t)$. The outcome was categorical, taking one of three levels ($J = 3$): 1, 2, and 3. In simulation studies, we compared four estimators:

- The estimated log-OR for exposure obtained directly from the POM model, with no adjustment ($\hat{\beta}_{POM}$);
- The estimated log-OR for exposure from a weighted POM with an IPT weight, where

the propensity score is a correctly specified function of the confounders ($\hat{\beta}_{IPTP}$);

- The estimated log-OR for exposure from a weighted POM with an IIV weight, where the intensity is a correctly specified function of the covariates affecting visit times ($\hat{\beta}_{IIVP}$); and
- The estimated log-OR for exposure from a doubly-weighted POM with both the IPT and the IIV weights, with both functions (corresponding to the exposure and visit models) correctly specified ($\hat{\beta}_{IPTMP}$).

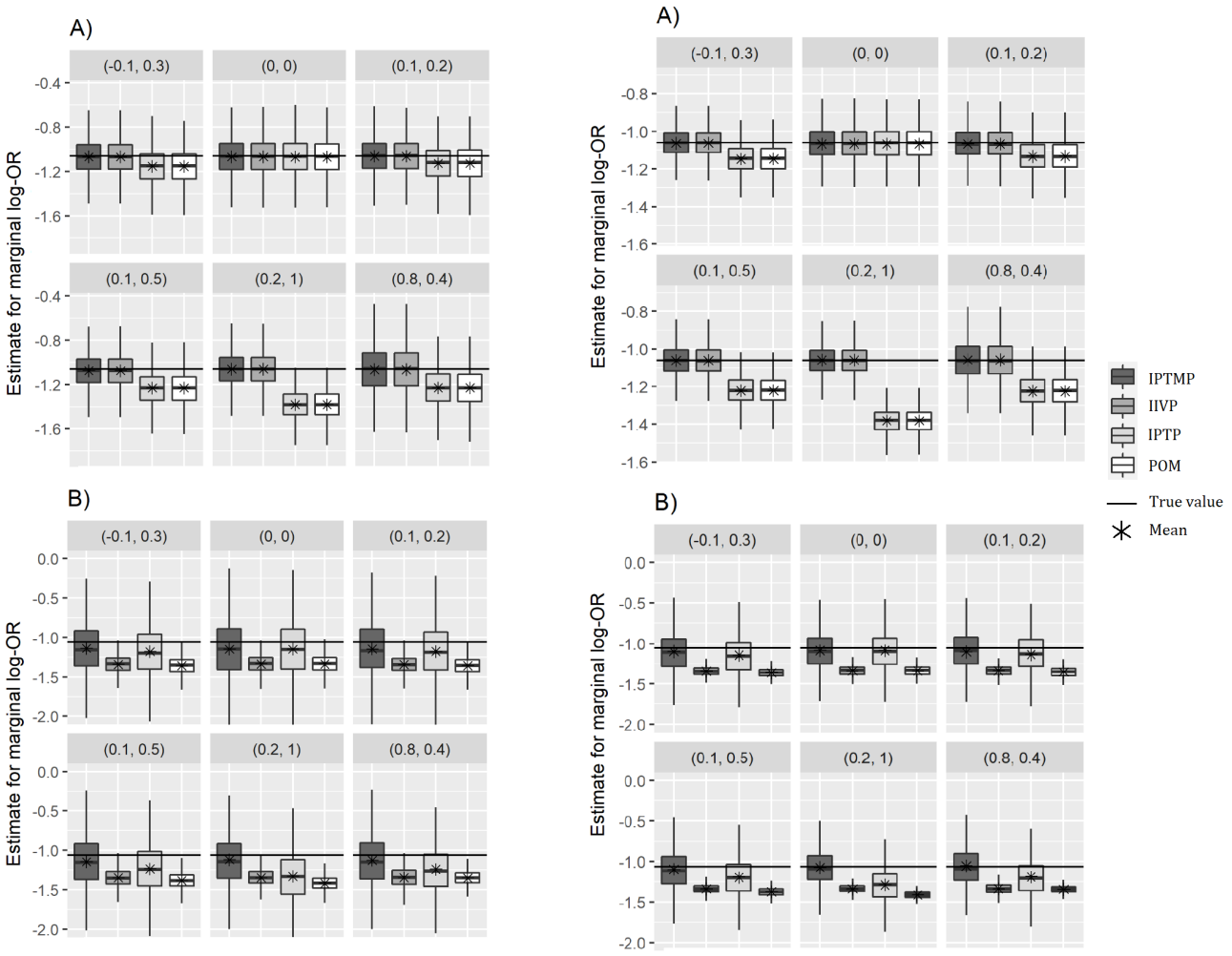
We used 1000 simulations per study and tested settings with either 250 or 1000 patients per simulated dataset. We assessed the case with, or without confounding. Confounders were defined (simulated) at time 0 ("cohort entry") for each individual, while the continuous exposure and a mediator of the relationship between the exposure and the outcome were simulated as time-varying. Visit times could vary across individuals and were driven by covariates. In the main analysis, only the exposure and the mediator affected the propensity of being monitored. In two sensitivity analyses, we also 1) assessed the performance of the four estimators when the exposure, the mediator, and confounders all affect the monitoring intensity (with the monitoring intensity model being correctly specified); and 2) changed the mean outcome model and tested different parameters in the monitoring intensity model so as to increase the bias due to covariate-driven monitoring times. In all analyses, the ordinal, categorical outcome was simulated as a function of the exposure, the mediator, and the confounders. The true marginal log-OR in the main analysis as well as that in the second sensitivity analysis (which differed, as the mean outcome models varied) were derived using Monte Carlo simulations. A detailed description of the simulation studies is presented in Appendix C.1.

5.3.1 Results

Figure 5.2 shows the results of the main simulation study for two settings (250 patients in the left panel, and 1000 patients in the right panel), and for the case where there is no confounding (top panel) and where there is confounding (bottom panel). As expected, the proposed doubly-weighted estimator $\hat{\beta}_{IPTMP}$ was the least biased across all four estimators. In the case with no confounding, its performance was equivalent to that of $\hat{\beta}_{IIVP}$, which also accounted for covariate-driven monitoring times. In the case with confounding, the data presented with biasing imbalances related to both the confounders and the covariate-driven monitoring times, and $\hat{\beta}_{IPTMP}$ was the least biased across the board (it was the only estimator to account for both of these features). When increasing the sample size, our proposed estimator converged to the true value (horizontal dark line in Figure 5.2) while all other estimators remained biased.

Table 5.1: Comparison of four estimators for the marginal log-OR for 1-unit increase in $D_i(t)$ in the POM, for a sample size of $n = 250$ patients and 1000 simulations per study. Study without confounding (Conf.=N) and with confounding (Conf.=Y).

Conf. (Y/N)	γ	Mean no. visits (min-max)	Absolute empirical bias				Empirical variance			
			IPTMP	IIVP	IPTP	POM	IPTMP	IIVP	IPTP	POM
N	(-0.3, 0.1)	3 (0-14)	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02
	(-0.1, 0.3)	3 (0-15)	0.01	0.01	0.09	0.09	0.03	0.03	0.03	0.03
	(0, 0)	2 (0-13)	0.01	0.01	0.01	0.01	0.03	0.03	0.03	0.03
	(0.1, 0.2)	3 (0-14)	0.00	0.00	0.07	0.07	0.03	0.03	0.03	0.03
	(0.1, 0.5)	3 (0-15)	0.01	0.01	0.17	0.17	0.03	0.03	0.03	0.03
	(0.2, 1)	5 (0-19)	0.00	0.00	0.32	0.32	0.02	0.02	0.02	0.02
	(0.8, 0.4)	2 (0-12)	0.00	0.00	0.17	0.17	0.05	0.05	0.03	0.03
Y	(-0.3, 0.1)	2 (0-13)	0.09	0.28	0.11	0.27	0.24	0.02	0.24	0.02
	(-0.1, 0.3)	3 (0-16)	0.08	0.28	0.12	0.30	0.19	0.01	0.19	0.01
	(0, 0)	2 (0-13)	0.09	0.28	0.09	0.28	0.23	0.02	0.23	0.02
	(0.1, 0.2)	3 (0-16)	0.09	0.28	0.12	0.30	0.21	0.01	0.20	0.01
	(0.1, 0.5)	3 (0-15)	0.08	0.28	0.17	0.32	0.17	0.01	0.17	0.01
	(0.2, 1)	5 (0-21)	0.06	0.29	0.27	0.36	0.16	0.01	0.14	0.01
	(0.8, 0.4)	5 (0-46)	0.07	0.28	0.18	0.29	0.16	0.02	0.14	0.01



250 patients

1000 patients

Figure 5.2: A) Simulation study with no confounding. B) Simulation study with confounding. Study with 250 patients (left panel) or 1000 patients (right panel). Distribution of the estimated marginal log-OR for a 1-unit increase in the continuous exposure $D_i(t)$ (1000 simulations per study) across different scenarios for the monitoring process ($\boldsymbol{\gamma}$ parameters at the top of each subgraph) and for the 4 estimators compared. The horizontal dark line represents the true value of the effect. The left panel values correspond to those reported in Table C.2.

We show a comparison of the empirical variance of all four estimators in Table C.2. There, we find that the proposed estimator $\hat{\beta}_{IPTMP}$ is approximately as variable as the IPT-weighted estimator, with both being more variable than other estimators compared. However, given

that the proposed estimator $\hat{\beta}_{IPTMP}$ is unbiased, the empirical mean squared error (MSE) for that estimator was smaller across the board, in settings with no confounding (results not shown). In settings with confounding, even if less biased, the MSE for $\hat{\beta}_{IPTMP}$ was typically greater than that of $\hat{\beta}_{IIVP}$, indicating that the variance due to the generalized IPT weight was quite important. As expected, the gap between the empirical MSE of $\hat{\beta}_{IPTMP}$ and $\hat{\beta}_{IIVP}$ in the confounded setting tended to decrease as the sample size grew (the MSE of $\hat{\beta}_{IPTMP}$ was 112% larger on average, in the case with $n = 250$, and 45% larger when $n = 1000$). In settings with more confounding, or if we increased the sample size in our studies, we therefore would expect smaller MSEs across the board with the proposed $\hat{\beta}_{IPTMP}$, while other estimators would remain biased.

We observed simular results in both sensitivity analyses (Appendix C.2), but the second sensitivity analysis also led to a greater bias due to covariate-driven monitoring times in the estimators that did not account for this type of imbalance, as expected.

5.4 Application to the *Add Health* study

The proposed methodology was applied to the *Add Health* study, a longitudinal study made of several waves, in which individuals were followed from their adolescence to their adulthood (Harris [2009]). Individuals' personal data were collected via several questionnaires. They included demographics, social and biological determinants, behavioral determinants, and others. For the purpose of this analysis, we used public-use data sets from the first four waves, which are free and available online (Harris and Udry [2018]). These data respectively correspond to years 1994-1995, 1996, 2001-2002, and 2008. We solely focused on the data obtained from the in-home questionnaires, which were available at all four waves but that did not contain all the same questions, as well as the parent questionnaire (available for the first wave only). The sample population contained 6504 individuals.

Our aim was to assess the (causal) marginal effect of the number of hours spent gaming with

video and/or computer games (further referred to as video games) per week, on the number of suicide attempts. In all four in-home questionnaires (corresponding to the four waves), the question *During the past 12 months, how many times did you actually attempt suicide?* was asked, and responses were categorized differently across waves (Waves 1-2: 0 time, 1 time, 2-3 times, 4-5 times, 6 or more times; Waves 3-4: 0 time, 1 time, 2 times, 3-4 times, 5 or more). To have a consistent outcome definition across all waves, the study outcome was further categorized as 0 attempts, 1 attempt, or 2 or more attempts at suicide. In all four in-home questionnaires, individuals were also asked *How many hours a week do you play video or computer games?*. This variable was used to define the exposure. Given that the variable distribution was highly skewed, in the main analysis, we log (base 2) transformed the exposure (after adding one unit), yielding a more symmetric and approximately normally distributed exposure. The log base 2 would also provide a straightforward interpretation for a 1-unit increase in the transformed variable, which then corresponded to a 2-fold increase in the former *skewed* variable, the number of hours spent playing video games. Four individuals who respectively answered that they played video games 120, 140, 168 and 168 hours per week at one of the four waves had that number truncated to 112 hours per week before conducting the log transformation, a maximum that corresponded to an average of 16 hours of gaming per day. We postulated that the effect of video games on suicide attempts is mediated by the depressive mood (see Figure 5.1, right panel). Information on individuals' mood was available via the question *How often was the following true during the past week? You felt depressed.* Possible answers to that question were consistent across all waves, and consisted in an ordinal scale from 0 to 3, with 0 being never or rarely, and 3 being most of the time or all the time.

For the causal question of interest, we assume that exposure to video games (possibly on a transformed scale) affects the marginal odds of suicide attempt both directly, and indirectly via the depressive mood (see e.g. Goldfield et al. [2016], Maras et al. [2015], and Johnson et al. [2013]). Depressive mood therefore lies on the causal path from the exposure to video

games, to the outcome. Other mediators, which are affected by exposure to video games and could affect suicide, such as the quality of being fearless about death or having elevated physical pain tolerance (Houtsma [2017]), were not available in this particular study. Further, we postulate that being depressed and/or spending time on video games, as well as other characteristics in the confounder set, may make a participant less willing to complete a questionnaire (therefore influencing their chances of being "monitored" according to our definition). These assumptions are depicted in the causal diagram in Figure 5.1, right panel.

The set of potential confounders included age, sex, socioeconomic status (SES) (defined using two questions asked to one of the participant's parent: *About how much total income, before taxes did your family receive in 1994?* and *How far did you go in school?*; the answers were transformed into quintiles and summed to give a score betwen 0 and 10, with 10 the highest SES), ethnicity, the frequency of having trouble relaxing (FHTR), the level of grooming of the respondent (LGR), seeing that the respondent seemed bored or impatient (RSBI), their most recent grades in mathematics (MATH), English or language arts (ENG), History or Social Sciences (HSS) and in Science (GS), the frequency with which they hang out with friends (HOF), their feeling that friends cared about them (FCA) and the number of days when they smoked cigarettes over the past month. The waves when these covariates were measured, and therefore could potentially be updated, are described in Table 5.2 under the *Exposure model* column.

In the study, visits (or monitoring times) consisted of the times or waves when the outcome related to suicide attempts were available. We chose a set of predictors for the monitoring times which reflected our beliefs about which individuals' characteristics can influence both their response availability and the number of suicide attempts. Our choice was also influenced by the work of Kalsbeek et al. [2020] on determinants of nonresponse in the *Add Health* study. For the monitoring model, we therefore selected age, sex, SES, ethnicity, the variables FHTR,

LGR, RSBI, MATH, ENG, HSS, GS, HOF, FCA, cigarette consumption in the past month, as well as the weekly number of hours spent playing video games, and, as defined earlier, the frequency of feeling depressed, which is a potential mediator of the effect under study. The waves when these covariates were measured are also described in Table 5.2, under the *Monitoring model* column.

Since the predictors of monitoring must be measured both when there is ($dN_i(t) = 1$) and there is no visit ($dN_i(t) = 0$) to properly model the monitoring indicators, and that confounders should also be available at all times to adequately adjust for confounding, multiple imputation with 5 replicated datasets (Rubin [1976]) was used to impute missing covariates on times when there was no visit, as well as on times when the outcome was *not* missing but that these variables were simply not recorded or assessed. We therefore assumed that covariates in the exposure and in the monitoring models were missing at random. Before conducting the imputation, some variables were merely replaced by sensible summary values: the grades (MATH, ENG, HSS, GS), which were only measured in the first two waves, were averaged over those two waves and used to replace missing values at any of the four waves. The sex, SES, and ethnicity, were defined only once at baseline and duplicated at all other waves. If they were missing in Wave 1, they were imputed separately at all four waves based on other characteristics. Age at all four waves could be inferred by using the years corresponding to each wave, and the year of birth for each individual. It was only missing if no information on date of birth or age was ever available. After imputation, all covariates (including the exposure and the mediator of interest) were completely filled, except for the outcome that was left as is. More details on the procedure we used for imputation, the rates of missing values in each covariates, and the performance of the imputation can be found in Appendix C.3.

Table 5.2: Variable definition for the analysis of the *Add Health* study, United States, 1994-2008, $n = 6504$ individuals. Waves 1, 2, 3 and 4 are respectively represented by acronyms W1, W2, W3, and W4. A column presents the times (waves) when these questions were asked to participants or their parents.

Variable (*question*)	Times of measurement	Exposure model	Monitoring model
Age	W1, W2, W3, W4	X	X
Sex	W1†	X	X
Socioeconomic status (SES)‡	W1†	X	X
Ethnicity	W1†	X	X
Frequency of having trouble relaxing (FHTR)$^{\zeta}$	W1, W2	X	X
Level of grooming of the respondent (LGR)$^{\nu}$	W1, W2, W3, W4	X	X
Respondent seemed bored or impatient (RSBI)$^{\nu}$	W1, W2, W3, W4	X	X
Most recent grade in Mathematics (MATH)	W1, W2	X	X
Most recent grade in English/language arts (ENG)	W1, W2	X	X
Most recent grade in History/Social sciences (HSS)	W1, W2	X	X
Most recent grade in Science (GS)	W1, W2	X	X
Frequency of hanging out with friends (HOF)$^{\iota}$	W1, W2, W3	X	X
Feeling that friends care about you (FCA)	W1, W2	X	X
How many days of smoking cigarettes over past month	W1, W2, W3, W4	X	X
Number of hours spent on video or computer games	W1, W2, W3, W4		X
Frequency of feeling depressed	W1, W2, W3, W4		X

† Considered as remaining fixed throughout the study; ‡ Defined as the decile of a combination between patients' salary and patients' education; ν Question answered by the researcher questioning the participant, rather than directly by the respondent; ι In the past week

Separate proportional intensity models for monitoring, and separate linear regression models for the exposure (one as a function of the confounders, and another for stabilization), were fitted on each imputed dataset. In those models, we used the covariates mentioned above and presented in Table 5.2. Time since cohort entry (thus, since Wave 1) was the time axis we considered in the Andersen and Gill model for the monitoring rates, and therefore the baseline monitoring rate $\lambda_0(t)$ canceled across individuals at each wave, such that it did not require estimation. We present in Appendix C.4 a comparison of the standardized mean differences (SMDs) for the covariates in the proportional intensity model, across observations that were monitored (outcome has been assessed), versus not monitored (outcome not assessed), and before or after reweighting the observations by the inverse intensity of visit weight. Those results are shown for only one imputed dataset. Checking that balance may help in assessing

the performance of the inverse intensity of visit weight. In this case, the weight clearly reduced the imbalances due to the monitoring process; none of the SMDs were greater than 0.10 after re-weighting, and all were reduced except for those corresponding to the variables *FHTR* and *FCA* which already were smaller or equal to 0.02 in the unweighted sample (Table C.3).

Table 5.3: Estimated rate ratios (95% CI) for the monitoring model, *Add Health* study, United States, 1994-2008, $n = 6504$ individuals.

Variable	Rate ratio (Bootstrap 95% CI)
Number of hours spent on video or computer games	1.00 (1.00, 1.00)
Frequency of feeling depressed (Ref.= Never or rarely)	
Sometimes	1.00 (0.97, 1.02)
A lof of the time	0.99 (0.94, 1.03)
Most of the time or all the time	1.01 (0.94, 1.08)
Age	0.93 (0.93, 0.94)
Sex (Female)	1.11 (1.09, 1.13)
SES	1.01 (1.01, 1.02)
Race (Ref.= White)	
Black/African American	0.93 (0.91, 0.95)
American Indian/Alaskan Native	0.96 (0.87, 1.04)
Asian/Pacific Islander	0.92 (0.88, 0.97)
Other	0.90 (0.86, 0.94)
FHTR	1.00 (0.99, 1.02)
LGR	0.99 (0.98, 1.00)
RSBI	0.98 (0.95, 1.02)
MATH	1.00 (0.99, 1.01)
ENG	1.00 (0.99, 1.01)
HSS	1.00 (0.99, 1.01)
GS	0.99 (0.98, 0.99)
HOF	1.00 (0.99, 1.02)
FCA	1.00 (0.98, 1.01)
How many days of smoking cigarettes over past month	1.00 (1.00, 1.00)

Average rate ratios and estimates for the marginal log-OR of exposure (the log base 2 transformed number of hours spent playing video games) were computed using Rubin's rule for multiply imputed datasets (Rubin [2004]). For computing 95% confidence intervals (CI), we used a nonparametric bootstrap. Data were sampled with replacement within each participant cluster, so that each individual had the same number of monitored outcomes in

each sample as in the original dataset and within-person correlation was maintained across bootstrap resamples.

We conducted three additional analyses to assess the sensitivity of the results to non-normality of the exposure distribution. First, in analysis S1, we assessed the original methodology with the generalized IPT weight presented in equation (5.4) to estimate the log-OR for a 1-unit or a 10-hour increases in the number of hours spent playing video games weekly (not transformed). Second (S2), the same exposure variable as in analysis S1 was used, but the IPT weight was computed by categorizing the exposure in five bins and using a POM model to fit the estimated probability of belonging to a respective category as a function of potential confounders. That probability was then used in an inverse weight in the weighted estimating equations. The five categories for the exposure were: 0 hours, 1 hour, and three other ranges based on the tertiles of the rest of the exposure distribution (the number of hours spent playing video games). Those categories were chosen for the high frequencies of 0 and 1 in the number of hours spent gaming in the dataset. Third (S3), the same analysis as in S2 was reproduced, but the exposure in the outcome model was the log base 2 transformed number of hours spent playing video games, rather than the number of hours itself. The categorical IPT weight was used in S3, with each observation being assigned the same category as in S2 (given there is a one-to-one mapping between the continuous and the log base 2 transformed variables). For S3, contrary to S1 and S2, we looked at the marginal effect of a 1-unit increase or 3-unit increase in the log base 2 number of hours, which respectively correspond to a 2-fold or 8-fold increases in the number of hours spent playing video games.

The estimated rate ratios for monitoring are shown in Table 5.3, along with bootstrap 95% CI. We found that characteristics such as being older, being male, lower SES, being Black/African American, Asian/Pacific Islander, or from another race than those listed, as well as having a greater grade in science, were statistically significantly associated with a lower

chance of having replied to the question on suicide attempts. Recall that the group that did not respond to the question was a mix of patients who did not respond to any questionnaire at a given wave (who were completely absent from the study), and those who simply did not respond to the question on suicide attempts in particular.

In the main analysis, the four estimators we compared for the marginal OR for a 1-unit increase or 3-unit increase in the log base 2 number of hours spent on video games are presented in Table 5.4, with their respective bootstrap CI. A 1-unit or a 3-unit increases in the log base 2 number of hours respectively correspond to a 2-fold or a 8-fold increases in the number of hours spent playing video games. We find in Table 5.4 that most estimators show a decrease in the number of suicide attempts for a greater number of hours spent on video games (2-fold OR: 0.91 ($\hat{\beta}_{POM}$), 0.95 ($\hat{\beta}_{IIVP}$)) while our doubly-weighted estimator $\hat{\beta}_{IPTMP}$ provides an estimate for a 2-fold increase in the number of hours of gaming that corresponds to a multiplicative effect of 1.05 (95% CI 0.92, 1.15) on the odds of passing to the next category of the categorical suicide attempts variable (i.e., going from 0 to 1 suicide attempt or from 1 to 2 or more attempts).

Table 5.4: Comparison of four estimators for the marginal OR for a two-fold or 8-fold increases in the time spent on video games per week, on the odds of suicide attempts (number of attempts categorized in 0, 1, or more), *Add Health* study, United States, 1994-2008, $n = 6504$ individuals. Confidence intervals computed via bootstrap resampling.

Estimator	2-fold increase OR (95% CI)	8-fold increase OR (95% CI)
$\hat{\beta}_{POM}$	0.91 (0.83, 0.98)	0.76 (0.57, 0.96)
$\hat{\beta}_{IPTP}$	0.99 (0.89, 1.08)	0.98 (0.69, 1.28)
$\hat{\beta}_{IIVP}$	0.95 (0.85, 1.03)	0.86 (0.61, 1.09)
$\hat{\beta}_{IPTMP}$	1.05 (0.92, 1.15)	1.15 (0.78, 1.53)

The coefficients (log-OR) were not statistically significantly different from 0 for the effect of video games. Our proposed doubly-weighted estimator $\hat{\beta}_{IPTMP}$ was the only estimator showing a (non-significant) increase in the probability of more suicide attempts when playing more video games. The three sensitivity analyses led to similar results (Appendix C.5, Tables

C.4 to C.6). Using the coefficients for the marginal OR of exposure and the outcome category-specific intercepts estimated with all four estimators, we can estimate the probability of 1 or more suicide attempt(s) for any value of the log base 2 transformed number of hours spent playing video games, or for the corresponding number of hours spent playing video games. For instance, in the main analysis, we estimated this probability for 5, 10, 30, 70, or 100 hours spent playing video games, and respectively obtained probabilities of (2.3, 2.1, 1.9, 1.7, 1.6) with the POM estimator, (2.0, 1.9, 1.8, 1.7, 1.6) with the IIVP estimator, (2.6, 2.6, 2.6, 2.6, 2.5) for the IPTP estimator, and (2.3, 2.4, 2.6, 2.7, 2.8) with the IPTMP estimator. In Figure 5.3, we plot and compare the estimated marginal probability of 1 or more suicide attempt given the time spent on video games, as well as the marginal probability of 2 or more attempts. The Figure also includes a rug plot on the X-axis, that shows the values that the exposure (the number of hours spent on video games) takes in the dataset, up to 110 hours per week. Four observations with respectively 120, 140, 168 and 168 hours of gaming per week are not included in that rug plot. In Appendix C.6, we present these estimated probabilities along with the corresponding 95% CI computed from the percentiles of the bootstrap distribution (Figure C.1). The bootstrap sampling procedure accounted for the variance of all respective fitted coefficients used to compute the probabilities. We also present the same plots that correspond to each of the three sensitivity analyses, either with or without the corresponding 95% CI – which tend to crowd the graphs – (Figures C.2 to C.7). We observed similar trends as in the main analysis in all these plots.

While it has been shown that the exposure to active or to serious video games could be beneficial to teenagers (see e.g. Zayeni et al. [forthcoming], Zurita-Ortega et al. [2018]), most studies on that topic were cross-sectional studies that used only one time point per individual to assess the effect of video games. Further, it is yet not clear that non-active video games are not detrimental to individuals' health and wellbeing (see e.g. Teismann et al. [2014], Messias et al. [2011], and Anderson and Ford [1986]). We hereby have found a non-significant detrimental effect of increasing the amount of time spent on video games

each week when accounting for bias due to confounding and selection to report. That effect is estimated marginally but there could be differential effects across sex (as suggested in Anderson and Ford [1986]) or by the type of video games. That latter feature, in addition to some other social behavior determinants or parental determinants, were not available to us, and therefore the marginal effect observed is one that may combine the effects in several different subgroups of video games users (active, non-active, addicted, etc.) which differ on characteristics. Finally, we could not find similar studies that were experimental and in which exposure to video games was randomized. More research is needed, with larger datasets, to understand whether there can be a significant detrimental effect of a large amount of time spent on video games weekly, on suicide attempts.

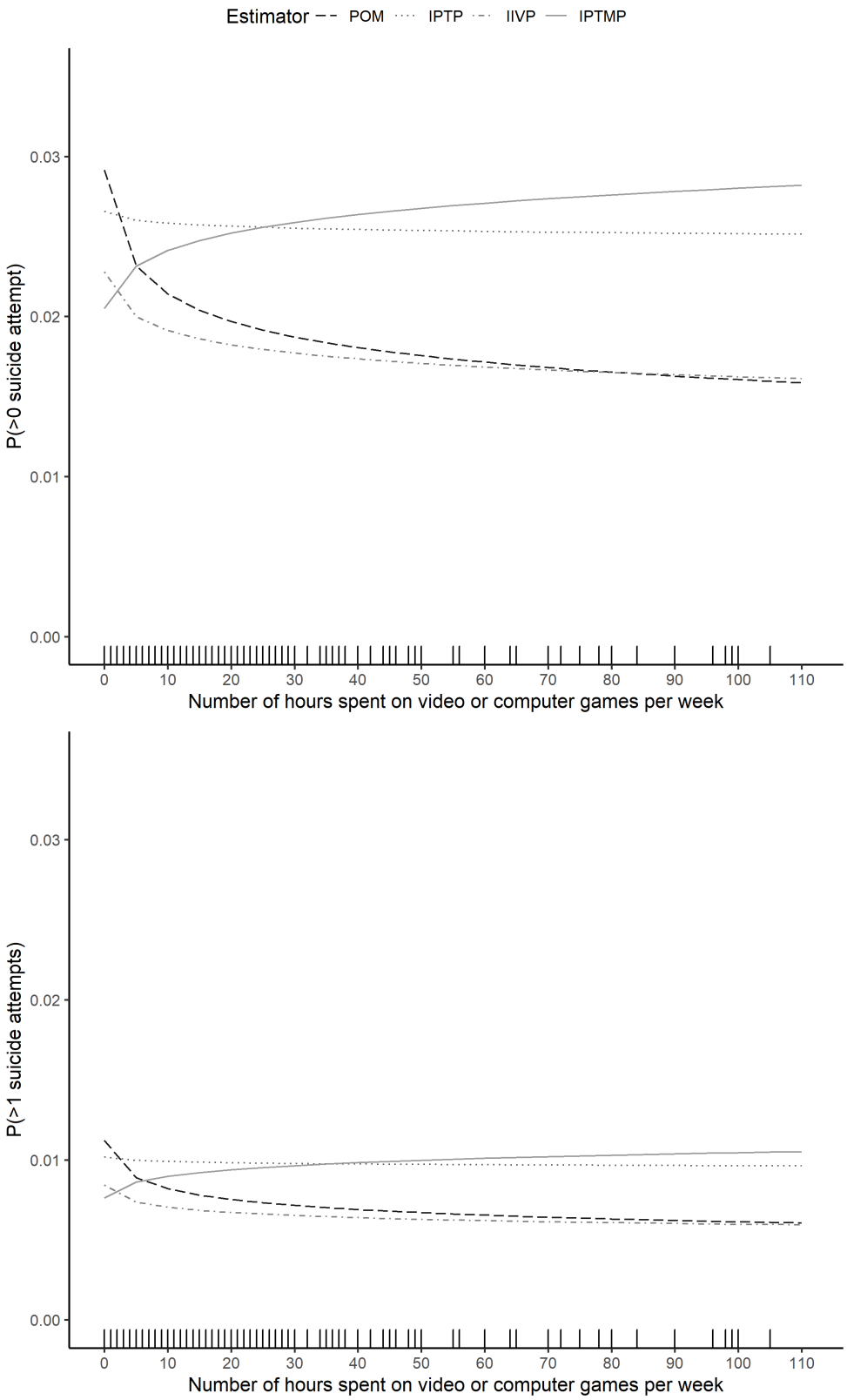


Figure 5.3: Probability of 1 or more suicide attempts (top panel) or of 2 or more suicide attempts (bottom) according to the number of hours spent on video games per week. Comparison of four estimators for the marginal log-OR in the main analysis. The rug plot on the X-axis shows the different values of the number of hours spent playing video games in the study cohort, up to 110 hours per week.

