UNIVERSITY OF COPENHAGEN
FACULTY OF HEALTH AND MEDICAL SCIENCES

PhD thesis

Causal inference and mediation analysis for longitudinal data

Marie Skov Breum

Academic advisors: Torben Martinussen and Thomas A. Gerds

Submitted: September 30th 2023

This thesis has been submitted to the Graduate School of the Faculty of Health and Medical Sciences, University of Copenhagen

Causal inference and mediation analysis for longitudinal data

PhD thesis

Marie Skov Breum

Section of Biostatistics
Department of Public Health
University of Copenhagen
Øster Farimagsgade 5, Entr. B
DK-1014 Copenhagen K
masb@sund.ku.dk

Main supervisor: Prof. Torben Martinussen, University of Copenhagen
Co-supervisor: Prof. Thomas A. Gerds, University of Copenhagen

This thesis has been submitted to the Graduate School of the Faculty of Health and Medical Sciences, University of Copenhagen on September 30th 2023.

Preface

This thesis has been submitted to the Graduate School of Health and Medical Sciences, University of Copenhagen. The work was mainly carried out at the Section of Biostatistics, Department of Public Health, University of Copenhagen under supervision of Prof. Torben Martinussen and Prof. Thomas Gerds. Four months were spent visiting Professor Mark van der Laan at the Center for Targeted Machine Learning and Causal Inference, School of Public Health, UC Berkeley.

Many people have contributed to the making of this thesis. First and foremost I would like to thank my supervisors Torben Martinussen and Thomas Gerds for introducing me to the topics explored in this thesis and for support and guidance through the project. I would also like to thank my other coauthors Anders Munch and Mads Sundby Palle for our collaboration. I also owe great thanks to the members of the Joint Initiative for Causal Inference (JICI) for providing a great forum for discussing and getting feedback on my research during my PhD, and to Mark van der Laan and the people at the Center for Targeted Machine Learning and Causal Inference for making my visit at Berkeley very pleasant. Finally I would like to thank friends, family and colleagues for the various ways you have contributed to this project.

Conflicts of interest

This PhD project was funded by a research gift from Novo Nordisk to the Section of Biostatistics, University of Copenhagen.

Summary

This thesis is broadly concerned with developing statistical methodology for causal inference based on longitudinal data. In particular it focuses on causal mediation analysis for longitudinal data, and on the application of causal inference methodology to time-to-event analysis.

The thesis consists of a synopsis containing five chapters, followed by three manuscripts. Chapters 1-3 of the synopsis provide the necessary background knowledge for better understanding the methodological contributions of the manuscripts. Chapter 4 contains a summary of the manuscripts. Chapter 5 discusses limitations of the proposed methods and outlines possible directions for future research. The contributions of the manuscripts can be summarized as follows:

- Manuscript I proposes a method for estimating the extent to which the effect of a baseline exposure on a terminal time-to-event outcome (e.g. death) is mediated through a non-terminal time-to-event outcome (e.g. onset of disease). The method extends the concept of 'separable' direct and indirect effects to the illness-death setting.

- Manuscript II is motivated by a data application from the NASH clinical trial conducted by Novo Nordisk. We propose a method for estimating the extent to which the effects of Semaglutide on NASH is mediated through weight loss which is a repeatedly measured covariate. The proposed method builds upon work on 'randomized interventional' direct and indirect effects.

- Manuscript III is concerned with the estimation of concordance measures for right censored time-to-event data. Many widely used estimators will depend on the censoring distribution under model misspecification. Our contribution is that we view the concordance measures as model-free estimands and propose non-parametric estimators.

Resumé

Denne afhandling beskæftiger sig i bred forstand med at udvikle statistiske metoder til kausal inferens baseret på longitudinelle data. I særdeleshed forkuseres der på kausal mediationsanalyse for longitudinelle data og på at anvende kausale metoder indenfor overlevelsesanalyse.

Afhandlingen består af en synopsis indeholdende fem kapitler efterfulgt af tre manuskriper. Kapitel 1-3 i synopsen giver den nødvendige baggrundsviden for bedre at forstå de metodiske bidrag i manuskripterne. Kapitel 4 indeholder et resumé af artiklerne. Kapitel 5 diskuterer begrænsningerne ved de forelåede metoder og skitserer mulige retninger for fremtidig forskning. Bidragene i manuskripterne kan opsummeres som følger:

- Manuskript I foreslår en metode til at estimere, i hvilket omfang effekten af en baseline eksponering på en terminal levetid (f.eks tid til død) medieres gennem en ikke-terminal levetid (f.eks. tid til sygdomsdiagnose). Metoden udvider begrebet 'separable' direkte og indirekte effekter til en 'illness-death' model.

- Manuskript II er motiveret af en dataapplikation fra det kliniske studie 'NASH' udført af Novo Nordisk. Vi foreslår en metode til at estimere, i hvilket omfang virkningerne af Semaglutid på NASH medieres gennem vægttab, som er en kovariat der er målt gentagne gange over tid. Den foreslåede metode bygger på arbejde omhandlene 'randomiserede interventionelle' direkte og indirekte effekter.

- Manuskript III beskæftiger sig med estimation af konkordans mål for højre censureret overlevelses data. Mange udbredte estimatorer afhænger af censurerings fordelingen hvis modellen er misspecificeret. Vores bidrag er, at vi formulerer 'concordance' målene som modelfri estimander og foreslår ikke-parametriske estimatorer.

Contents

Chapter 1

Introduction

The aim of this thesis is to develop statistical methods for causal inference and mediation analysis for longitudinal and time-to-event data that account intrinsic features such as censoring and time-dependent confounding. This chapter serves as a broad introduction to the thesis.

The chapter is organized as follows. In Section 1.1 we give some in-depth motivation behind the topics explored in this thesis. Section 1.2 provides an overview of the organization of the thesis and a reading guide. Section 1.3 introduces some notation that will be used throughout the synopsis.

1.1 Motivation

Causality lies at the heart of many scientific questions in public health research, particularly those that involve evaluating policies or exposures. Over the past decades a formal framework for causal inference has been developed pioneered by Robins, Pearl, van der Laan and many others [Robins, 1986, Robins et al., 2000, van der Laan and Robins, 2003, Pearl, 2000]. These methods, when integrated into applied research, makes it possible to be very precise about the causal question and the assumptions required to answer the causal questions with observed data. The process of applying formal causal methods to applied research in a systematic way is sometimes referred to as the 'roadmap' for causal inference [Petersen and van der Laan, 2014]. The roadmap can be boiled down to the following steps

(i) Translate the scientific question into a clearly defined target causal estimand stated in the language of potential outcomes [Neyman, 1923, Rubin, 1978].

(ii) Assess identifiability, i.e. derive the assumptions which are necessary to to identify the target causal estimand from the observed data. The identification result defines the target statistical estimand.

(iii) Given the target statistical estimand choose an estimator that respects the causal knowledge in the data. This may require flexible estimators that allow for parts of the data-generating process to be unrestricted. Such estimators can be constructed using tools from semiparametric theory and empirical process theory [van der Vaart, 2000, Tsiatis, 2006, van der Laan and Rose, 2011, 2018, Chernozhukov et al., 2018].

Causal inference is a large field of research. In this thesis we focus specifically on two areas of causal inference.

The first of these areas is mediation analysis for longitudinal data. Mediation analysis is a popular tool in epidemiological and medical research for describing the mechanisms by which an exposure or treatment affects an outcome. More specifically, the methods attempt to decompose a total treatment effect into a so-called 'indirect effect' which is mediated by a particular intermediate variable (or set of variables), and a remaining unmediated 'direct effect' [VanderWeele, 2015]. An early technique inspired by Baron and Kenny [1986] is to estimate mediation effects using parametric structural equation models without using explicitly causal language and notation. Robins and Greenland [1992] and Pearl [2001] introduced counterfactual based non-parametric definitions of the direct and indirect effects, which has led to an improved understanding of eg. confounding adjustment and allows for interactions and non-linearity. However the existence and identification of the so-called natural direct and indirect effects of Robins and Greenland [1992] and Pearl [2001] is somewhat controversial because they rely on independence assumptions 'across worlds', that may be unrealistic and are impossible to empirically verify. Particularly these cross-world assumptions imply that natural (in)direct effects are generally not identified in the presence of a confounder of the mediator-outcome relationship which is affected by the exposure. This limits their practical applicability, and means that they do not immediately generalize to the longitudinal setting. This have led several researchers to propose alternative causal mediation estimands that do not rely on cross-world assumptions. This topic will be explored further in Chapter 2 of the synopsis.

The second area that we will focus on is the use of causal inference methodology in survival analysis. The development of formal causal inference frameworks over the past decades has sparked an interest in clarifying the causal interpretation of well-known estimands from time-to-event analysis such as the Cox hazard ratio [Hernán, 2010, Aalen et al., 2015, Martinussen et al., 2020] and the 'net risk' and 'crude risk' from the competing event literature [Young et al., 2020]. Moreover the use of model-free (non-parametric) definitions of statistical target estimand is becoming increasingly popular in the survival literature. For instance nonparametric definitions of hazard ratio type estimands [Uno and Horiguchi, 2023, Vansteelandt et al., 2022] have been proposed as an alternative to the Cox HR which is known to depend on the study specific censoring distribution when the model is misspecified [Struthers and

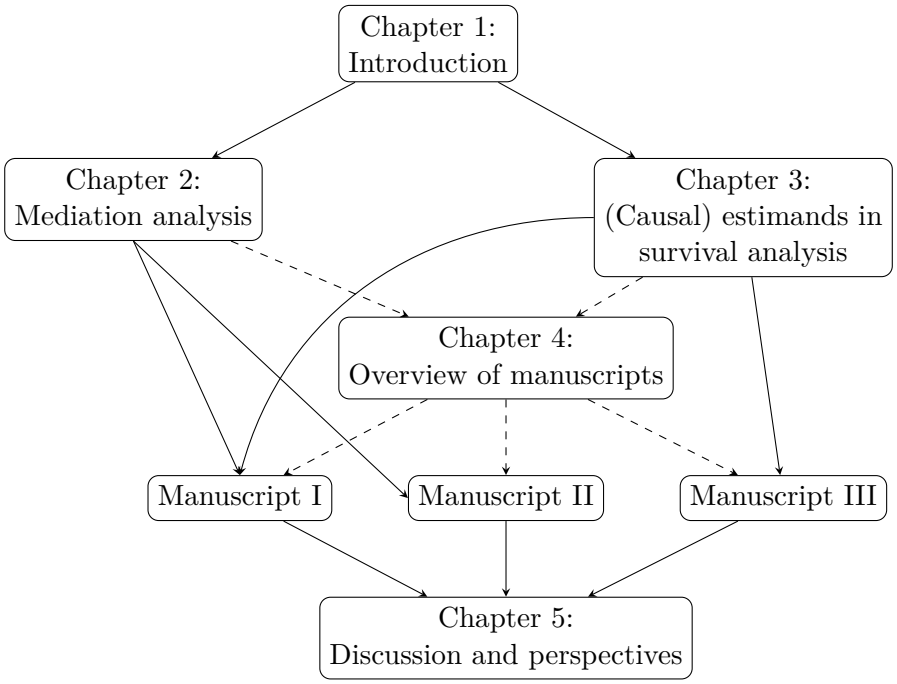


Figure 1.1: Suggested reading order.

Kalbfleisch, 1986]. These topics will be explored further in Chapter 3 of the synopsis.

1.2 Overview of thesis

This thesis consists of five introductory chapters and three article manuscripts. The manuscripts are:

Manuscript I: Breum, M. S., Munch, A., Gerds, T. A. and Martinussen, T. (2023). Estimation of separable direct and indirect effects in a continuous-time illness-death model. *Lifetime Data Analysis*, 1-38. `https://doi.org/10.1007/s10985-023-09601-y`

Manuscript II: Breum, M. S. and Palle, M. S. (2023). Estimation of data-dependent (in)direct effects with a repeatedly measured mediator and missing outcome data. To be submitted to *Statistics in Medicine*.

Manuscript III: Breum, M. S. and Martinussen, T. (2023). Efficient non-parametric estimators of discrimination measures with censored survival data. Under preparation.

In Chapter 2 we introduce causal mediation analysis with an emphasis on causal mediation estimands which do not rely on cross-world assumptions.

This serves as background for the contributions in manuscripts I and II. The topic of Chapter 3 is the use of (causal) estimands survival analysis which serves as background for manuscripts I and III. In Chapter 4 we provide a summary of the three manuscripts, and in Chapter 5 we discuss the challenges and limitations of the proposed methods and provide ideas for future research. The three manuscripts are printed in the back of the thesis. A suggested reading order of this thesis is given in Figure 1.1.
In addition to the background provided in the introductory chapters of this thesis we assume that the reader is familiar with basic concepts of semiparametric theory. Many excellent overviews of this topic exist e.g. Hines et al. [2022], Kennedy [2016], Kennedy [2023] and Fisher and Kennedy [2021], among others.

1.3 Notation

For any random variables X and Y we use $\mathcal{X}$ to denote the support of X and the Cartesian product $\mathcal{X} \times \mathcal{Y}$ to denote the support of (X, Y). We let $p_X(x)$ denote the marginal distribution of X and $p_{Y|X}(y \mid x)$ the conditional distribution of Y given X.

Throughout the synopsis we will use counterfactual or 'potential outcomes' notation where $Y(x)$ denotes the value of Y that would have been observed had X, possibly contrary to the fact, been set to the value x [Neyman, 1923, Rubin, 1978]. Unless otherwise stated we will assume consistency i.e. that the potential outcome $Y(x)$ corresponds to the observed outcome Y if the actual level of X is x. For further discussion on this assumption see Hernán and Taubman [2008].

Chapter 2

Mediation analysis

This chapter provides a review of causal mediation estimands. The chapter is organized as follows. In Section 2.1 we introduce the notion of natural direct and indirect effects and present the identification assumptions of Pearl [2001]. In Section 2.2 we will give examples of settings where the so-called cross-world assumption is violated. In Section 2.3 we introduce the notion of path specific effects which provides a more general definition of mediated effects than the notion of natural (in)direct effect, but still relies on cross-world assumptions. In Section 2.4 we review some alternative causal mediation estimands that have been proposed to circumvent the cross-world assumptions.

2.1 Natural direct and indirect effects

We consider a data structure given by $O = (W, A, M, Y)$ where W is a vector of baseline covariates, A is a binary treatment, M is the potential mediator and Y is the outcome of interest. We assume that W is not affected by treatment A, and that M is measured after the exposure and before the outcome. A DAG representing the assumed causal structure is given in Figure 2.1. The observed data is assumed to be a sample of independent observations $O_1, ..., O_n$ which are identically distributed according to some unknown probability distribution denoted P, which is assumed to lie in a statistical model $\mathcal{P}$.

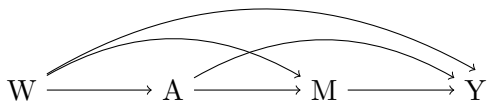


Figure 2.1: DAG representing the causal structure in the setting in Section 1.1.

We let $Y(a, M(a^*))$ the cross-world (nested) counterfactual outcome that

would have been oberved had A been set to the value a and M been set to the value that it would naturally have taken under $A = a^*$. In addition to the usual consistency assumptions we need the so-called composition assumption that $Y(a) = Y(a, M(a))$. We refer to VanderWeele and Vansteelandt [2009] for further discussion on this assumption.

Pearl [2001] define the natural direct effect (NDE) as the counterfactual contrast

$$NDE : E\left\{Y(a, M(a^*)) - Y(a^*, M(a^*)\right\}, \tag{2.1}$$

that is, as the average difference in the outcome Y under $A = a$ and $A = a^*$ when the mediator fixed to what it would have been had $A = a^*$.

The natural indirect effect (NIE) is defined as

$$NIE : E\left\{Y(a, M(a)) - Y(a, M(a^*))\right\}, \tag{2.2}$$

ie. as the average difference in outcome if the exposure was fixed at $A = a$, but mediator was set to what it would have been if exposure had been a versus if exposure had been a^*.

Together the NDE and NIE provide a decomposition of the total effect (TE)

$$TE : E\{Y(a)\} - E\{Y(a^*)\}$$

The natural direct and indirect effects are identifiable from the data under the following assumptions

A.1: $Y(a, m) \perp\!\!\!\perp A \mid W$

A.2: $Y(a, m) \perp\!\!\!\perp M \mid (A, W)$

A.3: $M(a) \perp\!\!\!\perp A \mid W$

A.4: $Y(a, m) \perp\!\!\!\perp M(a^*) \mid W$

Assumptions A.1 and A.2 state that there are no unmeasured confounders of the exposure-outcome and mediator outcome relationship, respectively. Assumption A.3 states that there is no unmeasured confounding of the exposure-mediator relationship. Note that if A is randomized then A.1 and A.3 hold automatically. Assumption A.4 is the so-called cross-world assumption, which states that the counterfactual outcome and mediator values are independent "across worlds" with one being a world in which the exposure is set to A = a for the outcome and the other being a world in which it is set to A = a for the mediator. This assumption is impossible to verify as this does not correspond to any feasible experiment. We will further discuss the implications of the cross-world assumption, and when it may be violated in the next section.

Let $\Psi^{a,a^*}(P) = E\{Y(a, M(a^*))\}$. If assumptions A.1-A.4 hold Pearl [2001] show that $\Psi^{a,a^*}(P)$ is (non-parametrically) identified via the so-called mediational G-formula

$$\begin{aligned}&\Psi^{a,a^*}(P)\\&= \int_{\mathcal{W}} \int_{\mathcal{M}} E\{Y \mid A = a, m, w\}\, p_{M|A,W}(m \mid a^*, w) p_W(w)\, \mathrm{d}\mu_M(m)\, \mathrm{d}\mu_W(w),\end{aligned} \tag{2.3}$$

where μ_M and μ_W are some dominating measures.

This leads to the following identifying formulas for the natural direct and indirect effects

$$\begin{aligned}NDE(a, a^*)(P) =& \Psi^{a,a^*}(P) - \Psi^{a^*,a^*}(P),\\NIE(a, a^*)(P) =& \Psi^{a,a}(P) - \Psi^{a,a^*}(P).\end{aligned}$$

Note that while the identification assumptions in (A1)-(A4) are sufficient for identification of natural (in)direct effects they are not necessary conditions. In Section 2.3 we give a brief description of complete identification strategy using so-called recanting district conditions [Avin et al., 2005].

Natural direct and indirect effect have many advantages including adding up to the total treatment effect and allowing for interactions and non-linearities. Moreover they allow for the use natural effect models which directly parameterize the nested counterfactuals using marginal structural models [Lange et al., 2012, Vansteelandt et al., 2012]. However the cross-world assumptions limits the practical applicability of natural (in)direct effects as we will demonstrate in the following section.

2.2 Examples of cross-world independence violations

Below we give some examples where the cross world assumption in A.4 is violated.

Example 1: Intermediate confounding

Under the assumption that the data is generated from a non-parametric structural equation model (NPSEM) [Pearl, 2000], the cross-world assumption is violated in the presence of a confounder of the mediator outcome relationship which is affected by exposure.

To see this let W, A, M, Y be defined as in Section 2.1 and let L be a post-treatment confounder. The corresponding DAG is depicted in Figure 2.2, and

the corresponding NPSEM is

$$\begin{aligned} W &= f_W(\varepsilon_W), \\ A &= f_A(W, \varepsilon_A), \\ L &= f_L(A, W, \varepsilon_L), \\ M &= f_M(A, W, L, \varepsilon_M), \\ Y &= f_Y(A, W, L, M, \varepsilon_Y), \end{aligned}$$

where f_W, f_A, f_L, f_M and f_Y are deterministic mappings and $\varepsilon_W, \varepsilon_A, \varepsilon_L, \varepsilon_M, \varepsilon_Y$ are exogenous random variables.

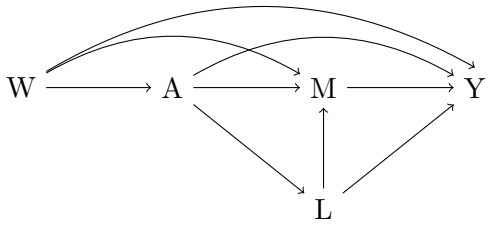


Figure 2.2: DAG representing example 1.

The corresponding counterfactuals can be represented as

$$\begin{aligned} L(a) &= f_L(a, W, \varepsilon_L), \\ L(a^*) &= f_L(a^*, W, \varepsilon_L), \\ M(a^*) &= f_M(a^*, W, L(a^*), \varepsilon_M) = f_M(a^*, W, f_L(a^*, W, \varepsilon_L), \varepsilon_M), \\ Y(a, m) &= f_Y(a, W, L(a), m, \varepsilon_Y) = f_Y(a, W, f_L(a, W, \varepsilon_L), m, \varepsilon_Y). \end{aligned}$$

We see that $Y(a, m) \not\perp M(a^*)$ since $Y(a, m)$ and $M(a^*)$ share the common error term ε_L.

Example 2: Mediation with a survival outcome

When the outcome is a time-to-event variable then natural (in)direct effects are only identified if the mediator is measured immediately after exposure [Lange and Hansen, 2011].

To illustrate this let T be survival time and assume no censoring. Let $N_t = I(T \leq t)$ be a counting process which jumps when an event occurs and let $dN_t = I(T = t)$ be the event indicator. Suppose we observe survival status at time $t = 1$ and $t = 2$, and that the mediator M is measured in between these time-points. The data structure is illustrated in Figure 2.3 below, and

the following NPSEM encodes the time-ordering of the variables

$$\begin{aligned} A &= f_A(\varepsilon_A), \\ dN_1 &= f_{dN_1}(A, \varepsilon_{dN_1}), \\ M &= f_{M_2}(A, dN_1, \varepsilon_M), \\ dN_2 &= f_{dN_2}(A, M, dN_1, \varepsilon_{dN_2}). \end{aligned}$$

Clearly N_1 acts as a post-treatment confounder of survival status at $t = 2$ and the mediator. It follows from Example 1 that the cross-world assumption is violated.

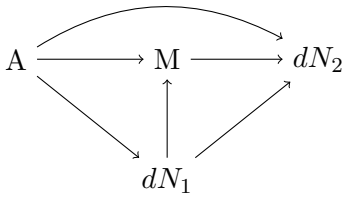


Figure 2.3: DAG representing example 2.

Another issue when the mediator is not measured immediately after exposure, as pointed out in Didelez [2019], is that the cross-world counterfactual $T(a, M(a^*))$ is not well-defined. This is because the patient may live longer under $A = a$ than under $A = a^*$, i.e. setting the mediator to what it would have been under $A = a^*$ is not a well defined intervention if the patient would not have lived long enough under $A = a^*$ for the mediator to be measured.

Example 3: Latent cross-world confounders

The cross-world assumption can also be violated by other causes than intermediate confounding. We will show this with an example due to Robins and Richardson [2011]. Suppose A is a binary treatment variable, M is a potential mediator and Y is the outcome. Let U be a latent variable that affects $M(0)$ and $Y(1, m)$, as depicted in Figure 2.4 below.

The corresponding structural equations are

$$\begin{aligned} U &= f_U(\varepsilon_U), \\ A &= f_A(\varepsilon_A), \\ M &= \begin{cases} f_{1,M}(\varepsilon_M) \text{ if } A = 1 \\ f_{0,M}(U, \varepsilon_M) \text{ if } A = 0 \end{cases}, \\ Y &= \begin{cases} f_{1,Y}(M, U, \varepsilon_Y) \text{ if } A = 1 \\ f_{0,Y}(M, \varepsilon_Y) \text{ if } A = 0 \end{cases}, \end{aligned}$$

where U is unobserved, f_U, f_A, $f_{0,M}$, $f_{1,M}$, $f_{0,Y}$ and $f_{1,Y}$ are deterministic mappings and $\varepsilon_U, \varepsilon_A, \varepsilon_M, \varepsilon_Y$ are exogenous random variables.

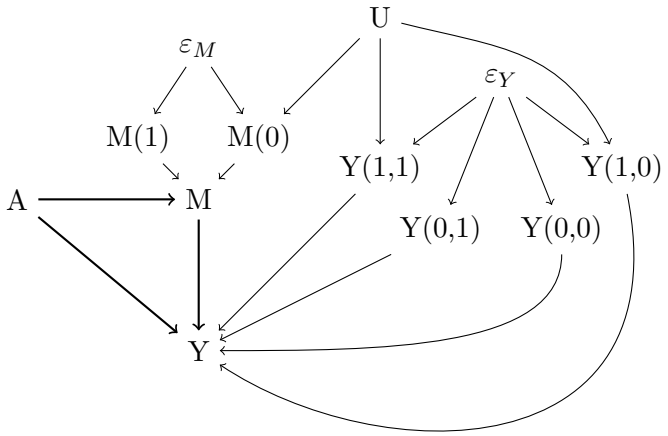


Figure 2.4: Causal diagram representing Example 3. This figure is based on Figure 3 in Andrews and Didelez [2021] and Figure 11 in Robins and Richardson [2011].

The counterfactuals are then

$$\begin{aligned} M(1) &= f_{1,M}(\varepsilon_M) \\ M(0) &= f_{0,M}(\tilde{\varepsilon}_M) \\ Y(0,m) &= f_{0,Y}(m, \varepsilon_Y) \\ Y(1,m) &= f_{1,Y}(m, \tilde{\varepsilon}_Y) \end{aligned}$$

where $\tilde{\varepsilon}_M$ is a combination of ε_M and the unobserved U, and similarly $\tilde{\varepsilon}_Y$ is a combination of ε_Y and U. It is seen that the cross-world assumption is violated due to the lack of independence between $\tilde{\varepsilon}_M$ and $\tilde{\varepsilon}_Y$.

2.3 Path-specific effects

The natural direct and indirect effects defined in Section 2.1 are special cases of path specific effects. Using standard graph theory notation, let $\mathbf{V}$ be a set of random variables with state space $\mathfrak{X}_{\mathbf{V}}$, and $\mathcal{G}(\mathbf{V})$ the corresponding DAG.

The literature on path specific effects defines the mediated effect as a set of directed paths π in $\mathcal{G}(\mathbf{V})$ from the exposure to the outcome [Steen and Vansteelandt, 2018, Shpitser, 2018]. If we let (W, A, M, Y) be defined as in Section 2.1 then the natural indirect effect can be fined as the set of paths from A to Y through M. That is $\pi = \{A \to M \to Y\}$. The natural direct

effect can be as the set of paths from A to Y not in π, i.e. the set of paths in $\bar{\pi} = \{A \to Y\}$.

Complete graphical conditions for identifiability of path-specific effects under NPSEMs are given in Avin et al. [2005] and Shpitser [2013]. These identifiability conditions rely on the notion of 'recantation'. Specifically a variable L is called a recanting witness for π if there exists a directed path from treatment A to outcome Y of the form $A \to L \to ... \to Y$ in both π and $\bar{\pi}$. Note that L in Example 1 of Section 2.2 is a recanting witness. It was shown in Avin et al. [2005] that if an only if there is no recanting witness then the path-specific effect can be identified from the NPSEM representation of a specific DAG. The result of Avin et al. [2005] holds for DAGs with no hidden variables. It was extended models with unmeasured confounding by Shpitser [2013] using a more general notion 'recantation' called a 'recanting district'. In a hidden variable DAG the complete graphical identification strategy states that when $E\{(Y(a)\}$ is identifiable, then every path specific effect for which there is no recanting district is also identifiable.

In addition to providing a complete identification strategy for natural (in)direct effect this graphical criterion can help identify alternative effect decompositions which are identified from the data under a NPSEM representation of a specific DAG. For instance in Example 1 of Section 2.2 while the natural (in)direct effects are not identified, the π-specific effect with $\pi = \{A \to M \to Y\}$ can still be identified [Vanderweele et al., 2014]. The identification strategy is also applicable to more complicated effect decompositions in settings with multiple mediators or a repeatedly measured mediator Vansteelandt et al. [2019]. However like the natural (in)direct effects the definition of path-specific effects rely on nested counterfactuals whose existence and identification remains controversial.

2.4 Alternative estimands

As we have seen in the section above, in addition to being impossible to empirically verify, the cross-world independence assumption is often problematic to justify, and in some situations the cross world quantity may not be well-defined since it is not possible to set the mediator to a specific value.

In the following section will consider alternative approaches to causal mediation analysis that redefine the causal estimand and thereby avoid cross-world counterfactuals.

Interventionist direct and indirect effects

The concept of interventionist effects is based on the extended graphical approach by Robins and Richardson [2011], which was revisited by Robins et al. [2021]. Instead of considering manipulations of the mediator independently of

the treatment given, the method 'extends the story' by considering an intervention that decomposes treatment into separate components.

More explicitly, suppose that treatment A has two separate components that act through different causal pathways

- one component A^M which only affects the outcome through it's effect on the mediator
- and one component A^Y which affects the outcome directly (not through the mediator)

These treatment component are not observed, but we assume that, at least in principle, an intervention exists where $A^Y \neq A^M$.

In the setting considered in Section 2.1 this corresponds to the assumptions

S.1 $M \perp A^Y \mid A^T = a, W$,

S.2 $Y \perp A^M \mid M = m, A^Y = a, W$,

which are illustrated in Figure 2.5.

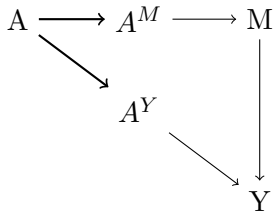


Figure 2.5: Extended graph illustrating assumptions S.1 and S.2. Thick arrows indicate deterministic relations.

Robins et al. [2021] define the separable direct effect (SDE) as

$$SDE : E\left\{Y(A^Y = a, A^M = a^*) - Y(A^Y = a^*, A^M = a^*)\right\}, \tag{2.4}$$

and the separable indirect effect (SIE) as

$$SIE : E\left\{Y(A^Y = a^*, A^M = a) - Y(A^Y = a^*, A^M = a^*)\right\}. \tag{2.5}$$

That is, the separable direct effect (2.4) is defined as the average difference in outcome under $A^Y = a$ and $A^Y = a^*$ when A^M is fixed at a^*. Similarly the separable indirect effect (2.5) is defined as the average difference in outcome under $A^M = a$ and $A^M = a^*$ when A^Y is fixed at a^*. Under assumptions (A.1)-(A.3) of Section 2.1 and assumptions (S.1) and (S.2) the SDE and the SIE are identified via the mediational g-formula in (2.3).

As opposed to the natural (in)direct effects this definition of mediated effects avoids reference to cross-world counterfactuals, and indentification requires only single-world assumptions. However this comes at the cost the separability assumption which may not always be appropriate.

This approach has been extended to settings with a survival outcome and a repeatedly measured mediator by Didelez [2019] and Aalen et al. [2020]. It was moreover extended to the competing risk setting by Stensrud et al. [2022, 2021] which we shall return to in Chapter 3.

Randomized interventional direct and indirect effects

A different approach defines the direct and indirect effect in terms of stochastic interventions on the mediator distribution [Robins, 2003, van der Laan and Petersen, 2008]. These types of effects have many names in the litterature e.g. 'standardized' [Didelez et al., 2006, Geneletti, 2007], 'randomized interventional' [Vanderweele et al., 2014, Vansteelandt and Daniel, 2017] or 'stochastic' [Rudolph et al., 2018] direct and indirect effects. We will use the term randomized interventional (in)direct effects to decribe them. There are subtle differences between the definitions that we will not dive into.

In general randomized interventional (in)direct effects are defined in terms of potential outcomes where the exposure A is assigned to a level a and the mediator is drawn from a distribution Γ.

Specifically consider the setting from Section 2.1 and let $\Gamma^{a^*}(m \mid w) = P(M(a^*) = m \mid W = w)$ be the conditional distribution of the mediator if the exposure had been set to $A = a^*$. Consider an intervention where A is set to a and we randomly draw $M \sim \Gamma^{a^*}$. Let $Y(a, M \sim \Gamma^{a^*})$ denote the resulting counterfactual outcome. Then the randomized interventional direct effect (RDE) may be defined as

$$RDE : E\left\{Y(a, M \sim \Gamma^{a^*}) - Y(a^*, M \sim \Gamma^{a^*})\right\}, \tag{2.6}$$

and the randomized interventional indirect effect (RIE) as

$$RIE : E\left\{Y(a^*, M \sim \Gamma^{a}) - Y(a^*, M \sim \Gamma^{a^*})\right\}. \tag{2.7}$$

The RDE and RIE are identified under assumption A.1-A.3 of Section 2.1 via the mediational g-formula (2.3).

Vanderweele et al. [2014] define and identify interventional (in)direct effects in the presence of an exposure induced mediator-outcome confounder. Vansteelandt and Daniel [2017] extended this to the setting with multiple mediators where the effects through each mediator are of interest.

VanderWeele and Tchetgen Tchetgen [2017] and Zheng and van der Laan [2017] both proposed extensions of the interventional effects of Vanderweele et al. [2014] to the setting with time-varying mediators and exposures. The

proposals differ in their definitions of the interventional mediator distributions, and thus capture different effects. VanderWeele and Tchetgen Tchetgen [2017] consider random draws from a marginal mediator distribution (conditional on only baseline covariates) at a certain exposure level, whereas Zheng and van der Laan [2017] consider random draws from a conditional mediator distribution at a certain exposure level. A limitation of the proposal of VanderWeele and Tchetgen Tchetgen [2017] is that it does not immediately generalize to time-to-event data because a person who is still alive would be allowed to draw the mediator value of someone who has died.

Organic direct and indirect effects

A different approach by Lok [2016, 2020], Lok and Bosch [2011] consider so-called 'organic' interventions that cause the mediator under treatment level a to have the same distribution as the mediator under treatment level a^* .

Specifically let I indicate an intervention on the mediator M. The intervention is organic if

O.1: $M(a, I = 1) \sim M(a^*)$,

O.2: $Y(a, I = 1) \mid M(a, I = 1) = m \sim Y(a) \mid M(a) = m.$

Assumption O.1 states that the distribution of the mediator under treatment level a combined with the intervention I is the same as the distribution of the mediator under $A = a^*$. Assumption O.2 says that the intervention I does not affect the outcome directly. Then Lok [2016] defines the organic direct effect (ODE) and organic indirect effects (OIE) as respectively

$$ODE : E\left\{Y(a, I = 1) - Y(a^*)\right\}, \tag{2.8}$$

and

$$OIE : E\left\{Y(a) - Y(a, I = 1)\right\}. \tag{2.9}$$

Under assumptions O.1-O.2 and A.1-A.3 the organic (in)direct effects are identifiable from the data, and their identifying functional will can be obtained via the mediational g-formula in (2.3).

In appendix F of Lok [2016] it is shown that the conditions in O.1-O.2 may also be expressed as conditional independence statements

O.1*: $M \perp R \mid R \neq 1$,

O.2*: $Y \perp R \mid M = m, R \neq 0$,

where R is an extended treatment variable described as follows: R = 0: treatment 0, R = 1: treatment 1 and R = 2: treatment 1 combined with an

organic intervention I on the mediator. Then the organic direct effect (ODE) and organic indirect effects (OIE) can be defined as respectively

$$ODE : E\left\{Y(R=1) - Y(R=2)\right\}, \tag{2.10}$$

and

$$OIE : E\left\{Y(R=2) - Y(R=0)\right\}. \tag{2.11}$$

As with the previous approaches organic effects avoid cross-world assumptions, and the cross world independence assumption is replaced by another structural assumption, namely that of the intervention being 'organic'. The organic effects are very similar to randomized interventionial effects in that they are defined in terms of interventions on the mediator distributions. They also share a lot of similarities with interventionist effects in that they can be interpreted in terms of interventions on a hypothetical treatment variable.

Other approaches

The literature on causal mediation analysis is large and ever expanding. Many interesting and useful approaches did not make it into this chapter.

Among those are the population intervention direct and indirect effects [Díaz and Hejazi, 2020, Hejazi et al., 2022]. This method allows for stochastic interventions on treatment such as incremental propensity score interventions [Kennedy, 2019] which alter the odds of receiving the treatment. Such stochastic interventions are sometimes more realistic than static interventions. Moreover this method is applicable to both categorical and continuous exposure variables.

Another interesting approach is causal bounds [Cai et al., 2008, Kaufman et al., 2005, Tchetgen Tchetgen and Phiri, 2014, Miles et al., 2015], which give a range of possible values for the NDE and NIE without relying on the cross-world assumption. Dukes et al. [2023] propose a proximal mediation method which uses proxy variables to identify natural (in)direct effect when the no unmeasured confounding assumptions in A.1-A.3 fail.

Chapter 3

(Causal) estimands in survival analysis

In this chapter we give a short overview of the use of estimands in survival analysis. This chapter is organized as follows. In Section 3.1 we give a brief introduction to survival analysis. In Section 3.2 we discuss the causal interpretation of the hazard ratio computed from the Cox regression model and alternative model-free hazard-ratio estimands. In Section 3.3 we give an introduction to competing risk analysis and in Section 3.4 we discuss the causal interpretation of common estimands from the competing risks literature. We also introduce the novel separable effects estimand of Stensrud et al. [2022] which solves some of the interpretational problems of the classical estimands.

3.1 Survival analysis

Let $T \in \mathbb{R}_+$ be the survival time, and assume that we only observe the minimum of T and C where $C \in \mathbb{R}_+$ is a censoring time. The survival time T can also be represented through the counting process $N(t) = I(T \leq t)$ [Andersen et al., 2012, Martinussen and Scheike, 2006, Aalen et al., 2008]. Let $\tilde{T} = T \wedge C$ denote the observed survival time and $\Delta = I(T \leq C)$ the failure indicator. Let $X \in \mathbb{R}^q$ be a vector of covariates.

Let $f_{T|X}$ be the conditional density of T and let $S(t \mid X) = P(T > t \mid X)$ denote the conditional survival function. The distribution of T given X can be characterized through the conditional hazard function

$$\alpha(t \mid X) = \frac{f_{T|X}(t \mid X)}{S(t \mid X)} = \lim_{dt \to 0} \frac{P(t \leq T < t + dt \mid T \geq t, X)}{dt},$$

which can be interpreted as the instantaneous risk of failure at time t, given that the individual survives up to t.

The most commonly used model in survival analysis is the proportional hazards model of Cox [1972], which assumes that the conditional hazard rate

has the form

$$\alpha(t \mid X) = \lambda_0(t) \exp(\beta^T X)$$

where the baseline hazard function is left unspecified. The underlying assumption is that the covariates have a proportional effect on the conditional hazard where the strength of the association is measured by β.

3.2 Interpretation of the hazard ratio

Suppose $X = (A, W)$ where A is a treatment indicator and W is a vector of baseline covariates. To evaluate treatment effects, the hazard ration (HR), which is the ratio of hazard rates in the treatment group and the placebo group, is typically reported

$$\frac{\lambda(t \mid A = 1, W)}{\lambda(t \mid A = 1, W)} = \exp(\beta_A).$$

The Cox HR is a convenient summary measure of the treatment effect. However there are several drawbacks regarding the interpretation of the Cox HR.

One issue is that when the proportional hazards assumption does not hold, the result depends on the underlying study-specific censoring distribution, which is of no scientific interest [Struthers and Kalbfleisch, 1986, Whitney et al., 2019]. To address this concern Vansteelandt et al. [2022] propose a model-free hazard ratio estimand using an assumption-lean approach Vansteelandt and Dukes [2022]. This estimand is defined as

$$\theta(P) = \frac{\operatorname{cov}\left[g\left\{S(t \mid X)\right\}, X\right]}{\operatorname{var}(X)}.$$

If we use the link function $g(x) = \log\{-\log(x)\}$ and $S(t \mid X)$ is estimated from a correctly specified Cox regression model then $\theta(P)$ reduces to the standard Cox HR, but it remains well-defined otherwise. Other model-free alternatives to the Cox HR are the 'average' hazard ratio [Kalbfleisch and Prentice, 1981, Schemper, 1992, Xu and O'Quigley, 2000, Uno and Horiguchi, 2023].

Another major drawback of the Cox HR is that, even in the absence of model misspecification, it cannot be causally interpreted as a hazard ratio. This was pointed out in Hernán [2010] and shown mathematically by Martinussen et al. [2020]. This is partly due to the non-collapsability of the Cox hazard ratio, and partly due to an underlying selection process making subjects more and more frail as time progresses.

While the Cox HR cannot be interpreted as a hazard ratio it can has other interpretations as shown in e.g. De Neve and Gerds [2020]. Moreover effect measures derived from survival curves, e.g. survival differences or restricted mean survival times, are easier to interpret causally.

3.3 Competing risk model

A competing event is an event whose occurrence precludes the occurrence of the primary event of interest. Let $T \in \mathbb{R}_+$ be an event time and let $\varepsilon \in \{1, ..., K\}$ be the event type. For simplicity we consider a competing risk setting with $K = 2$ causes where $\varepsilon = 1$ denotes the event of interest (Y) and $\varepsilon = 2$ the competing event (D). Because of loss to follow-up, we only observe the failure indicator $\Delta = I(T \leq C)$ and $\tilde{T} = T \wedge C$, where C denotes the right censoring time. We assume that (T, ε) and C are conditionally independent given the covariate vector X.

The competing risk model can be fully characterized by the cause-specific conditional hazard functions

$$\alpha_k(t \mid X) = \lim_{dt \to 0} \frac{P(t \leq T < t + dt, \varepsilon = k \mid T \geq t, X)}{dt}, \quad k \in \{0, 1\} \tag{3.1}$$

This is the instantaneous risk of experiencing an event of type k given that the individual has not yet experienced any event prior to time t.

A common summary measure is the so-called 'crude risk' or cumulative incidence function $P_1(t \mid X) = P(T \leq t, \varepsilon = 1 \mid X)$ which is related to the cause-specific hazard function trough

$$P_1(t \mid X) = \int_0^t \alpha_1(s \mid X) S(s- \mid X) \, \mathrm{d}s, \tag{3.2}$$

where $S(t \mid X) = \exp(-\int_0^t \sum_k \alpha_k(s \mid X) \, \mathrm{d}s)$.

We also have the relation

$$P_1(t \mid X) = 1 - \exp(-\int_0^t \lambda_1^*(s \mid X) \, \mathrm{d}s),$$

where λ_1^* is the subdistribution hazard which defined as

$$\lambda_1^*(s \mid X) = \lim_{dt \to 0} \frac{P(t \leq T < t + dt, \varepsilon = 1 \mid (T \geq t) \cup (T \leq t, \varepsilon \neq 1), X)}{dt}. \tag{3.3}$$

Another summary measure is the 'net risk' or marginal cumulative incidence defined as $S_1(t \mid X) = P(Y > t \mid X)$ which is related to the marginal hazard

$$\lambda_1(t \mid X) = \lim_{dt \to 0} \frac{P(t \leq Y < t + dt \mid Y \geq t, X)}{dt}, \tag{3.4}$$

through the relation $S_1(t \mid X) = \exp(-\int_0^t \lambda_1(d \mid X) \, \mathrm{d}s)$. The marginal hazard function describes the instantaneous risk of experiencing the event of interest given that a subject has not yet experienced the event of interest at time t. Note that identification of the 'net risk' and the marginal hazard requires that Y and D are conditionally independent given X.

3.4 Causal interpretation of statistical estimands in competing event settings

In a recent paper Young et al. [2020] discuss the causal interpretation of common estimands from the competing risk literature, and show that there are many similarities between the competing event setting and the mediation setting. In particular they clarify that the contrasts of the cumulative incidence in (3.2) under different treatment levels can be interpreted as the total effect of treatment on the event of interest. When the treatment affects the competing event then the total effect includes the effect mediated through the competing event which makes the interpretation difficult. Young et al. [2020] also clarify that the contrasts of the marginal cumulative incidence function under different treatment levels can be interpreted as a controlled direct effects. However this controlled direct effect is not well defined because it relies on an unspecified intervention which eliminates the competing event. Furthermore estimands derived from the cause-specific hazard (3.1) or the subdistribution hazard (3.3) do not have a causal interpretation [Young et al., 2020, Hernán, 2010], and thus do not solve this interpretational problem.

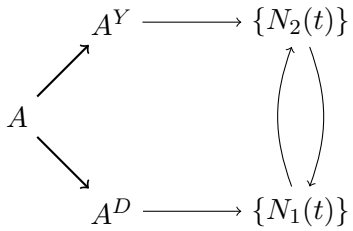


Figure 3.1: Causal diagram illustrating the separability assumptions. $N_1(t)$ is a counting process which jumps when a subject experiences the event of interest and $N_2(t)$ is a counting process which jumps at the competing event time. Thick lines indicate a deterministic relationship.

To address these limitations of classical competing risk estimands Stensrud et al. [2022] propose a novel estimand based on the concept of separable effects. Inspired by the interventionist approach to mediation analysis of Robins and Richardson [2011], Robins et al. [2021] decribed in Chapter 2 they propose to consider a decomposition of the baseline treatment A into two separable components A^Y and A^D such that A_Y is the component which affects the event of interest Y and A_D the the component which only affect Y through its effect on the competing event D.

Letting $T(a^Y, a^D)$ and $\varepsilon(a^Y, a^D)$ denote the counterfactual event time and event type respectively under an intervention that sets A^Y to a^Y and A^D to a^D we can define

$$P_1(t, a_Y, a_D) = P(T(a^Y, a^D) \leq t, \varepsilon(a^Y, a^D) = 1),$$

and we can define the separable direct effect (SDE) and separable indirect effecr (SIE) as

$$SDE = P_1(t, 1, 1) - P_1(t, 0, 1), \tag{3.5}$$

and

$$SIE = P_1(t, 0, 1) - P_1(t, 0, 0). \tag{3.6}$$

That is, the SDE is the effect of treatment on the event of interest outside of its effect on the competing event, and the separable indirect effect is the effect of treatment on the event time of interest only through the competing event.

This method was extended to continuous time by Martinussen and Stensrud [2023] and to the setting with time-varying confounders of the event of interest and the competing event in Stensrud et al. [2021]. Moreover it has been applied to settings with recurrent events [Janvin et al., 2023] and intercurrent events [Stensrud and Dukes, 2022].

Chapter 4

Overview of manuscripts

In this chapter we provide a brief overview of the three manuscripts, and relate them to the methods presented in Chapters 2 and 3.

4.1 Manuscript I

Manuscript I: Breum, M. S., Munch, A., Gerds, T. A. and Martinussen, T. (2023). Estimation of separable direct and indirect effects in a continuous-time illness-death model. *Lifetime Data Analysis*, 1-38.