5.5 Discussion

In this work, we proposed a methodology for causal inference on the marginal effect of a continuous exposure on a categorical, ordinal outcome. Our methodology utilizes the theory of estimating equations to derive an estimator for the marginal OR of an exposure. The methodology applies to settings in which monitoring times are not fixed across individuals (i.e., can occur at anytime, for anyone) and in which confounding and selection due to covariate-driven monitoring times can affect the estimator for the exposure effect. Our approach requires relatively weak assumptions on the monitoring model to allow inference in settings with both fixed observation times such as in our motivating example as well as irregular observation times, such as would be present in electronic health records data. We illustrated our methodology via an analysis that aimed to evaluate the causal effect of the time spent on video games, on the number of suicide attempts, in a cohort of individuals followed from adolescence to adulthood. We found a (non-significant) detrimental effect of a larger time spent on video games on the odds of suicide attempts in the study population. We used nonparametric bootstrap to compute variance estimates in the application. Although not included, semiparametric theory and influence function-based inference could be used to derive a variance estimator for the marginal OR, or to derive its asymptotic properties more generally, as has been done in the continuous outcome setting (Coulombe et al. [2021a]).

The methodology we proposed relies on important assumptions about the exposure and the monitoring model. First, although this was not discussed in depth, our estimator relies on the assumption of temporality, where the exposure, to cause the outcome, must have occurred before. In the longitudinal survey, the exposure related to the previous week, and the outcome, to the previous year. Some extrapolation is necessary to assume that the current weekly exposure to video games was relatively similar to that one year ago (for a given individual), and that the effect of interest is well-defined. Secondly, for translating our parametric inference to the causal framework, we made the assumption of conditional

exchangeability for the potential outcome across the different exposure levels, conditional on a certain set of confounders. As in all observational studies, it is possible that unmeasured confounders could not be captured. For instance, other social determinants than those we chose in our models could affect the exposure to video games and the suicide attempts. As discussed in Houtsma [2017], features such as previous aggressions, exposure to media violence, fearlessness to pain or about death are risk factors for suicide. These factors could possibly relate to the exposure to video games, either by confounding the effect or by being along the causal path from the exposure to the suicide outcome. Other behavioral characteristics, such as those influenced by the parents during adolescence, were not measured either and could influence both the exposure to gaming and the mood of the individuals. Furthermore, the proposed methodology relies on the normality assumption for the exposure distribution, and therefore, the generalized IPT weight could not adjust fully for confounding if the distribution is far from being normally distributed (even in the cases where we captured all potential confounders). If the distribution is very skewed or multimodals, then other generalized IPT weights such as those assessed or proposed in Naimi et al. [2014], and Schulz and Moodie [2021] should perform better. However, we are confident that our results were not unduly influenced by our assumptions regarding the distribution of the exposure as our sensitivity analyses all came to the same conclusions. Furthermore, correct specifications of the outcome model (the POM) as a function of the exposure, the monitoring model, and the exposure model can be assessed via standard diagnostic checks, such as residuals check, and by looking at the balance in covariates across different exposure levels (Austin [2019]) and across the observations corresponding to monitored and unmonitored outcomes (as we presented in Appendix C.4). Further work will focus on assessing the performance of such balance checks for the monitoring model, and developing new criteria for the monitoring model fit. We also assumed positivity of the exposure, hypothesizing that all individuals could be exposed to any level of time spent on video games after conditioning on their characteristics. It is unlikely that everyone had the chance of being exposed to more than

e.g. 30 to 40 hours of video game per week; this could possibly depend on unmeasured characteristics such as the parents' house rules or other personal individuals' characteristics (for instance, having a day job). Some methods have been proposed for causal inference on a continuous exposure that relax the strong assumptions on the exposure positivity (Haneuse and Rotnitzky [2013], Muñoz and van der Laan [2012]). Some authors also proposed to target the effect of a shift in the propensity of being treated rather than the effect of setting treatment to a fixed value, an approach that does not require positivity (Kim et al. [2019]). The study context is also one in which interference may be possible, while we assumed that it was not present. For instance, the exposure of an individual to video games can certainly affect their friend's exposure and, possibly, their friend's outcomes. This remains to study. However, given the sample size relative to the United States population, it is quite unlikely that the participants in the *Add Health* study knew or influenced one another.

This work is the first to propose a methodology for causal inference data subject to covariate-driven monitoring times and confounding when the exposure is continuous, and the outcome, ordinal. Several study outcomes are ordinal, such that our proposed methodology can be useful in several settings. Our study is also the first substantive study to look at the effect of the time spent weekly on video games on the suicide attempts, that accounts both for confounding and covariate-driven monitoring times. While other studies have considered potential confounders (without accounting for monitoring times), most presented only a modest adjustment for confounders (see e.g. Messias et al. [2011]). In studies of video game exposure, having covariate-driven monitoring times is plausible (see e.g. Khazaal et al. [2014] who discuss self-selection for online surveys on video games); furthermore, it is very likely that patients with more suicidal ideation may present with different monitoring patterns than others (Tylee [1999]). Thus, it is critical to think about those possible biases in observational studies, when looking at similar causes and their effect on suicide attempts.

5.5.1 Acknowledgements

This research was enabled in part by support from Compute Canada (www.computecanada.ca). Our work is supported by a doctoral scholarship from the Natural Sciences and Engineering Research Council (NSERC) of Canada (Ref. 401223940) to author JC. EEMM acknowledges support from a Discovery Grant from NSERC and a chercheur-boursier career award from the Fonds de recherche du Québec–Santé. RWP acknowledges support from a Discovery Grant from NSERC and a Foundation Scheme Grant from CIHR.

This research uses data from the *Add Health* program. These data are available in the Data Sharing for Demographic Research repository, at https://doi.org/10.3886/ICPSR21600.v21. The analysis was restricted to the *Add Health* public-use data.

Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill. The project was funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development from 1994-2021, with cooperative funding from 23 other federal agencies and foundations. *Add Health* is currently directed by Robert A. Hummer; it was previously directed by Kathleen Mullan Harris (2004-2021) and J. Richard Udry (1994-2004). More information on obtaining *Add Health* data is available on the project website (https://addhealth.cpc.unc.edu). The *Add Health* Parent Study/Parents (2015-2017) data collection was funded by a grant from the National Institute on Aging (RO1AG042794) to Duke University, V. Joseph Hotz (PI) and the Carolina Population Center at the University of North Carolina at Chapel Hill, Kathleen Mullan Harris (PI). The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the University of North Carolina at Chapel Hill.

Chapter 6

Conclusion

6.1 Summary

In this thesis, I have proposed a general framework for causal inference regarding the marginal effect of an exposure on a longitudinal outcome, when data are subject to confounding and covariate-driven monitoring times that are associated with the outcome.

In a first manuscript, in Chapter 3, two novel consistent estimators were derived and implemented. One of them, the IPCTM weighted estimator, incorporated a generalized inverse probability of treatment weight that accounted for the re-centered version of the design matrix in the estimating equations, and it did not require the specification of an intercept function in the outcome mean model. The simulation studies have shown that this extended estimator performs well both in settings with time-fixed or time-varying exposure and confounders, and that it was slightly less variable than the second estimator, the FIPTM, in studies with time-varying exposure. However, the FIPTM was found to be easier to implement in practice. In finite samples, both estimators presented with similar (small) bias, and both outperformed other estimators that did not account for confounding and covariate-driven monitoring times simultaneously.

The second manuscript, presented in Chapter 4 of this thesis, was motivated by an application to data from the CPRD in the UK, and the aim to compare the marginal effects of two antidepressant drugs, citalopram and fluoxetine, on BMI. In that manuscript, a thorough description of the bias due to covariate-driven monitoring times was provided. Stratification collider bias in a longitudinal monitoring context was depicted and its effect on the inference about the ATE was discussed. To answer the substantive research question, a novel cumulated and stabilized weight was proposed, that breaks the links between the monitoring and the covariate processes all throughout follow-up time, assuming that the monitoring and the outcome processes are independent conditional on that covariate process. This weight is the first to be proposed for the context where covariates interact with the monitoring process, inducing long-term dependencies. In simulation studies, the performance of that weight was demonstrated when using different stabilizers for the visit intensity function. One of the stabilizers outperformed others. The doubly-weighted estimator using that stabilizer was consistenly unbiased. The work presented in Chapter 4 also relaxed the assumption about the availability of the covariate process affecting the visit times. Sequential ignorability for subsequent monitoring indicators was assumed, conditional on the covariate process being assessed only at times when the outcome was observed. This provides more flexibility in the analysis of longitudinal data subject to covariate-driven monitoring times.

In Chapter 5, in a third manuscript, the methodology from Chapter 3 was extended to scenarios where the exposure of interest is continuous, and where the outcome is categorical and ordinal. A generalized IPT weight was used to break the links between the pretreatment covariates and the continuous exposure. Alternative weighting strategies were assessed for settings where the exposure distribution is not normally distributed (e.g., when it is skewed or multimodal). A few different inverse probability of treatment weighting methods were compared in the application to the *Add Health* study, and more insights were gained into the marginal effect of the time spent playing video games, on the number of suicide attempts.

6.2 Limitations and avenues for future work

Limitations of the work in this thesis relate mostly to the assumptions made about the exposure and the monitoring models, and to the availability of the data.

The proposed methodologies all rely on a set of assumptions that are necessary for the corresponding estimators to be consistent. The potential confounders and the predictors of the monitoring indicators must be identified correctly, and their form (e.g. linear, logged) in the models should be specified correctly. Further, for the exposure models, the usual causal inference assumptions of temporality, positivity, consistency of the potential outcome, a well-defined exposure and conditional exchangeability (on a set of potential confounders) must be met. Similar conditions must be met for the monitoring model. The IIVW only properly adjust for informative monitoring times if positivity and conditional exchangeability (on a set of predictors) are met. While some of these assumptions are empirically verifiable, the exchangeability conditions are hard to check in practice. That is especially true for the monitoring model, where, to my knowledge, no sensitivity analysis has been developed to assess conditional exchangeability specifically. However, as suggested by the authors Sun et al. [2005], their joint modeling method could potentially be used to assess independence between the monitoring and the outcome processes. This thesis also relied throughout on the assumption that the monitoring times coincide with the times when the outcome process is observed, and for Chapters 3 and 5, it relied on the covariate process which affects monitoring to be continuously measured. However, in most observational studies like those using EHR data, it is likely that covariates will be updated at irregular times during follow-up, and that a mix of observed and unobserved covariates will be available at each point in time. A very thorough analysis could potentially account for the monitoring probability of each single covariate, having for instance a monitoring model for the outcome, the exposure, and for each covariate that affects the outcome or the outcome monitoring.

When used in practice, the work presented in this thesis relies heavily on a correct specification of the data generating mechanism and corresponding causal diagram. In particular,

if a variable acts as a mediator but it is treated as a confounder, then including it in the set of potential confounders can bias the proposed estimators. This suggests that substantive knowledge about the research question is highly important. In the first and the third manuscripts of this thesis, the substantive research questions were used mostly as illustrations for the respective methods, and co-authors had modest knowledge about the more substantive theory (e.g., on biological mechanisms). For correct causal inferences, it would be important to include in the research team experts in the respective substantive fields, and to include in the corresponding work exhaustive literature reviews on the matter.

For future work, the proposed methods could be compared to imputation-type methods where missing values in the outcome are replaced, rather than being treated using IIV weights. It would be interesting to compare the variances of the different methods. Furthermore, the methods extended in this thesis were also extended outside of the causal framework to account for random effects (e.g., for contexts where the outcome is missing not at random). It would be interesting to describe the advantages of using random effects, and assess their use in the monitoring or in the exposure models, in a causal inference context.

6.3 Concluding remarks

In this thesis, I proposed novel methods that can be used in longitudinal observational studies to assess the causal marginal effect of an exposure, on a continuous or an ordinal longitudinal outcome. This thesis advances the statistical literature on causal inference through the development of the asymptotic theory of some of the proposed estimators, and through their demonstration via extensive simulation studies. In addition, I provided a comprehensive description of the bias due to covariate-driven monitoring times, via the use of causal diagrams. I applied the proposed methods to different types of datasets, and to answer varied substantive research questions concerning mental health in adolescents and adults. This provided step-by-step illustrations for other researchers who would like to apply

the proposed methods in other contexts.

Appendices

APPENDIX A

Appendix to Manuscript 1

A.1 Theoretical proofs and development of an estimator for the conditional effect of treatment

1) First, we show that

$$E[dP_i(t)|I_i(t), \mathbf{K_i(t)}] = 0$$

where $P_i(t) = \int_0^t \frac{1}{\rho_i(s;\boldsymbol{\gamma})} \{(Y_i(s) - \beta_I I_i(s) - \boldsymbol{\beta_K'}\mathbf{K_i(s)})\, dN_i(s) - \xi_i(s) \exp(\boldsymbol{\gamma_V'} \boldsymbol{V_i(s)})\, d\mathscr{A}(s)\}$, with

$$\mathscr{A}(t) = \int_0^t \alpha(s) d\Lambda(s). \tag{A0}$$

Assumptions:

$$E[Y_i(t)|I_i(t), \mathbf{K_i(t)}] = \alpha(t) + \beta_I I_i(t) + \boldsymbol{\beta_K'}\mathbf{K_i(t)} \tag{A1}$$

$$E[dN_i(t)|\mathbf{V_i(t)}] = \xi_i(t) \exp(\boldsymbol{\gamma}_V' \mathbf{V_i(t)})\, d\Lambda(t) \tag{A2}$$

$$N_i(t) \perp Y_i(t)|\mathbf{V_i(t)} \tag{A3}$$

We have that

$$
\begin{aligned}
E[dP_i(t)|I_i(t),\mathbf{K_i(t)}] &= E\left[\frac{\exp\gamma_I I_i(t)}{\exp\left(\boldsymbol{\gamma_1'Z_i(t)}+\gamma_2 I_i(t)\right)}\left(Y_i(t)-\beta_I I_i(t)-\boldsymbol{\beta_K'}\mathbf{K_i(t)}\right)dN_i(t)\Big|I_i(t),\mathbf{K_i(t)}\right] \\
&\quad -E\left[\frac{\exp\gamma_I I_i(t)}{\exp\left(\boldsymbol{\gamma_1'Z_i(t)}+\gamma_2 I_i(t)\right)}\xi_i(t)\exp\left(\boldsymbol{\gamma_1'Z_i(t)}+\gamma_2 I_i(t)\right)d\mathscr{A}(t)\Big|I_i(t),\mathbf{K_i(t)}\right] \\
&= \frac{\exp\left(\gamma_I I_i(t)\right)}{\exp(\gamma_2 I_i(t))}E\left[\frac{1}{\exp\boldsymbol{\gamma_1'Z_i(t)}}\left(Y_i(t)-\beta_I I_i(t)-\boldsymbol{\beta_K'}\mathbf{K_i(t)}\right)dN_i(t)\Big|I_i(t),\mathbf{K_i(t)}\right] \\
&\quad -\frac{\exp\left(\gamma_I I_i(t)\right)}{\exp(\gamma_2 I_i(t))}E\left[\frac{1}{\exp\left(\boldsymbol{\gamma_1'Z_i(t)}\right)}\xi_i(t)\exp\left(\boldsymbol{\gamma_1'Z_i(t)}+\gamma_2 I_i(t)\right)d\mathscr{A}(t)\Big|I_i(t),\mathbf{K_i(t)}\right] \\
&= \frac{\exp\left(\gamma_I I_i(t)\right)}{\exp(\gamma_2 I_i(t))}\alpha(t)E\left[\frac{1}{\exp\boldsymbol{\gamma_1'Z_i(t)}}dN_i(t)\Big|I_i(t),\mathbf{K_i(t)}\right] \\
&\quad -\frac{\exp\left(\gamma_I I_i(t)\right)}{\exp(\gamma_2 I_i(t))}\alpha(t)E\left[\frac{1}{\exp\left(\boldsymbol{\gamma_1'Z_i(t)}\right)}\xi_i(t)\exp\left(\boldsymbol{\gamma_1'Z_i(t)}+\gamma_2 I_i(t)\right)d\Lambda(t)\Big|I_i(t),\mathbf{K_i(t)}\right],
\end{aligned}
$$

where the first term in the final equality follows from (A1) and (A3), and the second term follows from (A0). Using the iterated expectation for

$$
E\left[dN_i(t)|I_i(t),\mathbf{K_i(t)}\right] = E\left[E\left[dN_i(t)|I_i(t),\mathbf{K_i(t)},\mathbf{Z_i(t)}\right]\right],
$$

it follows that

$$
\begin{aligned}
&E[dP_i(t)|I_i(t),\mathbf{K_i(t)}] \\
&= \frac{\exp\left(\gamma_I I_i(t)\right)}{\exp(\gamma_2 I_i(t))}\alpha(t)E\left[E\left[\frac{1}{\exp\boldsymbol{\gamma_1'Z_i(t)}}dN_i(t)\Big|I_i(t),\mathbf{K_i(t)},\mathbf{Z_i(t)}\right]\right] \\
&\quad -\frac{\exp\left(\gamma_I I_i(t)\right)}{\exp(\gamma_2 I_i(t))}\alpha(t)E\left[E\left[\frac{1}{\exp\left(\boldsymbol{\gamma_1'Z_i(t)}\right)}\xi_i(t)\exp\left(\boldsymbol{\gamma_1'Z_i(t)}+\gamma_2 I_i(t)\right)d\Lambda(t)\Big|I_i(t),\mathbf{K_i(t)},\mathbf{Z_i(t)}\right]\right] \\
&= \frac{\exp\left(\gamma_I I_i(t)\right)}{\exp(\gamma_2 I_i(t))}\alpha(t)E\left[\frac{1}{\exp\boldsymbol{\gamma_1'Z_i(t)}}\xi_i(t)\exp\left(\boldsymbol{\gamma_1'Z_i(t)}+\gamma_2 I_i(t)\right)d\Lambda(t)\right] \\
&\quad -\frac{\exp\left(\gamma_I I_i(t)\right)}{\exp(\gamma_2 I_i(t))}\alpha(t)E\left[\frac{1}{\exp\left(\boldsymbol{\gamma_1'Z_i(t)}\right)}\xi_i(t)\exp\left(\boldsymbol{\gamma_1'Z_i(t)}+\gamma_2 I_i(t)\right)d\Lambda(t)\right] \text{ from (A2)} \\
&= 0.
\end{aligned}
$$

2) Second, we derive the estimators for the conditional effects. The estimators are based on

the set of estimating equations:

$$\sum_{i=1}^{n} P_i(t;\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\mathscr{A}) = 0 \quad \forall t \in [0,\tau] \tag{A4}$$

$$\sum_{i=1}^{n} \int_0^{\tau} W(t) \begin{pmatrix} I_i(t) \\ \mathbf{K_i(t)} \end{pmatrix} dP_i(t;\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\mathscr{A}}) = 0. \tag{A5}$$

In equation (A4), we have that:

$$\sum_{i=1}^{n} P_i(t) = \sum_{i=1}^{n} \int_0^t \frac{1}{\rho_i(s;\boldsymbol{\gamma})} \left\{ \left(Y_i(s) - \beta_I I_i(s) - \boldsymbol{\beta}_{\boldsymbol{K}}' \mathbf{K_i(s)}\right) dN_i(s) - \xi_i(s) \exp\left(\boldsymbol{\gamma}_{\boldsymbol{V}}' \mathbf{V_i(s)}\right) d\mathscr{A}(s) \right\} = 0$$

with one set of solutions being

$$\frac{1}{\rho_i(s;\boldsymbol{\gamma})} \left\{ \left(Y_i(s) - \beta_I I_i(s) - \boldsymbol{\beta}_{\boldsymbol{K}}' \mathbf{K_i(s)}\right) dN_i(s) \right\} = \frac{1}{\rho_i(s;\boldsymbol{\gamma})} \left\{ \xi_i(s) \exp\left(\boldsymbol{\gamma}_{\boldsymbol{V}}' \mathbf{V_i(s)}\right) d\mathscr{A}(s) \right\}$$

$$\forall\, 0 < s < t, i = 1, ..., n,$$

which implies

$$\sum_{i=1}^{n} \frac{1}{\rho_i(s;\boldsymbol{\gamma})} \left\{ \left(Y_i(s) - \beta_I I_i(s) - \boldsymbol{\beta}_{\boldsymbol{K}}' \mathbf{K_i(s)}\right) dN_i(s) \right\} = d\mathscr{A}(s) \sum_{i=1}^{n} \frac{1}{\rho_i(s;\boldsymbol{\gamma})} \left\{ \xi_i(s) \exp\left(\boldsymbol{\gamma}_{\boldsymbol{V}}' \mathbf{V_i(s)}\right) \right\}$$

$$\forall\, 0 < s < t,$$

leading to the estimator

$$\hat{\mathscr{A}}(t) = \sum_{i=1}^{n} \int_0^t \frac{\frac{1}{\rho_i(s;\boldsymbol{\gamma})} \left(Y_i(s) - \beta_I I_i(s) - \boldsymbol{\beta}_{\boldsymbol{K}}' \mathbf{K_i(s)}\right) dN_i(s))}{\sum_{j=1}^{n} \frac{1}{\rho_j(s;\boldsymbol{\gamma})} \xi_j(s) \exp\left(\boldsymbol{\gamma}_{\boldsymbol{V}}' \mathbf{V_j(s)}\right)}. \tag{A6}$$

Now, substituting the equation from (A6) into equation (A5), we have

$$
\begin{aligned}
&\sum_{i=1}^{n}\int_0^{\tau} W(t)\begin{pmatrix} I_i(t) \\ \mathbf{K_i(t)} \end{pmatrix} dP_i(t;\beta,\hat{\boldsymbol{\gamma}},\hat{\mathscr{A}}) \\
&= \sum_{i=1}^{n}\int_0^{\tau} W(t)\begin{pmatrix} I_i(t) \\ \mathbf{K_i(t)} \end{pmatrix}\left[\frac{1}{\rho_i(t;\boldsymbol{\gamma})}\left[(Y_i(t)-\beta_I I_i(t)-\beta_{\mathbf{K}}{}'\mathbf{K_i(t)})dN_i(t)\right]\right] \\
&- \sum_{i=1}^{n}\int_0^{\tau} W(t)\begin{pmatrix} I_i(t) \\ \mathbf{K_i(t)} \end{pmatrix}\left[\frac{1}{\rho_i(t;\boldsymbol{\gamma})}\left[\xi_i(t)\exp^{\boldsymbol{\gamma}_{\boldsymbol{V}}'\mathbf{V_i(t)}}\left\{\frac{\sum_{d=1}^{n}\frac{1}{\rho_d(t;\boldsymbol{\gamma})}(Y_d(t)-\beta_I I_d(t)-\beta_{\mathbf{K}}{}'\mathbf{K_d(t)})dN_d(t)}{\sum_{j=1}^{n}\frac{1}{\rho_j(t;\boldsymbol{\gamma})}\xi_j(t)\exp^{\boldsymbol{\gamma}_{\boldsymbol{V}}'\mathbf{V_j(t)}}}\right\}\right]\right] \\
&= \sum_{i=1}^{n}\int_0^{\tau} W(t)\frac{1}{\rho_i(t;\boldsymbol{\gamma})}\left[(Y_i(t)-\beta_I I_i(t)-\boldsymbol{\beta}_{\boldsymbol{K}}'\mathbf{K_i(t)})dN_i(t)\right] \\
&\times\left[\begin{pmatrix} I_i(t) \\ \mathbf{K_i(t)} \end{pmatrix} - \begin{pmatrix} I_i(t) \\ \mathbf{K_i(t)} \end{pmatrix}\xi_i(t)\exp^{\boldsymbol{\gamma}_{\boldsymbol{V}}'\mathbf{V_i(t)}}\left[\frac{\frac{1}{\rho_i(t;\boldsymbol{\gamma})}}{\sum_{j=1}^{n}\frac{1}{\rho_j(t;\boldsymbol{\gamma})}\xi_j(t)\exp\boldsymbol{\gamma}_{\boldsymbol{V}}'\mathbf{V_j(t)}}\right]\right] \\
&= \sum_{i=1}^{n}\int_0^{\tau} W(t)\frac{1}{\rho_i(t;\boldsymbol{\gamma})}\left[(Y_i(t)-\beta_I I_i(t)-\beta_{\mathbf{K}}{}'\mathbf{K_i(t)})dN_i(t)\right] \\
&\times\left[\begin{pmatrix} I_i(t) \\ \mathbf{K_i(t)} \end{pmatrix} - \begin{pmatrix} I_i(t) \\ \mathbf{K_i(t)} \end{pmatrix}\xi_i(t)\frac{\exp^{\boldsymbol{\gamma}_{\boldsymbol{V}}'\mathbf{V_i(t)}}}{\rho_i(t;\boldsymbol{\gamma})}\frac{1}{\sum_{j=1}^{n}\frac{\exp\boldsymbol{\gamma}_{\boldsymbol{V}}'\mathbf{V_j(t)}}{\rho_j(t;\boldsymbol{\gamma})}\xi_j(t)}\right] \\
&= \sum_{i=1}^{n}\int_0^{\tau}\frac{W(t)}{\rho_i(t;\boldsymbol{\gamma})}\left[(Y_i(t)-\beta_I I_i(t)-\boldsymbol{\beta}_{\boldsymbol{K}}'\mathbf{K_i(t)})\begin{pmatrix} I_i(t)-\bar{I}(t;\hat{\gamma}_I) \\ \mathbf{K_i(t)}-\bar{\mathbf{K}}(\mathbf{t};\hat{\boldsymbol{\gamma}}_{\mathbf{I}}) \end{pmatrix} dN_i(t)\right]
\end{aligned}
$$

and the estimation equation is provided by the very last row above. Just as in Bůžková and Lumley [2009], and Lin and Ying [2001], we re-centered the outcome by $\bar{Y}^*(t;\hat{\gamma}_I)$ to reduce

the variance of the estimators, leading to the final estimating equations

$$U(\boldsymbol{\beta}, \alpha, \hat{\boldsymbol{\gamma}}) = \sum_{i=1}^{n} \int_0^{\tau} \frac{W(t)}{\rho_i(t; \hat{\boldsymbol{\gamma}})} \begin{pmatrix} I_i(t) - \bar{I}(t; \hat{\gamma}_I) \\ \mathbf{K_i(t)} - \bar{\mathbf{K}}(\mathbf{t}; \hat{\gamma}_\mathbf{I}) \end{pmatrix}' \times \left[Y_i(t) - \bar{Y}^*(t; \hat{\gamma}_I) - [\beta_I \, \boldsymbol{\beta}_{\boldsymbol{K}}]' \begin{pmatrix} I_i(t) - \bar{I}(t; \hat{\gamma}_I) \\ \mathbf{K_i(t)} - \bar{\mathbf{K}}(\mathbf{t}; \hat{\gamma}_\mathbf{I}) \end{pmatrix} \right] dN_i(t). \tag{A7}$$

Using these estimating equations, we find the following estimators for the conditional effect of $\begin{bmatrix} I_i(t) & \mathbf{K_i(t)} \end{bmatrix}'$ in (O2):

$$[\hat{\beta}_I \, \hat{\boldsymbol{\beta}}_{\boldsymbol{k}}]' = \left[\sum_{i=1}^{n} \int_0^{\tau} \frac{W(t)}{\rho_i(t; \hat{\boldsymbol{\gamma}})} \begin{pmatrix} I_i(t) - \bar{I}(t; \hat{\gamma}_I) \\ \mathbf{K_i(t)} - \bar{\mathbf{K}}(\mathbf{t}; \hat{\gamma}_\mathbf{I}) \end{pmatrix}^{\otimes 2} dN_i(t) \right]^{-1} \times \sum_{i=1}^{n} \int_0^{\tau} \frac{W(t)}{\rho_i(t; \hat{\boldsymbol{\gamma}})} \begin{pmatrix} I_i(t) - \bar{I}(t; \hat{\gamma}_I) \\ \mathbf{K_i(t)} - \bar{\mathbf{K}}(\mathbf{t}; \hat{\gamma}_\mathbf{I}) \end{pmatrix}' \left(Y_i(t) - \bar{Y}^*(t; \hat{\gamma}_I) \right) dN_i(t). \tag{A8}$$