In this manuscript we consider the problem of defining and estimating (in)direct effects when both the outcome and the potential mediator are time-to-event variables. In particular we assume that the mediator is a non-terminal event (T_1) corresponding to the time a subject enters the 'illness' state of the illness-death model, and that the outcome is a terminal event time (T_2) corresponding to the time a subject enter the 'death' state of an 'illness-death' model [Andersen et al., 2012]. The main challenge in defining causal mediation estimands in this setting is that the potential mediator is truncated by the outcome, meaning that manipulations of the mediator are inconceivable. The contribution of this manuscript is that we use the separable (interventionist) effects approach [Robins and Richardson, 2011, Stensrud et al., 2022, 2021] described in Chapters 2 and 3 to define meaningful causal estimands. Particularly we will assume that the baseline treatment indicator A can be separated into two binary components which we will denote A^I and A^D, where the component A^I only affects the terminal event through it's effect on the intermediate event, and the component A^D only affects the terminal event directly. This is depicted in Figure 4.1.

Letting $T_2(a^D, a^I)$ denote the counterfactual event times under an intervention that sets A^D to a^D and A^I to a^I we can define the separable direct

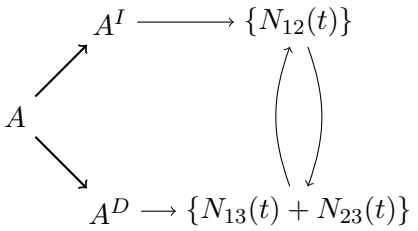


Figure 4.1: Causal diagram illustrating the separability assumptions. $N_{12}(t)$ is a counting process which jumps when a subject experiences the intermediate event and $N_{13}(t) + N_{23}(t)$ is a counting process which jumps at the terminal event time. Thick lines indicate a deterministic relationship. This is a simplified version of Fig 2 in manuscript I.

effect

$$E\left\{I(T_2(1,0) \leq \tau)\right\} - E\left\{I(T_2(0,0) \leq \tau)\right\},$$

and the separable indirect effect

$$E\left\{I(T_2(1,1) \leq \tau)\right\} - E\left\{I(T_2(1,0) \leq \tau)\right\}.$$

In the manuscript we derive the identification assumptions which are necessary to identify the separable (in)direct effects from the observed data. This builds upon work by Martinussen and Stensrud [2023] who consider similar causal estimands for the continous-time competing risk model. For estimation we propose a one-step estimator based on the efficient influence function. We show the robustness properties of the one-step estimator theoretically and we verify it in a simulation study. In addition to being multiply robust the one-step estimator is compatible with data-adaptive estimators of the transition intensity functions, provided certain rate conditions hold. This was not explored further in Manucript I. In Munch et al. [2023] a method is presented for estimating the transition intensity functions of the illness-death model based on penalized Poisson regression.

4.2 Manuscript II

Manuscript II: Breum, M. S. and Palle, M. S. (2023). Estimation of data-dependent (in)direct effects with a repeatedly measured mediator and missing outcome data. To be submitted to *Statistics in Medicine*.

In manuscript II we are motivated by the NASH phase II clinical trial conducted by Novo Nordisk. The NASH trial was a randomized trial constructed to evaluate the effect of Semaglutide on liver histology for patients suffering

from non-alcoholic steatohepatitis (NASH), which is an advanced form of non-alcoholic fatty liver disease. We were interested in decomposing the overall effect of Semaglutide into the part which arises indirecly by inducing body weight loss and the remaining direct effect of the drug. This is an example of a longitudinal mediation problem as the potential mediator is measured at multiple schedule visits during the study period. Particularly we expect feedback between certain time-varying covariates and the mediator, and these time-varying covariates may in turn be affected by treatment. As described in Chapter 2 there are many approaches to defining causal mediation estimands in the presence of post-treatment confounders. In this paper we take an approach based on interventional (in)direct effects which has been extended to the longitudinal setting by VanderWeele and Tchetgen Tchetgen [2017] and Zheng and van der Laan [2017]. We consider the effect decomposition proposed by Zheng and van der Laan [2017] since we want to define the mediated effect to include only paths directly from treatment to the mediator.

The contribution of our method compared to Zheng and van der Laan [2017] is that we consider data-adaptive versions of the randomized interventional (in)direct effects. That is we assume that the interventional mediator distribution is known and estimated from the data. This somewhat alters the interpretation of the (in)direct effects in the sense that it will depend on the estimated mediator distribution in the specific study. Moreover the direct and (in)direct effects are not guaranteed to provide a decomposition of the total treatment effect. An advantage of the data-adaptive method that we propose is that the effects are identified under slightly weaker identification assumptions than the non-data-adaptive version. Moreover the efficient influence function does not depend on the conditional densities of the covariates. For estimation we propose a longitudinal targeted minimum loss-based estimatior (LTMLE)[van der Laan and Rose, 2011, 2018]. We apply the method to the NASH trial where we find no evidence of a direct effect of Semaglutide on the primary endpoint not mediated through weight loss.

4.3 Manuscript III

Manuscript III: Breum, M. S. and Martinussen, T. (2023). Efficient nonparametric estimators of discrimination measures with censored survival data. Under preparation.

The topics of manuscript III is estimation of concordance measures for survival outcomes. Particularly we shall consider risk scores of the type $Y = \beta^T(P)X$ where X are the available markers and $\beta(P)$ some estimand, that we wish to evaluate in terms of their discriminatory power.

Some of the most widely used concordance measures in the medical literature are the C-index [Harrell et al., 1982, Harrell Jr et al., 1996], which, for

to random subjects indexed by i and j, can be informally defined as

$$\mathcal{C} = Pr(Y(i) \geq Y(j) \mid i \text{ has event before } j),$$

and concordance probability of Gönen and Heller [2005] which is informally defined as

$$K = Pr(i \text{ has event before } j \mid Y(i) \geq Y(j)).$$

If one is interested in t-year predicted risk rather than overall risk a more appropriate concordance measure is the t-year area under the receiver operating characteristic (ROC) curve (AUC_t) [Heagerty et al., 2000, Heagerty and Zheng, 2005, Blanche et al., 2013] which is informally defined as

$$\text{AUC}_t = Pr(Y_t(i) \geq Y_t(j) \mid i \text{ has event before } t \text{ and } j \text{ has event after } t),$$

where $Y_t = \beta_t^T(P)X$.

We note that these concordance measures are not causal per se, but this manuscript ties in with the rest of the thesis in that we use the 'estimand' framework from causal inference. Particularly the contribution of the manuscript is that we consider model free (non-parametric) definitions of the three concordance measures above and derive the efficient influence functions. We propose estimators based on the efficient influence function, which have advantages over existing estimators in that they do not assume independent censoring or rely on proportional hazards assumptions.

In addition to the concordance measures being defined non-parametrically so should $\beta_t(P)$. In the manuscript we define $\beta_t(P)$ using the assumption lean approach Vansteelandt and Dukes [2022], Vansteelandt et al. [2022] described in Chapter 3.

Chapter 5

Discussion and perspectives

In this chapter, we discuss limitations of the methods proposed in manuscripts I-III and we outline some possible topics for future research.

5.1 Separable direct and indirect effects with time-varying covariates

One of the main weaknesses in manuscript I is that we do not allow for time-varying covariates. Below we sketch how we believe that we can generalize our results to allow for covariates measured at random times on a continuous scale similar to the setting considerin in Rytgaard et al. [2022]:

Let $N_{\ell,1}(t)$ denote the counting process which jumps when we observe changes of a covariate vector $L(t)$ for patients in state 1, and similarly $N_{\ell,2}(t)$ the counting process which jumps when we observe changes of $L(t)$ for patients in state 2. Further we let $T_{\ell,1}^{(1)} < T_{\ell,1}^{(2)} < ... < T_{\ell,1}^{(N_{\ell,1}(\tau))}$ be the random times at which the covariate process changes for patients in state 1 and $T_{\ell,2}^{(1)} < T_{\ell,2}^{(2)} < ... < T_{\ell,2}^{(N_{\ell,2}(\tau))}$ the random times at which the covariate process changes for patients in state 2. Let $\mu_t()$ be the conditional density of covariates $L(t)$ at any time where $N_{\ell,1}(t)$ or $N_{\ell,2}(t)$ jumps, and let $\lambda_{\ell,1}()$ and $\lambda_{\ell,2}()$ be the intensities of times where the covariates change for patients in state 1 and state 2 respectively. To identify the target we will need the following modified dismissible components conditions

$$\lambda_{12}^{a^D=1,a^I}(t \mid \mathcal{L}_{t-}) = \lambda_{12}^{a^D=0,a^I}(t \mid \mathcal{L}_{t-}) \text{ for } a^I \in \{0,1\}$$
$$\lambda_{13}^{a^D,a^I=1}(t \mid \mathcal{L}_{t-}) = \lambda_{13}^{a^D,a^I=0}(t \mid \mathcal{L}_{t-}) \text{ for } a^D \in \{0,1\}$$
$$\lambda_{23}^{a^D,a^I=1}(t, t-r \mid \mathcal{L}_{t-}) = \lambda_{23}^{a^D,a^I=0}(t, t-r \mid \mathcal{L}_{t-}) \text{ for } a^D \in \{0,1\}$$

where $\mathcal{L}_{t-}$ is the history of the observed covariate process up to time t.

Some identification results for separable effects with time-varying confounding have been given in Stensrud et al. [2021] who define separable direct

and indirect effects in a discrete-time competing risk model with time-varying common causes of the event of interest and the competing event. We believe that similar identification results may be relevant to our setting.

Particularly in the case where the covariate vector $L(t)$ is affected by treatment we need to decide whether L is directly affected by component A^I or A^D or by something else like a third component. This means that the separable indirect effect will not necessarily reflect the effect mediated through the intermediate event and nothing else.

5.2 Mediation analysis without overlap

One of the weaknesses of the method presented in manuscript II is that when there no or limited overlap between the conditional mediator distributions in the control and treatment arms the weights/clever covariates for the TMLE algorithm may become very extreme.

In this case it might be of interest to estimate the following 'generalized' stochastic direct effect (GSDE)

$$GSDE(a, a') = \left\{Y(a, \mathbf{g}^*) - Y(a', \mathbf{g}^*)\right\},$$

where g_t^* is a stochastic mediator distribution that does not condition on A. For instance we could consider

$$g_t^*(M_t \mid \bar{L}_t, \bar{M}_{t-1}) = \sum_a P(M_t \mid A = a, \bar{C}_t = 0, \bar{L}_t, \bar{M}_{t-1}) P(A = a \mid L_0),$$

where we marginalize over A. The GSDE is the effect of treatment on the outcome under an intervention that assigns the mediator to a random draw from the same distribution in both treatment arms. Similar to the controlled direct effect the GSDE does not have a complementary indirect effect. When g^* is assumed to be known and estimated from the data, then the GSDE can be estimated using the framework developed in manuscript II.

A further extension of this is to identify the target with most support in the data by choosing g^* to minimize the dissimilarity

$$\text{var}\left\{D^*(P)(a, g^*)(P)(O) - D^*(P)(a, g^{obs})(P)(O)\right\},$$

where g^{obs} is the observed mediator distribution and $D^*(P)$ is the efficient influence function derived in manuscript II.

5.3 Alternative concordance measures

One of the main criticisms of the concordance index (c-index) [Harrell et al., 1982, Harrell Jr et al., 1996, Uno et al., 2011] apart from it not being proper

[Blanche et al., 2019], is that it is not a very useful measure for evaluating predictive accuracy when there are many low risk patients with similar risk scores. This is because these patients will attenuate the concordance value event when their difference in risk score is not clinically meaningful [Hartman et al., 2023].

A possible extension of the c-index that addresses this criticism is the modified concordance probability

$$\kappa_{\tau,\nu}(P) = P(T_2 > T_1, T_1 \leq \tau \mid F_Y(Y_1) \geq \nu > F_Y(Y_2)), \tag{5.1}$$

for $0 < \nu \leq 1$. By choosing a high value of ν we avoid making many comparisons between subjects with near average risk scores which could potentially deflate the value of the concordance probability. Since it is not evident how to best choose ν we suggest plotting $\kappa_{\tau,\nu}$ for different values of ν.

The modified concordance probability can be written as

$$\begin{aligned}\kappa_{\tau,\nu}(P) &= \frac{P(T_2 > T_1, T_1 \leq \tau, F_Y(Y_1) \geq \nu > F_Y(Y_2))}{P(F_Y(Y_1) \geq \nu > F_Y(Y_2)))} \\ &= \frac{\int_{y_2<c_\nu}\int_{y_1\geq c_\nu} h_\tau(y_1,y_2)dF_Y(y_1)dF_Y(y_2)}{\int_{y_2<c_\nu}\int_{y_1\geq c_\nu} dF_Y(y_1)dF_Y(y_2)},\end{aligned}$$

where $c_\nu = F^{-1}(\nu)$ and $h_\tau(y_1, y_2) = \int_0^\tau S(t \mid y_2)S(t \mid y_1)\,\mathrm{d}\Lambda(t \mid y_1)$.

Using that the efficient influence function of c_ν is $-f_Y^{-1}(c_\nu)\{I(Y \leq c_\nu) - \nu\}$ (see e.g. Hines et al. [2022] Appendix B) we can derive the efficient influence function of $\kappa_{\tau,\nu}(P)$ similar to the approach in manuscript III.

The usefulness of this estimator in practice remains to be explored, and is the topic of future research.

5.4 Mediation analysis in continous time

In important topic for future research is the extension of mediation analysis to a a more general setting where we allow for the values of the time-dependent exposure, mediator and confounders to be updated at random follow-up times. An example of this type of data structure is cohort studies based on the Danish Nationwide registries which contain daily updated vital status and medical records on all persons residing in Denmark. Such settings have been considered elsewhere (e.g. Rytgaard et al. [2022], Røysland et al. [2022] and Lok et al.) but to our knowledge not in a mediation context.

One of the main challenges in the continuous-time setting is that subjects potentially have different number of measurements of the mediator process. Subjects with many measurements are likely different than subjects with few measurements (e.g. more sick). More research is needed to explore which assumptions are necessary to be able to distinguish between the effect of having many measurements and the true mediated effect.

Bibliography

O. O. Aalen, . Borgan, and H. K. Gjessing. *Survival and Event History Analysis: A Process Point of View*. Springer Science & Business Media, 2008.

O. O. Aalen, R. J. Cook, and K. Røysland. Does Cox analysis of a randomized survival study yield a causal treatment effect? *Lifetime Data Analysis*, 21: 579–593, 2015.

O. O. Aalen, M. J. Stensrud, V. Didelez, R. Daniel, K. Roysland, and S. Strohmaier. Time-dependent mediators in survival analysis: Modeling direct and indirect effects with the additive hazards model. *Biometrical Journal*, 62(3):532—549, 2020.

P. K. Andersen, O. Borgan, R. D. Gill, and N. Keiding. *Statistical models based on counting processes*. Springer Science & Business Media, 2012.

R. M. Andrews and V. Didelez. Insights into the "cross-world" independence assumption of causal mediation analysis. *Epidemiology*, 32:209–219, 2021.

C. Avin, I. Shpitser, and J. Pearl. Identifiability of path-specific effects. In *Proceedings of the international joint conference on artificial intelligence*, volume 19, pages 357–363. Morgan Kaufmann, 2005.

R. Baron and D. Kenny. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51:1173–1182, 1986.

P. Blanche, J.-F. Dartigues, and H. Jacqmin-Gadda. Review and comparison of roc curve estimators for a time-dependent outcome with marker-dependent censoring. *Biometrical Journal*, 55(5):687–704, 2013.

P. Blanche, M. W. Kattan, and T. A. Gerds. The c-index is not proper for the evaluation of-year predicted risks. *Biostatistics*, 20(2):347–357, 2019.

Z. Cai, M. Kuroki, J. Pearl, and J. Tian. Bounds on direct effects in the presence of confounded intermediate variables. *Biometrics*, 64(3):695–701, 2008.

V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.

D. R. Cox. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society Series B*, 34:187–220, 1972.

I. Díaz and N. S. Hejazi. Causal Mediation Analysis for Stochastic Interventions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(3):661–683, 2020.

J. De Neve and T. A. Gerds. On the interpretation of the hazard ratio in cox regression. *Biometrical Journal*, 62(3):742–750, 2020.

V. Didelez. Defining causal mediation with a longitudinal mediator and a survival outcome. *Lifetime data analysis*, 25(4):593–610, 2019.

V. Didelez, P. Dawid, and S. Geneletti. Direct and indirect effects of sequential treatments. In *Proc. 22nd Annual Conference on Uncertainty in Artificial Intelligence*, pages 138–146, 2006.

O. Dukes, I. Shpitser, and E. J. Tchetgen Tchetgen. Proximal mediation analysis. *Biometrika*, 2023.

A. Fisher and E. H. Kennedy. Visually communicating and teaching intuition for influence functions. *The American Statistician*, 2(75):162–172, 2021.

S. Geneletti. Identifying direct and indirect effects in a non-counterfactual framework. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):199–215, 2007.

M. Gönen and G. Heller. Concordance probability and discriminatory power in proportional hazards regression. *Biometrika*, 92(4):965–970, 2005.

F. E. Harrell, R. M. Califf, D. B. Pryor, K. L. Lee, and R. A. Rosati. Evaluating the yield of medical tests. *Journal of the American Medical Association*, 247 (18):2543–2546, 1982.

F. E. Harrell Jr, K. L. Lee, and D. B. Mark. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine*, 15(4):361–387, 1996.

N. Hartman, S. Kim, K. He, and J. D. Kalbfleisch. Pitfalls of the concordance index for survival outcomes. *Statistics in Medicine*, 42(13):2179–2190, 2023.

P. J. Heagerty and Y. Zheng. Survival model predictive accuracy and roc curves. *Biometrics*, 61(1):92–105, 2005.

P. J. Heagerty, T. Lumley, and M. S. Pepe. Time-dependent roc curves for censored survival data and a diagnostic marker. *Biometrics*, 56(2):337–344, 2000.

N. S. Hejazi, K. E. Rudolph, M. J. V. D. Laan, and I. Díaz. Nonparametric causal mediation analysis for stochastic interventional (in)direct effects. *Biostatistics*, 24(3):686–707, 2022.

M. Hernán and S. Taubman. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *International Journal of Obesity*, 32(Suppl 3):S8–S14, 2008.

M. A. Hernán. The hazards of hazard ratios. *Lifetime Data Analysis*, 21(1): 13–15, 2010.

O. Hines, O. Dukes, K. Diaz-Ordaz, and S. Vansteelandt. Demystifying statistical learning based on efficient influence functions. *The American Statistician*, 76(3):292–304, 2022.

M. Janvin, J. Young, P. Ryalen, and M. J. Stensrud. Causal inference with recurrent and competing events. *Lifetime Data Analysis*, pages 1–60, 2023.

J. D. Kalbfleisch and R. L. Prentice. Estimation of the average hazard ratio. *Biometrika*, 68(1):105–112, 1981.

S. Kaufman, J. Kaufman, M. R.F., S. Greenland, and C. Poole. Improved estimation of controlled direct effects in the presence of unmeasured confounding of intermediate variables. *Statistics in Medicine*, 24(11):1683–1702, 2005.

E. Kennedy. Semiparametric theory and empirical processes in causal inference. In H. He, P. Wu, and D.-G. Chen, editors, *Statistical Causal Inferences and Their Applications in Public Health Research*, pages 141–167. 2016.

E. H. Kennedy. Nonparametric causal effects based on incremental propensity score interventions. *Journal of the American Statistical Association,*, 114 (2):645–656, 2019.

E. H. Kennedy. Semiparametric doubly robust targeted double machine learning: a review. arXiv preprint arXiv:2203.06469, 2023.

T. Lange and J. V. Hansen. Direct and Indirect Effects in a Survival Context. *Epidemiology*, 22(4):575–581, 2011.

T. Lange, S. Vansteelandt, and M. Bekaert. A Simple Unified Approach for Estimating Natural Direct and Indirect Effects. *American Journal of Epidemiology*, 176(3):190–195, 2012.

S. H. Lin, J. G. Young, R. Logan, and T. J. VanderWheele. Mediation analysis for a survival outcome with time-varying exposures, mediators, and confouncers. *Statistics in Medicine*, 36(26):4153–4166, 2017.

J. Lok, R. Gill, A. Van Der Vaart, and J. Robins. Estimating the causal effect of a time-varying treatment on time-to-event using structural nested failure time models. *Statistica Neerlandica*, 58(3):271–295.

J. J. Lok. Defining and estimating causal direct and indirect effects when setting the mediator to specific values is not feasible. *Statistics in Medicine*, 35:4008–4020, 2016.

J. J. Lok. Organic direct and indirect effects with post-treatment common causes of mediator and outcome. arXiv preprint arXiv:1510.02753, 2020.

J. J. Lok and R. J. Bosch. Causal organic indirect and direct effects: Closer to the original approach to mediation analysis, with a product method for binary mediators. *Epidemiology*, 32(3):412–420, 2011.

T. Martinussen and T. H. Scheike. *Dynamic Regression Models for Survival Data*. Springer Science & Business Media, 2006.

T. Martinussen and M. J. Stensrud. Estimation of separable direct and indirect effects in continuous time. *Biometrics*, 79(1):127–139, 2023.

T. Martinussen, S. Vansteelandt, and P. Andersen. Subtleties in the interpretation of hazard contrasts. *Lifetime Data Analysis*, 26:833–855, 2020.

C. Miles, P. Kanki, S. Meloni, and E. Tchetgen Tchetgen. On partial identification of the pure direct effect. arXiv preprint arXiv:1509.01652, 2015.

A. Munch, M. S. Breum, T. Martinussen, and T. A. Gerds. Targeted estimation of state occupation probabilities for the non-markov illness-death model. *Scandinavian Journal of Statistics*, 50(3):1532–1551, 2023.

J. Neyman. Sur les applications de la theorie des probabilites aux experiences agri- coles: Essai des principes. *Statistical Science*, 5:465–480, 1923.

J. Pearl. *Causality*. Cambridge University Press, 2000.

J. Pearl. Direct and indirect effects. In *Proceedings of the seventeenth conference on uncertainty in artificial intelligence*, pages 411–420. Morgan Kaufman, 2001.

M. L. Petersen and M. J. van der Laan. Causal models and learning from data: integrating causal modeling and statistical estimation. *Epidemiology*, 25(3):418–426, 2014.

J. Robins. Semantics of causal dag models and the identification of direct and indirect effects. In P. Green, N. Hjort, and S. Richardson, editors, *Highly Structured Stochastic System*, pages 70–81. Oxford University Press, Oxford, 2003.

J. Robins, M. Hernan, and B. Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11:550–560, 2000.

J. M. Robins. A new approach to causal inference in mortality studies with sustained exposure periods – application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7(9):1393–1512, 1986.

J. M. Robins and S. Greenland. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 3(2):143–155, 1992.

J. M. Robins and T. S. Richardson. Alternative graphical causal models and the identification of direct effects. In P. E. Shrout, K. M. Keyes, and K. Ornstein, editors, *Causality and psychopathology: Finding the determinants of disorders and their cures*, chapter 6, page 103–158. Oxford University Press, 2011.

J. M. Robins, T. S. Richardson, and I. Shpitser. An interventionist approach to mediation analysis. arXiv preprint arXiv:2008.06019, 2021.

D. B. Rubin. Bayesian inference for causal effects: The role of randomization. *Annals of statistics*, 6:34–58, 1978.

K. E. Rudolph, O. Sofrygin, W. Zheng, and M. J. van der Laan. Robust and Flexible Estimation of Stochastic Mediation Effects: A Proposed Method and Example in a Randomized Trial Setting. *Epidemiologic Methods*, 1(7), 2018.