A.2 Unbiasedness of the IPCTM estimator

The unbiasedness of the IPCTM estimator can be shown using arguments similar to those used for standard IPT weighting in a linear regression model (see Rosenbaum and Rubin [1983]; Horvitz and Thompson [1952]; Robins [1986]). The estimating equation

$$U^{cond}(\boldsymbol{\beta}, \alpha, \hat{\boldsymbol{\gamma}}) = \sum_{i=1}^{n} \int_0^{\tau} \frac{W(t)}{\rho_i(t; \hat{\boldsymbol{\gamma}})} \begin{pmatrix} I_i(t) - \bar{I}(t; \hat{\gamma}_I) \\ \mathbf{K_i(t)} - \bar{\mathbf{K}}(\mathbf{t}; \hat{\gamma}_\mathbf{I}) \end{pmatrix} \times \left[Y_i(t) - \bar{Y}^*(t; \hat{\gamma}_I) - [\beta_I \, \boldsymbol{\beta}_{\boldsymbol{K}}]' \begin{pmatrix} I_i(t) - \bar{I}(t; \hat{\gamma}_I) \\ \mathbf{K_i(t)} - \bar{\mathbf{K}}(\mathbf{t}; \hat{\gamma}_\mathbf{I}) \end{pmatrix} \right] dN_i(t)$$

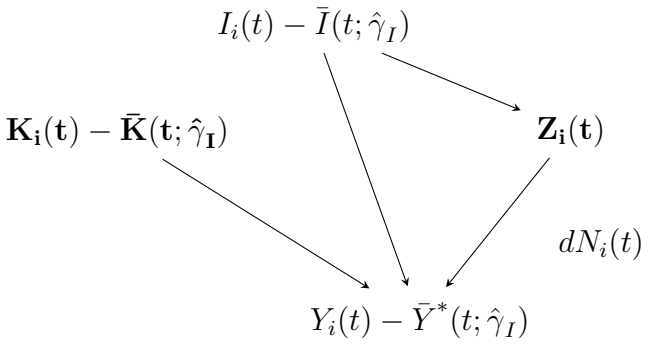
corresponds to the equation for the estimation of the conditional effect of intervention $I_i(t)$ in a model such as (O2) when the other assumptions mentioned in section 3.2 hold. The proof is presented in Appendix A.1. However, we can also see this as the estimating equation for a linear model,

$$E\left[Y_i(t) - \bar{Y}^*(t;\hat{\gamma}_I)|I_i(t) - \bar{I}(t;\hat{\gamma}_I), \mathbf{K_i(t)} - \bar{\mathbf{K}}(\mathbf{t};\hat{\boldsymbol{\gamma}}_\mathbf{I})\right] = \beta_I^*(I_i(t) - \bar{I}(t;\hat{\gamma}_I)) + \boldsymbol{\beta}_{\boldsymbol{K}}^*(\mathbf{K_i(t)} - \bar{\mathbf{K}}(\mathbf{t};\hat{\boldsymbol{\gamma}}_\mathbf{I})),$$

with weights $W(t)/\rho_i(t;\hat{\gamma})$. Re-weighting the observations by $W(t)/\rho_i(t;\hat{\boldsymbol{\gamma}})$ removes the arrow from $(I_i(t) - \bar{I}(t;\hat{\gamma}_I))$ and $\mathbf{Z_i(t)}$ to $dN_i(t)$, while further re-weighting the observations at time t, for individual i by the weight

$$\text{sgw}_i(t;\hat{\psi}_0,\hat{\boldsymbol{\psi}}_\mathbf{1},\hat{\psi}_m) = \frac{g^{-1}\left(\hat{\psi}_0 + \hat{\boldsymbol{\psi}}_\mathbf{1}'(\mathbf{K_i(t)} - \bar{\mathbf{K}}(\mathbf{t};\hat{\boldsymbol{\gamma}}_\mathbf{I}))\right)}{g^{-1}\left(\hat{\psi}_m\right)}$$

removes the arrow from $(\mathbf{K_i(t)} - \bar{\mathbf{K}}(\mathbf{t};\hat{\boldsymbol{\gamma}}_\mathbf{I}))$ to $(I_i(t) - \bar{I}(t;\hat{\gamma}_I))$ in the final DAG for the weighted pseudo-population, as shown below,



After re-weighting, one can estimate the marginal effect of $(I_i(t) - \bar{I}(t;\hat{\gamma}_I))$ on $(Y_i(t) - \bar{Y}^*(t;\hat{\gamma}_I))$ in the model

$$E\left[Y_i(t) - \bar{Y}^*(t;\hat{\gamma}_I)|I_i(t) - \bar{I}(t;\hat{\gamma}_I)\right] = \beta(I_i(t) - \bar{I}(t;\hat{\gamma}_I))$$

using the marginal estimating equation

$$U^{mar}(\beta, \alpha, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\psi}}) = \sum_{i=1}^{n} \int_0^{\tau} \frac{W(t)}{\rho_i(t; \hat{\boldsymbol{\gamma}})} \frac{1}{\text{sgw}_i(t; \hat{\boldsymbol{\psi}})} \left(I_i(t) - \bar{I}(t; \hat{\gamma}_I)\right) \times \left[Y_i(t) - \bar{Y}^*(t; \hat{\gamma}_I) - \beta \left(I_i(t) - \bar{I}(t; \hat{\gamma}_I)\right)\right] dN_i(t). \quad \text{(E3)}$$

Here, β is the same parameter as in

$$E\left[Y_i(t) | I_i(t)\right] = \alpha(t) + \beta I_i(t).$$

A.3 Asymptotic variance for the flexible inverse probability of treatment and monitoring estimator

Important notes:

- Here we show the variance for the case where the inverse probability of treatment weight $e_i(t; \hat{\boldsymbol{\omega}})$ varies in time. When it does not vary and is adjusted for confounders at baseline only, the formula has to be adjusted to take into account the fact that the logistic regression model incorporates only one data row per individual rather than one data row per (discretized) time unit for all individuals.
- In the calculation below, we consider that the confounders $\mathbf{K}$ do not affect the monitoring process. If they do, the calculation below should be updated by incorporating them into the matrix $Z(\mathbf{t})$.

Under assumptions (O2), (V1), (V2), (P1), (P2), (P3) and (C1), $\hat{\boldsymbol{\beta}}_{\boldsymbol{FIPTM}}$ is a two-step m-estimator. Two-step estimators are defined by Newey and McFadden [1994] as estimators that are based on some preliminary, first-step estimator of a parameter vector. Often, a first-step estimator is used to estimate the nuisance parameters (e.g., parameters used in a weight, which will further be incorporated into the two-step estimator). M-estimators, on

the other hand, are obtained by solving a sample average equation and often consist of the zero roots of an estimating equation.

For $\hat{\boldsymbol{\beta}}_{\boldsymbol{TSE}}$ a two-step semiparametric estimator and $\boldsymbol{\beta_0}$ the vector of true parameters, Newey and McFadden [1994] show that $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\boldsymbol{TSE}} - \boldsymbol{\beta_0}) \to N(\mathbf{0}, \boldsymbol{\Sigma})$, with

$$\boldsymbol{\Sigma} = \mathbf{G}_{\beta}^{-\mathbf{1}} \mathbb{E}\left[\left\{\mathbf{g}(\mathbf{o}; \beta_{\mathbf{0}}, \phi_{\mathbf{0}}) - \mathbf{G}_{\phi}\mathbf{M}^{-\mathbf{1}}\mathbf{m}(\mathbf{o}; \phi_{\mathbf{0}})\right\}^{\otimes 2}\right] \mathbf{G}_{\beta}^{-\mathbf{1}} \tag{A.1}$$

where

$$\begin{aligned} \mathbf{G}_{\beta} &= \mathbb{E}(\nabla_{\beta}\mathbf{g}(\mathbf{o}; \beta_{\mathbf{0}}, \phi_{\mathbf{0}})) \\ \mathbf{G}_{\phi} &= \mathbb{E}(\nabla_{\phi}\mathbf{g}(\mathbf{o}; \beta_{\mathbf{0}}, \phi_{\mathbf{0}})) \\ \mathbf{M} &= \mathbb{E}(\nabla_{\phi}\mathbf{m}(\mathbf{o}; \phi_{\mathbf{0}})) \end{aligned}$$

for $\mathbf{o}$ the data, and $\mathbf{m}(\mathbf{o}; \phi_{\mathbf{0}})$ and $\mathbf{g}(\mathbf{o}; \beta_{\mathbf{0}}, \phi_{\mathbf{0}})$ the estimating equations for the nuisance parameters $\boldsymbol{\phi}$ and the parameters of interest $\boldsymbol{\beta}$, respectively.

Recall

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{FIPTM}} = \left[\sum_{i=1}^{n} \int_0^{\tau} \frac{\mathrm{e}_i(t; \boldsymbol{\omega})}{\varphi_i(t; \hat{\boldsymbol{\gamma}})} \mathbf{S_i}(\mathbf{t})^{\otimes 2} dN_i(t)\right]^{-1} \sum_{i=1}^{n} \int_0^{\tau} \frac{\mathrm{e}_i(t; \boldsymbol{\omega})}{\varphi_i(t; \hat{\boldsymbol{\gamma}})} \mathbf{S_i}(\mathbf{t})' Y_i(t) dN_i(t),$$

where the weights and the outcome models are based on the following estimating equations:

$$\mathbf{m_i}(\mathbf{o(t)};\hat{\phi}) = \mathbf{m_i}(\mathbf{o(t)};\hat{\boldsymbol{\omega}},\hat{\boldsymbol{\gamma}}_{\mathbf{aug}}) = \begin{bmatrix} \left(I_i(t) - \frac{\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i^\dagger(t)})}{1+\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i^\dagger(t)})}\right) K_{i1}^\dagger(t) \\ \left(I_i(t) - \frac{\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i^\dagger(t)})}{1+\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i^\dagger(t)})}\right) K_{i2}^\dagger(t) \\ \left(I_i(t) - \frac{\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i^\dagger(t)})}{1+\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i^\dagger(t)})}\right) K_{i3}^\dagger(t) \\ \left(I_i(t) - \frac{\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i^\dagger(t)})}{1+\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i^\dagger(t)})}\right) K_{i4}^\dagger(t) \\ \left(dN_i(t) - \frac{\exp(\hat{\boldsymbol{\gamma}}'_{\boldsymbol{aug}}\mathbf{V_i^\ddagger(t)})}{1+\exp(\hat{\boldsymbol{\gamma}}'_{\boldsymbol{aug}}\mathbf{V_i^\ddagger(t)})}\right) V_{i1}^\ddagger(t) \\ \left(dN_i(t) - \frac{\exp(\hat{\boldsymbol{\gamma}}'_{\boldsymbol{aug}}\mathbf{V_i^\ddagger(t)})}{1+\exp(\hat{\boldsymbol{\gamma}}'_{\boldsymbol{aug}}\mathbf{V_i^\ddagger(t)})}\right) V_{i2}^\ddagger(t) \\ \left(dN_i(t) - \frac{\exp(\hat{\boldsymbol{\gamma}}'_{\boldsymbol{aug}}\mathbf{V_i^\ddagger(t)})}{1+\exp(\hat{\boldsymbol{\gamma}}'_{\boldsymbol{aug}}\mathbf{V_i^\ddagger(t)})}\right) V_{i3}^\ddagger(t) \end{bmatrix},$$

$$\mathbf{g_i}(\mathbf{o(t)};\hat{\beta},\hat{\phi}) = \mathbf{g_i}(\mathbf{o(t)};\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\omega}},\hat{\boldsymbol{\gamma}}) = \begin{bmatrix} \frac{e_i(t;\hat{\boldsymbol{\omega}})}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\left(Y_i(t) - \hat{\boldsymbol{\beta}}'\mathbf{S_i(t)}\right) S_{i1}(t) \\ \frac{e_i(t;\hat{\boldsymbol{\omega}})}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\left(Y_i(t) - \hat{\boldsymbol{\beta}}'\mathbf{S_i(t)}\right) S_{i2}(t) \\ \frac{e_i(t;\hat{\boldsymbol{\omega}})}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\left(Y_i(t) - \hat{\boldsymbol{\beta}}'\mathbf{S_i(t)}\right) S_{i3}(t) \\ \frac{e_i(t;\hat{\boldsymbol{\omega}})}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\left(Y_i(t) - \hat{\boldsymbol{\beta}}'\mathbf{S_i(t)}\right) S_{i4}(t) \\ \frac{e_i(t;\hat{\boldsymbol{\omega}})}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\left(Y_i(t) - \hat{\boldsymbol{\beta}}'\mathbf{S_i(t)}\right) S_{i5}(t) \\ \frac{e_i(t;\hat{\boldsymbol{\omega}})}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\left(Y_i(t) - \hat{\boldsymbol{\beta}}'\mathbf{S_i(t)}\right) S_{i6}(t) \\ \frac{e_i(t;\hat{\boldsymbol{\omega}})}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\left(Y_i(t) - \hat{\boldsymbol{\beta}}'\mathbf{S_i(t)}\right) S_{i7}(t) \end{bmatrix}$$

for $\mathbf{K_i(t)}^\dagger$ and $\mathbf{V_i(t)}^\ddagger$ respectively the confounders and the monitoring predictors matrices augmented with a leading column of 1. The parameters $\boldsymbol{\gamma_{aug}}$ correspond to parameters $\boldsymbol{\gamma}$ augmented with a leading value that is the intercept of the monitoring model with design matrix $\mathbf{V_i(t)}^\ddagger$. Note that these parameters can be estimated using the `glm` function in R. More generally, we denote by $P_{il}(t)$ the value at time t and for individual i that is in the l^{th} column of the matrix P.

For the estimation of $\mathbf{G}_\beta = \mathbb{E}(\nabla_\beta \mathbf{g}(\mathbf{o}; \beta_\mathbf{0}, \phi_\mathbf{0}))$, we obtain

$$\hat{\mathbf{G}}_{\beta,\mathbf{n}} = \begin{bmatrix} -\frac{1}{n_v}\sum_{i=1}^{n}\int_0^\tau \frac{e_i(t;\hat{\boldsymbol{\omega}})}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\mathbf{S_i(t)}' S_{i1}(t) dN_i(t) \\ -\frac{1}{n_v}\sum_{i=1}^{n}\int_0^\tau \frac{e_i(t;\hat{\boldsymbol{\omega}})}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\mathbf{S_i(t)}' S_{i2}(t) dN_i(t) \\ -\frac{1}{n_v}\sum_{i=1}^{n}\int_0^\tau \frac{e_i(t;\hat{\boldsymbol{\omega}})}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\mathbf{S_i(t)}' S_{i3}(t) dN_i(t) \\ -\frac{1}{n_v}\sum_{i=1}^{n}\int_0^\tau \frac{e_i(t;\hat{\boldsymbol{\omega}})}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\mathbf{S_i(t)}' S_{i4}(t) dN_i(t) \\ -\frac{1}{n_v}\sum_{i=1}^{n}\int_0^\tau \frac{e_i(t;\hat{\boldsymbol{\omega}})}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\mathbf{S_i(t)}' S_{i5}(t) dN_i(t) \\ -\frac{1}{n_v}\sum_{i=1}^{n}\int_0^\tau \frac{e_i(t;\hat{\boldsymbol{\omega}})}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\mathbf{S_i(t)}' S_{i6}(t) dN_i(t) \\ -\frac{1}{n_v}\sum_{i=1}^{n}\int_0^\tau \frac{e_i(t;\hat{\boldsymbol{\omega}})}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\mathbf{S_i(t)}' S_{i7}(t) dN_i(t) \end{bmatrix}$$

for n_v the number of rows in the dataset restricted to $dN_i(t) = 1$. Note that, for example, $\hat{\mathbf{G}}_{\beta,\mathbf{n}}$ is a 7×7 matrix in the case where we include two knots for the basis of the cubic splines, an intercept, and the intervention in $\mathbf{S(t)}$.

The matrix $\mathbf{G}_\phi = \mathbb{E}(\nabla_\phi \mathbf{g}(\mathbf{o}; \beta_\mathbf{0}, \phi_\mathbf{0}))$ is estimated by

$$\hat{\mathbf{G}}_{\phi,\mathbf{n}} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix}$$

where

$$\mathbf{A} = \begin{bmatrix} \frac{1}{n_r}\sum_{i=1}^{n}\int_0^\tau \frac{1}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\left(Y_i(t) - \hat{\boldsymbol{\beta}}'\mathbf{S_i(t)}\right) S_{i1}(t) \left\{\mathbb{I}_{I_i(t)=0}\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i(t)}^\dagger) + \mathbb{I}_{I_i(t)=1}\frac{-1}{\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i(t)}^\dagger)}\right\}\mathbf{K_i(t)}^{\dagger'} \\ \frac{1}{n_r}\sum_{i=1}^{n}\int_0^\tau \frac{1}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\left(Y_i(t) - \hat{\boldsymbol{\beta}}'\mathbf{S_i(t)}\right) S_{i2}(t) \left\{\mathbb{I}_{I_i(t)=0}\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i(t)}^\dagger) + \mathbb{I}_{I_i(t)=1}\frac{-1}{\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i(t)}^\dagger)}\right\}\mathbf{K_i(t)}^{\dagger'} \\ \frac{1}{n_r}\sum_{i=1}^{n}\int_0^\tau \frac{1}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\left(Y_i(t) - \hat{\boldsymbol{\beta}}'\mathbf{S_i(t)}\right) S_{i3}(t) \left\{\mathbb{I}_{I_i(t)=0}\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i(t)}^\dagger) + \mathbb{I}_{I_i(t)=1}\frac{-1}{\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i(t)}^\dagger)}\right\}\mathbf{K_i(t)}^{\dagger'} \\ \frac{1}{n_r}\sum_{i=1}^{n}\int_0^\tau \frac{1}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\left(Y_i(t) - \hat{\boldsymbol{\beta}}'\mathbf{S_i(t)}\right) S_{i4}(t) \left\{\mathbb{I}_{I_i(t)=0}\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i(t)}^\dagger) + \mathbb{I}_{I_i(t)=1}\frac{-1}{\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i(t)}^\dagger)}\right\}\mathbf{K_i(t)}^{\dagger'} \\ \frac{1}{n_r}\sum_{i=1}^{n}\int_0^\tau \frac{1}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\left(Y_i(t) - \hat{\boldsymbol{\beta}}'\mathbf{S_i(t)}\right) S_{i5}(t) \left\{\mathbb{I}_{I_i(t)=0}\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i(t)}^\dagger) + \mathbb{I}_{I_i(t)=1}\frac{-1}{\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i(t)}^\dagger)}\right\}\mathbf{K_i(t)}^{\dagger'} \\ \frac{1}{n_r}\sum_{i=1}^{n}\int_0^\tau \frac{1}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\left(Y_i(t) - \hat{\boldsymbol{\beta}}'\mathbf{S_i(t)}\right) S_{i6}(t) \left\{\mathbb{I}_{I_i(t)=0}\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i(t)}^\dagger) + \mathbb{I}_{I_i(t)=1}\frac{-1}{\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i(t)}^\dagger)}\right\}\mathbf{K_i(t)}^{\dagger'} \\ \frac{1}{n_r}\sum_{i=1}^{n}\int_0^\tau \frac{1}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\left(Y_i(t) - \hat{\boldsymbol{\beta}}'\mathbf{S_i(t)}\right) S_{i7}(t) \left\{\mathbb{I}_{I_i(t)=0}\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i(t)}^\dagger) + \mathbb{I}_{I_i(t)=1}\frac{-1}{\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i(t)}^\dagger)}\right\}\mathbf{K_i(t)}^{\dagger'} \end{bmatrix}$$

and

$$\mathbf{B} = \begin{bmatrix} -\frac{1}{n_r}\sum_{i=1}^{n}\int_0^\tau \frac{e_i(t;\hat{\boldsymbol{\omega}})}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\left(Y_i(t) - \hat{\boldsymbol{\beta}}'\mathbf{S_i(t)}\right) S_{i1}(t)\mathbf{V_i(t)}^{\ddagger\prime} \\ -\frac{1}{n_r}\sum_{i=1}^{n}\int_0^\tau \frac{e_i(t;\hat{\boldsymbol{\omega}})}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\left(Y_i(t) - \hat{\boldsymbol{\beta}}'\mathbf{S_i(t)}\right) S_{i2}(t)\mathbf{V_i(t)}^{\ddagger\prime} \\ -\frac{1}{n_r}\sum_{i=1}^{n}\int_0^\tau \frac{e_i(t;\hat{\boldsymbol{\omega}})}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\left(Y_i(t) - \hat{\boldsymbol{\beta}}'\mathbf{S_i(t)}\right) S_{i3}(t)\mathbf{V_i(t)}^{\ddagger\prime} \\ -\frac{1}{n_r}\sum_{i=1}^{n}\int_0^\tau \frac{e_i(t;\hat{\boldsymbol{\omega}})}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\left(Y_i(t) - \hat{\boldsymbol{\beta}}'\mathbf{S_i(t)}\right) S_{i4}(t)\mathbf{V_i(t)}^{\ddagger\prime} \\ -\frac{1}{n_r}\sum_{i=1}^{n}\int_0^\tau \frac{e_i(t;\hat{\boldsymbol{\omega}})}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\left(Y_i(t) - \hat{\boldsymbol{\beta}}'\mathbf{S_i(t)}\right) S_{i5}(t)\mathbf{V_i(t)}^{\ddagger\prime} \\ -\frac{1}{n_r}\sum_{i=1}^{n}\int_0^\tau \frac{e_i(t;\hat{\boldsymbol{\omega}})}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\left(Y_i(t) - \hat{\boldsymbol{\beta}}'\mathbf{S_i(t)}\right) S_{i6}(t)\mathbf{V_i(t)}^{\ddagger\prime} \\ -\frac{1}{n_r}\sum_{i=1}^{n}\int_0^\tau \frac{e_i(t;\hat{\boldsymbol{\omega}})}{\varphi_i(t;\hat{\boldsymbol{\gamma}})}\left(Y_i(t) - \hat{\boldsymbol{\beta}}'\mathbf{S_i(t)}\right) S_{i7}(t)\mathbf{V_i(t)}^{\ddagger\prime} \end{bmatrix},$$

for n_r the number of rows in the full dataset (counting all individuals and all times when covariates are recorded, even when the outcome is not). The matrix $\hat{\mathbf{G}}_{\phi,\mathbf{n}}$ has a dimension 7×7 when the matrix $\mathbf{S(t)}$ contains the basis of cubic splines with two knots and when $\boldsymbol{\phi}$ is made of 7 parameters. In this situation, the matrix $\mathbf{A}$ above is 7×4 and the matrix $\mathbf{B}$ is 7×3.

Finally, estimate $\mathbf{M} = \mathbb{E}(\nabla_\phi \mathbf{m}(\mathbf{o}; \phi_\mathbf{0}))$ using the blocks

$$\hat{\mathbf{M}}_n = \begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix}$$

where

$$\mathbf{C} = \begin{bmatrix} \frac{1}{n_r}\sum_{i=1}^{n}\int_0^\tau \frac{-\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i(t)}^\dagger)}{(1+\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i(t)}^\dagger))^2} K_{i1}(t)\mathbf{K_i(t)}^{\dagger\prime} \\ \frac{1}{n_r}\sum_{i=1}^{n}\int_0^\tau \frac{-\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i(t)}^\dagger)}{(1+\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i(t)}^\dagger))^2} K_{i2}(t)\mathbf{K_i(t)}^{\dagger\prime} \\ \frac{1}{n_r}\sum_{i=1}^{n}\int_0^\tau \frac{-\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i(t)}^\dagger)}{(1+\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i(t)}^\dagger))^2} K_{i3}(t)\mathbf{K_i(t)}^{\dagger\prime} \\ \frac{1}{n_r}\sum_{i=1}^{n}\int_0^\tau \frac{-\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i(t)}^\dagger)}{(1+\exp(\hat{\boldsymbol{\omega}}'\mathbf{K_i(t)}^\dagger))^2} K_{i4}(t)\mathbf{K_i(t)}^{\dagger\prime} \end{bmatrix},$$

and

$$\mathbf{D} = \begin{bmatrix} \frac{1}{n_r}\sum_{i=1}^{n}\int_0^\tau \frac{-\exp(\hat{\boldsymbol{\gamma}}'_{\boldsymbol{aug}}\mathbf{V_i}(\mathbf{t})^\ddagger)}{(1+\exp(\hat{\boldsymbol{\gamma}}'_{\boldsymbol{aug}}\mathbf{V_i}(\mathbf{t})^\ddagger))^2} V_{i1}(t)^\ddagger \mathbf{V_i}(\mathbf{t})^{\ddagger'} \\ \frac{1}{n_r}\sum_{i=1}^{n}\int_0^\tau \frac{-\exp(\hat{\boldsymbol{\gamma}}'_{\boldsymbol{aug}}\mathbf{V_i}(\mathbf{t})^\ddagger)}{(1+\exp(\hat{\boldsymbol{\gamma}}'_{\boldsymbol{aug}}\mathbf{V_i}(\mathbf{t})^\ddagger))^2} V_{i2}(t)^\ddagger \mathbf{V_i}(\mathbf{t})^{\ddagger'} \\ \frac{1}{n_r}\sum_{i=1}^{n}\int_0^\tau \frac{-\exp(\hat{\boldsymbol{\gamma}}'_{\boldsymbol{aug}}\mathbf{V_i}(\mathbf{t})^\ddagger)}{(1+\exp(\hat{\boldsymbol{\gamma}}'_{\boldsymbol{aug}}\mathbf{V_i}(\mathbf{t})^\ddagger))^2} V_{i3}(t)^\ddagger \mathbf{V_i}(\mathbf{t})^{\ddagger'} \end{bmatrix}.$$

A.4 Summary statistics for the simulation studies

Table A.1: Study with confounding and covariate-dependent monitoring times ($\tau = 2$, $n = 250$)

Treatment & confounders	$\alpha(t)$	(γ_1,γ_2)	Median no. visits (IQR)	Absolute bias				
				$\hat{\beta}_{OLS}$†	$\hat{\beta}_{VW}$‡	$\hat{\beta}_{IPT}$⋆	$\hat{\beta}_{FIPTM}$	$\hat{\beta}_{IPCTM}$
Time-fixed	3	$(0, 0)$	1 (1-2)	0.72	0.71	0.06	0.09	0.08
		$(-0.3, 0.2)$	2 (1-3)	1.05	0.72	1.77	0.04	0.01
		$(0.6, 0.3)$	5 (4-7)	1.98	0.76	2.65	0.00	0.02
	$2.5t$	$(0, 0)$	1 (1-2)	0.73	0.74	0.07	0.08	0.10
		$(-0.3, 0.2)$	2 (1-3)	1.07	0.75	1.77	0.03	0.01
		$(0.6, 0.3)$	5 (4-7)	1.97	0.76	2.64	0.04	0.03
	$\sin(t)$	$(0, 0)$	1 (1-2)	0.74	0.74	0.02	0.07	0.06
		$(-0.3, 0.2)$	2 (1-3)	1.08	0.71	1.75	0.04	0.04
		$(0.6, 0.3)$	5 (4-7)	1.98	0.77	2.53	0.01	0.00
	$\sqrt{t}$	$(0, 0)$	1 (1-2)	0.72	0.73	0.03	0.09	0.07
		$(-0.3, 0.2)$	2 (1-3)	1.06	0.74	1.71	0.01	0.02
		$(0.6, 0.3)$	5 (4-7)	1.96	0.78	2.62	0.01	0.01
	$\exp(2\lvert\sin(3t)\rvert)$	$(0, 0)$	1 (1-2)	0.72	0.73	0.01	0.09	0.11
		$(-0.3, 0.2)$	2 (1-3)	1.06	0.74	1.75	0.01	0.01
		$(0.6, 0.3)$	5 (4-7)	1.96	0.78	2.68	0.01	0.01
Time-varying	3	$(0, 0)$	1 (1-2)	0.54	0.52	0.04	0.06	0.02
		$(-0.3, 0.2)$	2 (1-3)	2.91	0.49	1.77	0.03	0.02
		$(0.6, 0.3)$	5 (4-7)	4.10	0.42	2.66	0.03	0.08
	$2.5t$	$(0, 0)$	1 (1-2)	0.49	0.48	0.09	0.01	0.07
		$(-0.3, 0.2)$	2 (1-3)	2.88	0.46	1.75	0.01	0.06
		$(0.6, 0.3)$	5 (4-7)	4.10	0.45	2.65	0.00	0.09
	$\sin(t)$	$(0, 0)$	1 (1-2)	0.52	0.51	0.00	0.05	0.11
		$(-0.3, 0.2)$	2 (1-3)	2.90	0.50	1.77	0.05	0.06
		$(0.6, 0.3)$	5 (4-7)	4.08	0.43	2.64	0.05	0.07
	$\sqrt{t}$	$(0, 0)$	1 (1-2)	0.52	0.51	0.04	0.06	0.10
		$(-0.3, 0.2)$	2 (1-3)	2.90	0.47	1.73	0.02	0.05
		$(0.6, 0.3)$	5 (4-7)	4.01	0.43	2.65	0.01	0.04
	$\exp(2\lvert\sin(3t)\rvert)$	$(0, 0)$	1 (1-2)	0.55	0.51	0.06	0.04	0.05
		$(-0.3, 0.2)$	2 (1-3)	2.91	0.48	1.77	0.03	0.04
		$(0.6, 0.3)$	5 (4-7)	4.11	0.43	2.65	0.03	0.06

† Ordinary least squares regression with outcome $Y_i(t)$ and exposure $I_i(t)$ with a constant intercept
‡ Weighted least squares regression with outcome $Y_i(t)$ and exposure $I_i(t)$ with a constant intercept and an inverse probability of monitoring weight computed from a proportional intensity model with $I_i(t)$ and $Z_i(t)$ as predictors
⋆ Weighted least squares regression with outcome $Y_i(t)$ and exposure $I_i(t)$ with a constant intercept and one an inverse probability of treatment weight computed from a logistic regression model with $\mathbf{K_i(t)}$ as predictors