K. Røysland, P. Ryalen, M. Nygård, and V. Didelez. Graphical criteria for the identification of marginal causal effects in continuous-time survival and event-history analyses. arXiv preprint arXiv:2202.02311, 2022.

H. C. Rytgaard, T. A. Gerds, and M. J. van der Laan. Continuous-time targeted minimum loss-based estimation of intervention-specific mean outcomes. *Annals of Statistics*, 2022.

M. Schemper. Cox analysis of survival data with non-proportional hazard functions. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 41(4):455–465, 1992.

I. Shpitser. Counterfactual graphical models for longitudinal mediation analysis with unobserved confounding. *Cognitive Science*, 37(6):1011–1035, 2013.

I. Shpitser. Identification in graphical causal models. In M. Maathuis, M. Drton, S. Lauritzen, and M. Wainwright, editors, *Handbook of Graphical Models*, chapter 16, pages 381–403. CRC Press, 2018.

J. Steen and S. Vansteelandt. Mediation analysis. In M. Maathuis, M. Drton, S. Lauritzen, and M. Wainwright, editors, *Handbook of Graphical Models*, chapter 17, pages 405–437. CRC Press, 2018.

M. J. Stensrud and O. Dukes. Translating questions to estimands in randomized clinical trials with intercurrent events. *Statistics in Medicine*, 41(16): 3211–3228, 2022.

M. J. Stensrud, M. Hernan, E. J. T. Tchetgen, J. M. Robins, and V. Didelez. A generalized theory of separable effects in competing event settings. *Lifetime Data Analysis*, 27(4):588–631, 2021.

M. J. Stensrud, J. G. Young, V. Didelez, J. M. Robins, and M. A. Hernán. Separable effects for causal inference in the presence of competing events. *Journal of the American Statistical Association*, 117(537):175–183, 2022.

C. A. Struthers and J. D. Kalbfleisch. Misspecified proportional hazard models. *Biometrika*, 73(2):363–369, 1986.

E. Tchetgen Tchetgen and K. Phiri. Bounds for pure direct effect. *Epidemiology*, 25(5):775–776, 2014.

A. Tsiatis. *Semiparametric theory and missing data*. Springer Science & Business Media, 2006.

H. Uno and M. Horiguchi. Ratio and difference of average hazard with survival weight: New measures to quantify survival benefit of new therapy. *Statistics in Medicine*, 42(7):936–952, 2023.

H. Uno, T. Cai, M. J. Pencina, R. B. D'agostino, and L. J. Wei. On the c-statistic for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in Medicine*, 30(10):1105–1117, 2011.

M. van der Laan and J. Robins. *Unified Methods for Censored Longitudinal Data and Causality*. Springer Science & Business Media, 2003.

M. J. van der Laan and M. L. Petersen. Direct effect models. *The International Journal of Biostatistics*, 4(1):Article 23, 2008.

M. J. van der Laan and S. Rose. *Targeted learning: Causal inference for observational and experimental Data*. Springer Science & Business Media, 2011.

M. J. van der Laan and S. Rose. *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies*. Springer Science & Business Media, 2018.

A. W. van der Vaart. *Asymptotic statistics*. Cambridge University Press, 3 edition, 2000.

T. J. VanderWeele. *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press, 2015.

T. J. VanderWeele and E. J. Tchetgen Tchetgen. Mediation analysis with time varying exposures and mediators. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 79(3):917–938, 2017.

T. J. VanderWeele and S. Vansteelandt. Conceptual issues concerning mediation, interventions and composition. *Statistics and Its Interface*, 2:457–468, 2009.

T. J. Vanderweele, S. Vansteelandt, and J. M. Robins. Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology*, 25(2):300–306, 2014.

S. Vansteelandt and R. Daniel. Interventional effects for mediation analysis with multiple mediators. *Epidemiology*, 28(2):258–265, 2017.

S. Vansteelandt and O. Dukes. Assumption-lean Inference for Generalised Linear Model Parameters. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(3):657–685, 2022.

S. Vansteelandt, M. Bekaert, and T. Lange. Imputation Strategies for the Estimation of Natural Direct and Indirect Effects. *Epidemiologic Methods*, 1(1):Article 7, 2012.

S. Vansteelandt, M. Linder, S. Vandenberghe, J. Steen, and J. Madsen. Mediation analysis of time-to-event endpoints accounting for repeatedly measured mediators subject to time-varying confounding. *Statistics in Medicine*, 38 (24):4828–4840, 2019.

S. Vansteelandt, O. Dukes, K. V. Lancker, and T. Martinussen. Assumption-lean Cox Regression. *Journal of the American Statistical Association*, pages 1–10, 2022.

D. Whitney, A. Shojaie, and M. Carone. Comment: Models as (deliberate) approximations. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 34(4):591–598, 2019.

R. Xu and J. O'Quigley. Estimating average regression effect under non-proportional hazards. *Biostatistics*, 1(4):423–39, 2000.

J. G. Young, M. J. Stensrud, E. J. Tchetgen Tchetgen, and M. A. Hernán. A causal framework for classical statistical estimands in failure-time settings with competing events. *Statistics in Medicine*, 39(8):1199–1236, 2020.

W. Zheng and M. J. van der Laan. Longitudinal mediation analysis with time-varying mediators and exposures, with application to survival outcomes. *Journal of Causal Inference*, 5(2), 2017.

Manuscript I

Estimation of separable direct and indirect effects in a continuous-time illness-death model

Marie Skov Breum, Anders Munch, Thomas A. Gerds and Torben Martinussen

Details:
The manuscript has been accepted for publication in Lifetime Data Analysis.

Lifetime Data Analysis
https://doi.org/10.1007/s10985-023-09601-y

Estimation of separable direct and indirect effects in a continuous-time illness-death model

Marie Skov Breum[1] · **Anders Munch**[1] · **Thomas A. Gerds**[1] · **Torben Martinussen**[1]

Received: 14 March 2022 / Accepted: 19 April 2023

Abstract

In this article we study the effect of a baseline exposure on a terminal time-to-event outcome either directly or mediated by the illness state of a continuous-time illness-death process with baseline covariates. We propose a definition of the corresponding direct and indirect effects using the concept of separable (interventionist) effects (Robins and Richardson in Causality and psychopathology: finding the determinants of disorders and their cures, Oxford University Press, 2011; Robins et al. in arXiv:2008.06019, 2021; Stensrud et al. in J Am Stat Assoc 117:175–183, 2022). Our proposal generalizes Martinussen and Stensrud (Biometrics 79:127–139, 2023) who consider similar causal estimands for disentangling the causal treatment effects on the event of interest and competing events in the standard continuous-time competing risk model. Unlike natural direct and indirect effects (Robins and Greenland in Epidemiology 3:143–155, 1992; Pearl in Proceedings of the seventeenth conference on uncertainty in artificial intelligence, Morgan Kaufmann, 2001) which are usually defined through manipulations of the mediator independently of the exposure (so-called cross-world interventions), separable direct and indirect effects are defined through interventions on different components of the exposure that exert their effects through distinct causal mechanisms. This approach allows us to define meaningful mediation targets even though the mediating event is truncated by the terminal event. We present the conditions for identifiability, which include some arguably restrictive structural assumptions on the treatment mechanism, and discuss when such assumptions are valid. The identifying functionals are used to construct plug-in estimators for the separable direct and indirect effects. We also present multiply robust and asymptotically efficient estimators based on the efficient influence functions. We verify the theoretical properties of the estimators in a simulation study, and we demonstrate the use of the estimators using data from a Danish registry study.

Keywords Separable effects · Illness-death model · Survival analysis · Mediation analysis · Causal inference

Extended author information available on the last page of the article

Published online: 04 June 2023

1 Introduction

Mediation analysis is an important tool in medical and epidemiological research for understanding the mechanisms that contribute to the overall effect of a treatment or exposure on an outcome of interest. Within the causal inference literature on mediation analysis, the target estimands of interest are often the natural (pure) direct and indirect effects (Robins and Greenland 1992; Pearl 2001), which together provide a nonparametric decomposition of the total treatment effect. A comprehensive overview of mediation analysis methods from a causal inference perspective can be found in VanderWeele (2015).

In this paper we study a continuous-time illness-death process where the potential mediator is the illness state. We are interested in the direct and indirect effect of a baseline exposure on the terminal event, adjusted for a set of pre-exposure covariates. This type of target estimand is often relevant when analysing real world data. We shall illustrate our method using a Danish registry study investigating the effects of dual antiplatelet therapy (DAPT) after myocardial infarction (MI) or stroke on mortality. DAPT is a treatment that combines aspirin and a second antiplatelet agent, which is often prescribed to MI or stroke patients to prevent blood clotting. It is well known that DAPT is associated with a lower risk of a recurrent cardiovascular event (Wallentin et al. 2009) which in turn is associated with increased mortality; this is the indirect effect of interest. At the same time DAPT has other effects that are associated with increased mortality, most notably it increases the risk of gastrointestinal bleeding (Kazi et al. 2015; Dinicolantonio et al. 2013); this is the direct effect of interest.

The conventional definition of natural direct and indirect effects is based on so-called cross-world quantities which require that we manipulate the mediator for each exposed individual to what would have occurred under non-exposure. Such quantities are not well defined in the illness-death setting since the mediator is effectively undefined when the terminal event occurs before the mediating event. This has implications for formulating the causal mediation targets of interest.

The term ‘semi-competing risks’ is often used in the literature when the outcome of interest is a non-terminal event that competes with a terminal time-to-event (Fine et al. 2001). We find that the definition of this term is unclear as discussed in Stensrud et al. (2021), and will refrain from using it in this paper. We will use the term "truncation" to describe the phenomenon when occurrence of the terminal event renders the intermediate event undefined, and the term "illness-death process" to describe the underlying data structure.

The challenges that arise when defining mediation targets for the illness-death models are similar to the well known challenges that arise when defining mediation targets for a survival outcome with a time-dependent mediator. Recent approaches in the literature redefine the target of interest beyond that of natural direct and indirect effects using randomized interventions (Zheng and van der 2017; Lin et al. 2017), path-specific effects (Vansteelandt et al. 2019) or separable effects (Didelez 2019; Aalen et al. 2020). While the setting in these papers is more general in that they allow for adjustment for time-varying covariates, they assume that the mediator

process is measured at discrete time-points, and are thus not directly applicable to our setting where we allow the mediator process to change in continuous time.

Similar to Didelez (2019) and Aalen et al. (2020) we propose a definition of the direct and indirect effects using a treatment separation approach which is commonly referred to as the 'separable effects' approach (Stensrud et al. 2022, 2021) or 'interventionist' approach (Robins and Richardson 2021) to causal mediation analysis. Based on an idea by Robins and Richardson (2011) this approach considers a hypothetical treatment decomposition under which it is possible to consider manipulations of the mediator independently of the treatment given. This is done by assuming that treatment has two binary components, a 'direct' one which is thought to affect the terminal event directly, and an 'indirect' one which only affects survival through it's effect on the intermediate event, and that the two components can be intervened upon separately. This makes it possible to define meaningful mediation targets even when the mediating event is truncated by death. The aim of this paper is to show how this approach can be applied to the continuous-time illness-death setting, and to derive estimators using semiparametric theory. In particular, the identifiability conditions and estimators we propose in this paper are an extension of Martinussen and Stensrud (2023), who consider similar causal targets and estimators in a continuous-time competing risk model.

The paper is organized as follows: In Sect. 2 we introduce the irreversible illness-death model as a stochastic process and describe the observed data structure. In Sect. 3 we formulate the targets of interest and present the identifiability conditions. In Sect. 4 we derive the efficient influence functions and establish their multiple robustness properties. We also suggest two estimators: a plug-in estimator based on the identifying functional and a one-step estimator based on using the efficient influence function as an estimating equation. We examine the performance of the estimators in a simulation study in Sect. 5. Section 6 illustrates the methods in the Danish registry data application. In Sect. 7 we provide further discussion. Proofs and additional technical details are given in the Appendices.

2 Setting and notation

2.1 Illness-death model

We consider an irreversible illness-death model, as depicted in Fig. 1. Following Andersen et al. (2012) the illness-death model is a stochastic process $\{X(t)\}_{t\in[0,\infty)}$ with right-continuous sample paths and state space $\{1, 2, 3\}$, where state 1 is the initial 'healthy' state, state 2 is the intermediate 'illness' state and state 3 corresponds to the absorbing state 'death'. We assume that $X(0) = 1$, i.e. all subjects start in the initial 'healthy' state. We further assume that $2 \rightarrow 1$ transitions are not possible, i.e. the process is irreversible. In our DAPT example a patient enters state 1 when experiencing a myocardial infarction (MI) for the first time. The patient stays in state 1 until they either die or experience a recurrent cardiovascular event. In the latter case the patient moves to state 2 where they remain until death.

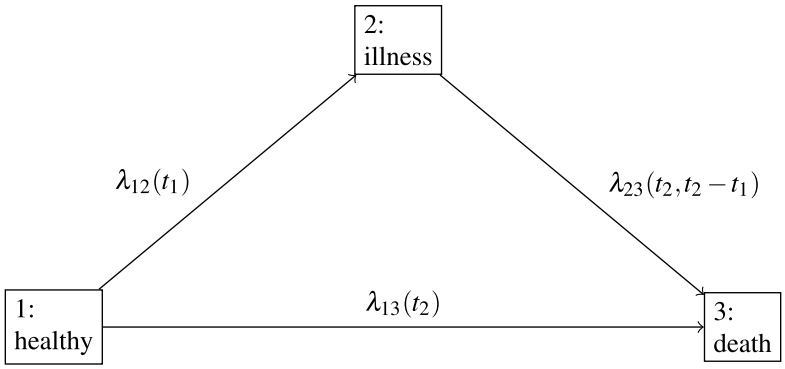


Fig. 1 Illness-death model without recovery

We define time until the subject leaves state 1 $T_1 = \inf_{t>0}\{X(t) \neq 1\}$ and time until death $T_2 = \inf_{t>0}\{X(t) = 3\}$. In addition to T_1 and T_2 we define the indicator

$$\eta = \begin{cases} 1 & \text{if } T_1 < T_2 \\ 0 & \text{if } T_1 = T_2 \end{cases}.$$

That is, $\eta = 1$ corresponds to $X(T_1) = 2$ and $X(T_2) = 3$, and $\eta = 0$ corresponds to $X(T_1) = X(T_2) = 3$.

Then the hazards for the transitions between states $1 \to 2$, $1 \to 3$ and $2 \to 3$, respectively, are defined as follows

$$\begin{aligned} \lambda_{12}(t) &= \lim_{\mathrm{d}t \to 0} \frac{Pr(T_1 \leq t + \mathrm{d}t, \eta = 1 \mid T_1 > t)}{\mathrm{d}t}, \\ \lambda_{13}(t) &= \lim_{\mathrm{d}t \to 0} \frac{Pr(T_1 \leq t + \mathrm{d}t, \eta = 0 \mid T_1 > t)}{\mathrm{d}t}, \\ \lambda_{23}(t, t-r) &= \lim_{\mathrm{d}t \to 0} I(r \leq t) \frac{Pr(T_2 \leq t + \mathrm{d}t \mid T_2 > t, T_1 = r)}{\mathrm{d}t}. \end{aligned}$$

2.2 Data structure

Let $A \in \{0, 1\}$ be a baseline treatment indicator, and $W \in \mathcal{W} = \mathbb{R}^d$ a vector of baseline covariates. The full uncensored data are $Z = \{T_2, T_1, \eta, A, W\} \sim Q$ where Q is a probability distribution belonging to a non-parametric statistical model $\mathcal{Q}$. Let μ be the density of W and $\pi(\cdot \mid W)$ be the conditional distribution of A given W which we will refer to as the propensity score. The underlying density q of the data Z under Q factorizes as follows

$$\begin{aligned} q(t, r, \eta, a, w) = &\left\{\lambda_{12}(r \mid a, w)\lambda_{23}(t, t-r \mid a, w)S_2(t- \mid r, a, w)\right\}^{\eta}\left\{\lambda_{13}(r \mid a, w)\right\}^{1-\eta} \\ &\times S_1(r- \mid a, w)\pi(a \mid w)\mu(w), \end{aligned} \tag{1}$$

where

$$S_1(t \mid a, w) = \exp\left\{-\Lambda_{12}(t \mid a, w) - \Lambda_{13}(t \mid a, w)\right\},$$
$$S_2(t \mid r, a, w) = \exp\left\{-\Lambda_{23}(t, t-r \mid a, w)\right\},$$

for

$$\Lambda_{12}(t \mid a, w) = \int_0^t \lambda_{12}(s \mid a, w)\,\mathrm{d}s, \quad \Lambda_{13}(t \mid a, w) = \int_0^t \lambda_{13}(s \mid a, w)\,\mathrm{d}s,$$
$$\Lambda_{23}(t, t-r \mid a, w) = \int_r^t \lambda_{23}(s, s-r \mid a, w)\,\mathrm{d}s.$$

That is, S_1 is the survival probability for the patients in state 1 and S_2 is the survival probability for patients in state 2.

We also let $N_{13}(s) = I(T_2 \leq s, \eta = 0)$, $N_{12}(s) = I(T_1 \leq s, \eta = 1)$ and $N_{23}(s) = I(T_2 \leq s, \eta = 1)$ denote the full-data counting processes corresponding to the transitions between states $1 \to 3$, $1 \to 2$ and $2 \to 3$, respectively. In our DAPT example N_{13} is the counting process which jumps when a patient in the study dies without having a recurrent cardiovascular event. Further N_{12} jumps when a patient experiences a recurrent cardiovascular event, and N_{23} jumps when when a patient in the study dies having experienced a recurrent cardiovascular event.

2.3 Right censoring

We allow for right censoring with $\tilde{C}$ denoting the censoring variable corresponding to the time that an individual would be lost to followup. Under right censoring we only observe $\tilde{T}_1 = T_1 \wedge \tilde{C}$, $\tilde{T}_2 = T_2 \wedge \tilde{C}$ and the indicators $\delta = I(T_2 < \tilde{C})$ and $\tilde{\eta} = I(\tilde{T}_1 < \tilde{T}_2)$. The observed data may then be represented as $O = \{\tilde{T}_2, \delta, \tilde{T}_1, \tilde{\eta}, A, W\} \sim P$ where P belongs to a non-parametric statistical model $\mathcal{P}$.

We may also define the observed-data counting processes $\tilde{N}_{13}(s) = I(\tilde{T}_2 \leq s, \tilde{\eta} = 0, \delta = 1)$, $\tilde{N}_{12}(s) = I(\tilde{T}_1 \leq s, \tilde{\eta} = 1)$ and $\tilde{N}_{23}(s) = I(\tilde{T}_2 \leq s, \tilde{\eta} = 1, \delta = 1)$ corresponding to the observed transitions between states $1 \to 3, 1 \to 2$ and $2 \to 3$, respectively.

We make the coarsening at random (CAR) assumption, i.e., we assume that the coarsening probabilities only depend on the data as a function of the observed data. This assumption is stated more formally in Appendix A. Under CAR we can define the increments of the censoring martingale

$$\mathrm{d}M_{\tilde{C}}\{u, O\} = \mathrm{d}N_{\tilde{C}}(u) - \lambda_{\tilde{C}}\{s; O\} I(\tilde{C} > s)\,\mathrm{d}s,$$

where

$$\mathrm{d}N_{\tilde{C}}(s) = I(s \leq \tilde{C} < s + \mathrm{d}s, T_1 > \tilde{C}) + I(s \leq \tilde{C} < s + \mathrm{d}s, T_1 < \tilde{C} \leq T_2)$$

is the censoring counting process corresponding to the observed censored observations up to and including time s, and

$$\lambda_{\tilde{C}}\{s; O\} I(\tilde{C} > s) = I(\tilde{T}_1 > s)\alpha_{\tilde{C},1}(s \mid A, W) + I(T_1 \leq s < \tilde{T}_2)\alpha_{\tilde{C},2}(s \mid T_1, A, W)$$

is the corresponding censoring intensity. We also define

$$K_{\tilde{C},1}(u \mid a, w) = \exp\left\{ -\int_0^u \alpha_{\tilde{C},1}(s \mid a, w)\, \mathrm{d}s \right\},$$
$$K_{\tilde{C},2}(v \mid u, a, w) = \exp\left\{ -\int_u^v \alpha_{\tilde{C},2}(s \mid u, a, w)\, \mathrm{d}s \right\},$$

which are the probabilities of being uncensored for patients in state 1 and state 2, respectively.

3 Separable direct and indirect effects

To define our estimand of interest we will use the concept of separable effects (Robins and Richardson 2011; Robins et al. 2021; Stensrud et al. 2022), which was briefly introduced in Sect. 1. This approach to mediation analysis moves the focus from intervening on the mediator process, which is conceptually problematic in the illness-death setting, to interventions on different components of the treatment A. To make the treatment decomposition more explicit we will think of the treatment A as having two binary components which we will denote A^I and A^D. As depicted in Fig. 2 we will assume that the component A^I only affects the terminal event through it's effect on the intermediate event, and that the component A^D only affects the terminal event directly. We will think of the corresponding four-arm trial as our 'target trial' and will define our target parameters based on the counterfactual variables defined by this target trial. In the observed data we have either $A^D = A^I = 1$ or $A^D = A^I = 0$, but we presume that an intervention is possible where $A^D \neq A^I$, i.e. the components could be set to different values. If such treatment components are assumed to exist and appropriate identification assumptions hold, then it is not necessary to conduct the four arm target trial. In fact the target parameters may be identified from the observed two-arm trial under the assumptions stated in Lemma 1 below.

This way of thinking about mediation analysis in terms of 'separable effects' can be useful when investigators want to know whether a specific mechanism of exposure is associated with the outcome. Often the hypothesis of interest concerns a specific 'active ingredient' of the exposure which may be difficult or impossible to measure.

In our example from Sect. 1 DAPT has been shown to have a protective effect on recurrent cardiovascular events, and is therefore often prescribed to MI or stoke

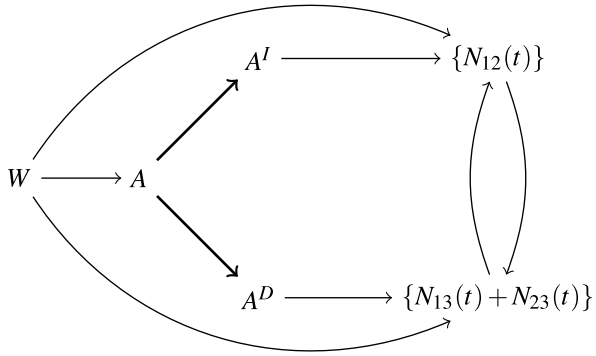


Fig. 2 An informal causal diagram illustrating the relationship between the treatment components and the counting processes. The thick edges indicate a deterministic relationship

patients. However DAPT is also associated with an increased risk of major bleeding (Wallentin et al. 2009). One of the primary forms of bleeding is gastrointestinal bleeding due to ulcers (Kazi et al. 2015; Dinicolantonio et al. 2013). We can then imagine a hypothetical treatment component A^D which has the same effect as DAPT on mortality, but lacks any effect on cardiovascular events, and a hypothetical treatment component A^I which has the same effect as DAPT on cardiovascular events but no direct effect on mortality. These treatment components do not necessarily correspond to meaningful real-world quantities. However, it can sometimes be useful to imagine them as hypothetical combination treatments. Assuming that gastrointestinal bleeding is the main effect of DAPT besides it's effect on cardiovascular outcomes, the A^I component would correspond to a modified treatment that does not promote ulcers. In practice, a drug that combines DAPT with an additional drug that promotes healing of ulcers and thereby nullifies the harmful effect DAPT may resemble this hypothetical treatment. For instance a recent Danish registry study has shown that proton pump inhibitors (PPI's) can induce ulcer healing among patients treated with DAPT (Sehested et al. 2019).

It is important to note that the validity of the approach does not depend on whether the treatment components correspond to meaningful real life quantities. The validity of the approach does however depend crucially on the assumption that the two treatment components can be manipulated separately which is a strong assumption.

3.1 Parameter of interest

For $j = 1, 2$ we let $T_j^{a^D,a^I}$ denote the counterfactual event times under an intervention that sets A^D to a^D and A^I to a^I and let T_j^a denote the counterfactual event times under an intervention that sets $A = a$ in the observed two-arm trial.

Then, the separable direct effect (SDE) and separable indirect effect (SIE) of the illness-death model are respectively defined as

$$SDE(\tau, a^I) = E\Big\{ I(T_2^{1,a^I} \leq \tau) \Big\} - E\Big\{ I(T_2^{0,a^I} \leq \tau) \Big\} \text{ for } a^I \in \{0, 1\}, \tag{2}$$

and

$$SIE(\tau, a^D) = E\Big\{ I(T_2^{a^D,1} \leq \tau) \Big\} - E\Big\{ I(T_2^{a^D,0} \leq \tau) \Big\} \text{ for } a^D \in \{0, 1\}. \tag{3}$$

where $E(\cdot)$ denotes expectations computed under the data-generating distribution.

That is, the SDE is the counterfactual contrast under $A^D = 1$ and $A^D = 0$ when A^I is fixed at some level a^I. The SIE is the counterfactual contrast under $A^I = 1$ and $A^I = 0$ when A^D is fixed at a^D.

Note that the separable direct and indirect effect add up to the total treatment effect

$$\begin{aligned} TE(\tau, a) =& SDE(\tau, a) + SIE(\tau, 1 - a) = E\big\{ I(T_2^{1,1} \leq \tau) \big\} \\ & - E\big\{ I(T_2^{0,0} \leq \tau) \big\} \text{ for } a \in \{0, 1\}. \end{aligned} \tag{4}$$

3.2 Identifiability conditions

In order to identify the parameters of the target trial given in Eqs. (2)–(3) from the observed two-arm trial we need the following assumptions

Lemma 1 (Identifiability) *Suppose the following assumptions hold*

A.0 We assume that the interventions are such that

$$T_j^{A^D=a,A^I=a} = T_j^a \text{ for } j = 1, 2$$

A.1 Conditional exchangeability:

$$(T_2^a, T_1^a) \perp\!\!\!\perp A \mid W \text{ for } a \in \{0, 1\}.$$

A.2 Consistency: If an individual is observed to receive treatment $A = a$, then

$$T_j^a = T_j \text{ for } j = 1, 2$$

A.3 Positivity:

$$\mu(w) > 0 \Rightarrow \pi(a \mid W = w) > 0 \text{ for } a \in \{0, 1\} \text{ and } w \in \mathcal{W},$$

and

$$P(T_1 > t \mid W = w) > 0 \Rightarrow$$
$$P(\tilde{T}_1 > t, A = a \mid W = w) > 0 \text{ for } a \in \{0, 1\}, t < \tau \text{ and } w \in \mathcal{W},$$

and

$$P(T_2 > t > T_1, T_1 = r \mid W = w) > 0 \Rightarrow$$
$$P(\tilde{T}_2 > t > T_1, T_1 = r, A = a \mid W = w) > 0 \text{ for } a \in \{0, 1\}, r < t < \tau \text{ and } w \in \mathcal{W}.$$

A.4 Dismissible components conditions: for all $t \in \mathbb{R}, r \in \mathbb{R}$

$$\lambda_{12}^{a^D=1,a^I}(t \mid W = w) = \lambda_{12}^{a^D=0,a^I}(t \mid W = w) \text{ for } a^I \in \{0, 1\}, \quad \Delta 1$$
$$\lambda_{13}^{a^D,a^I=1}(t \mid W = w) = \lambda_{13}^{a^D,a^I=0}(t \mid W = w) \text{ for } a^D \in \{0, 1\}, \quad \Delta 2$$
$$\lambda_{23}^{a^D,a^I=1}(t, t - r \mid W = w) = \lambda_{23}^{a^D,a^I=0}(t, t - r \mid W = w) \text{ for } a^D \in \{0, 1\}. \quad \Delta 3$$

where $\lambda_{ij}^{a^D,a^I}(\cdot)$ *denotes the transition hazards of the counterfactual illness-death process under an intervention that sets* $A^D = a^D$ *and* $A^I = a^D$.

Under assumptions A.1–A.4 we have,

$$E\Big\{ I(T_2^{a^D,a^I} \leq \tau) \Big\} = \psi(P; \tau, a^D, a^I) := E\Big\{ P_{13}(\tau, a^D, a^I, W) \Big\}, \quad (5)$$

where

$$P_{13}(\tau, a^D, a^I, W) = 1 - \Omega_\tau(a^D, a^I, w) - \int_0^\tau S_2(\tau|r, a^D, W)\Omega_r(a^D, a^I, w)\, \mathrm{d}\Lambda_{12}(r|a^I, W),$$

for

$$\Omega_r(a^D, a^I, w) = \exp\left\{-\Lambda_{12}(r \mid a^I, w) - \Lambda_{13}(r \mid a^D, w)\right\}.$$

Consequently, the separable direct and indirect effects are identified to

$$SDE(\tau, a^I) = \psi(P;\tau, 1, a^I) - \psi(P;\tau, 0, a^I), \tag{6}$$

and

$$SIE(\tau, a^D) = \psi(P;\tau, a^D, 1) - \psi(P;\tau, a^D, 0). \tag{7}$$

Proof In Appendix B. □

Assumption A.0 is a separable effects analog of the consistency assumption. Assumption A.1–A.3 are standard assumptions for causal inference. Assumption A.4 is the so-called dismissible components conditions, which is an extension of the dismissible components conditions in Martinussen and Stensrud (2023) to the illness-death setting. In particular, assumption ($\Delta 1$) states that the counterfactual hazards of the $1 \to 2$ transition are equal under all values of a^D, and assumption ($\Delta 2$) states that the counterfactual hazards of the $1 \to 3$ transition are equal under all values of a^I. Lastly assumption ($\Delta 3$) states that the counterfactual hazards of the $2 \to 3$ transition are equal under all values of a^I. When the treatment components correspond to meaningful real-world treatments, the dismissible components conditions are empirically verifiable in future trials.

The dismissible components conditions are violated if the A^D and A^I components cannot be manipulated separately. In our DAPT example this would be the case if the biological pathways through which the medication affects MI or stroke is intertwined with the pathways through which it affects bleeding. The dismissible components conditions are also violated if there is an unmeasured common cause of the risk the intermediate and the terminal event. This is similar to the classical 'no unmeasured mediator-outcome confounding' assumption which is needed to identify natural (in-)direct effects. In our DAPT example this would be the case if there is an unmeasured common cause of cardiovascular events such as MI or stroke, and death.

4 Estimation

In this section we address the question of how to construct estimators of the estimand in Eq. (5). Efficient influence functions (EIFs) are an important concept in statistical theory for constructing estimators of causal parameters with desirable properties. In particular estimators based on the EIF are locally efficient (Bickel

et al. 1993). Moreover they often exhibit multiple robustness properties in the sense that consistency of the estimator is preserved under misspecification of one or more components of the data distribution. Further, they are compatible with data adaptive estimation of nuisance parameters provided certain rate conditions hold.

In this paper we focus on the first two properties and assume (semi-)parametric models for the nuisance parameters. In particular, in what follows, we let $\hat{\Lambda}_{12,n}$, $\hat{\Lambda}_{13,n}$, $\hat{\Lambda}_{23,n}$, $\hat{\Lambda}_{\tilde{C},n}$, $\hat{\pi}_n$ denote (semi-)parametric estimators for the relevant components of the data distribution, and we let Λ^*_{12}, Λ^*_{13}, Λ^*_{23}, $\Lambda^*_{\tilde{C}}$, π^* and denote the large sample limits in probability of the (possibly misspecified) estimators. We let Q^* and P^* denote the corresponding distributions of Z and O respectively. If our working model for Λ_{12} is correctly specified then $\Lambda^*_{12} = \Lambda_{12}$ and the same holds for Λ_{13}, Λ_{23}, $\Lambda_{\tilde{C}}$ and π.

In Sect. 4.1 we derive the efficient influence function. In Sect. 4.2 we propose two types of estimators. The first is a 'plug-in' type estimator constructed by substituting estimators for the relevant part of the data distribution directly into (5). The second is a multiply robust estimator which uses the efficient influence function as an estimating equation. In Sect. 4.3 we provide details on how to construct estimators of their asymptotic variance.

4.1 Efficient influence function

Below we derive the EIF of the separable direct and indirect effects under a nonparametric model. We first derive the full-data efficient influence function and then, assuming CAR and Assumptions A.0–A.4 hold, map it to the observed data efficient influence function using results given in Tsiatis (2006). We also establish general multiple robustness properties that will be satisfied by any estimator which solves the EIF estimating function.

Full-data efficient influence function

Let $\psi : \mathcal{Q} \to \mathbb{R}$, where $Q^* \to \psi(Q^*;\tau,a^D,a^I) = E^*\left\{ I(T_2^{a^D,a^I} \leq \tau) \right\}$ and $E^*(\cdot)$ denotes the expectation computed under Q^*. In Appendix C we show that the efficient influence function for ψ at Q^* is given by

$$
\begin{aligned}
\tilde{\psi}(Q^*)(Z;\tau,a^D,a^I) = {} & \frac{I(A=a^I)}{\pi^*(a^I \mid W)} \int_0^\tau \frac{h^*_{12,\tau}(s,a^D,a^I,W)\,\mathrm{d}M^{F*}_{12}(s,a^I,W)}{S^*_1(s \mid a^I,W)} \\
& + \frac{I(A=a^D)}{\pi^*(a^D \mid W)} \left\{ \int_0^\tau \frac{h^*_{13,\tau}(s,a^D,a^I,W)\,\mathrm{d}M^{F*}_{13}(s,a^D,W)}{S^*_1(s \mid a^D,W)} \right. \\
& \left. + \frac{\eta h^*_{23,\tau}(T_1,a^D,a^I,W)}{S^*_1(T_1 \mid a^D,W)} \int_{T_1}^\tau \frac{\mathrm{d}M^{F*}_{23}(s,T_1,a^D,W)}{S^*_2(s \mid T_1,a^D,W)} \right\} \\
& + P^*_{13}(\tau,a^D,a^I,W) - \psi(Q^*;\tau,a^D,a^I),
\end{aligned}
\tag{8}
$$

with

$$h^*_{1j,\tau}(s,a^D,a^I,w) = \Omega^*_\tau(a^D,a^I,w) + \int_s^\tau S^*_2(\tau \mid r,a^D,w)\Omega^*_r(a^D,a^I,w)\,\mathrm{d}\Lambda^*_{12}(r \mid a^I,w) + \begin{cases} -S^*_2(\tau \mid s,a^D,w)\Omega^*_s(a^D,a^I,w), & \text{when } j=2 \\ 0, & \text{when } j=3 \end{cases},$$

and

$$h^*_{23,\tau}(s,a^D,a^I,w) = \frac{\lambda^*_{12}(s \mid a^I,w)}{\lambda^*_{12}(s \mid a^D,w)}\Omega^*_s(a^D,a^I,w)S^*_2(\tau \mid s,a^D,w),$$

and where $\mathrm{d}M^{F*}_{ij}$ denote the full-data martingale increments under Q^*

$$\begin{aligned} \mathrm{d}M^{F*}_{13}(s,A,W) &= \mathrm{d}N_{13}(s) - \lambda^*_{13}(s \mid A,W)I(T_1 > s)\,\mathrm{d}s, \\ \mathrm{d}M^{F*}_{12}(s,A,W) &= \mathrm{d}N_{12}(s) - \lambda^*_{12}(s \mid A,W)I(T_1 > s)\,\mathrm{d}s, \\ \mathrm{d}M^{F*}_{23}(s,T_1,A,W) &= \mathrm{d}N_{23}(s) - \lambda^*_{23}(s,s-T_1 \mid A,W)I(T_1 \le s < T_2)\,\mathrm{d}s. \end{aligned}$$

Lemma 2 (Multiple robustness) *The full-data efficient influence function admits a multiple robust structure in the sense that* $E\{\tilde{\psi}(Q^*)(Z)\} = \psi(Q) - \psi(Q^*)$ *if one of the following holds*

(i) $\pi^*(a \mid w) = \pi(a \mid w), \Lambda^*_{12}(r \mid a,w) = \Lambda_{12}(r \mid a,w)$ *and* $\Lambda^*_{23}(t,t-r \mid a,w) = \Lambda_{23}(t,t-r \mid a,w)$ *for all* $t,r \in [0,\tau], a \in \{0,1\}$ *and almost all* w,

(ii) $\pi^*(a \mid w) = \pi(a \mid w), \Lambda^*_{13}(t \mid a,w) = \Lambda_{13}(t \mid a,w)$ *and* $\Lambda^*_{23}(t,t-r \mid a,w) = \Lambda_{23}(t,t-r \mid a,w)$ *for all* $t,r \in [0,\tau], a \in \{0,1\}$ *and almost all* w,

(iii) $\pi^*(a \mid w) = \pi(a \mid w), \Lambda^*_{12}(t \mid a,w) = \Lambda_{12}(t \mid a,w)$ *and* $\Lambda^*_{13}(t \mid a,w) = \Lambda_{13}(t \mid a,w)$ *for all* $t \in [0,\tau]$, $a \in \{0,1\}$ *and almost all* w.

Proof In Appendix E. □

The multiple robustness properties stated in the lemma above imply that the full-data influence function $\tilde{\psi}(Q^*)(Z)$ is a consistent estimating function of $\psi(Q)$ when at most one of the transition intensities is inconsistently estimated.

Observed-data efficient influence function Let $\psi : \mathcal{P} \to \mathbb{R}$, where $P^* \to \psi(P^*;\tau,a^D,a^I) = E^*\left\{I(T_2^{a^D,a^I} \le \tau)\right\}$. In Appendix D we show that the observed data efficient influence function is given by

$$
\begin{aligned}
&\tilde{\psi}(P^*)(O;\tau,a^D,a^I)\\
&\quad=\frac{I(A=a^I)}{\pi^*(a^I\mid W)}\int_0^\tau \frac{h^*_{12,\tau}(s,a^D,a^I,W)}{K^*_{\tilde{C},1}(s\mid a^I,W)}\frac{\mathrm{d}M^*_{12}(s,a^I,W)}{S^*_1(s\mid a^I,W)}\\
&\qquad+\frac{I(A=a^D)}{\pi^*(a^D\mid W)}\Bigg\{\int_0^\tau \frac{h^*_{13,\tau}(s,a^D,a^I,W)}{K^*_{\tilde{C},1}(s\mid a^D,W)}\frac{\mathrm{d}M^*_{13}(s,a^D,W)}{S^*_1(s\mid a^D,W)}\\
&\qquad+\frac{\tilde{\eta}h^*_{23,\tau}(\tilde{T}_1,a^D,a^I,W)}{S^*_1(\tilde{T}_1\mid a^D,W)K^*_{\tilde{C},1}(\tilde{T}_1\mid a^D,W)}\int_{\tilde{T}_1}^\tau \frac{\mathrm{d}M^*_{23}(s,\tilde{T}_1,a^D,W)}{S^*_2(s\mid \tilde{T}_1,a^D,W)K^*_{\tilde{C},2}(s\mid \tilde{T}_1,a^D,W)}\Bigg\}\\
&\qquad+P^*_{13}(t,a^D,a^I,W)-\psi(P^*;t,a^D,a^I),
\end{aligned}
\tag{9}
$$

with $\mathrm{d}M^*_{ij}(\cdot)$ denoting the observed-data martingale increments under P^*

$$
\begin{aligned}
\mathrm{d}M^*_{12}(s,A,W)&=\mathrm{d}\tilde{N}_{12}(s)-\lambda^*_{12}(s\mid A,W)I(\tilde{T}_1>s)\,\mathrm{d}s,\\
\mathrm{d}M^*_{13}(s,A,W)&=\mathrm{d}\tilde{N}_{13}(s)-\lambda^*_{13}(s\mid A,W)I(\tilde{T}_1>s)\,\mathrm{d}s,\\
\mathrm{d}M^*_{23}(s,\tilde{T}_1,A,W)&=\mathrm{d}\tilde{N}_{23}(s)-\lambda^*_{23}(s,s-\tilde{T}_1\mid A,W)I(\tilde{T}_1\le s<\tilde{T}_2)\,\mathrm{d}s.
\end{aligned}
$$

Lemma 3 (Multiple robustness) *The observed-data efficient influence function admits a multiple robust structure in the sense that* $E\{\tilde{\psi}(P^*)(O)\}=\psi(P)-\psi(P^*)$ *if one of the following holds*

(i) $\pi^*(a\mid w)=\pi(a\mid w), \Lambda^*_{\tilde{C}}\{t\mid G_s(z)\}=\Lambda_{\tilde{C}}\{t\mid G_s(z)\}, \Lambda^*_{12}(r\mid a,w)=\Lambda_{12}(r\mid a,w)$ *and* $\Lambda^*_{23}(t,t-r\mid a,w)=\Lambda_{23}(t,t-r\mid a,w)$ *for all* $t,r\in[0,\tau]$, $a\in\{0,1\}$ *and almost all* w,

(ii) $\pi^*(a\mid w)=\pi(a\mid w), \Lambda^*_{\tilde{C}}\{t\mid G_s(z)\}=\Lambda_{\tilde{C}}(t\mid G_s(z)), \Lambda^*_{13}(t\mid a,w)=\Lambda_{13}(t\mid a,w)$ *and* $\Lambda^*_{23}(t,t-r\mid a,w)=\Lambda_{23}(t,t-r\mid a,w)$ *for all* $t,r\in[0,\tau]$, $a\in\{0,1\}$ *and almost all* w,

(iii) $\pi^*(a\mid w)=\pi(a\mid w), \Lambda^*_{\tilde{C}}\{t\mid G_s(z)\}=\Lambda_{\tilde{C}}\{t\mid G_s(z)\}, \Lambda^*_{12}(t\mid a,w)=\Lambda_{12}(t\mid a,w)$ *and* $\Lambda^*_{13}(t\mid a,w)=\Lambda_{13}(t\mid a,w)$ *for all* $t\in[0,\tau]$, $a\in\{0,1\}$ *and almost all* w,

(iv) $\Lambda^*_{12}(r\mid a,w)=\Lambda_{12}(r\mid a,w), \Lambda^*_{13}(t\mid a,w)=\Lambda_{13}(t\mid a,w)$ *and* $\Lambda^*_{23}(t,t-r\mid a,w)=\Lambda_{23}(t,t-r\mid a,w)$ *for all* $t,r\in[0,\tau]$, $a\in\{0,1\}$ *and almost all* w.

Proof In Appendix F. □

This means that when the censoring distribution is correctly specified the same multiple robustness properties hold as in the full-data case. The censoring model and propensity score are allowed to be misspecified when all three transition intensites are correctly specified.

Efficient influence functions of the separable direct and indirect effects

Consider the mappings $P^* \to \psi^{\text{SDE}}(P^*;\tau,a^D,a^I) = \psi(P^*;\tau,1,a^I) - \psi(P^*;\tau,0,a^I)$ for $a^I \in \{0,1\}$ and $P \to \psi^{\text{SIE}}(P^*;\tau,a^D,a^I) = \psi(P^*;\tau,a^D,1) - \psi(P^*;\tau,a^D,0)$ for $a^D \in \{0,1\}$. It follows by the functional delta method that the efficient influence functions of the separable direct and indirect effects in (6) and (7) are given by respectively

$$\tilde{\psi}^{\text{SDE}}(P^*)(O;\tau,a^I) = \tilde{\psi}(P^*)(O;\tau,1,a^I) - \tilde{\psi}(P^*)(O;\tau,0,a^I), \quad \text{for } a^I \in \{0,1\},$$

and

$$\tilde{\psi}^{\text{SIE}}(P^*)(O;\tau,a^D) = \tilde{\psi}(P^*)(O;\tau,a^D,1) - \tilde{\psi}(P^*)(O;\tau,a^D,0), \quad \text{for } a^D \in \{0,1\},$$

and will inherit the multiple robustness properties established in Lemma 3.

4.2 Estimators

Plug-in (G-computation) estimator

A plug-in estimator estimates the relevant part of the distribution of O, in this case the empirical distribution of W and appropriate estimators $\hat{\Lambda}_{12,n}, \hat{\Lambda}_{13,n}$ and $\hat{\Lambda}_{23,n}$ of the transition intensities, and substitutes them in place of the unknown quantities in Eq. (5). Then one obtains the estimator

$$\hat{\Psi}_n^{\text{Plug-in}}(\tau,a^D,a^I) = n^{-1}\sum_{i=1}^{n} \hat{P}_{13}(\tau,a^D,a^I,W_i), \tag{10}$$

where

$$\hat{P}_{13}(\tau,a^D,a^I,W) = P_{13}(\tau,a^D,a^I,W;\hat{\Lambda}_{12,n},\hat{\Lambda}_{13,n},\hat{\Lambda}_{23,n}).$$

Equation (5) is also known as the G-computation formula (Robins 1986), and the estimator in (10) is also referred to as a G-computation estimator. Note that consistency of $\hat{\Psi}_n^{\text{Plug-in}}(t,a^D,a^I)$ depends on consistency of the estimators of all three transition intensities.

One-step estimator

As mentioned above the efficient influence function is useful for constructing multiply robust efficient estimators. One way of doing this is to use the influence function directly as an estimating equation (van der and Robins 2003). Since the EIF in equation (9) is linear in the parameter of interest, this results the estimator:

$$\hat{\Psi}_n^{\text{one-step}}(\tau,a^D,a^I) = n^{-1}\sum_{i=1}^{n} \varphi(\hat{\pi}_n,\hat{\Lambda}_{12,n},\hat{\Lambda}_{13,n},\hat{\Lambda}_{23,n},\hat{\Lambda}_{\tilde{C},n})(O_i;\tau,a^D,a^I), \tag{11}$$

where

$$\varphi(P)(O;\tau,a^D,a^I) = \tilde{\psi}(P)(O;\tau,a^D,a^I) + \psi(P;t,a^D,a^I)$$

The estimator in (11) is multiply robust. In particular it is consistent under misspecification of (i) Λ_{12}, (ii) Λ_{13}, (iii) Λ_{23} or (iv) π and $\Lambda_{\tilde{C}}$ as shown in Lemma 3.