Table A.2: Study with confounding and covariate-dependent monitoring times ($\tau = 2$, $n = 500$)

Treatment & confounders	$\alpha(t)$	(γ_1,γ_2)	Median no. visits (IQR)	Absolute bias				
				$\hat{\beta}_{OLS}$†	$\hat{\beta}_{VW}$‡	$\hat{\beta}_{IPT}$⋆	$\hat{\beta}_{FIPTM}$	$\hat{\beta}_{IPCTM}$
Time-fixed	3	$(0, 0)$	1 (1-2)	0.74	0.73	0.01	0.05	0.11
		$(-0.3, 0.2)$	2 (1-3)	1.08	0.74	1.78	0.02	0.03
		$(0.6, 0.3)$	5 (4-7)	1.97	0.78	2.66	0.04	0.05
	$2.5t$	$(0, 0)$	1 (1-2)	0.69	0.69	0.04	0.05	0.06
		$(-0.3, 0.2)$	2 (1-3)	1.06	0.75	1.78	0.01	0.02
		$(0.6, 0.3)$	5 (4-7)	1.96	0.79	2.67	0.05	0.02
	$\sin(t)$	$(0, 0)$	1 (1-2)	0.72	0.72	0.00	0.05	0.06
		$(-0.3, 0.2)$	2 (1-3)	1.07	0.73	1.78	0.01	0.02
		$(0.6, 0.3)$	5 (4-7)	1.96	0.79	2.66	0.06	0.02
	$\sqrt{t}$	$(0, 0)$	1 (1-2)	0.73	0.72	0.01	0.04	0.07
		$(-0.3, 0.2)$	2 (1-3)	1.07	0.74	1.73	0.00	0.01
		$(0.6, 0.3)$	5 (4-7)	1.96	0.80	2.67	0.05	0.03
	$\exp(2\lvert\sin(3t)\rvert)$	$(0, 0)$	1 (1-2)	0.71	0.71	0.02	0.03	0.11
		$(-0.3, 0.2)$	2 (1-3)	1.05	0.75	1.78	0.00	0.00
		$(0.6, 0.3)$	5 (4-7)	1.95	0.80	2.67	0.07	0.03
Time-varying	3	$(0, 0)$	1 (1-2)	0.49	0.48	0.03	0.03	0.07
		$(-0.3, 0.2)$	2 (1-3)	2.90	0.48	1.76	0.03	0.00
		$(0.6, 0.3)$	5 (4-7)	4.10	0.41	2.67	0.06	0.05
	$2.5t$	$(0, 0)$	1 (1-2)	0.49	0.49	0.06	0.03	0.12
		$(-0.3, 0.2)$	2 (1-3)	2.89	0.47	1.73	0.01	0.04
		$(0.6, 0.3)$	5 (4-7)	4.08	0.42	2.63	0.04	0.06
	$\sin(t)$	$(0, 0)$	1 (1-2)	0.49	0.50	0.00	0.01	0.06
		$(-0.3, 0.2)$	2 (1-3)	2.88	0.47	1.79	0.00	0.01
		$(0.6, 0.3)$	5 (4-7)	4.09	0.40	2.69	0.05	0.04
	$\sqrt{t}$	$(0, 0)$	1 (1-2)	0.47	0.49	0.06	0.03	0.05
		$(-0.3, 0.2)$	2 (1-3)	2.89	0.46	1.75	0.01	0.02
		$(0.6, 0.3)$	5 (4-7)	4.12	0.44	2.68	0.03	0.05
	$\exp(2\lvert\sin(3t)\rvert)$	$(0, 0)$	1 (1-2)	0.49	0.49	0.05	0.03	0.09
		$(-0.3, 0.2)$	2 (1-3)	2.92	0.49	1.74	0.03	0.03
		$(0.6, 0.3)$	5 (4-7)	4.10	0.41	2.66	0.06	0.04

† Ordinary least squares regression with outcome $Y_i(t)$ and exposure $I_i(t)$ with a constant intercept
‡ Weighted least squares regression with outcome $Y_i(t)$ and exposure $I_i(t)$ with a constant intercept and an inverse probability of monitoring weight computed from a proportional intensity model with $I_i(t)$ and $Z_i(t)$ as predictors
⋆ Weighted least squares regression with outcome $Y_i(t)$ and exposure $I_i(t)$ with a constant intercept and one an inverse probability of treatment weight computed from a logistic regression model with $\mathbf{K_i(t)}$ as predictors

Table A.3: Empirical variance, relative efficiency and coverage of $\hat{\beta}_{FIPTM}$ in simulation studies with τ=2, $n = 250$ and $I(t)$ and $\mathbf{K(t)}$ time-fixed

$\alpha(t)$	(γ_1,γ_2)	Empirical variance	RE† w.r.t. empirical variance		Coverage‡		RE† of empirical var$(\hat{\beta}_{FIPTM})$ w.r.t. var$(\hat{\beta}_{IPCTM})$
			Asymptotic	Bootstrapped	Asymptotic	Bootstrapped	
3	$(0, 0)$	0.766	0.83	0.99	92.3%	95.3%	0.77
	$(-0.3, 0.2)$	0.393	1.04	1.06	93.1%	95.4%	0.89
	$(0.6, 0.3)$	0.376	0.90	0.95	95.9%	95.5%	0.81
$2.5t$	$(0, 0)$	0.719	0.90	1.11	92.6%	93.8%	0.88
	$(-0.3, 0.2)$	0.418	0.95	1.03	96.1%	95.8%	0.90
	$(0.6, 0.3)$	0.343	0.98	1.15	94.4%	95.3%	0.81
$\sin(t)$	$(0, 0)$	0.811	0.77	0.92	93.9%	96.0%	1.02
	$(-0.3, 0.2)$	0.409	1.00	1.16	94.3%	96.3%	0.80
	$(0.6, 0.3)$	0.356	0.97	1.04	93.7%	96.2%	0.76
$\sqrt{t}$	$(0, 0)$	0.739	0.82	1.06	92.8%	96.3%	1.01
	$(-0.3, 0.2)$	0.411	0.99	1.02	94.6%	95.1%	0.81
	$(0.6, 0.3)$	0.348	0.95	1.01	96.4%	95.8%	0.71
exp(2abs(sin(3t)))	$(0, 0)$	0.717	0.88	1.14	93.4%	93.2%	0.79
	$(-0.3, 0.2)$	0.397	1.06	1.11	94.7%	94.9%	0.78
	$(0.6, 0.3)$	0.372	0.95	1.14	95.7%	94.6%	0.83

† RE stands for relative efficiency, the ratio of variances

‡ Coverage is computed using a Normal approximation

Table A.4: Sensitivity analysis 1: Summary statistics for the simulation study with time-fixed confounders and intervention ($\tau = 2$, $n = 250$) when all 3 confounders are generated from one multivariate Normal distribution

$\alpha(t)$	(γ_1,γ_2)	Absolute bias					Mean Squared Error				
		$\hat{\beta}_{OLS}{}^{\dagger}$	$\hat{\beta}_{VW}{}^{\ddagger}$	$\hat{\beta}_{IPT}{}^{\star}$	$\hat{\beta}_{FIPTM}$	$\hat{\beta}_{IPCTM}$	$\hat{\beta}_{OLS}{}^{\dagger}$	$\hat{\beta}_{VW}{}^{\ddagger}$	$\hat{\beta}_{IPT}{}^{\star}$	$\hat{\beta}_{FIPTM}$	$\hat{\beta}_{IPCTM}$
$2.5t$	$(0, 0)$	0.97	0.98	0.13	0.17	0.18	1.35	1.27	1.09	0.91	1.01
	$(-0.3, 0.2)$	0.81	0.98	1.72	0.07	0.03	0.85	1.17	3.45	0.52	0.52
	$(0.6, 0.3)$	1.72	1.03	2.63	0.14	0.03	3.06	1.22	7.27	0.45	0.50
$\sqrt{t}$	$(0, 0)$	1.02	1.01	0.11	0.14	0.17	1.44	1.29	1.17	0.82	0.98
	$(-0.3, 0.2)$	0.84	0.96	1.73	0.04	0.03	0.91	1.11	3.53	0.49	0.58
	$(0.6, 0.3)$	1.71	1.03	2.62	0.15	0.04	3.03	1.23	7.21	0.44	0.56
$\exp(2\lvert\sin(3t)\rvert)$	$(0, 0)$	0.93	0.95	0.08	0.10	0.21	1.27	1.23	1.13	0.85	1.11
	$(-0.3, 0.2)$	0.82	0.98	1.69	0.07	0.07	0.89	1.19	3.48	0.55	0.60
	$(0.6, 0.3)$	1.72	1.03	2.60	0.12	0.04	3.06	1.24	7.15	0.46	0.51

† Ordinary least squares regression with outcome $Y_i(t)$ and exposure $I_i(t)$ with a constant intercept

‡ Weighted least squares regression with outcome $Y_i(t)$ and exposure $I_i(t)$ with a constant intercept and an inverse probability of monitoring weight computed from a proportional intensity model with $I_i(t)$ and $Z_i(t)$ as predictors

⋆ Weighted least squares regression with outcome $Y_i(t)$ and exposure $I_i(t)$ with a constant intercept and one an inverse probability of treatment weight computed from a logistic regression model with $\mathbf{K_i(t)}$ as predictors

Table A.5: Sensitivity analysis 2: Summary statistics for the simulation study with time-fixed confounders and intervention ($\tau = 2$, $n = 250$) and with parameters $\boldsymbol{\gamma_V} = (\gamma_1, \gamma_2, -0.4, 0.2, 0.2)$ in the proportional rate model, with the first two parameters pertaining to $\mathbf{I}$ and $\mathbf{Z(t)}$, respectively, and the last 3 parameters to $\mathbf{K}$

$\alpha(t)$	(γ_1,γ_2)	Absolute bias					Mean Squared Error				
		$\hat{\beta}_{OLS}{}^{\dagger}$	$\hat{\beta}_{VW}{}^{\ddagger}$	$\hat{\beta}_{IPT}{}^{\star}$	$\hat{\beta}_{FIPTM}$	$\hat{\beta}_{IPCTM}$	$\hat{\beta}_{OLS}{}^{\dagger}$	$\hat{\beta}_{VW}{}^{\ddagger}$	$\hat{\beta}_{IPT}{}^{\star}$	$\hat{\beta}_{FIPTM}$	$\hat{\beta}_{IPCTM}$
$2.5t$	(0,0)	0.65	0.66	0.03	0.06	0.13	0.81	0.73	0.88	1.21	1.78
	(-0.3,0.2)	1.08	0.74	1.76	0.02	0.05	1.37	0.76	3.51	0.72	1.01
	(0.6,0.3)	1.97	0.79	2.67	0.10	0.03	3.99	0.80	7.38	0.62	1.03
$\sqrt{t}$	(0,0)	0.72	0.71	0.06	0.13	0.19	0.92	0.77	0.83	1.14	1.61
	(-0.3,0.2)	1.10	0.70	1.79	0.01	0.04	1.40	0.69	3.59	0.74	0.89
	(0.6,0.3)	2.00	0.76	2.67	0.07	0.03	4.09	0.75	7.36	0.61	1.01
$\exp(2\lvert\sin(3t)\rvert)$	(0,0)	0.70	0.71	0.06	0.08	0.25	0.94	0.84	0.90	1.15	1.90
	(-0.3,0.2)	1.09	0.74	1.75	0.01	0.02	1.40	0.77	3.47	0.70	0.86
	(0.6,0.3)	1.98	0.78	2.67	0.10	0.02	4.03	0.80	7.38	0.62	1.19

† Ordinary least squares regression with outcome $Y_i(t)$ and exposure $I_i(t)$ with a constant intercept

‡ Weighted least squares regression with outcome $Y_i(t)$ and exposure $I_i(t)$ with a constant intercept and an inverse probability of monitoring weight computed from a proportional intensity model with $I_i(t)$ and $Z_i(t)$
as predictors

⋆ Weighted least squares regression with outcome $Y_i(t)$ and exposure $I_i(t)$ with a constant intercept and one an inverse probability of treatment weight computed from a logistic regression model with $\mathbf{K_i(t)}$ as predictors

Table A.6: Sensitivity analyses 3 and 4: Summary statistics for the simulation study with a modified weight for the IPCTM estimator, whose weight is computed using a cumulative logistic model for the quantiles of the distribution of the re-centered treatment variable ($\tau = 2$, $n = 250$). Time-fixed treatment and confounders.

$\alpha(t)$	(γ_1, γ_2)	Median no. visits (IQR)	Absolute bias				Mean squared error			
			$\hat{\beta}_{FIPTM}$	$\hat{\beta}_{IPCTM}$	$\hat{\beta}_{IPCTM10}$†	$\hat{\beta}_{IPCTM20}$‡	$\hat{\beta}_{FIPTM}$	$\hat{\beta}_{IPCTM}$	$\hat{\beta}_{IPCTM10}$†	$\hat{\beta}_{IPCTM20}$‡
3	$(0, 0)$	1 (1-2)	0.09	0.08	0.63	0.63	0.77	1.00	0.84	0.83
	$(-0.3, 0.2)$	2 (1-3)	0.04	0.01	0.21	0.21	0.39	0.44	0.28	0.28
	$(0.6, 0.3)$	5 (4-7)	0.00	0.02	0.02	0.01	0.38	0.47	0.25	0.25
$2.5t$	$(0, 0)$	1 (1-2)	0.08	0.10	0.59	0.58	0.73	0.83	0.74	0.74
	$(-0.3, 0.2)$	2 (1-3)	0.03	0.01	0.19	0.18	0.42	0.46	0.26	0.26
	$(0.6, 0.3)$	5 (4-7)	0.04	0.03	0.04	0.03	0.34	0.43	0.24	0.24
$\sin(t)$	$(0, 0)$	1 (1-2)	0.07	0.06	0.49	0.48	0.82	0.80	0.61	0.61
	$(-0.3, 0.2)$	2 (1-3)	0.04	0.04	0.17	0.16	0.41	0.51	0.29	0.29
	$(0.6, 0.3)$	5 (4-7)	0.01	0.00	0.02	0.02	0.36	0.47	0.23	0.23
$\sqrt{t}$	$(0, 0)$	1 (1-2)	0.09	0.07	0.50	0.49	0.75	0.74	0.63	0.63
	$(-0.3, 0.2)$	2 (1-3)	0.01	0.02	0.15	0.14	0.41	0.51	0.27	0.27
	$(0.6, 0.3)$	5 (4-7)	0.01	0.01	0.01	0.01	0.35	0.49	0.25	0.25
$\exp(2\lvert\sin(3t)\rvert)$	$(0, 0)$	1 (1-2)	0.01	0.11	0.70	0.69	0.72	0.92	0.95	0.94
	$(-0.3, 0.2)$	2 (1-3)	0.01	0.01	0.24	0.23	0.40	0.51	0.31	0.31
	$(0.6, 0.3)$	5 (4-7)	0.03	0.01	0.00	0.01	0.37	0.45	0.26	0.26

† IPCTM estimator with the IPT weight computed from a cumulative logistic model in which the outcome is the re-centered treatment variable binned into 10 quantiles
‡ IPCTM estimator with the IPT weight computed from a cumulative logistic model in which the outcome is the re-centered treatment variable binned into 20 quantiles

Table A.7: Sensitivity analyses 5, 6, 7 and 8: Summary statistics for the simulation studies for testing the effect of model misspecification, with time-fixed confounders and intervention ($\tau = 2$, $n = 250$)

Model misspecification	$\alpha(t)$	(γ_1,γ_2)	Absolute bias					Mean Squared Error				
			$\hat{\beta}_{OLS}{}^{\dagger}$	$\hat{\beta}_{VW}{}^{\ddagger}$	$\hat{\beta}_{IPT}{}^{\star}$	$\hat{\beta}_{FIPTM}$	$\hat{\beta}_{IPCTM}$	$\hat{\beta}_{OLS}{}^{\dagger}$	$\hat{\beta}_{VW}{}^{\ddagger}$	$\hat{\beta}_{IPT}{}^{\star}$	$\hat{\beta}_{FIPTM}$	$\hat{\beta}_{IPCTM}$
Log-normal errors in the outcome model[a]	$\sqrt{t}$	(0,0)	0.72	0.72	0.02	0.06	0.06	0.93	0.84	0.96	0.74	0.88
		(-0.3,0.2)	1.06	0.74	1.76	0.02	0.01	1.33	0.73	3.50	0.41	0.46
		(0.6,0.3)	1.98	0.76	2.70	0.07	0.01	4.06	0.75	7.55	0.39	0.48
Non-linear terms for covariates in the outcome model[b]	$\sqrt{t}$	(0,0)	0.26	0.27	0.03	0.01	0.01	0.50	0.41	0.94	0.77	0.81
		(-0.3,0.2)	1.56	0.23	1.82	0.07	0.08	2.64	0.25	3.81	0.47	0.71
		(0.6,0.3)	2.44	0.30	2.71	0.02	0.07	6.09	0.27	7.64	0.40	0.65
Non-linear terms for covariates in the visit model[c]	$\sqrt{t}$	(0,0)	0.69	0.71	0.00	0.05	0.05	0.88	0.79	0.90	0.72	0.81
		(-0.3,0.2)	0.69	0.59	1.39	10.05	1.22	0.50	1.34	1.99	115.83	2.86
		(0.6,0.3)	0.54	2.43	0.16	11.50	1.70	0.31	6.35	0.06	147.14	4.43
Varying $d\Delta_0(t)$ across individuals[d]	$\sqrt{t}$	(0,0)	0.76	0.75	0.09	0.10	0.12	1.04	0.89	1.13	0.86	1.01
		(-0.3,0.2)	1.11	0.70	1.77	0.02	0.01	1.41	0.69	3.56	0.41	0.45
		(0.6,0.3)	1.97	0.76	2.64	0.09	0.02	4.01	0.78	7.26	0.37	0.47

† Ordinary least squares regression with outcome $Y_i(t)$ and exposure $I_i(t)$ with a constant intercept

‡ Weighted least squares regression with outcome $Y_i(t)$ and exposure $I_i(t)$ with a constant intercept and an inverse probability of monitoring weight computed from a proportional intensity model with $I_i(t)$ and $Z_i(t)$ as predictors

⋆ Weighted least squares regression with outcome $Y_i(t)$ and exposure $I_i(t)$ with a constant intercept and one an inverse probability of treatment weight computed from a logistic regression model with $\mathbf{K_i(t)}$ as predictors

a Same simulation scenario as presented in Section 3.3, but with $\epsilon_i^{\dagger}(t) \sim \text{Log-Normal}(\phi_i, 0.01)$, $\phi_i \sim \text{N}(0, 0.04)$ and $\epsilon_i(t) = \epsilon_i^{\dagger}(t) - \overline{\epsilon^{\dagger}(t)}$

b Same simulation scenario as presented in Section 3.3, but with $Y_i(t) = \alpha(t) + 1\, I_i + 3\, [Z_i(t) - E\,[Z_i(t)|I_i]] + 0.4\, \sin(K_{1i}) + 0.05\, K_{2i} - 0.6\, (K_{3i})^2 + \epsilon_i(t)$

c Same simulation scenario as presented in Section 3.3, but with $\lambda_i(t|I_i, Z_i(t)) = \eta_i \exp\left(\gamma_1 I_i + \gamma_2 Z_i^2(t)\right)$

d Same simulation scenario as presented in Section 3.3, but with $d\Delta_0(t)$ drawn in $\{1; 1.5t; \sin t\}$ for each patient, with respective probabilities 0.5, 0.25, 0.25.

Table A.8: Sensitivity analysis 9: Summary statistics for the simulation studies where we compare the main models (top) to models where we incorporated/conditioned on all confounders as covariates (bottom part) (time-fixed treatment and confounders, $\tau = 2$, $n = 250$)

Models cond. on confounders	$\alpha(t)$	(γ_1,γ_2)	Median no. visits (IQR)	Absolute bias					Empirical variance				
				$\hat{\beta}_{OLS}{}^{\dagger}$	$\hat{\beta}_{VW}{}^{\ddagger}$	$\hat{\beta}_{IPT}{}^{\star}$	$\hat{\beta}_{FIPTM}$	$\hat{\beta}_{IPCTM}$	$\hat{\beta}_{OLS}{}^{\dagger}$	$\hat{\beta}_{VW}{}^{\ddagger}$	$\hat{\beta}_{IPT}{}^{\star}$	$\hat{\beta}_{FIPTM}$	$\hat{\beta}_{IPCTM}$
No	3	$(0, 0)$	1 (1-2)	0.72	0.71	0.06	0.09	0.08	0.41	0.30	1.06	0.77	0.99
		$(-0.3, 0.2)$	2 (1-3)	1.05	0.72	1.77	0.04	0.01	0.19	0.18	0.40	0.39	0.44
		$(0.6, 0.3)$	5 (4-7)	1.98	0.76	2.65	0.00	0.02	0.12	0.19	0.30	0.38	0.47
	$\sin(t)$	$(0, 0)$	1 (1-2)	0.74	0.74	0.02	0.07	0.06	0.40	0.29	0.92	0.81	0.80
		$(-0.3, 0.2)$	2 (1-3)	1.08	0.71	1.75	0.04	0.04	0.20	0.20	0.40	0.41	0.51
		$(0.6, 0.3)$	5 (4-7)	1.98	0.77	2.53	0.01	0.00	0.10	0.17	0.27	0.36	0.47
Yes	3	$(0, 0)$	1 (1-2)	0.04	0.03	0.04	0.04	0.09	0.47	0.36	0.80	0.67	1.06
		$(-0.3, 0.2)$	2 (1-3)	1.81	0.01	1.77	0.08	0.01	0.23	0.24	0.34	0.42	0.65
		$(0.6, 0.3)$	5 (4-7)	2.69	0.04	2.71	0.06	0.00	0.12	0.18	0.21	0.29	0.60
	$\sin(t)$	$(0, 0)$	1 (1-2)	0.01	0.00	0.02	0.02	0.10	0.45	0.35	0.82	0.70	1.07
		$(-0.3, 0.2)$	2 (1-3)	1.79	0.01	1.83	0.07	0.04	0.21	0.22	0.36	0.41	0.62
		$(0.6, 0.3)$	5 (4-7)	2.70	0.03	2.65	0.07	0.02	0.12	0.19	0.21	0.31	0.56

† Ordinary least squares regression with outcome $Y_i(t)$ and exposure $I_i(t)$ with a constant intercept

‡ Weighted least squares regression with outcome $Y_i(t)$ and exposure $I_i(t)$ with a constant intercept and an inverse probability of monitoring weight computed from a proportional intensity model with $I_i(t)$ and $Z_i(t)$
as predictors

⋆ Weighted least squares regression with outcome $Y_i(t)$ and exposure $I_i(t)$ with a constant intercept and one an inverse probability of treatment weight computed from a logistic regression model with $\mathbf{K_i(t)}$ as predictors

A.5 Details of the second simulation study with time-varying confounders and treatment

In a second simulation study, $I_i(t)$ and the confounders $\{K_{1i}(t), K_{2i}(t), K_{3i}(t)\}$ could vary in time. All confounders and the intervention at baseline ($t = 0$) were simulated as follows:

$$\begin{aligned} K_{1i}(0) &\sim \text{N}(3, 1), \\ K_{2i}(0) &\sim \text{Bernoulli}(0.55), \\ K_{3i}(0) &\sim \text{N}(-1.2, 1), \\ I_i(0) &\sim \text{Bernoulli}(p_{Ii}(0)), \end{aligned}$$

with $p_{Ii}(0) = \text{expit}\left(0.5 + 0.1K_{1i}(0) + 0.05K_{2i}(0) - 1K_{3i}(0)\right)$. The mediator at baseline was simulated as $Z_i(0)|I_i(0) = 1 \sim \text{N}(2, 1)$, $Z_i(0)|I_i(0) = 0 \sim \text{N}(4, 4)$. Then, the same variables were simulated at time t for $0 < t < \tau$, with

$$\begin{aligned} K_{1i}(t) &= K_{1i}(t-1) + 0.01, \\ K_{2i}(t) &= K_{2i}(t-1), \\ K_{3i}(t) &\sim N(K_{3i}(t-1), 0.05), \\ I_i(t) &\sim \text{Bernoulli}(p_{Ii}(t)), \end{aligned}$$

with $p_{Ii}(t) = \text{expit}\left(0.5 + 0.1K_{1i}(t) + 0.05K_{2i}(t) - 1K_{3i}(t) - 1.5I_i(t-1)\right)$. For the mediator, we used $Z_i(t)|I_i(t) = 1 \sim \text{N}(2, 1)$ and $Z_i(t)|I_i(t) = 0 \sim \text{N}(4, 4)$. For all time $t > 0$, the outcome was simulated such that $Y_i(t) = \mu_i + \alpha(t) + 2I_i(t) - 4\left[Z_i(t) - E\left[Z_i(t)|I_i(t)\right]\right] + 0.4K_{1i}(t) + 0.05K_{2i}(t) - 0.6K_{3i}(t) + \epsilon_i(t)$ with $\mu_i \sim N(0, 1.8)$, $\epsilon_i(t) \sim \text{N}(\phi_i, 0.01)$ and $\phi_i \sim \text{N}(0, 0.04)$.

In this second simulation study, the outcomes of the same individual i were correlated over time. Assignment to exposure was also dependent in time. However, no previous exposure predicted current confounders. The quantities above were once again simulated in continuous time, and then discretized over a grid of 0.01 units. Monitoring times were simulated according to a nonhomogeneous Poisson process, as in simulation study 1, with the same parameters in the intensity model but now accounting for the fact that $I_i(t)$ varies in time. Again, τ was set to 2, and the censoring time was simulated in the same way as in the time-fixed treatment study. We also tested the same combinations of (γ_1, γ_2) parameters and the same five functions for $\alpha(t)$ as in the time-fixed treatment study, with two different sample sizes of $n = 250$ and $n = 500$, and a total of 1000

simulations. Once again, our estimators were compared with the estimators $\widehat{\beta}_{OLS}$, a weighted least squares estimator in which a correctly-specified time-dependent monitoring weight was incorporated ($\widehat{\beta}_{VW}$), and a weighted least squares estimator that incorporated a correctly specified inverse probability of treatment weight ($\widehat{\beta}_{IPT}$). For the estimators $\widehat{\beta}_{IPCTM}$ and $\widehat{\beta}_{FIPTM}$, the model for the treatment and the model for the monitoring weight were both correctly specified.

A.6 Brief summary of the results for the sensitivity analyses

The results (distributions of biases and MSEs) for all 9 sensitivity analyses can be found in Tables A.4 (i), A.5 (ii), A.6 (iii and iv), A.7 (v, vi, vii, viii), and A.8 (ix) in Appendix A.4. In the first two scenarios, we observed similar results, with absolute biases that tended towards 0 for both proposed estimators. The empirical mean squared errors are typically smaller for the FIPTM estimator as compared to all four other estimators. In scenarios iii) and iv), we observe that using a different generalized IPT weight (computed from a cumulative logistic regression) in the IPCTM estimator sometimes provides better results, however this is not consistent across scenarios. Indeed, there is a bias-variance trade-off that occurs when using linear regression versus categorical regression for quantile-binned treatment models used to compute the (generalized) IPT weight. Indeed, the bias of the IPCTM estimator using linear regression for the treatment model was typically smaller than the one of the quantile-binned approaches, except when the number of visits per patient increased, in which case the bias was similar for all three modelling approaches (see Table A.6). In terms of mean squared error, on the other hand, the quantile-binning approach provided smaller values than the IPCTM estimator using linear regression for the treatment model, and even out-performed the FIPTM estimator. For the four sensitivity analyses (v, vi, vii, viii) where we assessed the behavior of our estimators in settings where different models are misspecified, we first note in Table A.5 that our methods were not too sensitive to misspecification of the outcome model. In particular, simulating log-Normal errors in the outcome model or generating an outcome that depended on non-linear functions of the confounder variables had little impact on the performance our two proposed estimators. On the other hand, we found important bias in our proposed estimators, especially the FIPTM, when generating data such that the visit process depended on non-linear functions of the covariates. In that seventh analysis, the proportional intensity model for the visits depended on $Z^2(t)$ rather than $Z(t)$, hence the coefficient for $Z(t)$ was greatly inflated and the inverse probability of visit weight used in the FIPTM estimator exploded in such setting, which strongly biased the FIPTM estimator. The IPCTM estimator was less affected by the non-linear form of $Z(t)$ in the

visit model, which could be due to the way the inverse probability of visit weight is used in that estimator. Instead of directly re-weighting the weighted least squares formula with that weight, the IPCTM uses this weight for re-centering the variables, which could make it less sensitive to misspecification of the function of each term in the weight model. Moreover, our proposed estimators were not much affected by variation of the baseline intensity function $d\Delta_0(t)$ across individuals in the sensitivity analysis (viii) that we performed. Finally, the last sensitivity analysis has shown that in the current context where the mean outcome model is linear in the covariates, and that there is no interaction in that model, the estimates for the marginal effect and the conditional effect (with respect to confounders $K(t)$) are very similar when adjusting for the same confounders via the IPT weight (marginal effect) or the conditional effect (conditioning on the variables) (Table A.8).

APPENDIX B

Appendix to Manuscript 2

B.1 Causal diagrams and biasing paths due to the monitoring process

Causal diagrams and biasing paths due to the monitoring process

In this section, we review the causal diagrams from the main manuscript, draw the biasing paths due to conditioning on the visit process, and show how models for the visit can be used to break the dependence between the monitoring and the outcome processes.