Note that we can write:

$$\begin{aligned}&\varphi(\hat{\pi}_n,\hat{\Lambda}_{12,n},\hat{\Lambda}_{13,n},\hat{\Lambda}_{23,n},\hat{\Lambda}_{\tilde{C},n})(O;\tau,a^D,a^I)\\&\quad=\tilde{\psi}(\hat{\pi}_n,\hat{\Lambda}_{12,n},\hat{\Lambda}_{13,n},\hat{\Lambda}_{23,n},\hat{\Lambda}_{\tilde{C},n})(O;\tau,a^D,a^I)+\hat{\Psi}_n^{\text{G-comp}}(\tau,a^D,a^I).\end{aligned}$$

This approach is also referred to as a so-called 'one-step' bias correction approach (Ibragimov and Has'minskii 1981; Pfanzagel and Wefelmeyer 1985), and we will refer to the estimator in (11) as a 'one-step' estimator.

4.3 Asymptotic variance

If all nuisance models are correctly specified, then a consistent estimator of the asymptotic variance can be obtained from the variance of the influence function. However if one or more of the nuisance models are misspecified then this variance estimator is no longer consistent, and other techniques must be used.

Suppose we are willing to assume fully parametric models for all nuisance parameters. Then we can derive the asymptotic distribution of the estimators in (10) and (11) by stacking the corresponding unbiased estimating equations for the target and nuisance parameters, and applying standard estimating equation theory (Stefanski and Boos 2002). In particular, let $\hat{\theta}_n$ be the estimators of the parameters of interest and nuisance parameters that solves

$n^{-1}\sum_{i=1}^n m(O_i,\hat{\theta}_n) = 0$ where $m(O,\theta)$ are the stacked estimating equations of both the parameter of interest and nuisance parameters. For the plug-in estimator in (10) this would be $\hat{\theta}_n = (\hat{\Psi}^{\text{Plug-in}},\hat{\Lambda}_{12,n},\hat{\Lambda}_{13,n},\hat{\Lambda}_{23,n})$ and $m(O,\theta) = (P_{13},S_{\Lambda_{12}},S_{\Lambda_{13}},S_{\Lambda_{23}})$ where $S_{\Lambda_{12}},S_{\Lambda_{13}},S_{\Lambda_{23}}$ are appropriate estimating equations for the transition hazards. Under suitable regularity conditions (Newey and McFadden 1994; van der Vaart 2000; Tsiatis 2006), we have

$$n^{1/2}(\hat{\theta}-\theta^*) \rightsquigarrow N\left(0, E\left\{-\frac{\partial m(O,\theta^*)}{\partial\theta^T}\right\}^{-1}\text{var}\{m(O,\theta^*)\}E\left\{-\frac{\partial m(O,\theta^*)}{\partial\theta^T}\right\}^{-1^T}\right).$$

It is then possible to derive an analytic expression for the asymptotic variance of the estimators in (10) and (11) using the sandwich variance estimator.

When the nuisance models are e.g. Cox regression models we need to take into account the variability of the baseline hazards which may be nonparametrically estimated. Then the asymptotic distribution can be derived using the functional delta method (van der Vaart 2000). This expression becomes very complicated, especially for the one-step estimator, and deriving an explicit estimator of the variance goes beyond the scope of this paper.

5 Simulation study

5.1 Simulation study 1: empirical performance

Below, we report the results from a simulation study where the aim is to compare the finite sample performance of the plug-in estimator and the one-step estimator.

The data was generated by the following simulation procedure:

$$
\begin{aligned}
&W \sim \mathit{Uniform}(0,1)\\
&A \mid W \sim \mathit{Bernoulli}(\text{expit}(-0.5 + W + \zeta W^2))\\
&T_1 \mid A, W \sim \mathit{Exponential}(\lambda_{12} + \lambda_{13}) \text{with} \, \lambda_{12} = 0.039 \cdot \exp(\log(2)W + A + \gamma^{12}AW)\\
&\text{and} \, \lambda_{13} = 0.026 \cdot \exp(\log(2)W + 0.5A + \gamma^{13}(1-A)W)\\
&\eta \mid A, W \sim \mathit{Bernoulli}(\lambda_1/(\lambda_1 + \lambda_2))\\
&T_2 = T_1 + \eta \cdot U \;\; \text{with} \;\; U \sim \mathit{Exponential}(\lambda_{23}) \;\; \text{where} \;\; \lambda_{23} = 0.052 \cdot \exp\,(\log(2)W +\\
&0.5A + \gamma^{23}(1-A)W)\\
&\tilde{C} \mid W \sim \mathit{Exponential}(\lambda_{\tilde{C}}) \text{ with } \lambda_{\tilde{C}} = 0.035 \cdot \exp(\theta W)
\end{aligned}
$$

where $\text{expit}(x) = \{1 + \exp(x)\}^{-1}$. Note that this corresponds to a scenario where treatment has a protective effect on both disease and death, and where the treatment effect on death is the same in diseased and disease-free subjects.

An estimator for the propensity score was constructed using a logistic regression model with main effects only. For the transition hazards we constructed estimators using a Cox regression model with main effects only and for the censoring hazard we used a Cox model with no covariate effects. The dependency of Λ_{23} on the time of reaching state 2 was handled by delayed entry. We considered 8 different scenarios: in scenario (i) all nuisance models were correctly specified which is the case when $(\zeta, \gamma^{12}, \gamma^{13}, \gamma^{23}, \theta) = 0$, and in scenarios (ii)–(viii) we considered misspecifications of different combinations of the nuisance models by varying the values of $(\zeta, \gamma^{12}, \gamma^{13}, \gamma^{23}, \theta)$ accordingly. Additional details on the misspecified scenarios are given in Appendix G.

For each scenario we generated 1000 datasets from the simulation procedure with a sample size of 400. For each dataset we computed the plug-in estimator and the one-step estimator for the SDE along with the bootstrap variance for each estimator based on 250 replicates. The results of our simulation study are summarized in Figs. 3 and 4 where for all scenarios we report bias, empirical standard error, coverage of the 95 % Wald confidence interval and accuracy of the standard error estimator computed at time points $t \in \{1, 5, 10, 15, 20, 25\}$.

As expected both the plug-in estimator and the one-step estimator are consistent in scenario (i) where all nuisance models are correctly specified and scenario (ii) were the propensity score and censoring models are misspecified. Moreover the coverages are close the nominal level. In scenarios (iii)–(v) where we consider misspecifications of at most one of the transition hazard models the one-step estimator provides a bias reduction over the plug-in estimator, as predicted by the multiple robustness properties in Lemma 3. In scenarios (vi)–(viii) where we consider misspecifications that go beyond the robustness properties of lemma 3 both the plug-in

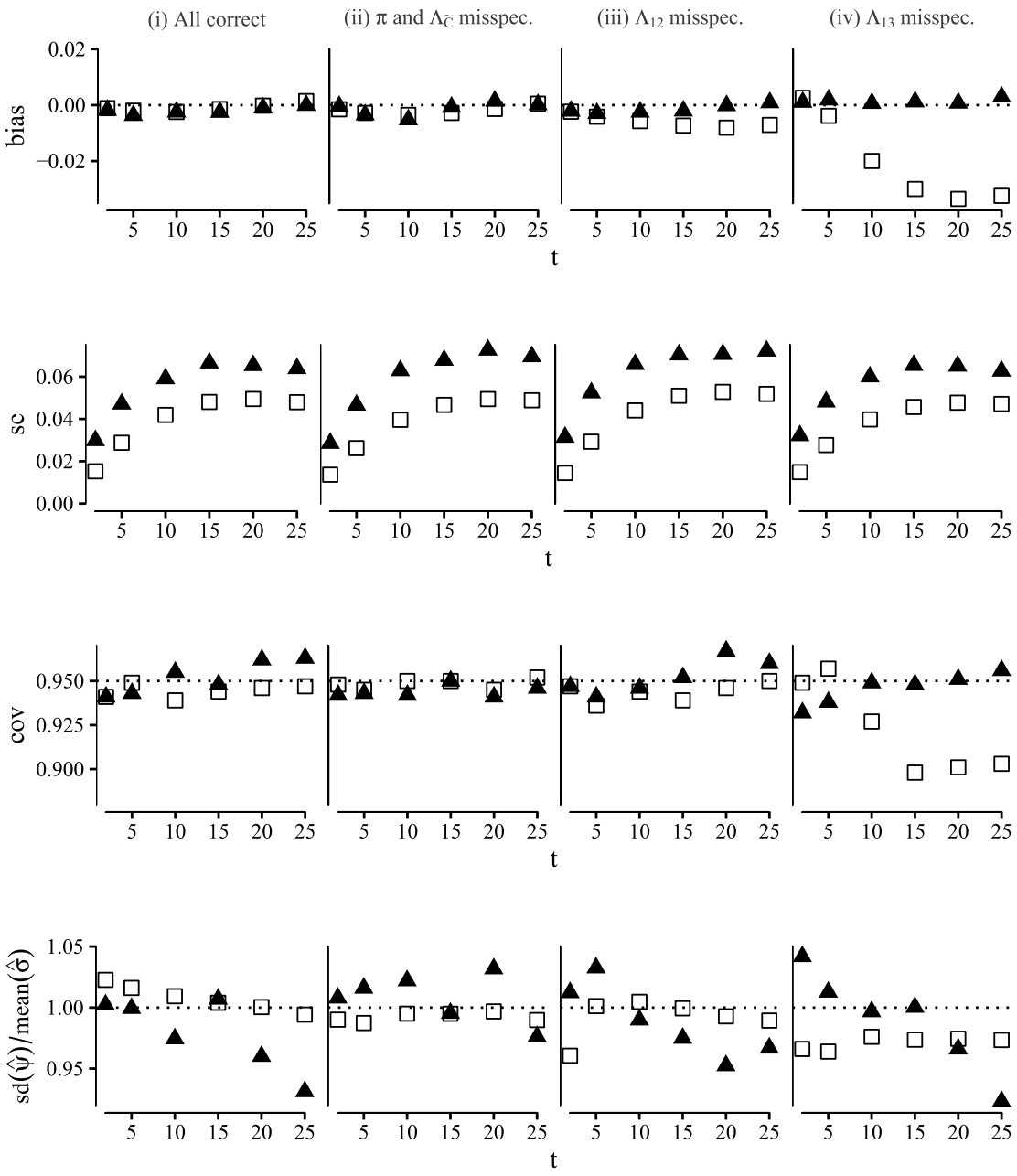


Fig. 3 Comparison of the G-computation (white rectangles) and one-step (black triangles) estimators of the SDE computed at time points $t \in \{2, 5, 10, 15, 20, 25\}$ in terms of bias, empirical standard error, coverage of 95% confidence intervals and accuracy of the standard error estimator. This figure contains scenarios (i)–(iv) (Color figure online)

estimator and the one-step estimator are biased, except in scenario (vi) where the plug-in estimator surprisingly appears unbiased. The one-step estimator is more variable than the plug-in estimator throughout all scenarios.

This simulation study confirms the double robustness properties of the one-step estimator derived in Sect. 4.1, which, along with the potential compatibility with data-adaptive estimation of nuisance parameters, highlights the real-word utility of the one-step estimator.

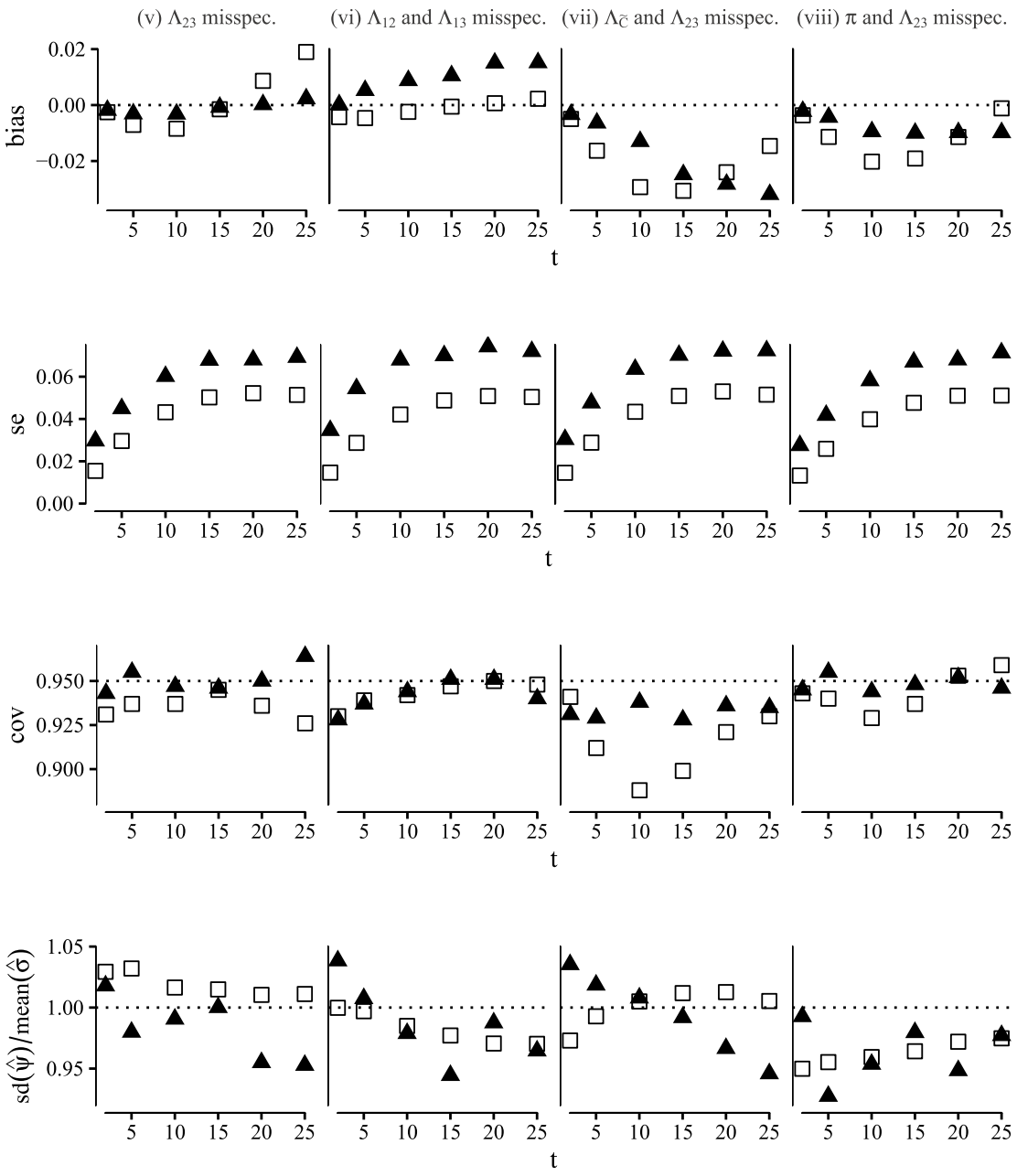


Fig. 4 Comparison of the G-computation (white rectangles) and one-step (black triangles) estimators of the SDE computed at time points $t \in \{2, 5, 10, 15, 20, 25\}$ in terms of bias, empirical standard error coverage of 95% confidence intervals and accuracy of the standard error estimator. This figure contains scenarios (v)–(vii) (Color figure online)

5.2 Simulation study 2: violation of assumptions

The dismissible components conditions in Lemma 1 are violated in the presence of an unmeasured common risk factor for illness and death. Below, we study such violations in a simulation study.

The data was generated by the following simulation procedure:

$W \sim Bernoulli(0.5)$
$A^D \mid W \sim Bernoulli(\text{expit}(-0.5 + W))$
$A^I \mid W \sim Bernoulli(\text{expit}(-0.5 + W))$
$U \sim Bernoulli(0.6)$
$T_1 \mid A^D, A^I, W \sim Exponential(\lambda_{12} + \lambda_{13})$ with
$\lambda_{12} = 0.039 \cdot \exp(\log(2)W + \beta_A^{12} A^I + \gamma_U U)$ and
$\lambda_{13} = 0.026 \cdot \exp(\log(2)W + \beta_A^{13} A^D + \gamma_U U)$
$\eta \mid A^D, A^I, W \sim Bernoulli(\lambda_1/(\lambda_1 + \lambda_2))$
$T_2 = T_1 + \eta \cdot V$ with $V \sim Exponential(\lambda_{23})$ with
$\lambda_{23} = 0.052 \cdot \exp(\log(2)W + \beta_A^{13} A^D + \gamma_U U)$
$\tilde{C} \sim Exponential(\lambda_{\tilde{C}})$ with $\lambda_{\tilde{C}} = 0.035$

We varied γ_U along the grid $\{-1, -0.9, ..., 0.9, 1\}$ and considered the four cases: (I) Protective treatment effect on disease and death, (II) Protective effect on disease and harmful effect on death, (III) Harmful effect on disease and protective effect on death and (IV) Harmful treatment effect on disease and death.

We constructed an estimator for the propensity score using a correctly specified logistic regression model. The censoring hazard was estimated using a Cox model with no covariate effects. The remaining nuisance models were estimated using Cox regression models adjusted for main effects of the observed variables. We generated 1000 datasets with a sample size of $n = 1000$. For each dataset we computed the plug-in estimator and the one-step estimator for the SDE evaluated at time point $t = 15$. The results are depicted in Fig. 5. It is seen that the bias increases with the magnitude of the association with the unmeasured common risk factor U. The direction of the bias depends on the effect of treatment on illness: when the treatment has a protective effect on disease the estimator is downwards biased, and when the treatment has a harmful effect on disease the bias is positive.

6 Real data application

Using data from the Danish nationwide registries we identified all hospital admissions for first time acute myocardial infarction (MI) between 2010 and 2014. To get a more homogeneous study population we only included patients who were treated with a Percutaneous Coronary Intervention (PCI). We also excluded patients with a preexisting alcohol abuse diagnosis or chronic kidney disease diagnosis and patients younger than 30 years or older than 100 years of age. We set the index date for inclusion at 30 days following discharge and excluded patients who died prior to the index data. We defined the treatment arm as those patients who picked up a prescription for DAPT before the index date and the placebo group as those who did not. Patients who were still alive by the end of 2019 were administratively censored. Among the 16,081 patients in the study population 3856 patients had a recurrent cardiovascular event (defined as a hospital diagnosis of MI, stroke or heart failure) and were subsequently censored, 968 patients died within follow-up without having a recurrent cardiovascular event

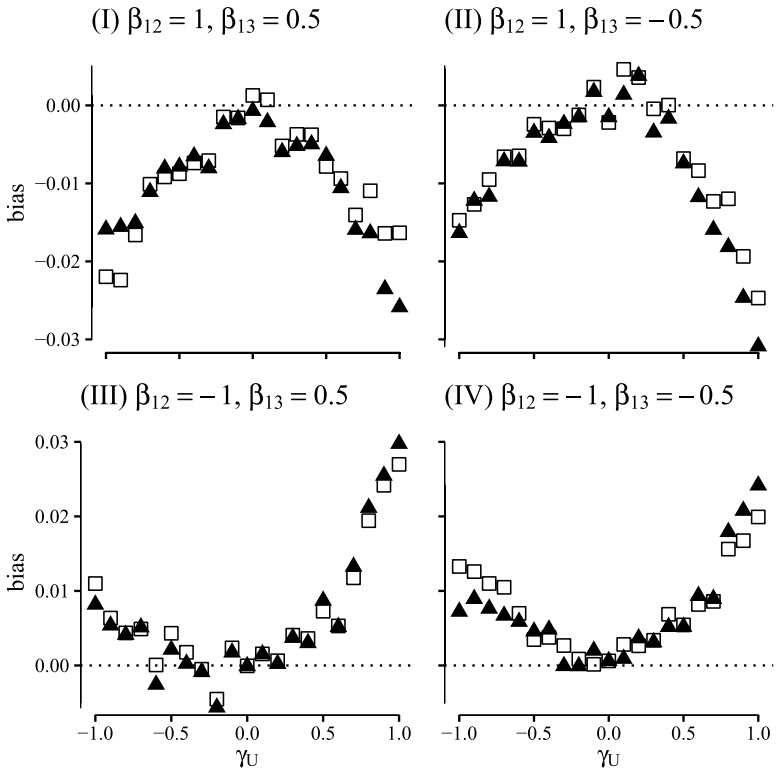


Fig. 5 Bias of the plug-in (white rectangles) and one-step (black triangles) estimators of the SDE computed at time points $t = 15$ under violation of the identification assumption

and 1385 patients experienced a recurrent cardiovascular event and subsequently died within followup.

The cumulative hazard curves in Fig. 6 suggest that treatment reduces both risk of recurrent cardiovascular event, overall mortality and death without recurrent cardiovascular event. To access how much of the effect of DAPT on mortality was mediated through recurrent cardiovascular events we estimated the separable direct and indirect effects. That is, we assume that the treatment has two components that could in principle be manipulated separately: one component A^I which only affects the risk of recurrent cardiovascular event directly and another component A^D which affects mortality through other pathways. A possible interpretation of these treatment components was discussed in Sect. 3. We can then define the separable indirect effect as the effect under an intervention that fixes the treatment component affecting affecting mortality through other pathways than recurrent cardiovascular events but varies the treatment component affecting cardiovascular events. Similarly we can define the separable direct effect as the effect that fixes the treatment component affecting cardiovascular events and varies the component affecting mortality through other pathways.

We estimated the separable effects using the plug-in estimator and the one-step estimator presented in Sect. 4.2. Both estimators used semi-parametric working models for the nuisance parameters. In particular, we used Cox regression models for the three transition hazards. The models were adjusted for baseline age, sex, hypertension diagnosis, prior gastrointestinal bleeding, diabetes, chronic liver disease, cancer, atrial fibrillation, Anemia, prior heart failure or stroke. We computed Wald-type point-wise confidence intervals based on 500 bootstrap data sets.

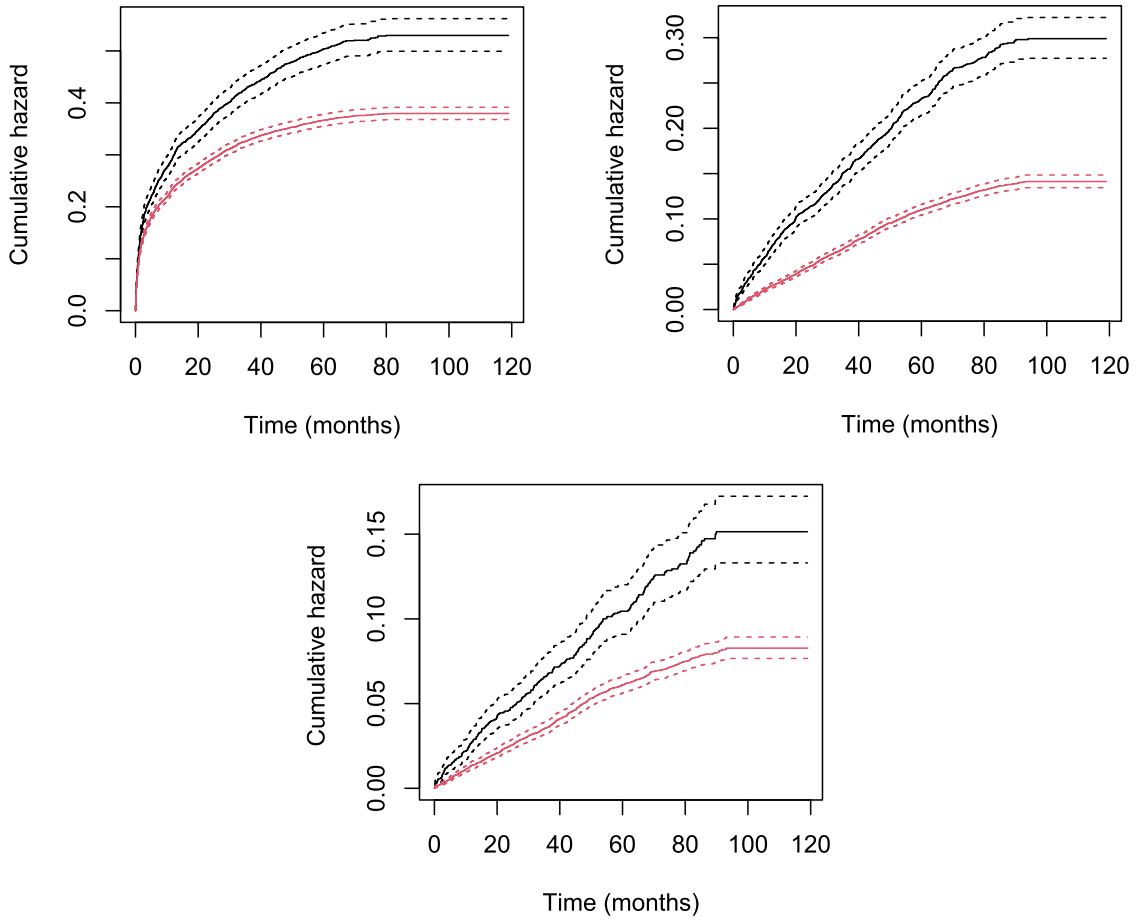


Fig. 6 Nelson–Aalen estimates of the cumulative hazards of MI (top left), overall mortality (top right) and death without recurrent MI (bottom) in our cohort. The red curves are the treatment arm and the black curves are the placebo arm. Along with the hazards (solid lines) are shown 95% confidence intervals (dashed lines) (Color figure online)

The results of our analysis are presented in Figs. 7 and 8. In addition to the separable direct and indirect effects we have also depicted the total effect, c.f., Eq. (4).

Our results suggest that the treatment reduces mortality both through recurrent cardiovascular events and through other pathways. That is, within the limitations of our study, we can conclude that the modified treatment that fixes the component affecting mortality through other pathways than recurrent cardiovascular events does not capture the entire protective effect of the treatment. In fact a substantial fraction of the protective effect of DAPT on mortality is a direct effect.

We recognize several potential limitations with our study. First, we likely have confounding by indication in that frail individuals are less likely to be prescribed the treatment. Therefore the drug will appear more effective than it actually is, also on non-cardiovascular mortality. This phenomenon is notoriously difficult to adjust for because of unmeasured confounding. Second, comorbidities such as diabetes status are essentially time-varying covariates. It is a major limitation of our method that we only adjust for baseline covarites. Third, a potential issue is that many cardiovascular events go undetected or are not entered into

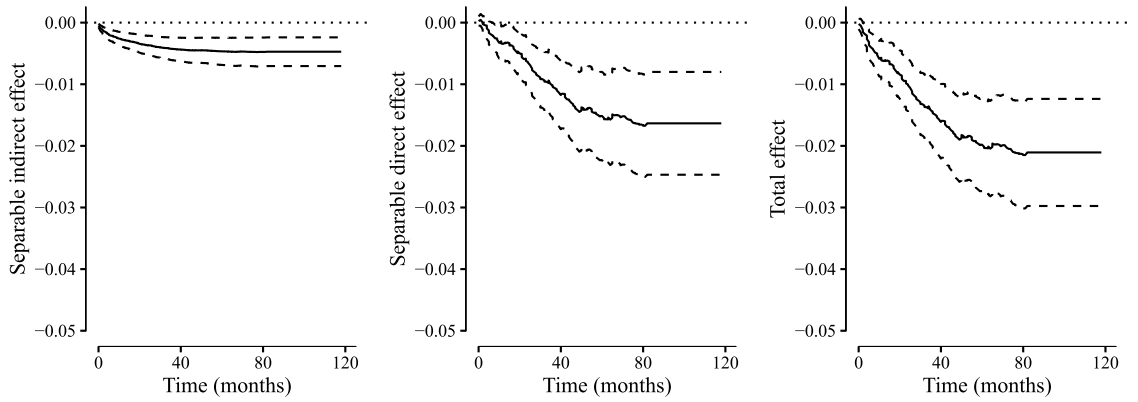


Fig. 7 Estimates of the separable direct effect (SDE), separable indirect effect (SIE) and total effect (TE) using the one-step estimator. Solid lines represent effect estimates and dashed lines the corresponding 95 % point-wise confidence intervals

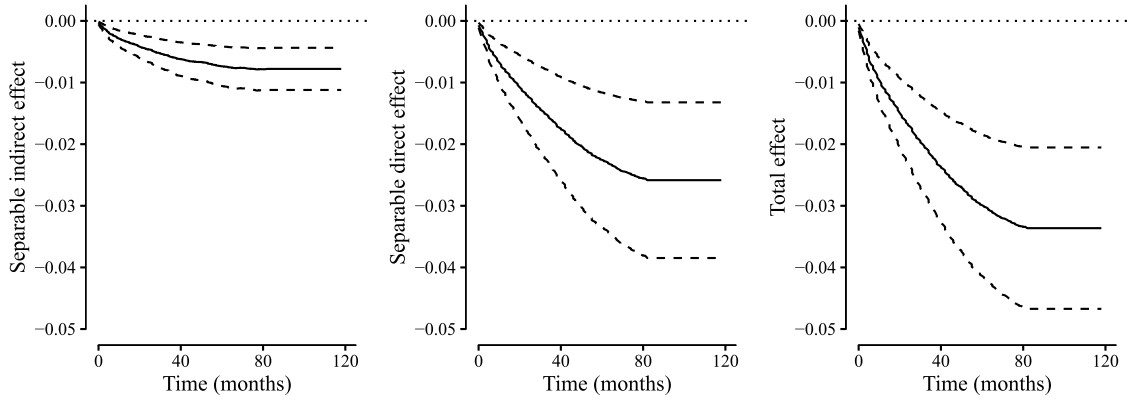


Fig. 8 Estimates of the separable direct effect (SDE), separable indirect effect (SIE) and total effect (TE) using the plug-in estimator. Solid lines represent effect estimates and dashed lines the corresponding 95 % point-wise confidence intervals

the registries e.g. when a patient dies suddenly without prior hospital admission. Finally, the overall risk of bleeding, which is the main side effect of DAPT, is very low.

7 Discussion

7.1 Relation to other approaches

The main difficulty when formulating causal mediation targets in the illness-death model is that the mediating event is truncated by the terminal event. In this paper we proposed causal mediation estimands using the concept of separable effects, which considers interventions on separate components of the treatment instead of interventions on the mediator. This approach avoids the conceptual issues that arise when the terminal event occurs before the mediator, rendering the mediator undefined.

However, this comes at the cost of assuming that the treatment components can be manipulated separately, which may not always be appropriate.

Depending on the causal question at hand there are other approaches in the literature that may be useful for defining mediation targets in the illness-death model.

Valeri et al. (2021) propose randomized interventional direct and indirect effects. Instead of considering manipulations of the mediator, they consider stochastic interventions on the intermediate time-to-event distribution conditional on baseline covariates. The authors then define the 'stochastic direct effect' as the difference in survival across exposure groups under a stochastic intervention that fixes the intermediate time-to-event distribution to be the same in both exposure groups. The 'stochastic indirect effect' is defined as the difference in survival within an exposure group when the intermediate time-to-event distribution is varied. Their approach result in the same identifying functionals as in our paper, but under different identifiability conditions. Thus the target parameter in our paper can also be interpreted as an interventional effect.

A different alternative is principal stratification which has often been advocated in the presence of truncation (Zhang and Rubin 2003; Comment et al. 2019). A recent paper by Gao et al. (2021) proposes a principal stratification approach for defining causal mediation effects in the subgroup where the intermediate event will happen before the potential terminal event when given either of two treatment options. This strata corresponds to a multistate model where only the transition from the 'healthy' state to the 'illness' state and from the 'illness' state to 'death' are involved, an thus their approach leads to a different identifying functional than the one in our paper. This method avoids the issues that arise when death occurs prior to the non-terminal event. However a limitation is that the empirical usefulness of the estimand is debatable since the subgroup for which the estimand is defined can never be observed.

Huang (2021) proposes a method for causal mediation with 'semicompeting risk data', based on counterfactual counting processes for the latent intermediate event and the terminal event. To circumvent the undefinability of the intermediate event the author assumes that if the intermediate event does not occur before the terminal event it would never occur within follow-up. The paper was accompanied by a number of commentaries (Stensrud et al. 2021; Fulcher et al. 2021; Chan et al. 2021) which argue that the identification assumptions are too restrictive for most practical contexts. As the authors do not use a classical illness-death model framework, it is not clear to us how their identifying functional is connected to ours.

7.2 Conclusion and possible extensions

In this paper we proposed causal estimands for the separable direct and indirect effects of a baseline exposure on a terminal time-to-event outcome mediated by the illness state of a continuous-time illness-death process. We proposed a plug-in estimator based on the identifying functional, and a one-step estimator which solves the efficient influence function. We showed that the one-step estimator is multiply robust under appropriate regularity conditions, and we confirmed these theoretical properties in a simulation study which showed an impressive performance of the

one-step estimator. To illustrate our method we applied the estimators to a Danish registry data set to study how much of the effect of DAPT on mortality was mediated through stroke or MI.

As mentioned in Sect. 6 a major limitation of our work is that we have only considered the case of baseline covariates, ignoring the possibility of changes during the followup period. Such changes are present in most real-word data including the DAPT example. Extending the method to handle time-varying covariates is thus an important topic for future research. We believe that our method can be generalized to allow for covariates measured at random times on a continuous scale similar to the setting considered in Rytgaard et al. (2022). Since the time-varying confounders are potentially affected by treatment one would need to consider a more general notion of separable effects as described in Stensrud et al. (2021) who define separable direct and indirect effects in a discrete-time competing risk model with time-varying common causes of the event of interest and the competing event.

Another important topic for future research is the use of data-adaptive estimation of the nuisance parameters in the illness-death model. Our focus in this paper has been on (semi-)parametric models for the nuisance parameters. However, to avoid misspecification of the nuisance models one may wish to use estimators that are more flexible. The one-step estimator derived in this article is compatible with such data-adaptive estimators, provided certain rate conditions hold.

Finally, in this paper we have limited our attention to the illness-death model where the terminal event death is the outcome of interest. Often the outcome of interest is not a terminal event, and both the mediator and the outcome are subject to competing risk from death. A natural extension of our work is to consider an 'extended illness-death model' with a fourth state representing the competing event.

Appendices

Appendix A: CAR
Appendix B: Proof of Lemma 1 (identifiability)
Appendix C: Derivation of full-data EIF
Appendix D: Derivation of observed-data EIF
Appendix E: Proof of Lemma 2 (multiple robustness of full-data EIF)
Appendix F: Proof of Lemma 3 (multiple robustness of observed-data EIF)
Appendix G: Simulation details

A CAR

This censoring mechanism induces monotone coarsening (Gill et al. 1997; Tsiatis 2006). Building upon the continuous-time monotone coarsening framework of Tsiatis (2006) Chapter 9.3 we introduce a so-called coarsening variable $\mathcal{C}$. The coarsening variable is a continuous random variable which is equal to the censoring time when $\tilde{C} < T_1$ or $T_1 < \tilde{C} \leq T_2$, and equal to ∞ when the data is uncensored.

Let τ be a time horizon chosen such that there exists $\epsilon > 0$ with $P(\tilde{C} > \tau) > \epsilon > 0$. For any time $r \in [0, \tau]$ we define the set

$$\{r \leq \mathcal{C} < r + \mathrm{d}r\} = \{r \leq \tilde{C} < r + \mathrm{d}r, \tilde{C} < T_1\} \cup \{r \leq \tilde{C} < r + \mathrm{d}r, \eta = 1, T_1 < \tilde{C} \leq T_2\}.$$

In particular, when $\mathcal{C} = r$ we observe the many-to-one mapping

$$G_r(Z) = \begin{cases} (T_1 \geq r, T_2 \geq r, A, W) & \text{if } r < T_1 \\ (\eta = 1, T_1, T_1 < r, T_2 \geq r, A, W) & \text{if } \eta = 1 \text{ and } T_1 < r \leq T_2 \\ (T_1, \eta, T_2, A, W) & \text{if } r = \infty \end{cases},$$

and the observed data may be expressed as

$$O = \left\{\mathcal{C}, G_{\mathcal{C}}(Z)\right\}.$$

The coarsening mechanism is monotone since $G_r(Z) \subseteq G_{r'}(Z)$ for $r > r'$. Following Tsiatis (2006) Chapter 9.3 the CAR assumption is formally defined by

$$\lambda_{\mathcal{C}}(r; G_r(Z)) = \lambda_{\mathcal{C}}(r; Z), \tag{12}$$

where the coarsening hazard may be written

$$\begin{aligned}
\lambda_{\mathcal{C}}(r;Z) &= \lim_{\mathrm{d}r \to 0} \frac{P(r \leq \mathcal{C} \leq r + \mathrm{d}r \mid \mathcal{C} \geq r, Z)}{\mathrm{d}r} \\
&= \lim_{\mathrm{d}r \to 0} \frac{P(\tilde{C} \leq r + \mathrm{d}r, T_1 > \tilde{C} \mid (\tilde{C} \geq r, T_1 > \tilde{C}) \cup (\tilde{C} \geq r, T_1 < \tilde{C} \leq T_2) \cup (T_2 < \tilde{C}), Z)}{\mathrm{d}r} \\
&\quad + \lim_{\mathrm{d}r \to 0} \frac{P(\tilde{C} \leq r + \mathrm{d}r, T_1 < \tilde{C} \leq T_2 \mid (\tilde{C} \geq r, T_1 > \tilde{C}) \cup (\tilde{C} \geq r, T_1 < \tilde{C} \leq T_2) \cup (T_2 < \tilde{C}), Z)}{\mathrm{d}r} \\
&= I(T_1 > r) \underbrace{\lim_{\mathrm{d}r \to 0} \frac{P(\tilde{C} \leq r + \mathrm{d}r, T_1 > \tilde{C} \mid \tilde{C} \geq r, Z)}{\mathrm{d}r}}_{:= \alpha_{\tilde{C},1}(r;Z)} \\
&\quad + I(T_1 < r \leq T_2) \underbrace{\lim_{\mathrm{d}r \to 0} \frac{P(\tilde{C} \leq r + \mathrm{d}r, T_1 < \tilde{C} \leq T_2 \mid T_1 < r \leq \tilde{C}, Z)}{\mathrm{d}r}}_{:= \alpha_{\tilde{C},2}(r;Z)}.
\end{aligned}$$

That is, if we assume (12),

$$\begin{aligned}
I(T_1 > r)\alpha_{\tilde{C},1}(r;Z) &= I(T_1 > r)\alpha_{\tilde{C},1}(r \mid A, W), \\
I(T_1 < r \leq T_2)\alpha_{\tilde{C},2}(r;Z) &= I(T_1 < r \leq T_2)\alpha_{\tilde{C},2}(r \mid T_1, A, W),
\end{aligned}$$

where

$$\alpha_{\tilde{C},1}(r \mid A, W) = \lim_{\mathrm{d}r \to 0} \frac{P(\tilde{C} \leq r + \mathrm{d}r, T_1 > \tilde{C} \mid \tilde{T}_2 \geq r, A, W)}{\mathrm{d}r},$$
$$\alpha_{\tilde{C},2}(r \mid T_1, A, W) = \lim_{\mathrm{d}r \to 0} \frac{P(\tilde{C} \leq r + \mathrm{d}r, T_1 < \tilde{C} \leq T_2 \mid T_1 < r \leq \tilde{T}_2, T_1, A, W)}{\mathrm{d}r}.$$

B Proof of Lemma 1

All transition probabilities of the illness-death model can be expressed in terms of the hazards for the transitions (see e.g. Putter et al. (2007)). For instance, the probability of going from state 1 directly to state 3, within a time interval $(s, t]$, can be expressed as

$$Pr(T_2 \leq t, \eta = 0 \mid T_1 > s) = \int_s^t \exp\left[-\int_s^r \left\{\lambda_{12}(u) + \lambda_{13}(u)\right\} \mathrm{d}u\right] \lambda_{13}(r)\, \mathrm{d}r,$$

where we have omitted the baseline covariates for now.

The probability of going from state 1 to state 3 moving through state 2, within a time interval $(s, t]$, can be expressed as

$$Pr(T_2 \leq t, \eta = 1 \mid T_1 > s) = \int_s^t \left[\int_r^t \exp\left\{-\int_r^s \lambda_{23}(u, u-r)\, \mathrm{d}u\right\} \lambda_{23}(s, s-r)\, \mathrm{d}s\right] \exp\left[-\int_s^r \left\{\lambda_{12}(u) + \lambda_{13}(u)\right\} \mathrm{d}u\right] \lambda_{12}(r)\, \mathrm{d}r.$$

Then

$$\begin{aligned} Pr(T_2 \leq t) =& Pr(T_2 \leq t, \eta = 0 \mid T_1 > 0) + Pr(T_2 \leq t, \eta = 1 \mid T_1 > 0) \\ =& \int_0^t \exp\left[-\int_0^r \left\{\lambda_{12}(u) + \lambda_{13}(u)\right\} \mathrm{d}u\right] \left\{\lambda_{12}(r) + \lambda_{13}(r)\right\} \mathrm{d}r \\ & - \int_0^t \exp\left\{-\int_r^t \lambda_{23}(u, u-r)\, \mathrm{d}u\right\} \exp\left[-\int_0^r \left\{\lambda_{12}(u) + \lambda_{13}(u)\right\} \mathrm{d}u\right] \lambda_{12}(r)\, \mathrm{d}r \\ =& 1 - \exp\left[-\int_0^t \left\{\lambda_{12}(u) + \lambda_{13}(u)\right\} \mathrm{d}u\right] \\ & - \int_0^t \exp\left\{-\int_r^t \lambda_{23}(u, u-r)\, \mathrm{d}u\right\} \exp\left[-\int_0^r \left\{\lambda_{12}(u) + \lambda_{13}(u)\right\} \mathrm{d}u\right] \lambda_{12}(r)\, \mathrm{d}r. \end{aligned} \tag{13}$$

where the last equality follows using $\int_r^t \exp\left\{-\int_r^s \lambda_{23}(u, u-r)\, \mathrm{d}u\right\} \lambda_{23}(s, s-r)\, \mathrm{d}s = 1 - \exp\left\{-\int_r^t \lambda_{23}(u, u-r)\, \mathrm{d}u\right\}$.

Then

$$
\begin{aligned}
E\left\{I(T_2^{a^D,a^I} \le \tau)\right\} =& E\left[E\left\{I(T_2^{a^D,a^I} \le \tau) \mid W\right\}\right] \\
=& E\Bigg[1 - \exp\left\{-\Lambda_{12}^{a^D,a^I}(\tau|W) - \Lambda_{13}^{a^D,a^I}(\tau|W)\right\} \\
& - \int_0^\tau \exp\left\{-\Lambda_{23}^{a^D,a^I}(\tau, \tau - r|W)\right\} \exp\left\{-\Lambda_{12}^{a^D,a^I}(r|W) - \Lambda_{13}^{a^D,a^I}(r|W)\right\} \mathrm{d}\Lambda_{12}^{a^D,a^I}(r|W)\Bigg] \\
=& E\Bigg[1 - \exp\left\{-\Lambda_{12}^{a^I,a^I}(\tau|W) - \Lambda_{13}^{a^D,a^D}(\tau|W)\right\} \\
& - \int_0^\tau \exp\left\{-\Lambda_{23}^{a^D,a^D}(\tau, \tau - r|W)\right\} \exp\left\{-\Lambda_{12}^{a^I,a^I}(r|W) - \Lambda_{13}^{a^D,a^D}(r|W)\right\} \mathrm{d}\Lambda_{12}^{a^I,a^I}(r|W)\Bigg] \\
=& E\Bigg[1 - \exp\left\{-\Lambda_{12}(\tau|a^I, W) - \Lambda_{13}(\tau|a^D, W)\right\} \\
& - \int_0^\tau S_2(\tau \mid r, a^D, W) \exp\left\{-\Lambda_{12}(r|a^I, W) - \Lambda_{13}(r|a^D, W)\right\} \mathrm{d}\Lambda_{12}(r|a^I, W)\Bigg].
\end{aligned}
$$

The first equality is by the law of iterated expectations. The second equality follows by using the representation in Eq. (13) under an intervention that sets $A^D = a^D$ and $A^I = a^I$. The third equality follows by applying the dismissible components conditions. The last equality follows by applying A.0–A.3.

C Full data EIF

Let Q_ε be a parametric submodel with parameter $\varepsilon \in \mathbb{R}$ which passes through Q at $\varepsilon = 0$. The corresponding tangent space $\mathcal{T}^F$ is the closure of the linear span of the scores of the parametric submodels. Due to the factorization of the probability distribution of the full-data density in (1) we can write this as the orthogonal sum

$$\mathcal{T}^F = \mathcal{T}_1^F \oplus \mathcal{T}_2^F \oplus \mathcal{T}_3^F \oplus \mathcal{T}_4^F$$

where

$$
\begin{aligned}
\mathcal{T}_1^F &= \left\{ \int \alpha(u, A, W) dM_{13}^F(u, A, W) \text{ for all functions } \alpha(u, a, w) \right\} \\
\mathcal{T}_2^F &= \left\{ \int \alpha(u, A, W) dM_{12}^F(u, A, W) \text{ for all functions } \alpha(u, a, w) \right\} \\
\mathcal{T}_3^F &= \left\{ \eta \int \alpha(u, T_1, A, W) dM_{23}^F(u, T_1, A, W) \text{ for all functions } \alpha(u, r, a, w) \right\} \\
\mathcal{T}_4^F &= \{\alpha(A, W) \in \mathcal{H} : E[\alpha(A, W)] = 0\}
\end{aligned}
$$

In particular the score on the parametric submodel can be written

$$
\begin{aligned}
\ell'_Z(z;0) = \partial \log q(z;\varepsilon)/\partial\varepsilon \mid_{\varepsilon=0} =& \ell'_W(w;0) + \ell'_{A|W}(a \mid w;0) + \ell'_{13}(t_1, \eta \mid a, w;0) \\
& + \ell'_{12}(t_1, \eta \mid a, w;0) + \ell'_{23}(t_2, t_2 - t_1, \eta \mid a, w;0)
\end{aligned}
$$

where $\ell'_W(w;\varepsilon) = \partial/\partial\varepsilon \log \mu(w;\varepsilon)$, $\ell'_{A|W}(w \mid a;\varepsilon) = \partial/\partial\varepsilon \log \pi(a \mid w;\varepsilon)$ and

$$\begin{aligned}\ell'_{12}(T_1, \eta \mid A, W;0) =& \frac{\partial}{\partial\varepsilon}\left\{\eta \log \lambda_{12}(T_1 \mid A, W;\varepsilon) - \int_0^{T_1} \lambda_{12}(u \mid A, W;\varepsilon)\,\mathrm{d}u\right\}\Big|_{\varepsilon=0} \\ =& \eta \frac{\frac{\partial}{\partial\varepsilon}\left\{\lambda_{12}(T_1 \mid A, W, \varepsilon)\right\}|_{\varepsilon=0}}{\lambda_{12}(T_1 \mid A, W;0)} \\ & - \int \frac{\frac{\partial}{\partial\varepsilon}\left\{\lambda_{12}(T_1 \mid A, W, \varepsilon)\right\}|_{\varepsilon=0}}{\lambda_{12}(T_1 \mid A, W;0)} \lambda_{12}(u \mid A, W) I(T_1 > u)\,\mathrm{d}u \\ =& \int \frac{\frac{\partial}{\partial\varepsilon}\left\{\lambda_{12}(s \mid A, W, \varepsilon)\right\}|_{\varepsilon=0}}{\lambda_{12}(s \mid A, W;0)} \,\mathrm{d}M^F_{12}(s, A, W),\end{aligned}$$

and

$$\begin{aligned}\ell'_{13}(T_1, \eta \mid A, W;0) =& \frac{\partial}{\partial\varepsilon}\left\{(1-\eta) \log \lambda_{13}(T_1 \mid A, W;\varepsilon) - \int_0^{T_1} \lambda_{13}(u \mid A, W;\varepsilon)\,\mathrm{d}u\right\}\Big|_{\varepsilon=0} \\ =& \int \frac{\frac{\partial}{\partial\varepsilon}\left\{\lambda_{13}(s \mid A, W, \varepsilon)\right\}|_{\varepsilon=0}}{\lambda_{13}(s \mid A, W;0)} \,\mathrm{d}M^F_{13}(s, A, W),\end{aligned}$$

and

$$\begin{aligned}\ell'_{23}(T_2, T_2 - T_1, \eta \mid A, W;0) =& \frac{\partial}{\partial\varepsilon}\Big\{\eta \log \lambda_{23}(T_2, T_2 - T_1 \mid A, W;\varepsilon) \\ & - \eta \int_{T_1}^{T_2} \lambda_{23}(u, u - T_1 \mid A, W;\varepsilon)\,\mathrm{d}u\Big\}\Big|_{\varepsilon=0} \\ =& \eta \int \frac{\frac{\partial}{\partial\varepsilon}\left\{\lambda_{23}(s, s - T_1 \mid A, W, \varepsilon)\right\}|_{\varepsilon=0}}{\lambda_{23}(s, s - T_1 \mid A, W;0)} \,\mathrm{d}M^F_{23}(s, T_1.A, W).\end{aligned}$$

By Riesz' representation theorem the efficient influence function can be characterized as any element $\tilde{\psi} \in \mathcal{T}^F$ which is a pathwise derivative of the target parameter in the sense that

$$\frac{\partial\psi(Q_\varepsilon)}{\partial\varepsilon}\Big|_{\varepsilon=0} = E\left[\tilde{\psi}, \ell'_Z\right] \tag{14}$$

for any one-dimensional submodel Q_ε with corresponding score ℓ'_Z.