We first review the causal diagram that corresponds to Figure 4.1 in the corresponding manuscript. We depict that diagram in Figure B.1, before intervening on it in any way. In Figure B.2, for the same causal diagram, we depict the biasing paths (in bold) due to the conditioning on the visit indicator, $dN(t)$ for $t \in \{0, 1, 2\}$, which acts as a collider.

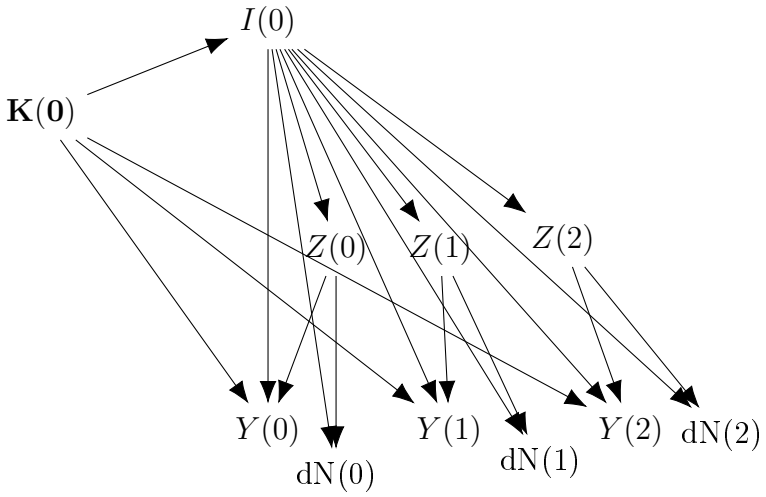


Figure B.1: Causal diagram for the first data generating mechanism (DGM) (patient index i removed)

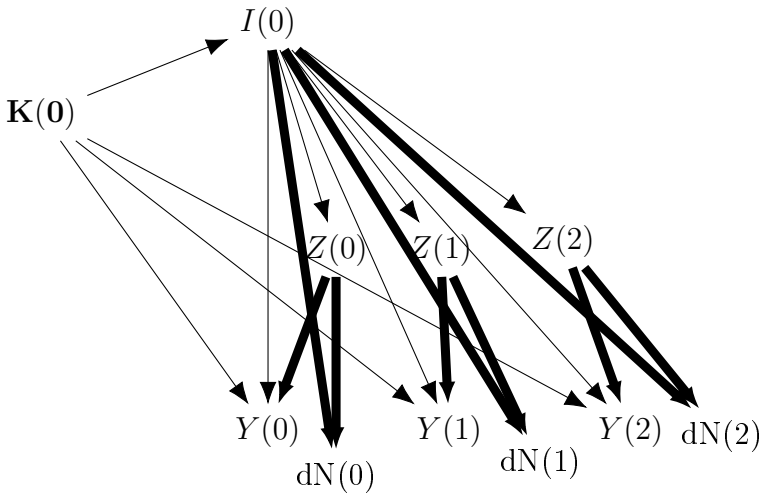


Figure B.2: Causal diagram for the first DGM (patient index i removed), biaising paths in bold

Finally, we depict in Figure B.3 what remains from the biasing paths (in bold) after adjusting for the monitoring rate via an inverse monitoring rate conditional on covariates $\mathbf{Z}$ and $\mathbf{I}$. We find that there is no more unblocked path from the exposure to the outcome due to the monitoring process, that is not due to the marginal effect of treatment. A proper adjustment for confounding factors $\mathbf{K}(\mathbf{0})$ must also be done to obtain unbiased estimates of the marginal effect of treatment (e.g. via inverse probability of treatment weights).

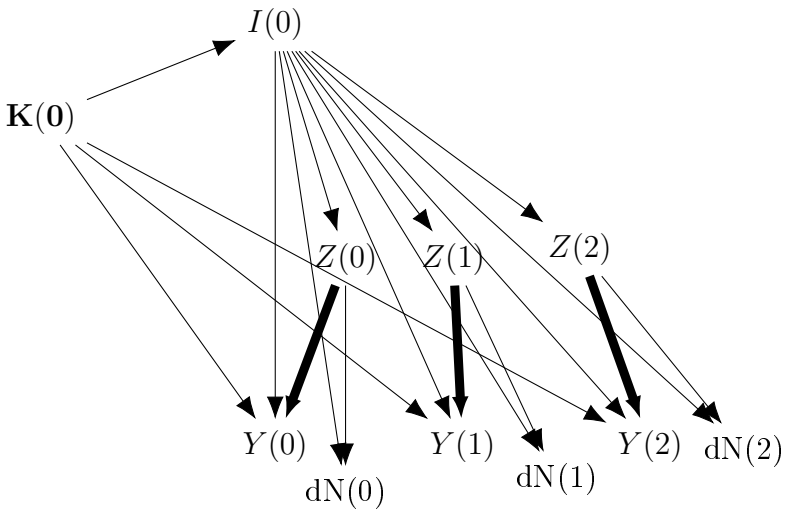


Figure B.3: Causal diagram for the first DGM (patient index i removed) after adjusting for the visit process

We now review the causal diagram that corresponds to Figure 4.2 in the main manuscript. That diagram is depicted in Figure B.4, before intervening on it in any way. In Figure B.5, for the same causal diagram, we depict the biasing paths (in bold) due to the conditioning on the visit indicator, $dN(t)$ for $t \in \{1, 2, 3\}$, which acts as a collider.

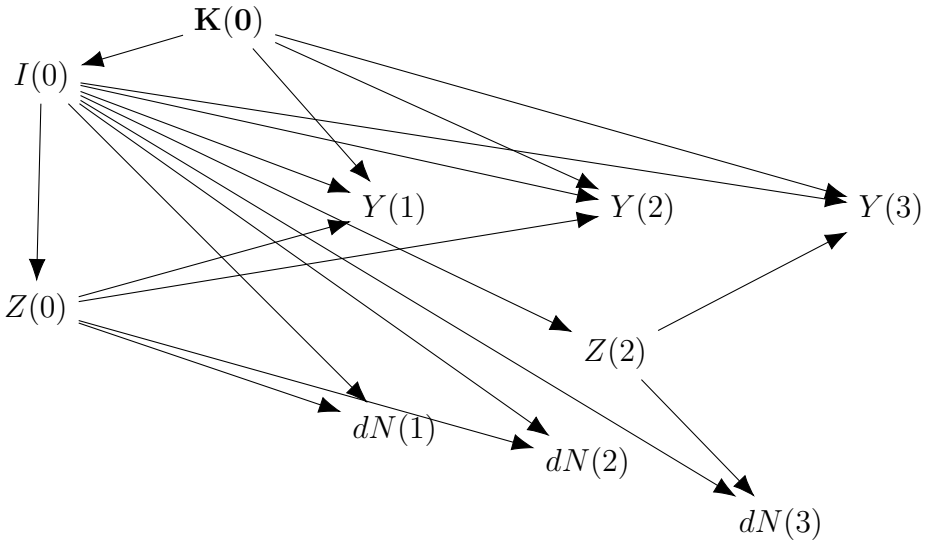


Figure B.4: Causal diagram for the second DGM (patient index i removed)

We depict in Figure B.6 what remains from the biasing paths (in bold) after adjusting for the monitoring rate via an inverse monitoring rate conditional on covariates $\mathbf{Z}$ (for the last values of them, which we assume affect

the monitoring indicator and the outcome) and $\mathbf{I}$. We find that there is no more unblocked path from the exposure to the outcome due to the monitoring process, that is not due to the marginal effect of treatment. In that case too, a proper adjustment for confounding factors $\mathbf{K}(\mathbf{0})$ must be done to obtain unbiased estimates of the marginal effect of treatment (e.g. via inverse probability of treatment weights).

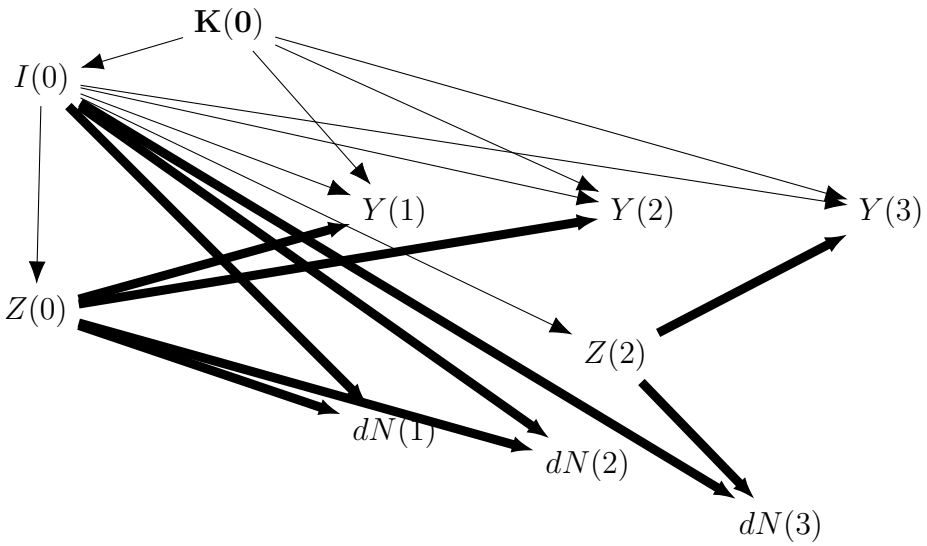


Figure B.5: Causal diagram for the second DGM (patient index i removed), biaising paths in bold

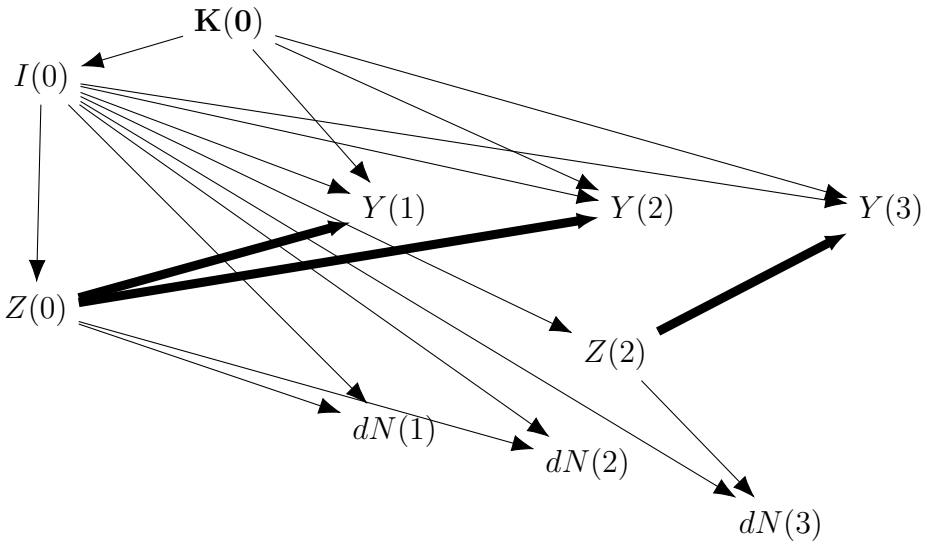


Figure B.6: Causal diagram for the second DGM (patient index i removed), remainings of the biaising paths in bold

Now, we review the causal diagram that corresponds to Figure 4.3 in the main manuscript. That diagram

is depicted in Figure B.7, where we show the diagram before intervening on it. In Figure B.8, we depict for the same diagram the biasing paths (in bold) due to the conditioning on the visit indicator, $dN(t)$ for $t \in \{1, 2\}$, which acts as a collider (Note: in this document, dashed edges are used to make it clearer which distinct paths can bias the estimate of the marginal effect of treatment, but they do not bear any special or different meaning as compared to other paths or causal arrows).

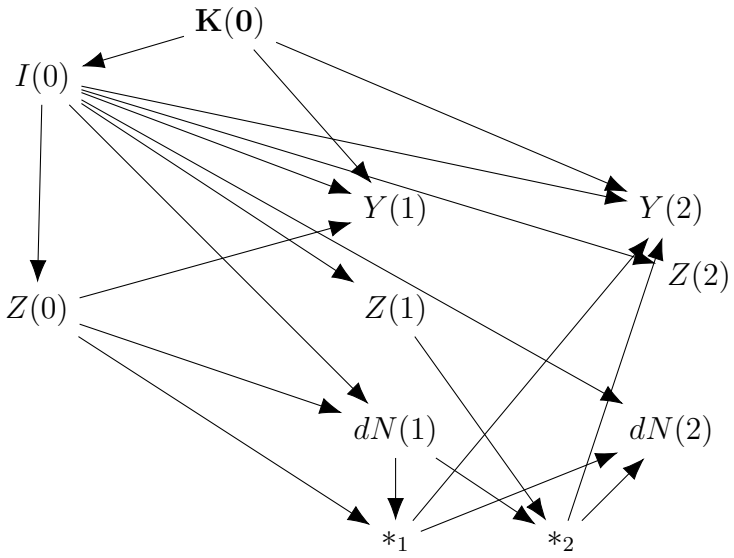


Figure B.7: Causal diagram for the third DGM (patient index i removed)

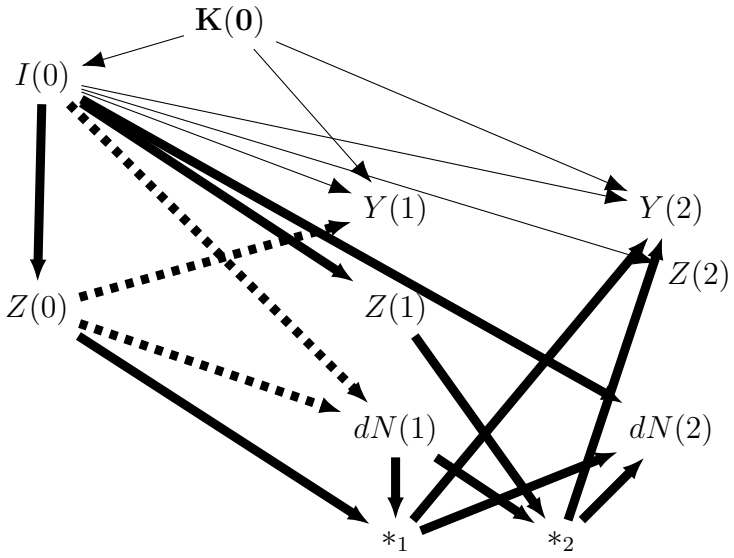


Figure B.8: Causal diagram for the third DGM (patient index i removed), biaising paths in bold. Dashed edges are used to make it clearer which distinct paths can bias the estimate of the marginal effect of treatment

We now make a dinstinction between two scenarios for the visit pattern, again assuming that we are in the setting depicted in Figure B.8: (a) There is a visit at time 1 and $dN(1) = 1$, and (b) there is no visit at time 1 and $dN(1) = 0$. In the former case (a), suppose we only adjust for an inverse intensity weight as a function of the last covariates observed. In Figure B.9, we depict what remains from the biasing paths after adjusting for the monitoring rate via an inverse monitoring rate conditional on the last covariates $\mathbf{Z}$ and $\mathbf{I}$ observed (in bold). We find that there is yet at least one unblocked path from the exposure to the outcome $Y(2)$ that is not due to the marginal effect of treatment and that is due to conditioning on collider $dN(2)$ (the path is given by $I(0) \rightarrow Z(0) \rightarrow *_{\mathbf{1}} - dN(2) - *_{\mathbf{2}} \rightarrow Y(2)$).

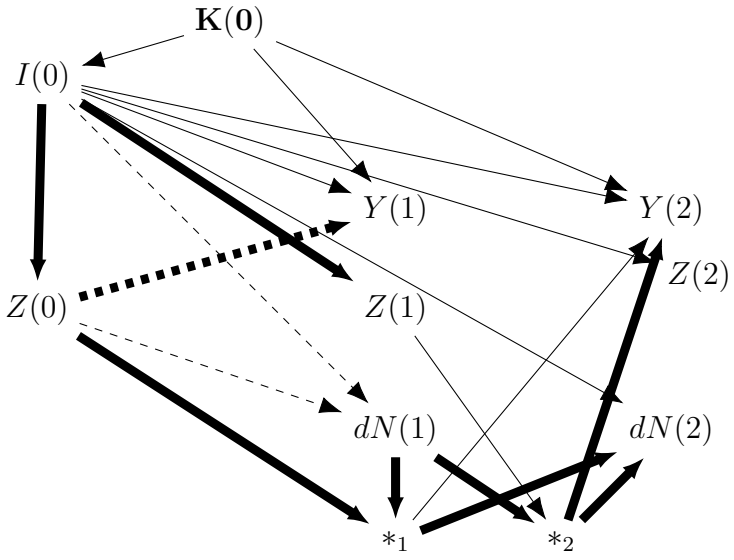


Figure B.9: Causal diagram for the third DGM (patient index i removed), remainings of the biaising paths in bold in scenario (a) when adjusting for the visit process using an inverse intensity weight as a function of the last covariates observed

In scenario (b), where there is no visit at time 1, suppose we only adjust for an inverse intensity weight as a function of the last covariates observed. What remains from the biasing paths after adjusting for the monitoring rate via an inverse monitoring rate conditional on the last covariates $\mathbf{Z}$ and $\mathbf{I}$ observed is depicted in Figure B.10. Here again, there is yet an unblocked path from the exposure to the outcome $Y(2)$ that is not due to the marginal effect of treatment (given by $I(0) \rightarrow Z(1) \rightarrow *_{\mathbf{2}} - dN(2) - *_{\mathbf{1}} \rightarrow Y(2)$).

Now, suppose that we use a cumulated weight that accounts for the full history of covariates and their interaction with the monitoring process (such as the weight $sw_{i,j}(\cdot)$ proposed in the manuscript). Figure B.11 depicts, for scenario (a), the remaining parts of the biasing paths in bold in such case; the adjustment for the whole monitoring process effectively blocks the unblocked path between subsequent monitoring indicators,

which paths were "due" to the interaction terms. We depicted this by removing the bold from the arrows leaving interaction terms $*_{\mathbf{1}}$ and $*_{\mathbf{2}}$, and entering the node $dN(2)$ but this could probably be depicted differently too (e.g. by removing the whole path between $dN(1)$ and $dN(2)$). The former unblocked path is now blocked by e.g. the interaction term $*_1$ which is a collider.

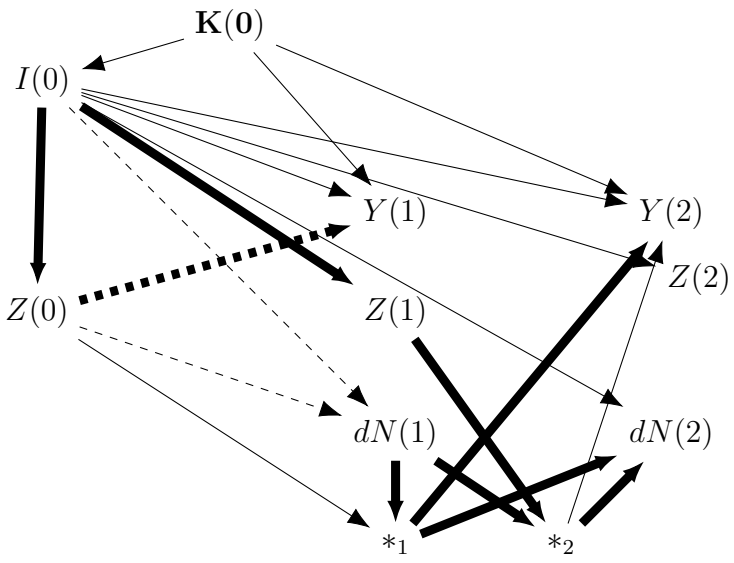


Figure B.10: Causal diagram for the third DGM (patient index i removed), remainings of the biaising path in bold in scenario (b) when adjusting for the visit process using an inverse intensity weight as a function of the last covariates observed

For scenario (b), Figure B.12 depicts the remaining biasing paths after using the proposed cumulated weight. There again, we depicted the impact of adjusting for the whole monitoring process by removing the bold from the arrows leaving interaction terms $*_{\mathbf{1}}$ and $*_{\mathbf{2}}$, and entering the node $dN(2)$; the formerly unblocked biasing path is now blocked by e.g. the collider $*_2$. In the two figures (Figures B.11 and B.12), there is no more biasing path from the exposure to the outcome that would be due to colliders $dN(\cdot)$ after using the proposed cumulated weight.

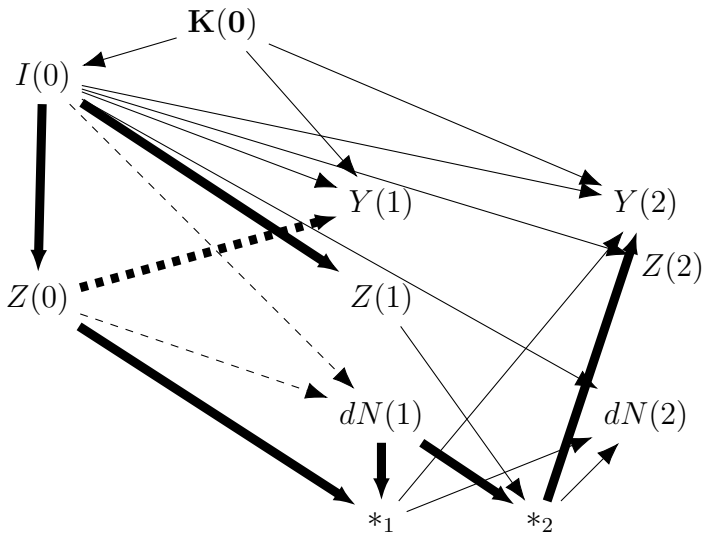


Figure B.11: Causal diagram for the third DGM (patient index i removed), remainings of the biaising paths in bold in scenario (a) when using the proposed cumulated weight

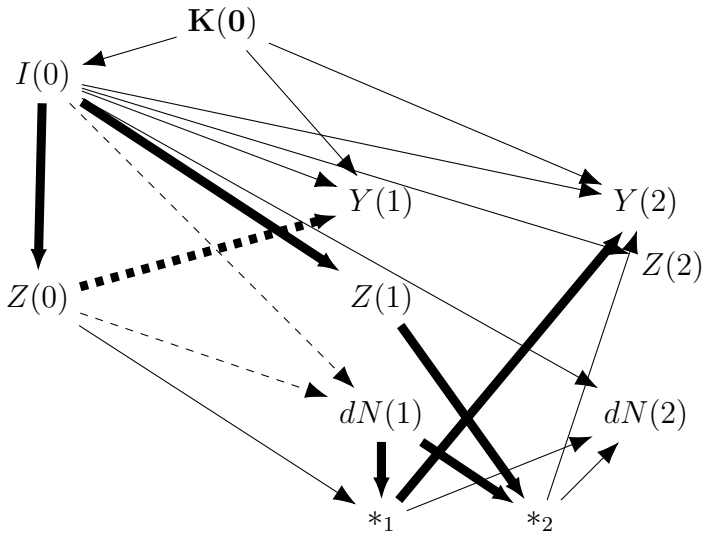


Figure B.12: Causal diagram for the third DGM (patient index i removed), remainings of the biaising paths in bold in scenario (b) when using the proposed cumulated weight

Finally, we present the last scenario, corresponding to Figure 4.4 in the main manuscript, which is similar to that depicted in Figure B.8, but where a previous outcome affects the next outcome and monitoring time. In Figure B.13, we show the causal diagram corresponding to that scenario, before intervening on it. In Figure B.14, we depict the (potential) biasing paths (in bold) due to the conditioning on colliders dN(t), $t \subset 1, 2$. That setting (Figures B.13 and B.14) is similar to that from Figure B.7, except for the collider $dN(2)$ that opens another path between $I(0)$ and $Y(1)$. By including the outcome as a predictor in the intensity model,

a similar adjustment using the proposed cumulated weight with the intensity modelled conditionnally on $\mathbf{I}$, $\mathbf{Z}$ and $\mathbf{Y}$ will adjust properly for the visit process (not shown).

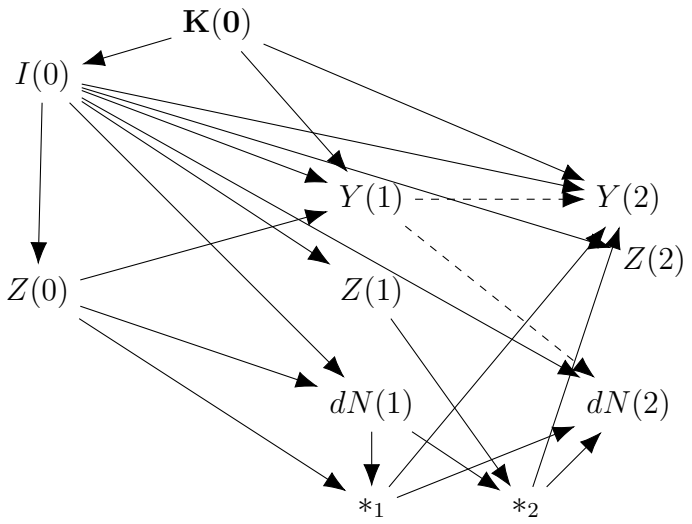


Figure B.13: Causal diagram for the fourth DGM (patient index i removed)

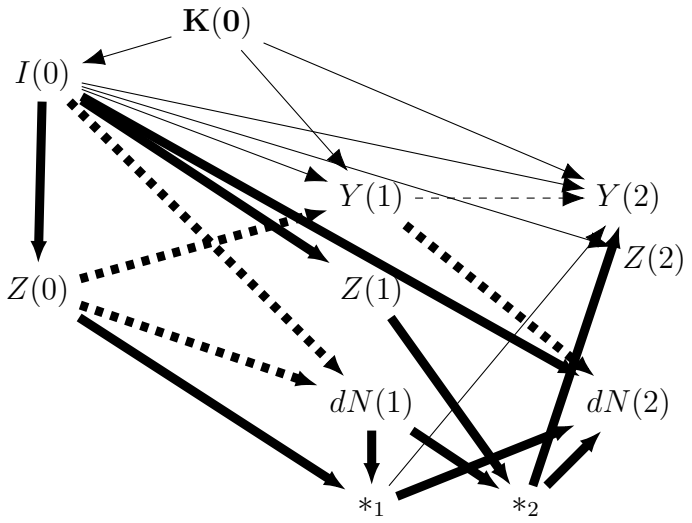


Figure B.14: Causal diagram for the fourth DGM (patient index i removed), biaising paths in bold

B.2 Estimating equation for the marginal effect of treatment on a continuous longitudinal outcome

In the main manuscript we assumed

$$I_i(0) \perp \{Y_{i0}(t), Y_{i1}(t)\} \,|\mathcal{H}_{\mathbf{i}}^{\mathbf{o}}(\mathbf{t}-), dN_i(t), \mathbf{K_i}(\mathbf{0}), \tag{I3}$$

$$dN_i(t) \perp Y_i(t)|\mathcal{H}_{\mathbf{i}}^{\mathbf{o}}(\mathbf{t}-), \text{ and} \tag{I4}$$

$$dN_i(t) \perp dN_i(t-)|\mathbf{Z_i}(\mathbf{l_i}(\mathbf{t})), I_i(0), B_i(t-), \tag{I5}$$

and proposed to use the following *partial* model for the monitoring intensity:

$$\lambda_i(t|\mathbf{Z_i}(\mathbf{l}(\mathbf{t_i})), I_i(0), B_i(t-)) = \lambda_0(B_i(t)) \exp(\gamma_I I_i(0) + \boldsymbol{\gamma_Z} \mathbf{Z_i}(\mathbf{l_i}(\mathbf{t}))). \tag{5}$$

We also cited Theorem 1 of Pearl (2009), which we recall:

Theorem 1 (The Causal Markov Condition). *Any distribution generated by a Markovian model M can be factorized as:*

$$P(v_1, v_2, ..., v_n) = \prod_i P(v_i|pa_i)$$

where $V_1, V_2, ...V_n$ are the endogenous variables in M, and pa_i are (values of) the endogenous "parents" of V_i in the causal diagram associated with M.

We assume that continuous time can be discretized in units of length 1 (e.g. days) over which a visit can or cannot occur, so $t \in 1, 2, 3, ..., \lfloor \tau \rfloor$. We denote by $\overline{P(t)}$ the history from time 0 to time t of the covariate process P, and by $P(t) \setminus J(t)$ the process $P(t)$ minus the $J(t)$ process. Using the assumptions above, we

have:

$$
\begin{aligned}
\mathbb{E}\left[dN_i(t)|\mathcal{H}_{\mathbf{i}}^{\mathbf{o}}(\mathbf{t}-)\right] =& \mathbb{P}(dN_i(t)|\mathcal{H}_{\mathbf{i}}^{\mathbf{o}}(\mathbf{t}-)) \\
=& \mathbb{P}(dN_i(t)|\mathcal{H}_{\mathbf{i}}^{\mathbf{o}}(\mathbf{t}-) \setminus \overline{dN_i(t-)}, \overline{dN_i(t-)}) \\
=& \frac{\mathbb{P}(dN_i(t), \overline{dN_i(t-)}|\mathcal{H}_{\mathbf{i}}^{\mathbf{o}}(\mathbf{t}-) \setminus \overline{dN_i(t-)}) P(\mathcal{H}_{\mathbf{i}}^{\mathbf{o}}(\mathbf{t}-) \setminus \overline{dN_i(t-)})}{P(\mathcal{H}_{\mathbf{i}}^{\mathbf{o}}(\mathbf{t}-))} \\
\propto& \mathbb{P}(\overline{dN_i(t)}|\mathcal{H}_{\mathbf{i}}^{\mathbf{o}}(\mathbf{t}-) \setminus \overline{dN_i(t-)}) \\
=& \mathbb{P}(dN_i(t)|\mathbf{Z_i}(\mathbf{l_i}(\mathbf{t})), I_i(0), B_i(t-1)) \\
& \times \mathbb{P}(dN_i(t-1)|\mathbf{Z_i}(\mathbf{l_i}(\mathbf{t-1})), I_i(0), B_i(t-2)) \\
& \times ... \\
& \times \mathbb{P}(dN_i(2)|\mathbf{Z_i}((\mathbf{l_i}(\mathbf{1})), I_i(0), B_i(1)) \\
& \times \mathbb{P}(dN_i(1)|\mathbf{Z_i}(\mathbf{0}), I_i(0)) \quad \text{using assumption (I5) and Theorem 1} \\
=& \prod_{s=0}^{t} \left\{\xi_i(s) \exp\left(\gamma_I I_i(0) + \gamma_Z \mathbf{Z_i}(\mathbf{l_i}(\mathbf{s}))\right) \lambda_0(B_i(s)) ds\right\}^{\mathbb{I}(dN_i(s)=1)} \\
& \times \left\{1 - \xi_i(s) \exp\left(\gamma_I I_i(0) + \gamma_Z \mathbf{Z_i}(\mathbf{l_i}(\mathbf{s}))\right) \lambda_0(B_i(s)) ds\right\}^{\mathbb{I}(dN_i(s)=0)}, \qquad \text{(B.1)}
\end{aligned}
$$

with the last term equal to the weight $usw_i(t|\mathcal{H}_{\mathbf{i}}^{\mathbf{o}}(\mathbf{t}-))$, and using assumption (5) for the visit model. As in the main manuscript, we use a product integral to emphasize that the product must be taken over continuous time. We use the following estimating equation for the coefficients $\boldsymbol{\beta_S}$, with the marginal effect of treatment consisting in e.g. the last coefficient of $\boldsymbol{\beta_S}$:

$$
\mathbb{E}\left(\int_0^\tau \frac{\mathbf{Y(t)} - [\boldsymbol{\beta}_{\boldsymbol{s}}' \mathbf{S(t)}]}{\mathbf{w(t|K(t))usw}(\mathbf{t}|\mathcal{H}^{\mathbf{o}}(\mathbf{t}-))} \mathbf{dN(t)}\right) = \mathbf{0}.
$$

The matrix $\mathbf{S(t)}$ may, for instance, incorporate a column of 1 for estimating a constant intercept, or several columns as the basis of a cubic spline for accounting for the effect of time.

Using iterated expectation (similarly to Lin et al. [2004]), we have that

$$
\begin{aligned}
&\mathbb{E}\left(\int_0^\tau \frac{\mathbf{Y(t)} - [\boldsymbol{\beta}_s'\mathbf{S(t)}]}{\mathbf{w(t|K(t))usw(t|\mathcal{H}^o(t-))}}\mathbf{dN(t)}\right) \\
&\qquad = \mathbb{E}\left[\mathbb{E}\left(\int_0^\tau \frac{\mathbf{Y(t)} - [\boldsymbol{\beta}_s'\mathbf{S(t)}]}{\mathbf{w(t|K(t))usw(t|\mathcal{H}^o(t-))}}\mathbf{dN(t)}\right) \mid \mathcal{H}^o(\mathbf{t}-)\right] \\
&\qquad = \mathbb{E}\left(\int_0^\tau \frac{\mathbf{Y(t)} - [\boldsymbol{\beta}_s'\mathbf{S(t)}]}{\mathbf{w(t|K(t))usw(t|\mathcal{H}^o(t-))}}\mathbb{E}\left[\mathbf{dN(t)}|\mathcal{H}^o(\mathbf{t}-)\right]\right) \text{ using assumption (I4)} \\
&\qquad \propto \mathbb{E}\left(\int_0^\tau \frac{\mathbf{Y(t)} - [\boldsymbol{\beta}_s'\mathbf{S(t)}]}{\mathbf{w(t|K(t))usw(t|\mathcal{H}^o(t-))}}\mathbf{usw(t|\mathcal{H}^o(t-))}dt\right) \\
&\qquad = \mathbb{E}\left(\int_0^\tau \frac{\mathbf{Y(t)} - [\boldsymbol{\beta}_s'\mathbf{S(t)}]}{\mathbf{w(t|K(t))}}dt\right).
\end{aligned}
$$

Under assumption (I3) and correct model specifications, we have that the last expression in the final line is equal to 0, that is,

$$\mathbb{E}\left(\int_0^\tau \frac{\mathbf{Y(t)} - [\boldsymbol{\beta}_s'\mathbf{S(t)}]}{\mathbf{w(t|K(t))}}dt\right) = \mathbf{0}$$

and the estimating equation for the marginal effect of treatment is unbiased.

B.3 Asymptotic properties of the proposed estimator

Under the assumptions on the exposure and the monitoring models that follow:

$$I_i(t) \perp \{Y_{i0}(t), Y_{i1}(t)\} \,|\mathcal{H}_{\mathbf{i}}^{\mathbf{o}}(\mathbf{t}-), dN_i(t), \mathbf{K_i(t)} \tag{I3}$$

$$dN_i(t) \perp Y_i(t)|\mathcal{H}_{\mathbf{i}}^{\mathbf{o}}(\mathbf{t}-) \tag{I4}$$

$$dN_i(t) \perp dN_i(t-)|\mathbf{Z_i(l_i(t))}, I_i(0), B_i(t-), dN_i(t-) \tag{I5}$$

$$0 < P(dN_i(t) = 1|\mathbf{K_i(t)}, \mathcal{H}_{\mathbf{i}}^{\mathbf{o}}(\mathbf{t}-)), P(dN_i(t) = 0|\mathbf{K_i(t)}, \mathcal{H}_{\mathbf{i}}^{\mathbf{o}}(\mathbf{t}-)) < 1 \tag{P1}$$

$$0 < P(I_i(t) = 1|\mathbf{K_i(t)}, \mathcal{H}_{\mathbf{i}}^{\mathbf{o}}(\mathbf{t}-), dN_i(t)), P(I_i(t) = 0|\mathbf{K_i(t)}, \mathcal{H}_{\mathbf{i}}^{\mathbf{o}}(\mathbf{t}-), dN_i(t)) < 1, \tag{P2}$$

as well as no interference, and correct model specifications for the exposure, the outcome and the visit models, the proposed estimator resulting from the following estimating equation

$$\mathbb{E}\left(\int_0^\tau \frac{\mathbf{Y(t)} - [\boldsymbol{\beta}_s'\mathbf{S(t)}]}{\mathbf{w(t|K(t))sw_j(t|\mathcal{H}^o(t-))}}\mathbf{dN(t)}\right) = \mathbf{0} \tag{B.2}$$

is a two-step m-estimator. Two-step estimators often rely on substituting an estimate of a nuisance parameter in the estimating function for the parameter of interest (Newey and McFadden [1994]). One can use a first-step estimator for the nuisance parameters (e.g., here, the parameters from the IPT weights and from the

monitoring weights). M-estimators of the parameters of interest are obtained by solving a sample average equation and often consist of the zero roots of an estimating equation.

For $\hat{\boldsymbol{\beta}}_{\boldsymbol{TSE}}$ a two-step semiparametric estimator and $\boldsymbol{\beta_0}$ the vector of true parameters, Newey and McFadden [1994] show that $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\boldsymbol{TSE}} - \boldsymbol{\beta_0}) \to N(\mathbf{0}, \boldsymbol{\Sigma})$, with

$$\boldsymbol{\Sigma} = \mathbf{G}_{\beta}^{-\mathbf{1}} \mathbb{E}\left[\left\{\mathbf{g}(\mathbf{o}; \beta_{\mathbf{0}}, \phi_{\mathbf{0}}) - \mathbf{G}_{\phi}\mathbf{M}^{-\mathbf{1}}\mathbf{m}(\mathbf{o}; \phi_{\mathbf{0}})\right\}^{\otimes 2}\right] \mathbf{G}_{\beta}^{-\mathbf{1}} \tag{B.3}$$

where

$$\begin{aligned} \mathbf{G}_{\beta} &= \mathbb{E}(\nabla_{\beta}\mathbf{g}(\mathbf{o}; \beta_{\mathbf{0}}, \phi_{\mathbf{0}})) \\ \mathbf{G}_{\phi} &= \mathbb{E}(\nabla_{\phi}\mathbf{g}(\mathbf{o}; \beta_{\mathbf{0}}, \phi_{\mathbf{0}})) \\ \mathbf{M} &= \mathbb{E}(\nabla_{\phi}\mathbf{m}(\mathbf{o}; \phi_{\mathbf{0}})) \end{aligned}$$

for $\mathbf{o}$ the data, and $\mathbf{m}(\mathbf{o}; \phi_{\mathbf{0}})$ and $\mathbf{g}(\mathbf{o}; \beta_{\mathbf{0}}, \phi_{\mathbf{0}})$ the estimating equations for the nuisance parameters $\boldsymbol{\phi}$ and the parameters of interest $\boldsymbol{\beta}$, respectively.

In the case of interest, the nuisance parameters consist in the parameters from the intervention model and those from the intensity model. The estimating equations for the former are those corresponding to a logistic regression model for the exposure, and those for the latter correspond to equations from the Cox model for the visit intensity.

We do not show how to compute the variance component due to the inverse monitoring weight but one could potentially develop that component of variance by using extensions of the Greenwood formula for the survivor function (Greenwood [1926]). In particular, the product integral used in our cumulated weight can be approximated by an exponential function since that we take the product of small quantities, in continuous. The delta method could be used in the development of the variance. However in practice, a nonparametric bootstrap (resampling on individuals) will provide a good estimate of the estimator variance. Alternatively, a "robust" variance may serve as a conservative estimate of the variance, similar to the case in marginal structural models (Robins et al. [2000a]).

B.4 Details of the simulation studies

In the main study, we first simulated for each patient i three baseline confounders $\{K_{1i}, K_{2i}, K_{3i}\}$ with $K_{1i} \sim \mathrm{N}(1,1), K_{2i} \sim$ Bernoulli(0.55), and $K_{3i} \sim \mathrm{N}(0,1)$. The intervention $I_i(t)$ was binary and time-

fixed: $I_i \sim$ Bernoulli(p_{Ii}) with $p_{Ii} = \text{expit}\,(0.5 + 0.8K_{1i} + 0.05K_{2i} - 1K_{3i})$. One time-varying mediator $Z_i(\cdot)$ was generated, conditional on I_i. It was only updated whenever there was a new visit ($dN_i(\cdot) = 1$), and was simulated as $Z_i(t)|I_i = 1 \sim \text{N}(2,1)$ and $Z_i(t)|I_i = 0 \sim \text{N}(4,2^2)$ on those visit days. On other (non-visit) days, we denote the process by $Z_i(l_i(t))$, simply carrying forward the last observed value. Time was discretized over a grid of 0.01 units, from 0 to τ. The intensity of monitoring at each time point over that grid was simulated as $\lambda_i(t|I_i, Z_i(l_i(t))) = 0.02B_i(t)\exp(\gamma_1 I_i + \gamma_2 Z_i(l_i(t)))$. We used Bernoulli draws with probabilities proportional to these intensities to assign monitoring times (one draw per time point, for each time point over the grid). Whenever a new monitoring time occurred, the endogenous covariate process $\mathbf{Z}(\mathbf{l}(\mathbf{t}))$ was updated according to the simulation scheme given above (i.e. depending on the value of the baseline intervention I_i), the outcome was simulated dependent on the gap time as detailed below, and then the gap time $B_i(t)$ was reset to 0. On each subsequent day when there was no visit, the gap time cumulated a value of 0.01 according to our discrete time grid.

We considered several different combinations of the parameters (γ_1, γ_2) (see Table 1), so as to vary the strength of the selection bias due to the visit process. The outcome $Y_i(t)$ was generated according to $Y_i(t) = 0.2B_i(t) + 1\,I_i - 0.8\,(Z_i(l_i(t)) - E\,[Z_i(l_i(t))|I_i]) + 0.4\,K_{1i} + 0.05\,K_{2i} - 0.6\,K_{3i} + \epsilon_i(t)$ with $\epsilon_i(t) \sim \text{N}(0, 0.5^2)$. The re-centering of $Z_i(l_i(t))$ in the outcome model ensures that we estimate the target marginal intervention effect. Monitoring times were drawn up until the maximum follow-up time τ, which we fixed to $\tau = 5$. Data were simulated to correspond to a study cohort of 500 patients. For each patient, the follow-up time was "censored" (stopped) at time C_i, with $C_i \sim \text{Uniform}(\tau/2, \tau)$; the censoring was non-informative. A total of 1000 replicate datasets were simulated for each simulation study scenario.

B.5 Results of the main simulation study, including the average number of visits and estimated parameters in the visit model

Table B.1: Main analysis: Estimated parameters, average number of visits and mean absolute bias for the estimators compared for 1000 simulations with $\tau = 5$, $n = 500$.

$\boldsymbol{\gamma}$	$\hat{\boldsymbol{\gamma}}$	$\overline{N}(\tau)$	Mean absolute bias of the estimator					
		$I = 0, 1$	$\hat{\beta}_{LS}$	$\hat{\beta}_{IPT}$	$\hat{\beta}_{IH}$	$\hat{\beta}_{USW}$	$\hat{\beta}_{SW1}$	$\hat{\beta}_{SW2}$
-0.3; 0.1	-0.3; 0.1	1.9, 2.9	0.35	0.37	0.12	0.04	0.14	0.03
-0.2; 0.2	-0.2; 0.2	2.5, 3.5	0.49	0.24	0.11	0.00	0.09	0.01
-0.1; 0.2	-0.1; 0.2	3.0, 3.9	0.64	0.08	0.09	0.14	0.00	0.02
-0.1; -0.3	-0.1; -0.3	3.1, 2.9	0.69	0.01	0.11	0.07	0.01	0.03
0; 0	0; 0	3.9, 3.9	0.73	0.01	0.03	0.16	0.01	0.01
0.1; -0.3	0.1; -0.3	4.8, 3.7	0.69	0.03	0.11	0.19	0.04	0.02
0.2; -0.2	0.2; -0.2	6.0, 4.3	0.64	0.12	0.26	0.18	0.19	0.02
0.3; 0.2	0.3; 0.2	7.1, 5.9	0.67	0.08	0.34	0.24	0.30	0.05

B.6 Results of all sensitivity analyses

Table B.2: Simulation study results for sensitivity analysis 1 with $\tau = 5$, $n = 500$ patients, 1000 simulations, $\alpha_i(t) = 0.2\,B_i(t)$. A constant intercept was fitted in the outcome model, rather than a cubic spline as a function of gap time

$\boldsymbol{\gamma}^{\dagger}$	$\overline{N(\tau)}$	Mean absolute bias $\hat{\beta}$						Empirical variance $\hat{\beta}$					
	$I = 0, 1$	$\hat{\beta}_{LS}$	$\hat{\beta}_{IPT}$	$\hat{\beta}_{IH}$	$\hat{\beta}_{USW}$	$\hat{\beta}_{SW1}$	$\hat{\beta}_{SW2}$	$\hat{\beta}_{LS}$	$\hat{\beta}_{IPT}$	$\hat{\beta}_{IH}$	$\hat{\beta}_{USW}$	$\hat{\beta}_{SW1}$	$\hat{\beta}_{SW2}$
a	1.9, 2.9	0.45	0.25	0.18	0.08	0.16	0.04	0.03	0.08	0.17	1.56	0.13	0.09
b	3.0, 3.9	0.65	0.05	0.11	0.09	0.01	0.01	0.02	0.05	0.10	1.16	0.06	0.05
c	3.9, 3.9	0.72	0.03	0.04	0.14	0.03	0.03	0.02	0.06	0.10	0.59	0.06	0.06
d	4.8, 3.7	0.79	0.08	0.06	0.17	0.02	0.02	0.02	0.04	0.07	0.42	0.05	0.05
e	7.1, 5.9	0.79	0.08	0.31	0.25	0.25	0.05	0.02	0.03	0.06	0.29	0.08	0.06

†: a. (-0.3, 0.1); b. (-0.1, 0.2); c. (0, 0); d. (0.1, -0.3); e. (0.3, 0.2).

Table B.3: Simulation study results for sensitivity analysis 2 with $\tau = 10$, $n = 500$ patients, 1000 simulations, $\alpha_i(t) = 0.2\,B_i(t)$

$\boldsymbol{\gamma}^\dagger$	$\overline{N(\tau)}$	Mean absolute bias $\hat{\beta}$						Empirical variance $\hat{\beta}$					
	$I = 0, 1$	$\hat{\beta}_{LS}$	$\hat{\beta}_{IPT}$	$\hat{\beta}_{IH}$	$\hat{\beta}_{USW}$	$\hat{\beta}_{SW1}$	$\hat{\beta}_{SW2}$	$\hat{\beta}_{LS}$	$\hat{\beta}_{IPT}$	$\hat{\beta}_{IH}$	$\hat{\beta}_{USW}$	$\hat{\beta}_{SW1}$	$\hat{\beta}_{SW2}$
a	4.2, 6.2	0.39	0.35	0.14	0.02	0.15	0.02	0.01	0.03	0.06	1.40	0.08	0.03
b	6.6, 8.1	0.65	0.07	0.10	0.15	0.01	0.01	0.01	0.02	0.04	0.67	0.03	0.02
c	8.1, 8.1	0.72	0.02	0.02	0.13	0.02	0.02	0.01	0.02	0.03	0.25	0.02	0.02
d	10.0, 7.7	0.68	0.05	0.12	0.12	0.04	0.02	0.01	0.02	0.03	0.17	0.02	0.02
e	13.5, 11.2	0.64	0.11	0.38	0.06	0.30	0.05	0.01	0.02	0.03	0.15	0.04	0.04

†: a. (-0.3, 0.1); b. (-0.1, 0.2); c. (0, 0); d. (0.1, -0.3); e. (0.3, 0.2).

Table B.4: Simulation study results for sensitivity analysis 3 with $\tau = 5$, $n = 500$ patients, 1000 simulations, $\alpha_i(t) = 0.2\,B_i(t)$, and with the process $Z(\cdot)$ depending on the cumulative number of previous visits

$\boldsymbol{\gamma}^\dagger$	$\overline{N(\tau)}$	Mean absolute bias $\hat{\beta}$						Empirical variance $\hat{\beta}$					
	$I = 0, 1$	$\hat{\beta}_{LS}$	$\hat{\beta}_{IPT}$	$\hat{\beta}_{IH}$	$\hat{\beta}_{USW}$	$\hat{\beta}_{SW1}$	$\hat{\beta}_{SW2}$	$\hat{\beta}_{LS}$	$\hat{\beta}_{IPT}$	$\hat{\beta}_{IH}$	$\hat{\beta}_{USW}$	$\hat{\beta}_{SW1}$	$\hat{\beta}_{SW2}$
a	1.9, 2.8	0.35	0.38	0.12	0.05	0.16	0.04	0.02	0.05	0.10	1.18	0.09	0.05
b	3.1, 3.8	0.63	0.09	0.09	0.09	0.02	0.00	0.01	0.04	0.06	0.74	0.04	0.04
c	3.9, 3.9	0.72	0.02	0.04	0.13	0.02	0.02	0.01	0.02	0.05	0.33	0.02	0.02
d	4.9, 3.7	0.69	0.04	0.11	0.19	0.05	0.01	0.01	0.02	0.04	0.24	0.02	0.03
e	8.1, 6.6	0.65	0.10	0.36	0.18	0.32	0.05	0.01	0.02	0.04	0.19	0.04	0.04

†: a. (-0.3, 0.1); b. (-0.1, 0.2); c. (0, 0); d. (0.1, -0.3); e. (0.3, 0.2).

Table B.5: Simulation study results for sensitivity analysis 4 with $\tau = 5$, $n = 500$ patients, 1000 simulations, $\alpha(t) = \alpha = 0.02$.

$\boldsymbol{\gamma}^\dagger$	$\overline{N(\tau)}$	Mean absolute bias $\hat{\beta}$						Empirical variance $\hat{\beta}$					
	$I = 0, 1$	$\hat{\beta}_{LS}$	$\hat{\beta}_{IPT}$	$\hat{\beta}_{IH}$	$\hat{\beta}_{USW}$	$\hat{\beta}_{SW1}$	$\hat{\beta}_{SW2}$	$\hat{\beta}_{LS}$	$\hat{\beta}_{IPT}$	$\hat{\beta}_{IH}$	$\hat{\beta}_{USW}$	$\hat{\beta}_{SW1}$	$\hat{\beta}_{SW2}$
a	1.9, 2.9	0.36	0.34	0.17	0.02	0.10	0.00	0.02	0.04	0.07	1.35	0.06	0.04
b	3.1, 3.9	0.63	0.09	0.09	0.10	0.01	0.01	0.01	0.04	0.07	0.77	0.04	0.04
c	3.9, 3.9	0.73	0.02	0.03	0.13	0.02	0.02	0.01	0.03	0.04	0.36	0.03	0.03
d	4.8, 3.7	0.70	0.03	0.11	0.18	0.04	0.01	0.01	0.03	0.04	0.26	0.03	0.03
e	7.1, 5.9	0.66	0.09	0.34	0.22	0.30	0.06	0.01	0.02	0.04	0.20	0.04	0.04

†: a. (-0.3, 0.1); b. (-0.1, 0.2); c. (0, 0); d. (0.1, -0.3); e. (0.3, 0.2).

B.7 Comparison of the bootstrap and the empirical variance of the estimators

Table B.6: Comparison of bootstrap and empirical variance for all simulation studies in the main analysis (studies with $\tau = 5$, $n = 500$, 1000 simulations)

Intercept fitted	$\boldsymbol{\gamma}$ no.†	Empirical variance of $\hat{\beta}$						Bootstrap variance $\hat{\beta}$					
		$\hat{\beta}_{LS}$	$\hat{\beta}_{IPT}$	$\hat{\beta}_{IH}$	$\hat{\beta}_{USW}$	$\hat{\beta}_{SW1}$	$\hat{\beta}_{SW2}$	$\hat{\beta}_{LS}$	$\hat{\beta}_{IPT}$	$\hat{\beta}_{IH}$	$\hat{\beta}_{USW}$	$\hat{\beta}_{SW1}$	$\hat{\beta}_{SW2}$
Constant	1	0.01	0.04	0.09	1.25	0.07	0.05	0.03	0.08	0.17	1.56	0.13	0.09
	2	0.01	0.04	0.08	0.98	0.06	0.05	0.03	0.06	0.11	1.19	0.09	0.07
	3	0.01	0.03	0.06	0.81	0.03	0.03	0.02	0.05	0.10	1.16	0.06	0.05
	4	0.01	0.03	0.06	0.43	0.04	0.03	0.02	0.06	0.10	0.63	0.07	0.06
	5	0.01	0.03	0.05	0.38	0.03	0.03	0.02	0.06	0.10	0.59	0.06	0.06
	6	0.01	0.02	0.05	0.25	0.03	0.03	0.02	0.04	0.07	0.42	0.05	0.05
	7	0.01	0.02	0.05	0.23	0.04	0.03	0.02	0.03	0.07	0.38	0.06	0.05
	8	0.01	0.02	0.05	0.20	0.05	0.04	0.02	0.03	0.06	0.29	0.08	0.06
Cubic	1	0.02	0.04	0.10	1.21	0.07	0.04	0.03	0.07	0.16	1.56	0.13	0.08
spline	2	0.01	0.04	0.08	1.02	0.05	0.04	0.03	0.06	0.11	1.23	0.08	0.06
	3	0.01	0.03	0.06	0.72	0.03	0.03	0.02	0.05	0.09	1.11	0.06	0.05
	4	0.01	0.03	0.06	0.39	0.04	0.03	0.02	0.05	0.08	0.59	0.06	0.05
	5	0.01	0.03	0.06	0.35	0.03	0.03	0.02	0.05	0.09	0.56	0.05	0.05
	6	0.01	0.03	0.04	0.24	0.03	0.03	0.02	0.04	0.06	0.40	0.05	0.05
	7	0.01	0.03	0.05	0.24	0.03	0.04	0.02	0.03	0.06	0.36	0.05	0.05
	8	0.01	0.02	0.04	0.18	0.05	0.04	0.02	0.03	0.06	0.26	0.07	0.05

†. **1.** (-0.3, 0.1); **2.** (-0.2, 0.2); **3.** (-0.1, 0.2); **4.** (-0.1, -0.3); **5.** (0, 0); **6.** (0.1, -0.3); **7.** (0.2, -0.2); **8.** (0.3, 0.2).

B.8 Table of baseline characteristics stratified by intervention group, in the analysis of CPRD data

Table B.7: Baseline characteristics of the study cohort stratified by treatment at cohort entry (n=246,503), Clinical Practice Research Datalink, United Kingdom, 1998-2017

	Treatment	
Variable[1]	Citalopram	Fluoxetine
BMI, mean	26.8	26.8
Age, mean	43.4	40.7
Female sex	63.6	66.1
IMD quintile[2], mean	3.01	3.08
Smoking status		
Ever	50.8	48.7
Never	34.0	31.3
Unknown	15.2	20.1
Diabetes	5.3	4.3
Alcohol abuse	8.0	6.6
Anxiety or GAD	30.5	22.1
Other psychiatric diseases		
Schizophrenia	1.4	1.1
Bipolar disorder	0.8	0.7
Autism spectrum disorder	0.2	0.1
Obsessive compulsive disorder	0.5	0.6
Antipsychotic drugs	11.7	10.6
Benzodiazepine drugs	19.7	17.1
Lipid lowering drugs	7.6	5.0

Abbreviations: BMI, Body mass index; IMD, Index of multiple deprivation; GAD, Generalized anxiety disorder.
1. In % unless otherwise stated.
2. The IMD was available in the format of quintiles, with the greater quintile being the most deprived

B.9 Multivariate outcome model in the analysis of the CPRD data

We present below the coefficients for each covariate in a linear multivariate outcome model for the outcome BMI (no inverse weight is incorporated in this model).

Table B.8: Coefficients for each covariate included in the conditional outcome mean model (except the intercept and spline on time), Clinical Practice Research Datalink, United Kingdom, 1998-2017

Variable	Coefficient	Robust 95% CI
Citalopram (Ref.: Fluoxetine)	-0.55	-0.67, -0.43*
Age at baseline	-0.02	-0.03, -0.02*
Sex (Ref.: Female)	-0.58	-0.70, -0.46*
IMD at baseline	0.40	0.35, 0.44*
Smoking (Ref.: Never)		
Ever	-0.53	-0.68, -0.38*
Missing	0.03	-0.12, 0.18
Diabetes	4.21	4.03, 4.39*
Alcohol abuse	-0.66	-1.13, -0.20*
Anxiety or GAD	-0.22	-0.43, -0.01*
Psychiatric diagnosis	0.24	-0.49, 0.96
Number of hospitalisations in prior month	-0.69	-0.86, -0.52*
Antipsychotic drugs	-0.05	-0.43, 0.33
Benzodiazepine drugs	-1.55	-1.77, -1.33*
Lipid lowering drugs	1.66	1.48, 1.85*

Abbreviations: IMD, Index of Multiple Deprivation; GAD, Generalized Anxiety Disorder.
* Confidence interval does not contain 0.

APPENDIX C

Appendix to Manuscript 3

C.1 Details of the simulation studies

We conducted several simulation studies to assess the proposed methodology in a setting where, in contrast to the *Add Health* study, monitoring times can occur at any time during follow-up, for every individual. Our aim was to estimate the causal marginal OR for a 1-unit increase in the exposure $D_i(t)$ on a categorical and ordinal outcome $Y_i(t)$. The outcome was categorical, taking one of three levels ($J = 3$): 1, 2, and 3. In simulation studies, we compared four estimators:

- The estimated log-OR for exposure obtained directly from the POM model, with no adjustment ($\hat{\beta}_{POM}$);
- The estimated log-OR for exposure from a weighted POM with an IPT weight, where the propensity score is a correctly specified function of the confounders ($\hat{\beta}_{IPTP}$);
- The estimated log-OR for exposure from a weighted POM with an IIV weight, where the intensity is a correctly specified function of the covariates affecting visit times ($\hat{\beta}_{IIVP}$); and
- The estimated log-OR for exposure from a doubly-weighted POM with both the IPT and the IIV weights, with both functions (corresponding to the exposure and visit models) correctly specified ($\hat{\beta}_{IPTMP}$).

In the following description of the data generating mechanism that we used, the individual index is omitted for ease of exposition. We used 1000 simulations per study and tested settings with either 250 or 1000

patients per simulated dataset. First, three confounders were simulated at time 0 ("cohort entry") and for each individual, as $K_1 \sim N(1,1)$, $K_2 \sim Bernoulli(0.55)$, and $K_3 \sim N(0,1)$. The time-varying, continuous exposure $D(t)$ was simulated at each time t from a Normal distribution with a mean that depended on the confounders, as $D(t) \sim N(-0.5 + 0.5K_1 + 1K_2 - 0.05K_3; 0.5^2)$ in the study with confounding, and $D(t) \sim N(-0.5; 0.5^2)$ in the simulation study with no confounding. The mediator of the relationship between $D(t)$ and $Y(t)$ was binary and time-varying, and simulated as $Z(t) \sim Bernoulli(p_D(t))$ with $p_D(t) = 0.3$ if $D(t) > 0.5$ and $p_D(t) = 0.8$ otherwise.

In the study, time was continuous, and discretized over a grid of 0.01. Visit times could vary across individuals, and could occur at anytime during follow-up. Monitoring times were simulated according to the monitoring intensity, as a function of $D(t)$ and $Z(t)$ and with an individual random effect η, such that $\lambda(t|D(t), Z(t)) = 0.01\eta \exp(\gamma_D D(t) + \gamma_Z Z(t))$. We varied the parameters $\boldsymbol{\gamma}$ to obtain different strengths for the dependence between the covariates and the monitoring times. The random effect was simulated as a random Gamma variable with mean 1 and variance 0.01. Monitoring indicators at each time were simulated according to a random Bernoulli draw, with a probability proportional to the intensity above.

The categorical outcome $Y(t)$ was simulated as a function of the exposure $D(t)$, the mediator $Z(t)$, and the confounders $\mathbf{K}(\mathbf{t})$. To simulate the outcome, we followed a methodology similar to Thomas (2014). We first simulated the linear mean function as $\mu(t|D(t), Z(t), \mathbf{K}(\mathbf{t})) = -2D(t) + 5Z(t) + 0.4K_1 + 0.05K_2 - 0.6K_3$. Then, a random draw from the logistic distribution was performed, according to the mean $\mu(t|D(t), Z(t), \mathbf{K}(\mathbf{t}))$ for the distribution. If that drawn value was smaller or equal to 5, the categorical outcome was set to 1. If the value was greater than 5 and smaller or equal to 8, it was set to 2. It was set to 3, otherwise. That choice of thresholds led to a good distribution across all three levels.

Given that we used the POM to estimate the marginal effect of exposure (which assumes that the relation between the linear predictors and the outcome, the link function, is the expit function), and given that some mediators make up part of the total effect of exposure on the outcome, the true marginal log-OR cannot be analytically derived solely by knowing the simulation parameter for the exposure in the outcome model. To know the true value (or *target* for an estimator) we conducted a Monte Carlo simulation in which we simulated the data of 10,000 patients a total of 1000 times where all parameters in the outcome model were kept as above, but there was no covariate-driven treatment or visit process such that there was no selection or confounding bias. We then computed the log-OR of exposure each time. In that very large study, no imbalances due to confounding or due to covariate-driven monitoring times were present. For that, we merely set all the parameters corresponding to the predictors in the exposure and in the monitoring models to zero. We obtained the target marginal log-OR by averaging the 1000 log-OR for exposure across those simulations,

which equalled to -1.061.

In a first sensitivity analysis (results presented in Appendix C.2), the confounders K_1 and K_2 along with the exposure and the mediator predicted the monitoring intensity. The monitoring intensity was simulated as $\lambda(t|D(t), Z(t), K_1, K_2) = 0.01\eta \exp(\gamma_D D(t) + \gamma_Z Z(t) + 0.05K_1 - 0.1K_2)$ with different sets of (γ_D, γ_Z) parameters, as tested in the main simulation study, and with all other parameters remaining the same.

In a second sensitivity analysis (results presented in Appendix C.2), we modified the mean outcome model to strenghten the bias due to covariate-driven monitoring times, using the linear mean function $\mu(t|D(t), Z(t), \mathbf{K(t)}) = -2D(t) - 6Z(t) + 0.4K_1 + 0.05K_2 - 0.6K_3$ instead of $\mu(t|D(t), Z(t), \mathbf{K(t)}) = -2D(t) + 5Z(t) + 0.4K_1 + 0.05K_2 - 0.6K_3$ (note the difference in coefficients for the mediator $Z(t)$). We also strenghtened the dependence of the monitoring intensity on the covariates by increasing the absolute values of the $\boldsymbol{\gamma}$ parameters. The target marginal log-OR was once again obtained via Monte Carlo simulations and equalled -1.394.

References

Thomas, A M. (2014). "The proportional odds model: Simulations studies and predictive accuracy," Research project, School of Statistics, University of Minnesota.

C.2 Results of the simulation study sensitivity analyses

Results of the simulation study sensitivity analysis 1 in which the confounders, the exposure and the mediator all affect the monitoring intensity

Table C.1: Comparison of four estimators for the marginal log-OR for 1-unit increase in $D_i(t)$ in the POM, for a sample size of $n = 250$ patients and 1000 simulations per study. Sensitivity analysis with confounding (Conf.=Y) in which the confounders also affect visit intensity.

Conf. (Y/N)	$\boldsymbol{\gamma}$	Mean no. visits (min-max)	Absolute empirical bias				Empirical variance			
			IPTMP	IIVP	IPTP	POM	IPTMP	IIVP	IPTP	POM
Y	(-0.3, 0.1)	2 (0-14)	0.08	0.28	0.08	0.28	0.27	0.02	0.27	0.02
	(-0.1, 0.3)	3 (0-13)	0.10	0.29	0.15	0.30	0.20	0.01	0.21	0.01
	(0, 0)	2 (0-13)	0.10	0.29	0.10	0.29	0.21	0.02	0.21	0.02
	(0.1, 0.2)	3 (0-14)	0.10	0.29	0.13	0.30	0.20	0.01	0.20	0.01
	(0.1, 0.5)	3 (0-16)	0.10	0.29	0.19	0.32	0.18	0.01	0.17	0.01
	(0.2, 1)	5 (0-22)	0.05	0.28	0.29	0.35	0.17	0.01	0.14	0.01
	(0.8, 0.4)	5 (0-49)	0.05	0.28	0.17	0.29	0.18	0.02	0.17	0.01

Results of the simulation study sensitivity analysis 2 in which we changed the mean outcome model and tested another set of parameters in the monitoring intensity model so as to increase the bias due to covariate-driven monitoring times

Table C.2: Comparison of four estimators for the marginal log-OR for 1-unit increase in $D_i(t)$ in the POM, for a sample size of $n = 250$ patients and 1000 simulations per study. Sensitivity analysis with confounding (Conf.=Y) in which we changed the mean outcome model.

Conf. (Y/N)	$\boldsymbol{\gamma}$	Mean no. visits (min-max)	Absolute empirical bias				Empirical variance			
			IPTMP	IIVP	IPTP	POM	IPTMP	IIVP	IPTP	POM
Y	(-1,2)	10 (0-142)	0.04	0.41	0.64	0.77	0.47	0.10	0.38	0.05
	(1, -1)	4 (0-58)	0.10	0.47	0.29	0.06	0.73	0.16	0.38	0.06
	(1, -2)	4 (0-58)	0.02	0.36	0.40	0.10	0.80	0.28	0.43	0.06
	(0, 0)	2 (0-13)	0.10	0.47	0.10	0.47	0.45	0.07	0.45	0.07
	(-1.5, -0.5)	2 (0-63)	0.02	0.41	0.33	0.63	0.64	0.16	0.79	0.05
	(0.5, 0.5)	4 (0-24)	0.16	0.50	0.08	0.48	0.64	0.07	0.46	0.05
	(-2, 2)	15 (0-247)	0.17	0.22	0.39	0.27	0.65	0.26	0.79	0.05

C.3 Imputation models for missing covariates

In the application to the *Add Health* study, covariates were imputed using multivariate imputations by chained equations (MICE), using fully conditional specification, and starting with the covariates with the

least missing values, to the most missing values in the condition in the imputation models. The rates of missing values, calculated as the number of missing values in the vector, divided by the length of each covariate vector (for all 4 waves stacked together) are: Sex: 0.0%; Age: 0.1%; Race: 0.3%; ENG: 1.6%; MATH: 3.6%; HSS: 5.4%; GS: 6.2%; LGR: 18.1%; RSBI: 18.1% Feeling depressed: 18.1%; Video games: 18.9%; SES: 24.6%; Smoking: 43.2%; HOF: 56.4%; FHTR: 56.5 %; FCA 56.7%. To assess the validity of the imputation and model fits, we used strip plots and density plots. The imputation models provided a good fit, with similar distributions between the observed covariates and the fitted covariates distributions. The strip plots and density plots looked very similar between the observed and fitted distributions for all covariates (except for the variable sex, which fitted distribution was based on only 4 missing values, and for the variable age which presented with slight deviations between the observed and fitted distributions).

C.4 Balance in covariates across monitored and unmonitored observations, before and after IIV-weighting (one imputed dataset used)

In Table C.3 we present the mean (SD) or the proportions for each covariate, stratified by the outcome being monitored ($dN = 1$) or not being monitored ($dN = 0$). The standardized means differences (SMD) are shown. The results are shown before and after weighting the observations with the fitted inverse intensity of visit weight. Observations corresponding to the outcome not being monitored have been reweighted by the inverse of the intensity of not being monitored, and the observations corresponding to the outcome being monitored have been reweighted by the inverse of the intensity of being monitored.

Table C.3: Comparison of covariates after imputation, stratified by monitored and unmonitored observations, before and after inverse intensity of visit weighting. Continuous variables (or numerical categorical variables with several categories) are described using the mean (SD) and other variables are described with proportions. Standardized mean differences (SMD) are shown.

Variable	No IIV-weighting			After IIV-weighting		
	$dN = 0$	$dN = 1$	SMD	$dN = 0$	$dN = 1$	SMD
Number of hours spent on video or computer games	3.9 (8.5)	3.2 (7.4)	0.08	3.8 (8.4)	3.5 (8.1)	0.04
Frequency of feeling depressed			0.06			0.02
Never or rarely	69.1	66.8		68.7	68.5	
Sometimes	23.9	25.3		24.1	24.3	
A lof of the time	5.2	5.4		5.2	5.0	
Most of the time or all the time	1.9	2.5		1.9	2.2	
Age	22.8 (4.8)	20.3 (5.8)	0.46	22.4 (4.8)	22.9 (6.1)	0.10
Sex (Female)	0.50	0.50	0.15	0.50	0.50	0.08
SES	6.4 (2.1)	6.6 (2.1)	0.10	6.5 (2.1)	6.6 (2.1)	0.05
Race			0.12			0.07
White	61.9	67.3		62.4	65.5	
Black/African American	26.3	23.0		26.0	24.1	
American Indian/Alaskan Native	1.2	1.2		1.2	1.2	
Asian/Pacific Islander	3.8	3.2		3.7	3.3	
Other	6.8	5.3		6.7	5.9	
FHTR	0.89	0.91	0.01	0.90	0.91	0.04
LGR	0.76	0.77	0.01	0.76	0.77	0.00
RSBI	0.30	0.30	0.01	0.30	0.30	0.01
MATH	0.92	0.96	0.03	0.92	0.93	0.02
ENG	0.84	0.87	0.05	0.84	0.84	0.03
HSS	0.92	0.94	0.04	0.93	0.91	0.02
GS	0.93	0.94	0.07	0.93	0.92	0.04
HOF	3.0 (1.0)	3.0 (1.0)	0.02	3.0 (1.0)	3.0 (1.0)	0.01
FCA	0.88	0.87	0.02	0.88	0.90	0.03
Days smoking cigarettes	8.0 (10.4)	8.2 (10.8)	0.02	8.1 (10.4)	8.1 (10.9)	0.00

C.5 Results from the sensitivity analyses: Comparison of the marginal effect of the time spent playing video games on the number of suicide attempts

Table C.4: Analysis S1. Comparison of four estimators for the marginal OR for 1-hour or 10-hour increases in the time spent playing video games per week, on the odds of suicide attempts (number of attempts categorized in 0, 1, or more), *Add Health* study, United States, 1994-2008, $n = 6504$.

Estimator	1-hour increase OR (Bootstrap 95% CI)	10-hour increase OR (Bootstrap 95% CI)
$\hat{\beta}_{POM}$	0.99 (0.97, 1.01)	0.93 (0.70, 1.06)
$\hat{\beta}_{IPTP}$	1.00 (0.97, 1.01)	1.01 (0.77, 1.15)
$\hat{\beta}_{IIVP}$	1.00 (0.97, 1.01)	0.97 (0.75, 1.09)
$\hat{\beta}_{IPTMP}$	1.01 (0.99, 1.02)	1.11 (0.88, 1.26)

Table C.5: Analysis S2. Comparison of four estimators for the marginal OR for 1-hour or 10-hour increases in the time spent playing video games per week, on the odds of suicide attempts (number of attempts categorized in 0, 1, or more), *Add Health* study, United States, 1994-2008, $n = 6504$.

Estimator	1-hour increase OR (Bootstrap 95% CI)	10-hour increase OR (Bootstrap 95% CI)
$\hat{\beta}_{POM}$	0.99 (0.97, 1.01)	0.93 (0.70, 1.07)
$\hat{\beta}_{IPTP}$	1.01 (0.98, 1.02)	1.07 (0.84, 1.20)
$\hat{\beta}_{IIVP}$	1.00 (0.97, 1.01)	0.97 (0.76, 1.09)
$\hat{\beta}_{IPTMP}$	1.01 (0.99, 1.02)	1.08 (0.87, 1.20)

Table C.6: Analysis S3. Comparison of four estimators for the marginal OR for a two-fold or 8-fold increases in the time spent playing video games per week, on the odds of suicide attempts (number of attempts categorized in 0, 1, or more), *Add Health* study, United States, 1994-2008, $n = 6504$.

Estimator	2-fold OR (Bootstrap 95% CI)	8-fold OR (Bootstrap 95% CI)
$\hat{\beta}_{POM}$	0.91 (0.82, 0.99)	0.76 (0.55, 0.98)
$\hat{\beta}_{IPTP}$	1.00 (0.88, 1.09)	0.99 (0.69, 1.30)
$\hat{\beta}_{IIVP}$	0.95 (0.85, 1.04)	0.86 (0.61, 1.11)
$\hat{\beta}_{IPTMP}$	1.03 (0.91, 1.13)	1.09 (0.74, 1.44)

C.6 Results from the main analysis and the sensitivity analyses: Graphs of the estimated probability of 1 or more, or of 2 or more suicide attempts, by the number of hours spent playing video games weekly

Main analysis with the log-OR estimated for a 1-unit increase in $\log_2$ (number of hours spent playing video games+1) and the use of a generalized inverse probability of treatment weight for modelling $\log_2$ (number of hours spent playing video games+1) as a continuous variable, along with the 95% confidence interval bands added

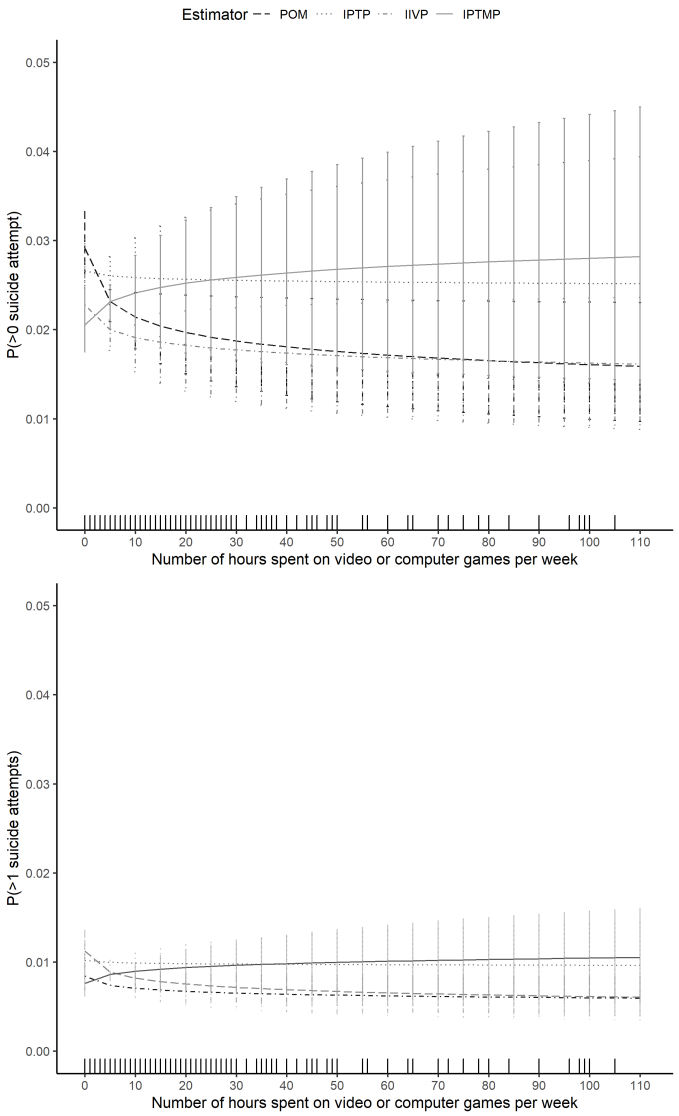


Figure C.1: Main analysis. Probability of 1 or more suicide attempts (top panel) or of 2 or more suicide attempts (bottom) according to the number of hours spent playing video games per week. Comparison of four estimators for the marginal log-OR. The bands around the point estimates correspond to 95% CIs computed using the bootstrap percentiles. The rug plot on the X-axis shows the different values of the number of hours spent playing video games in the study cohort, up to 110 hours per week.

Sensitivity analysis 1 (S1): The number of hours spent playing video games as a continuous exposure incorporated in the outcome model, and the use of a generalized inverse probability of treatment weight for the continuous exposure (under the normality assumption for the number of hours spent playing video games weekly)

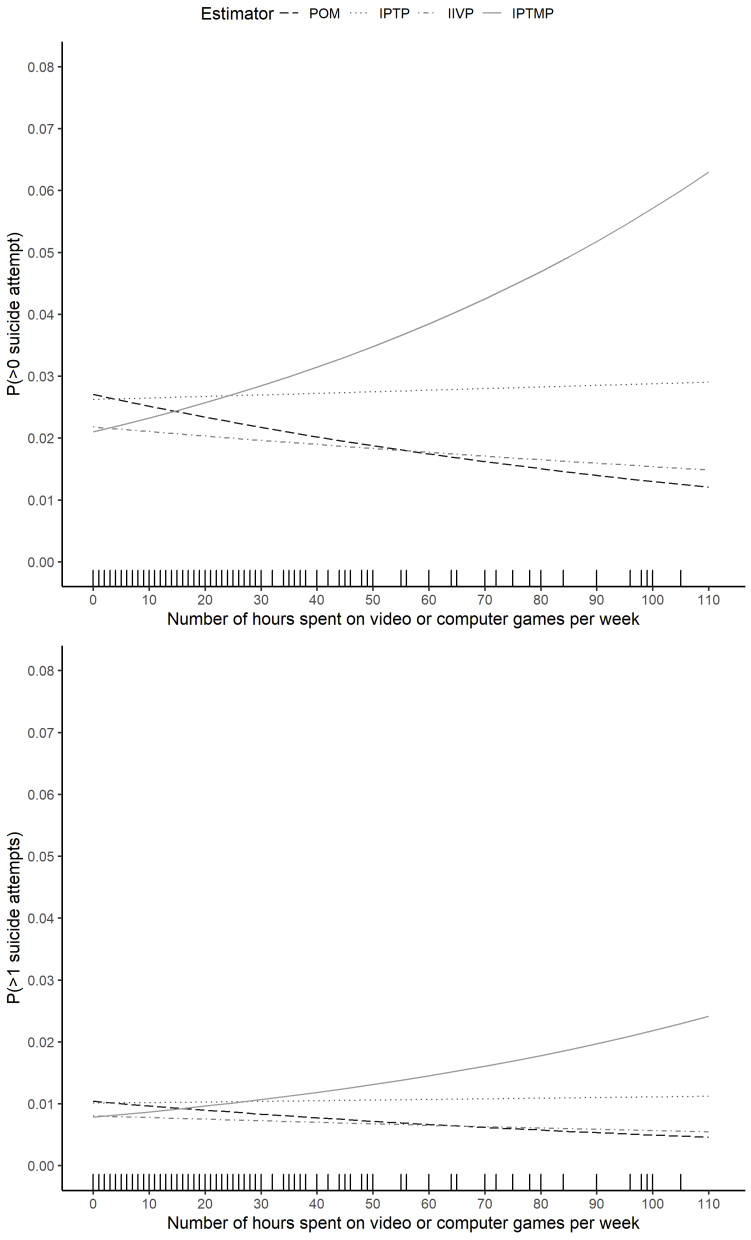


Figure C.2: Analysis S1. Probability of 1 or more suicide attempts (top panel) or of 2 or more suicide attempts (bottom) according to the number of hours spent playing video games per week. Comparison of four estimators for the marginal log-OR. The rug plot on the X-axis shows the different values of the number of hours spent playing video games in the study cohort, up to 110 hours per week.

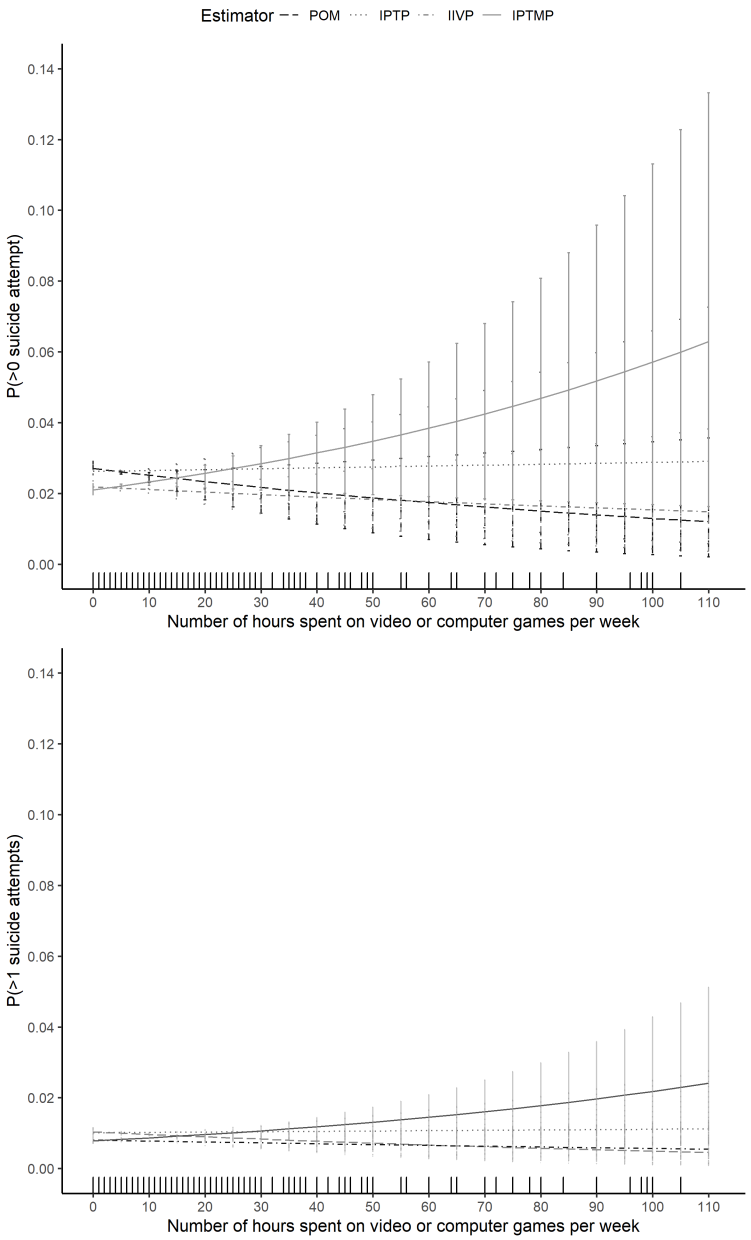


Figure C.3: Analysis S1. Probability of 1 or more suicide attempts (top panel) or of 2 or more suicide attempts (bottom) according to the number of hours spent playing video games per week. Comparison of four estimators for the marginal log-OR. The bands around the point estimates correspond to 95% CIs computed using the bootstrap percentiles. The rug plot on the X-axis shows the different values of the number of hours spent playing video games in the study cohort, up to 110 hours per week.

Sensitivity analysis 2 (S2): The number of hours spent playing video games as a continuous exposure incorporated in the outcome model, and the use of a generalized inverse probability of treatment weight for the categorical exposure (five categories)

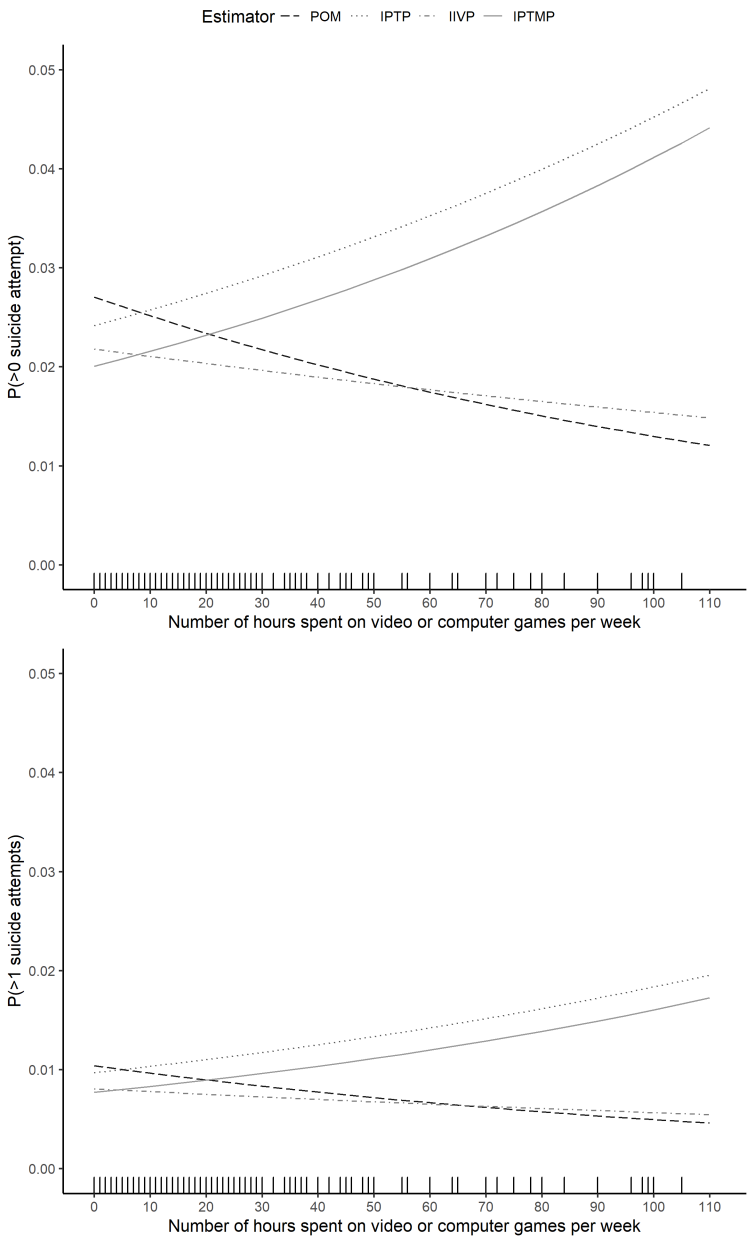


Figure C.4: Analysis S2. Probability of 1 or more suicide attempts (top panel) or of 2 or more suicide attempts (bottom) according to the number of hours spent playing video games per week. Comparison of four estimators for the marginal log-OR. The rug plot on the X-axis shows the different values of the number of hours spent playing video games in the study cohort, up to 110 hours per week.

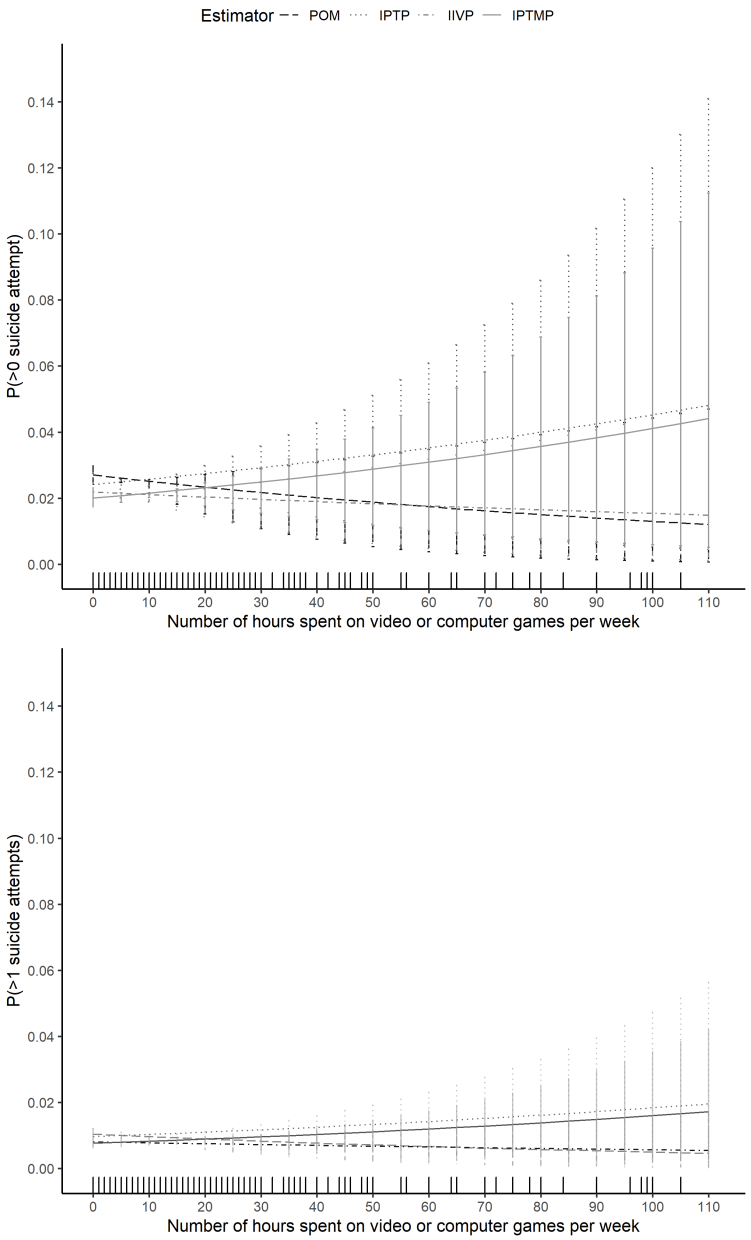


Figure C.5: Analysis S2. Probability of 1 or more suicide attempts (top panel) or of 2 or more suicide attempts (bottom) according to the number of hours spent playing video games per week. Comparison of four estimators for the marginal log-OR. The bands around the point estimates correspond to 95% CIs computed using the bootstrap percentiles. The rug plot on the X-axis shows the different values of the number of hours spent playing video games in the study cohort, up to 110 hours per week.

Sensitivity analysis 3 (S3): The $\log_2$(number of hours spent playing video games+1) as a continuous exposure incorporated in the outcome model, and the use of a generalized inverse probability of treatment weight for the categorical exposure (five categories)

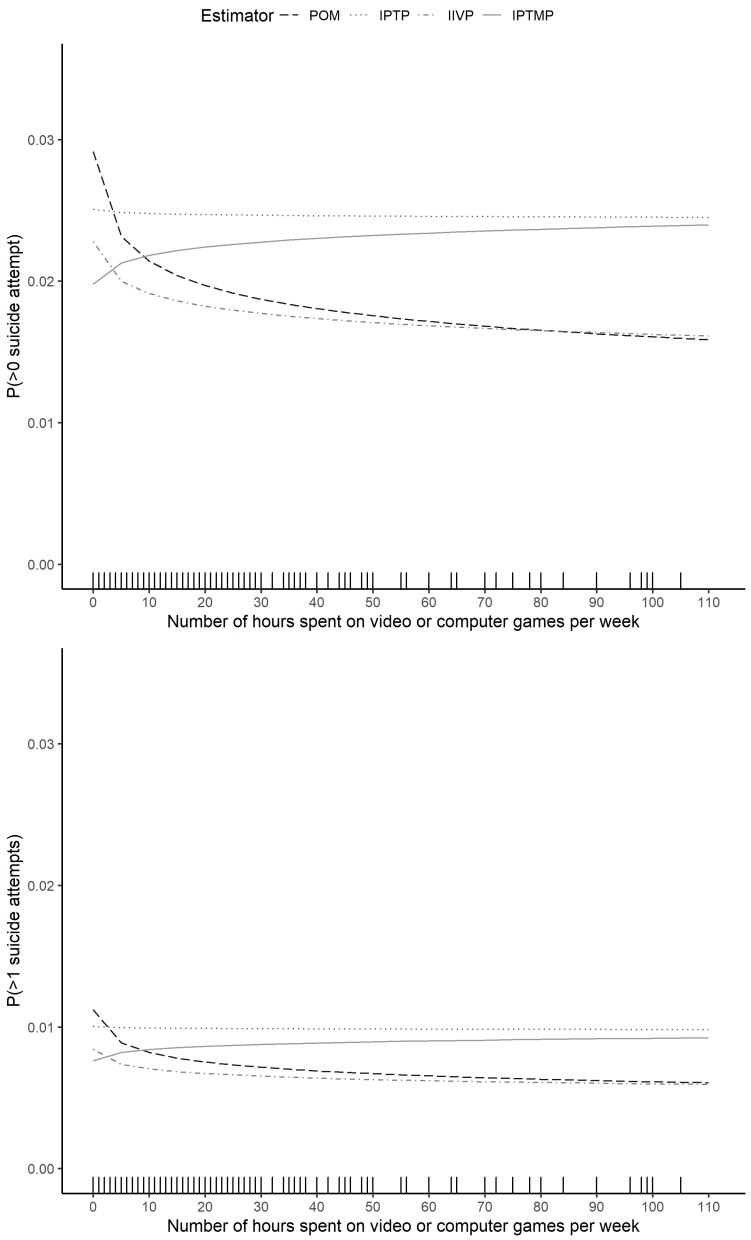


Figure C.6: Analysis S3. Probability of 1 or more suicide attempts (top panel) or of 2 or more suicide attempts (bottom) according to the number of hours spent playing video games per week. Comparison of four estimators for the marginal log-OR. The rug plot on the X-axis shows the different values of the number of hours spent playing video games in the study cohort, up to 110 hours per week.

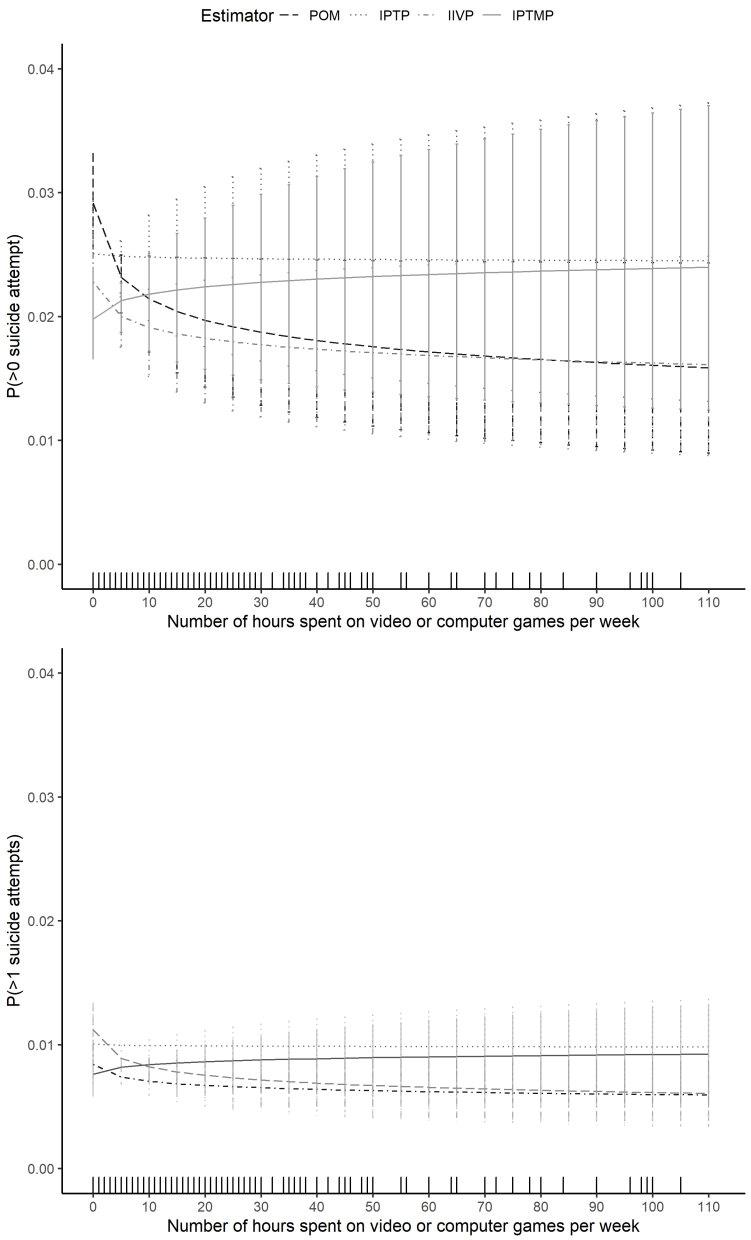


Figure C.7: Analysis S3. Probability of 1 or more suicide attempts (top panel) or of 2 or more suicide attempts (bottom) according to the number of hours spent playing video games per week. Comparison of four estimators for the marginal log-OR. The bands around the point estimates correspond to 95% CIs computed using the bootstrap percentiles. The rug plot on the X-axis shows the different values of the number of hours spent playing video games in the study cohort, up to 110 hours per week.

References

Jennifer Ahern, Alan Hubbard, and Sandro Galea. Estimating the effects of potential public health interventions on population disease burden: a step-by-step illustration of causal inference methods. *American Journal of Epidemiology*, 169(9):1140–1147, 2009.

Leila D Amorim and Jianwen Cai. Modelling recurrent events: a tutorial for analysis in epidemiology. *International Journal of Epidemiology*, 44(1):324–333, 2015.

Per Kragh Andersen and Richard D Gill. Cox's regression model for counting processes: a large sample study. *The Annals of Statistics*, 10(4):1100–1120, 1982.

Craig A Anderson and Catherine M Ford. Affect of the game player: short-term effects of highly and mildly aggressive video games. *Personality and Social Psychology Bulletin*, 12 (4):390–402, 1986.

Peter M Aronow. A general method for detecting interference between units in randomized experiments. *Sociological Methods & Research*, 41(1):3–16, 2012.

Bruno Arpino and Alessandra Mattei. Assessing the causal effects of financial aids to firms in Tuscany allowing for interference. *The Annals of Applied Statistics*, 10(3):1170–1194, 2016.

Peter C Austin. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46(3):399–424, 2011.

Peter C Austin. Assessing covariate balance when using the generalized propensity score with quantitative or continuous exposures. *Statistical Methods in Medical Research*, 28(5): 1365–1377, 2019.

Hailey R Banack and Jay S Kaufman. From bad to worse: collider stratification amplifies confounding bias in the "obesity paradox". *European Journal of Epidemiology*, 30(10): 1111–1114, 2015.

Petra Bůžková and Thomas Lumley. Marginal regression modeling under irregular, biased sampling. 2005. UW Biostatistics Working Paper Series. Working Paper 261. http://biostats.bepress.com/uwbiostat/paper261.

Petra Bůžková and Thomas Lumley. Semiparametric modeling of repeated measurements under outcome-dependent follow-up. *Statistics in Medicine*, 28(6):987–1003, 2009.

Na Cai, Wenbin Lu, and Hao Helen Zhang. Time-varying latent effect model for longitudinal data with informative observation times. *Biometrics*, 68(4):1093–1102, 2012.

Hongyuan Cao, Jialiang Li, and Jason P Fine. On last observation carried forward and asynchronous longitudinal regression analysis. *Electronic Journal of Statistics*, 10(1):1155–1180, 2016.

Vincent J Carey. *gee: Generalized Estimation Equation Solver*, 2019. Ported to R by Thomas Lumley and Brian Ripley. https://CRAN.R-project.org/package=gee. R package version 4.13-20.

Gary Chamberlain. Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34(3):305–334, 1987.

Richard J Cook and Jerald F Lawless. *The statistical analysis of recurrent events*. New York: Springer Science & Business Media, 2007.

Richard J Cook and Jerald F Lawless. *Multistate models for the analysis of life history data*. Boca Raton: CRC Press, 2018.

Janie Coulombe, Erica EM Moodie, and Robert W Platt. Weighted regression analysis to correct for informative monitoring times and confounders in longitudinal studies. *Biometrics*, 77(1):162–174, 2021a.

Janie Coulombe, Erica EM Moodie, Susan Shortreed, and Christel Renoux. Can the risk of severe depression-related outcomes be reduced by tailoring the antidepressant therapy to patient characteristics? *American Journal of Epidemiology*, 190(7):1210–1219, 2021b.

Janie Coulombe, Erica EM Moodie, and Robert W Platt. 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*, forthcoming.

David R Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972.

Richard K Crump, V Joseph Hotz, Guido W Imbens, and Oscar A Mitnik. Dealing with limited overlap in estimation of average treatment effects. *Biometrika*, 96(1):187–199, 2009.

Hongsheng Dai and Jianxin Pan. Joint modelling of survival and longitudinal data with informative observation times. *Scandinavian Journal of Statistics*, 45(3):571–589, 2018.

Carlos De las Cuevas, Wenceslao Peñate, and Emilio J Sanz. Risk factors for non-adherence to antidepressant treatment in patients with mood disorders. *European Journal of Clinical Pharmacology*, 70(1):89–98, 2014.

Iain Deas, Brian Robson, Cecilia Wong, and Michael Bradford. Measuring neighbourhood deprivation: a critique of the Index of Multiple Deprivation. *Environment and Planning C: Government and Policy*, 21(6):883–903, 2003.

Laura Forastiere, Edoardo M Airoldi, and Fabrizia Mealli. Identification and estimation of treatment and interference effects in observational studies on networks. *Journal of the American Statistical Association*, forthcoming.

Justin Gagnon, Marie-Thérèse Lussier, Brenda MacGibbon, Stella S Daskalopoulou, and Gillian Bartlett. The impact of antidepressant therapy on glycemic control in Canadian primary care patients with Diabetes Mellitus. *Frontiers in Nutrition*, 5(47):1–10, 2018.

Melissa M Garrido, Amy S Kelley, Julia Paris, Katherine Roza, Diane E Meier, R Sean Morrison, and Melissa D Aldridge. Methods for constructing and assessing propensity scores. *Health Services Research*, 49(5):1701–1720, 2014.

Richard D Gill and Soren Johansen. A survey of product-integration with a view toward application in survival analysis. *The Annals of Statistics*, 18(4):1501–1555, 1990.

Gary S Goldfield, Marisa Murray, Danijela Maras, Angela L Wilson, Penny Phillips, Glen P Kenny, Stasia Hadjiyannakis, Angela Alberga, Jameason D Cameron, Heather Tulluch, et al. Screen time is associated with depressive symptomatology among obese adolescents: a HEARTY study. *European Journal of Pediatrics*, 175(7):909–919, 2016.

Sander Greenland. Randomization, statistics, and causal inference. *Epidemiology*, 1(6): 421–429, 1990.

Sander Greenland. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*, 14(3):300–306, 2003.

Sander Greenland and Hal Morgenstern. Confounding in health research. *Annual Review of Public Health*, 22(1):189–212, 2001.

Sander Greenland and James M Robins. Identifiability, exchangeability, and epidemiological confounding. *International Journal of Epidemiology*, 15(3):413–419, 1986.

Sander Greenland, Judea Pearl, and James M Robins. Causal diagrams for epidemiologic research. *Epidemiology*, 10(1):37–48, 1999.

Brandon Greenwell, Andrew McCarthy, Brad Boehmke, Dungang Liu, and Maintainer Brandon Greenwell. *Package 'sure'*, 2017. URL https://mran.microsoft.com/snapshot/2017-11-25/web/packages/sure/sure.pdf. R package version 3.1.

Brandon M Greenwell, Andrew J McCarthy, Bradley C Boehmke, and Dungang Liu. Residuals and diagnostics for binary and ordinal regression models: an introduction to the sure package. *The R Journal.*, 10(1):381–394, 2018.

Major Greenwood. The natural duration of cancer. *Reports on Public Health and Medical Subjects, London: His Majesty's Stationery Office*, (33):1–26, 1926.

Rolf HH Groenwold, Jonathan AC Sterne, Debbie A Lawlor, Karel GM Moons, Arno W Hoes, and Kate Tilling. Sensitivity analysis for the effects of multiple unmeasured confounders. *Annals of Epidemiology*, 26(9):605–611, 2016.

Neil E Grunberg. Nicotine, cigarette smoking, and body weight. *British Journal of Addiction*, 80(4):369–377, 1985.

Ulrich Halekoh, Søren Højsgaard, and Jun Yan. The R package geepack for generalized estimating equations. *Journal of Statistical Software*, 15(2):1–11, 2006.

Sebastian Haneuse and Andrea Rotnitzky. Estimation of the effect of interventions that modify the received treatment. *Statistics in Medicine*, 32(30):5260–5277, 2013.

Kathleen Mullan Harris. The National Longitudinal Study of Adolescent to Adult Health (Add Health), Waves I & II, 1994–1996; Wave III, 2001–2002; Wave IV, 2007-2009 [machine-readable data file and documentation]. *Chapel Hill: Carolina Population Center, University of North Carolina at Chapel Hill*, 2009.

Kathleen Mullan Harris. The Add Health study: design and accomplishments. *Chapel Hill: Carolina Population Center, University of North Carolina at Chapel Hill*, pages 1–22, 2013.

Kathleen Mullan Harris and J Richard Udry. The National Longitudinal Study of Adolescent to Adult Health (Add Health), Waves I & II, 1994–1996; Wave III, 2001–2002; Wave IV, 2007-2009 [machine-readable data file and documentation]. *Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/ICPSR21600.v21 (accessed May 6, 2018)*, 2018.

Sean Hennessy. Use of health care databases in pharmacoepidemiology. *Basic & Clinical Pharmacology & Toxicology*, 98(3):311–313, 2006.

Miguel A Hernán and James M Robins. Estimating causal effects from epidemiological data. *Journal of Epidemiology & Community Health*, 60(7):578–586, 2006.

Miguel A Hernán and James M Robins. *Causal Inference*. Boca Raton: CRC Press, 2016.

Miguel A Hernán, Babette Brumback, and James M Robins. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*, 11(5):561–570, 2000.

Miguel A Hernán, Sonia Hernández-Díaz, and James M Robins. A structural approach to selection bias. *Epidemiology*, 15(5):615–625, 2004.

Emily Herrett, Arlene M Gallagher, Krishnan Bhaskaran, Harriet Forbes, Rohini Mathur, Tjeerd Van Staa, and Liam Smeeth. Data resource profile: Clinical Practice Research Datalink (CPRD). *International Journal of Epidemiology*, 44(3):827–836, 2015.

Paul W Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960, 1986.

Daniel G Horvitz and Donovan J Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260): 663–685, 1952.

Claire Houtsma. A longitudinal investigation of the effect of violent video game play on capability for suicide. 2017. Master's thesis, University of Southern Mississippi.

Peter J Huber. Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35:73–101, 1964.

Peter J Huber. The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 221–233. University of California Press, 1967.

Michael G Hudgens and M Elizabeth Halloran. Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842, 2008.

Kosuke Imai and David A Van Dyk. Causal inference with general treatment regimes: generalizing the propensity score. *Journal of the American Statistical Association*, 99 (467):854–866, 2004.

Daniel Johnson, Christian Jones, Laura Scholes, and Michelle Colder Carras. Videogames and wellbeing: a comprehensive review. 2013. Gaming Research Group, Young and Well Cooperative Reseach Centre, Melbourne, 2013.

William D Kalsbeek, Juan Yang, and Robert P Agans. Predictors of nonresponse in a longitudinal survey of adolescents. 2020. Presented at: Annual Meeting of the American Statistical Association, Joint Statistical Meetings, 2002, Alexandria, Virginia, United States.

Edward L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, 1958.

Yasser Khazaal, Mathias Van Singer, Anne Chatton, Sophia Achab, Daniele Zullino, Stephane Rothen, Riaz Khan, Joel Billieux, and Gabriel Thorens. Does self-selection affect samples' representativeness in online surveys? An investigation in online video game research. *Journal of Medical Internet Research*, 16(7):e164, 2014.

Kwangho Kim, Edward H Kennedy, and Ashley I Naimi. Incremental intervention effects in studies with many timepoints, repeated outcomes, and dropout. *arXiv preprint arXiv:1907.04004*, 2019.

Timothy L Lash and Aliza K Fink. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology*, 14(4):451–458, 2003.

Jerald F Lawless and Claude Nadeau. Some simple robust methods for the analysis of recurrent events. *Technometrics*, 37(2):158–168, 1995.

David Lewis. Counterfactuals and comparative possibility. *Journal of Philosophical Logic*, 2, 1973.

Kung-Yee Liang and Scott L Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

Yu Liang, Wenbin Lu, and Zhiliang Ying. Joint modeling and analysis of longitudinal data with informative observation times. *Biometrics*, 65(2):377–384, 2009.

Danyu Y Lin and Zhiliang Ying. Semiparametric and nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association*, 96(453):103–126, 2001.

Danyu Y Lin, Lee-Jen Wei, I Yang, and Zhiliang Ying. Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):711–730, 2000.

Haiqun Lin, Daniel O Scharfstein, and Robert A Rosenheck. Analysis of longitudinal data

with irregular, outcome-dependent follow-up. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):791–813, 2004.

James K Lindsey. *Statistical Analysis of Stochastic Processes in Time*, volume 14. Cambridge University Press, 2004.

Stuart R Lipsitz, Garrett M Fitzmaurice, Joseph G Ibrahim, Richard Gelber, and Steven Lipshultz. Parameter estimation in longitudinal studies with outcome-dependent follow-up. *Biometrics*, 58(3):621–630, 2002.

Armend Lokku, Lily S Lim, Catherine S Birken, and Eleanor M Pullenayegum. Summarizing the extent of visit irregularity in longitudinal data. *BMC Medical Research Methodology*, 20(1):1–9, 2020.

Danijela Maras, Martine F Flament, Marisa Murray, Annick Buchholz, Katherine A Henderson, Nicole Obeid, and Gary S Goldfield. Screen time is associated with depression and anxiety in Canadian youth. *Preventive Medicine*, 73:133–138, 2015.

Lawrence C McCandless, Paul Gustafson, and Adrian Levy. Bayesian sensitivity analysis for unmeasured confounding in observational studies. *Statistics in Medicine*, 26(11):2331–2347, 2007.

Peter McCullagh. Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2):109–127, 1980.

Peter McCullagh. Quasi-likelihood functions. *The Annals of Statistics*, 11(1):59–67, 1983.

Charles E McCulloch, John M Neuhaus, and Rebecca L Olin. Biased and unbiased estimation in longitudinal studies with informative visit processes. *Biometrics*, 72(4):1315–1324, 2016.

Erick Messias, Juan Castro, Anil Saini, Manzoor Usman, and Dale Peeples. Sadness, suicide, and their association with video game and internet overuse among teens: results from the

youth risk behavior survey 2007 and 2009. *Suicide and Life-Threatening Behavior*, 41(3): 307–315, 2011.

Erica EM Moodie and David A Stephens. Comment: Clarifying endogeneous data structures and consequent modelling choices using causal graphs. *Statistical Science*, 35(3):391–393, 2020.

Iván Díaz Muñoz and Mark van der Laan. Population intervention causal effects based on stochastic interventions. *Biometrics*, 68(2):541–549, 2012.

Ashley I Naimi, Erica EM Moodie, Nathalie Auger, and Jay S Kaufman. Constructing inverse probability weights for continuous exposures: a comparison of methods. *Epidemiology*, 25 (2):292–299, 2014.

KW Newey and D McFadden. Large sample estimation and hypothesis. *Handbook of Econometrics*, 4:2112–2245, 1994.

Whitney K Newey. Adaptive estimation of regression models via moment restrictions. *Journal of Econometrics*, 38(3):301–339, 1988.

Jerzy S Neyman. On the application of probability theory to agricultural experiments. Essay on principles. Section 9 (translation published in 1990). *Statistical Science*, 5(4):472–480, 1923.

Ann-Kathrin Ozga, Meinhard Kieser, and Geraldine Rauch. A systematic comparison of recurrent event models for application to composite endpoints. *BMC Medical Research Methodology*, 18(1):1–12, 2018.

Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.

Judea Pearl. Graphs, causality, and structural equation models. *Sociological Methods & Research*, 27(2):226–284, 1998.

Judea Pearl. Causal inference in statistics: an overview. *Statistics Surveys*, 3:96–146, 2009.

Margaret Sullivan Pepe and Jianwen Cai. Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *Journal of the American Statistical Association*, 88(423):811–820, 1993.

Ross L Prentice, Benjamin J Williams, and Arthur V Peterson. On the regression analysis of multivariate failure time data. *Biometrika*, 68(2):373–379, 1981.

Eleanor M Pullenayegum. Multiple outputation for the analysis of longitudinal data subject to irregular observation. *Statistics in Medicine*, 35(11):1800–1818, 2016.

Eleanor M Pullenayegum. Meeting the assumptions of inverse-intensity weighting for longitudinal data subject to irregular follow-up: suggestions for the design and analysis of clinic-based cohort studies. *Epidemiologic Methods*, 9(1):1–17, 2020.

Eleanor M Pullenayegum and Brian M Feldman. Doubly robust estimation, optimally truncated inverse-intensity weighting and increment-based methods for the analysis of irregularly observed longitudinal data. *Statistics in Medicine*, 32(6):1054–1072, 2013.

Eleanor M Pullenayegum and Lily SH Lim. 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):2992–3014, 2016.

James M Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7(9-12):1393–1512, 1986.

James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90(429):106–121, 1995.

James M Robins, Miguel A Hernán, and Babette Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560, 2000a.

James M Robins, Andrea Rotnitzky, and Daniel O Scharfstein. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, pages 1–94. New York: Springer, 2000b.

Paul R Rosenbaum. The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society: Series A (General)*, 147(5):656–666, 1984.

Paul R Rosenbaum. Model-based direct adjustment. *Journal of the American Statistical Association*, 82(398):387–394, 1987.

Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974.

Donald B Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

Donald B Rubin. *Multiple imputation for nonresponse in surveys*, volume 81. Hoboken New Jersey: John Wiley & Sons, 2004.

Donald B Rubin. Causal inference using potential outcomes: design, modeling, decisions. *Journal of the American Statistical Association*, 100(469):322–331, 2005.

Duchwan Ryu, Debajyoti Sinha, Bani Mallick, Stuart R Lipsitz, and Steven E Lipshultz. Longitudinal studies with outcome-dependent follow-up: models and Bayesian regression. *Journal of the American Statistical Association*, 102(479):952–961, 2007.

Martin Saveski, Jean Pouget-Abadie, Guillaume Saint-Jacques, Weitao Duan, Souvik Ghosh, Ya Xu, and Edoardo M Airoldi. Detecting network effects: randomizing over randomized experiments. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1027–1035, 2017.

Sebastian Schneeweiss. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiology and Drug Safety*, 15(5):291–303, 2006.

Juliana Schulz and Erica EM Moodie. Doubly robust estimation of optimal dosing strategies. *Journal of the American Statistical Association*, 116(533):256–268, 2021.

Alessandro Serretti and Laura Mandelli. Antidepressants and body weight: a comprehensive review and meta-analysis. *The Journal of Clinical Psychiatry*, 71(10):1259–1272, 2010.

Zach Shahn, Ying Li, Zhaonan Sun, Amrita Mohan, Cristina Sampaio, and Jianying Hu. G-computation and hierarchical models for estimating multiple causal effects from observational disease registries with irregular visits. *AMIA Summits on Translational Science Proceedings*, 2019:789, 2019.

Ian Shrier and Robert W Platt. Reducing bias through directed acyclic graphs. *BMC Medical Research Methodology*, 8(1):1–15, 2008.

Leonard A Stefanski and Dennis D Boos. The calculus of M-estimation. *The American Statistician*, 56(1):29–38, 2002.

Lenka Stepankova, Eva Kralikova, Kamila Zvolska, Alexandra Pankova, Petra Ovesna, Milan Blaha, and Leonie S Brose. Depression and smoking cessation: evidence from a smoking cessation clinic with 1-year follow-up. *Annals of Behavioral Medicine*, 51(3):454–463, 2017.

Adam J Streeter, Nan Xuan Lin, Louise Crathorne, Marcela Haasova, Christopher Hyde, David Melzer, and William E Henley. Adjusting for unmeasured confounding in nonran-

domized longitudinal studies: a methodological review. *Journal of Clinical Epidemiology*, 87:23–34, 2017.

Margaret Sullivan Pepe and Garnet L Anderson. A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics - Simulation and Computation*, 23(4):939–951, 1994.

Jianguo Sun, Do-Hwan Park, Liuquan Sun, and Xingqiu Zhao. Semiparametric regression analysis of longitudinal data with informative observation times. *Journal of the American Statistical Association*, 100(471):882–889, 2005.

Jianguo Sun, Xingwei Tong, and Xin He. Regression analysis of panel count data with dependent observation times. *Biometrics*, 63(4):1053–1059, 2007.

Liuquan Sun, Xinyuan Song, and Jie Zhou. 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):2902–2919, 2011.

Liuquan Sun, Xinyuan Song, Jie Zhou, and Lei Liu. Joint analysis of longitudinal data with informative observation times and a dependent terminal event. *Journal of the American Statistical Association*, 107(498):688–700, 2012.

Norman Sussman and David Ginsberg. Rethinking side effects of the selective serotonin reuptake inhibitors: sexual dysfunction and weight gain. *Psychiatric Annals*, 28(2):89–97, 1998.

Kay See Tan, Benjamin French, and Andrea B Troxel. Regression modeling of longitudinal data with outcome-dependent observation times: extensions and comparative evaluation. *Statistics in Medicine*, 33(27):4770–4789, 2014.

Tobias Teismann, Eva-Maria AD Förtsch, Patrick Baumgart, Serkan Het, and Johannes

Michalak. Influence of violent video gaming on determinants of the acquired capability for suicide. *Psychiatry Research*, 215(1):217–222, 2014.

Terry M Therneau. *A Package for Survival Analysis in R*, 2020. URL https://CRAN.R-project.org/package=survival. R package version 3.1-12.

Andre Tylee. Depression in the community: physician and patient perspective. *The Journal of Clinical Psychiatry*, 60(Suppl. 7):12–16, 1999.

Tatjana van Strien, Hanna Konttinen, Judith R Homberg, Rutger CME Engels, and Laura HH Winkens. Emotional eating as a mediator between depression and weight gain. *Appetite*, 100:216–224, 2016.

William N Venables and Brian David Ripley. *Modern Applied Statistics with S*. New York: Springer, fourth edition, 2002.

Robert WM Wedderburn. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, 61(3):439–447, 1974.

Lee-Jen Wei, Danyu Y Lin, and Lisa Weissfeld. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association*, 84(408):1065–1073, 1989.

Brian W Whitcomb, Enrique F Schisterman, Neil J Perkins, and Robert W Platt. Quantification of collider-stratification bias and the birthweight paradox. *Paediatric and Perinatal Epidemiology*, 23(5):394–402, 2009.

Judith J Wurtman. Depression and weight gain: the serotonin connection. *Journal of Affective Disorders*, 29(2-3):183–192, 1993.

Darius Zayeni, Jean-Philippe Raynaud, and Alexis Revet. Therapeutic and preventive use of video games in child and adolescent psychiatry: a systematic review. *Frontiers in Psychiatry*, forthcoming.

Scott L Zeger and Kung-Yee Liang. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42(1):121–130, 1986.

Yayuan Zhu, Jerald F Lawless, and Cecilia A Cotton. Estimation of parametric failure time distributions based on interval-censored data with irregular dependent follow-up. *Statistics in Medicine*, 36(10):1548–1567, 2017.

Félix Zurita-Ortega, Ramón Chacón-Cuberos, Manuel Castro-Sánchez, Francisco Luis Gutiérrez-Vela, and Gabriel González-Valero. Effect of an intervention program based on active video games and motor games on health indicators in university students: a pilot study. *International Journal of Environmental Research and Public Health*, 15(7): 1329, 2018.