Note that under the nonparametric model we have that the full-data tangent space is the entire Hilbert space $L^2_0(Q)$ of measurable, mean-zero functions of Z equipped with the covariance inner product. Then any pathwise derivate will trivially be contained in $\mathcal{T}^F$. Hence we only need to check that the proposed EIF in (8) satisfies (14).

Consider first the left-hand side of (14). We may write

$$
\begin{aligned}
&\left.\frac{\partial \psi(Q_\varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=0} \\
&= \int_{\mathscr{W}} P_{13}(\tau, a^D, a^I, w)\ell'_W(w;0)d\mu(w) \\
&\quad + \int_{\mathscr{W}} \Omega_t(a^D, a^I, w)\left\{\int_0^\tau \frac{\partial}{\partial \varepsilon}\, \mathrm{d}\Lambda_{12}(r \mid a^I, w;\varepsilon)\mid_{\varepsilon=0}\right\} \mathrm{d}\mu(w) \\
&\quad + \int_{\mathscr{W}} \Omega_t(a^D, a^I, w)\left\{\int_0^\tau \frac{\partial}{\partial \varepsilon}\, \mathrm{d}\Lambda_{13}(r \mid a^D, w;\varepsilon)\mid_{\varepsilon=0}\right\} \mathrm{d}\mu(w) \\
&\quad - \int_{\mathscr{W}}\int_0^\tau S_2(\tau \mid r, a^D, w)\Omega_r(a^D, a^I, w)\frac{\partial}{\partial \varepsilon}\, \mathrm{d}\Lambda_{12}(r \mid a^I, w;\varepsilon)\mid_{\varepsilon=0} \mathrm{d}\mu(w) \\
&\quad + \int_{\mathscr{W}}\int_0^\tau S_2(\tau \mid r, a^D, w)\Omega_r(a^D, a^I, w)\left\{\int_0^r \frac{\partial}{\partial \varepsilon}\, \mathrm{d}\Lambda_{12}(s \mid a^I, w;\varepsilon)\mid_{\varepsilon=0}\right\} \mathrm{d}\Lambda_{12}(r \mid a^I, w)\, \mathrm{d}\mu(w) \\
&\quad + \int_{\mathscr{W}}\int_0^\tau S_2(\tau \mid r, a^D, w)\Omega_r(a^D, a^I, w)\left\{\int_0^r \frac{\partial}{\partial \varepsilon}\, \mathrm{d}\Lambda_{13}(s \mid a^D, w;\varepsilon)\mid_{\varepsilon=0}\right\} \mathrm{d}\Lambda_{12}(r \mid a^I, w)\, \mathrm{d}\mu(w) \\
&\quad - \int_{\mathscr{W}}\int_0^\tau S_2(\tau \mid r, a^D, w)\left\{\int_r^\tau \frac{\partial}{\partial \varepsilon}\, \mathrm{d}\Lambda_{23}(s, s-r \mid a^D, w;\varepsilon)\mid_{\varepsilon=0}\right\}\Omega_r(a^D, a^I, w)\, \mathrm{d}\Lambda_{12}(r \mid a^I, w)\, \mathrm{d}\mu(w) \\
&= \int_{\mathscr{W}} P_{13}(\tau, a^D, a^I, w)\ell'_W(w;0)d\mu(w) \\
&\quad + \int_{\mathscr{W}}\int_0^\tau h_{12,\tau}(r, a^D, a^I, w)\frac{\partial}{\partial \varepsilon}\, \mathrm{d}\Lambda_{12}(r \mid a^I, w;\varepsilon)\mid_{\varepsilon=0} \mathrm{d}\mu(w) \\
&\quad + \int_{\mathscr{W}}\int_0^\tau h_{13,\tau}(r, a^D, a^I, w)\frac{\partial}{\partial \varepsilon}\, \mathrm{d}\Lambda_{13}(r \mid a^D, w;\varepsilon)\mid_{\varepsilon=0} \mathrm{d}\mu(w) \\
&\quad + \int_{\mathscr{W}}\int_0^\tau h_{23,\tau}(r, a^D, a^I, w)\int_r^\tau \mathrm{d}\Lambda_{12}(r \mid a^D, w)\frac{\partial}{\partial \varepsilon}\, \mathrm{d}\Lambda_{23}(u, u-r \mid a^D, w;\varepsilon)\mid_{\varepsilon=0} \mathrm{d}\mu(w)
\end{aligned}
$$

where the second equality follows by changing the order of integration.

Consider now the right-hand side of (14). We have by iterated expectations, and the properties of score functions that

$$
E\left\{\tilde{\psi}(Z;\tau, a^D, a^I)\ell'_Z(Z;0)\right\} = E\left\{P_{13}(\tau, a^D, a^I, W)\ell'_W(w;0)\right\} \tag{15}
$$

$$
\begin{aligned}
&+ E\left[\frac{I(A = a^I)}{P(A = a^I \mid W)}\int_0^\tau h_{12,\tau}(s, a^D, a^I, W)\frac{dM^F_{12}(s, a^I, W)}{S_1(s \mid a^I, W)}\right. \\
&\left.\times\left\{\ell'_{12}(T_1 \mid A, W;0) + \ell'_{13}(T_1, \eta \mid A, W;0)\right\}\right]
\end{aligned} \tag{16}
$$

$$
\begin{aligned}
&+ E\left[\frac{I(A = a^D)}{P(A = a^D \mid W)}\int_0^\tau h_{13,\tau}(s, a^D, a^I, W)\frac{dM^F_{13}(s, a^D, W)}{S_1(s \mid a^D, W)}\right. \\
&\left.\times\left\{\ell'_{12}(T_1 \mid A, W;0) + \ell'_{13}(T_1, \eta \mid A, W;0)\right\}\right]
\end{aligned} \tag{17}
$$

$$
+ E\left\{\frac{I(A = a^D)}{P(A = a^D \mid W)}\frac{\eta h_{23,\tau}(T_1, a^D, a^I, W)}{S_1(T_1 \mid a^D, W)}\int_{T_1}^\tau \frac{dM^F_{23}(s, T_1, a^I, W)}{S_2(s \mid a^D, W)} \times \ell'_{23}(T_2, \eta \mid T_1, A, W;0)\right\} \tag{18}
$$

Du to the representation of the scores ℓ_{12} and ℓ_{13} in terms of the full-data martingales, the expectations in (16)–(18) are the covariances of martingale stochastic integrals. They be computed by finding the expectation of the corresponding predictable covariation processes (Fleming and Harrington 1991). In particular, the predictable covariation process of M_{ij}^F with itself is the compensator part of the martingale. The predictable covariation process of M_{12}^F and M_{13}^F is zero because the counting processes N_{12} and N_{13} by definition do not jump simultaneously.

Then we may write (16) as

$$
\begin{aligned}
&E\left[\frac{I(A=a^I)}{\pi(a^I \mid W)}\int_0^\tau \frac{h_{12,\tau}(s,a^D,a^I,W)}{S_1(s \mid a^I,W)}\frac{\frac{\partial}{\partial\varepsilon}\left\{\lambda_{12}(s \mid a^I,W,\varepsilon)\right\}\mid_{\varepsilon=0}}{\lambda_{12}(s \mid a^I,W;0)}I(T_1>s)\lambda_{12}(s \mid a^I,W)\,\mathrm{d}s\right]\\
&=E\left[\frac{E\left\{I(A=a^I)\mid W\right\}}{\pi(a^I \mid W)}\int_0^\tau \frac{h_{12,\tau}(s,a^D,a^I,W)}{S_1(s \mid a^I,W)}\frac{\partial}{\partial\varepsilon}\left\{\lambda_{12}(s \mid a^I,W,\varepsilon)\right\}\mid_{\varepsilon=0}E\left\{I(T_1>s)\mid a^I,W\right\}\mathrm{d}s\right]\\
&=E\left[\int_0^\tau h_{12,\tau}(s,a^D,a^I,W)\frac{d}{d\varepsilon}\left\{\mathrm{d}\Lambda_{12}(s \mid a^I,W;\varepsilon)\right\}\mid_{\varepsilon=0}\right],
\end{aligned}
$$

and similarly for (17)

$$
\begin{aligned}
&E\left[\frac{I(A=a^D)}{\pi(a^D \mid W)}\int_0^\tau \frac{h_{13,\tau}(s,a^D,a^I,W)}{S_1(s \mid a^D,W)}\frac{\frac{\partial}{\partial\varepsilon}\left\{\lambda_{13}(T_1 \mid A,W,\varepsilon)\right\}\mid_{\varepsilon=0}}{\lambda_{13}(T_1 \mid A,W;0)}I(T_1>s)\lambda_{13}(s \mid a^D,W)\,\mathrm{d}s\right]\\
&=E\left[\int_0^\tau h_{13,\tau}(s,a^D,a^I,W)\frac{d}{d\varepsilon}\left\{\mathrm{d}\Lambda_{13}(s \mid a^D,W;\varepsilon)\right\}\mid_{\varepsilon=0}\right].
\end{aligned}
$$

Finally we may rewrite (18) as

$$
\begin{aligned}
&E\left[\frac{I(A=a^D)}{\pi(a^D \mid W)}\frac{\eta h_{23,\tau}(T_1,a^D,a^I,W)}{S_1(T_1 \mid a^D,W)}\right.\\
&\left.\int_{T_1}^\tau \frac{I(T_1<s<T_2)}{S_2(s \mid T_1,a^D,W)}\frac{\frac{\partial}{\partial\varepsilon}\left\{\lambda_{23}(s,s-T_1 \mid a^D,W,\varepsilon)\right\}\mid_{\varepsilon=0}}{\lambda_{23}(s,s-T_1 \mid a^D,W;0)}\lambda_{23}(s,s-T_1 \mid a^D,W)\,\mathrm{d}s\right]\\
&=E\left[\frac{E\left\{I(A=a^D)\mid W\right\}}{\pi(a^D \mid W)}E\left\{\frac{\eta h_{23,\tau}(T_1,a^D,a^I,W)}{S_1(T_1 \mid a^D,W)}\right.\right.\\
&\int_{T_1}^\tau \frac{E\left\{I(T_1<s<T_2)\mid T_1,\eta,a^D,W\right\}}{S_2(s \mid T_1,a^D,W)}\\
&\left.\left.\times\frac{\partial}{\partial\varepsilon}\left\{\lambda_{23}(s,s-T_1 \mid a^D,W,\varepsilon)\right\}\mid_{\varepsilon=0}\mathrm{d}s\mid a^D,W\right\}\right]\\
&=\int_0^\tau h_{23,\tau}(r,a^D,a^I,W)\int_r^\tau \frac{d}{d\varepsilon}\left\{\mathrm{d}\Lambda_{23}(u,u-r \mid a^D,w;\varepsilon)\right\}\mid_{\varepsilon=0}\mathrm{d}\Lambda_{12}(r \mid a^D,W).
\end{aligned}
$$

Hence we have shown that the proposed influence function is in fact the efficient full-data influence function.

D Observed-data EIF

By Tsiatis (2006) theorem 10.1 and 10.4 we can map the full-data EIF to the observed-data EIF using the linear operator $\mathscr{J} : L_0^2(Q) \to L_0^2(P)$ which is defined by

$$\mathscr{J}(\tilde{\psi}(Z;\tau,a^D,a^I)) = \frac{\delta\tilde{\psi}(Z;\tau,a^D,a^I)}{K_{\tilde{C}}\{\tilde{T}_2, G_{\tilde{T}_2}(Z)\}} + \int \frac{E\{\tilde{\psi}(Z;,a^D,a^I) \mid G_u(Z)\}}{K_{\tilde{C}}\{u, G_u(Z)\}} \mathrm{d}M_{\tilde{C}}\{u, G_u(Z)\}. \quad (19)$$

where

$$K_{\tilde{C}}\{u, G_u(Z)\} := \exp\left[-\int_0^u \lambda_{\tilde{C}}\left\{u; G_u(Z)\right\} \mathrm{d}u\right].$$

Using the following lemma we can rewrite the efficient influence function in terms of the observed-data martingales

Lemma A.1 *for any element* $\int h(u,Z)\,\mathrm{d}M^F(u,Z) \in \mathscr{T}_i^F$, *for* $i = 1,2,3$, *it holds that*

$$\mathscr{J}\left\{\int h(u,Z)\,\mathrm{d}M^F(u,Z)\right\} = \int h(u,Z)\frac{\mathrm{d}M(u,Z)}{K_{\tilde{C}}(u,Z)}.$$

Proof We prove the lemma for $i = 1$. The remaining cases follow by similar calculations.

First note that given $(T_1 \wedge T_2 \geq u, A, W)$, $M_{12}^F(v,A,W)$ is zero-mean martingale for $v \geq u$, and hence $E\left[\mathrm{d}M_{12}^F(v,A,W) \,\middle|\, T_1 \wedge T_2 \geq u, A, W\right] = 0$ for $v \geq u$. For $v < u$ we have that, given $(T_1 \wedge T_2 \geq u, A, W)$, $N_{12}(v) = 0$ and $E[I(T_1 > v, T_2 > v) \mid T_1 \wedge T_2 \geq u, A, W] = 1$, and hence we can write

$$E\left[\mathrm{d}M_{12}^F(v,X) \,\middle|\, T_1 \wedge T_2 \geq u, A, W\right] = -I(v < u)\lambda_{12}(v \mid A, W)\,\mathrm{d}v. \quad (20)$$

Also, for the second term we note that M_{12}^F is fixed given T_1 and A, W, so

$$E\left[\mathrm{d}M_{12}^F(v,X) \,\middle|\, \eta = 1, T_1 < u < T_2, T_1, A, W\right] = \mathrm{d}M_{12}(v,A,W). \quad (21)$$

Then, using (20) and (21), we can write

$$\begin{aligned}
&\int \frac{E\left[\int h(v,Z)\,\mathrm{d}M_{12}^F(v,Z) \,\middle|\, G_u(Z)\right] \mathrm{d}M_{\tilde{C}}(u,Z)}{K_{\tilde{C}}(u,Z)} \\
&= \int_0^{\tilde{T}_1} \frac{E\left[\int h(v,Z)\,\mathrm{d}M_{12}^F(v,Z) \,\middle|\, G_u(Z)\right] \mathrm{d}M_{\tilde{C}}(u,Z)}{K_{\tilde{C}}(u, G_u(Z))} + \eta \int_{T_1}^{\tilde{T}_2} \frac{E\left[\int h(v,Z)\,\mathrm{d}M_{12}^F(v,Z) \,\middle|\, G_u(Z)\right] \mathrm{d}M_{\tilde{C}}(u,Z)}{K_{\tilde{C}}(u, G_u(Z))} \\
&= -\int_0^{\tilde{T}_1} \int_0^v h(v,X)\,\mathrm{d}\Lambda_{12}(v,X) \frac{\mathrm{d}M_{\tilde{C}}(u,Z)}{K_{\tilde{C}}(u, G_u(Z))} + \tilde{\eta} \int h(u,X)\,\mathrm{d}M_{12}^F(u,Z) \int_{T_1}^{\tilde{T}_2} \frac{\mathrm{d}M_{\tilde{C}}(u,Z)}{K_{\tilde{C}}(u, G_u(Z))}.
\end{aligned} \quad (22)$$

With similar calculations as in Lu and Tsiatis (2008) Lemma A.2 the first term equals

$$
\begin{aligned}
&-\frac{(1-\tilde{\eta})(1-\delta)}{K_{\tilde{C}}(\tilde{T}_1 G_{\tilde{T}_1}(Z))}\int_0^{\tilde{T}_1} h(u,X)\,\mathrm{d}\Lambda_{12}(u,X)+\int_0^{\tilde{T}_1}\int_u^{\tilde{T}_1}\frac{\mathrm{d}\Lambda_{\tilde{C}}(s;G_s(Z))}{K_{\tilde{C}}(s,G_s(Z))}h(u,X)\,\mathrm{d}\Lambda_{12}(u,X)\\
&=-\frac{(1-\tilde{\eta})(1-\delta)}{K_{\tilde{C}}(\tilde{T}_1 G_{\tilde{T}_1}(Z))}\int_0^{\tilde{T}_1} h(u,X)\,\mathrm{d}\Lambda_{12}(u,X)\\
&\quad+\frac{1}{K_{\tilde{C}}(\tilde{T}_1,G_{\tilde{T}_1}(Z))}\int_0^{\tilde{T}_1} h(u,X)\,\mathrm{d}\Lambda_{12}(u,X)-\int_0^{\tilde{T}_1}\frac{h(u,X)}{K_{\tilde{C}}(u,G_u(Z))}\,\mathrm{d}\Lambda_{12}(u,X)\\
&=\frac{\tilde{\eta}}{K_{\tilde{C}}(\tilde{T}_1,G_{\tilde{T}_1}(Z))}\int_0^{\tilde{T}_1} h(u,X)\,\mathrm{d}\Lambda_{12}(u,X)-\int_0^{\tilde{T}_1}\frac{h(u,X)}{K_{\tilde{C}}(u,G_u(Z))}\,\mathrm{d}\Lambda_{12}(u,X)\\
&\quad+\frac{\delta(1-\tilde{\eta})}{K_{\tilde{C}}(\tilde{T}_1,G_{\tilde{T}_1}(Z))}\int_0^{\tilde{T}_1} h(u,X)\,\mathrm{d}\Lambda_{12}(u,X)\\
&=-\frac{\tilde{\eta}}{K_{\tilde{C}}(\tilde{T}_1,G_{\tilde{T}_1}(Z))}\int_0^{\tilde{T}_1} h(u,X)\,\mathrm{d}M_{12}^F(u,X)+\int_0^{\tilde{T}_1}\frac{h(u,X)}{K_{\tilde{C}}(u,G_u(Z))}\,\mathrm{d}M_{12}(u,X)\\
&\quad+\frac{\delta(1-\tilde{\eta})}{K_{\tilde{C}}(\tilde{T}_1,G_{\tilde{T}_1}(Z))}\int_0^{\tilde{T}_1} h(u,X)\,\mathrm{d}\Lambda_{12}(u,X)
\end{aligned}
$$

The second term in (22) equals

$$
\begin{aligned}
&\tilde{\eta}\int h(u,X)\,\mathrm{d}M_{12}^F(u,Z)\int_{T_1}^{\tilde{T}_2}\frac{\mathrm{d}M_{\tilde{C}}(u,Z)}{K_{\tilde{C}}(u,G_u(Z))}\\
&=\tilde{\eta}\int h(u,X)\,\mathrm{d}M_{12}^F(u,Z)\left[\frac{1-\delta}{K_{\tilde{C}}(\tilde{T}_2,G_{\tilde{T}_2}(Z))}-\left(\frac{1}{K_{\tilde{C}}(\tilde{T}_2,G_{\tilde{T}_2}(Z))}-\frac{1}{K_{\tilde{C}}(T_1,G_{T_1}(Z))}\right)\right]\\
&=\tilde{\eta}\int h(u,X)\,\mathrm{d}M_{12}^F(u,Z)\left[\frac{1}{K_{\tilde{C}}(T_1,G_{T_1}(Z))}-\frac{\delta}{K_{\tilde{C}}(\tilde{T}_2,G_{\tilde{T}_2}(Z))}\right].
\end{aligned}
$$

Using that $\tilde{T}_1=\tilde{T}_2$ and $\Lambda_{12}=-M_{12}^F$ when $\delta(1-\tilde{\eta})=1$ and $T_1=\tilde{T}_1$ when $\eta=1$, adding the two final lines of the two previous displays gives

$$
\int_0^{\tilde{T}_1}\frac{h(u,X)}{K_{\tilde{C}}(u,G_u(Z))}\,\mathrm{d}M_{12}(u,X)-\frac{\delta}{K_{\tilde{C}}(\tilde{T}_2,G_{\tilde{T}_2}(Z))}\int_0^{\tilde{T}_2} h(u,X)\,\mathrm{d}M_{12}^F(u,X),
$$

which is the desired result. □

E Proof of Lemma 2

By iterated expectations, it follows that

$$
\begin{aligned}
&E\left[\frac{I(A=a^I)}{\pi^*(a^I \mid W)}\int_0^\tau \frac{h_{12}^*(s,a^D,a^I,W)dM_{12}^{F*}(s,a^I,W)}{S_1^*(s \mid a^I,W)}\right] \\
&\quad = E\left[\frac{\pi(a^I \mid W)}{\pi^*(a^I \mid W)}\right. \\
&\quad \left.\int_0^\tau \frac{h_{12}^*(s,a^D,a^I,W)S_1(s \mid a^I,W)}{S_1^*(s \mid a^I,W)}\left\{\mathrm{d}\Lambda_{12}(s \mid a^I,W)-\mathrm{d}\Lambda_{12}^*(s \mid a^I,W)\right\}\right],
\end{aligned}
\tag{23}
$$

$$
\begin{aligned}
&E\left[\frac{I(A=a^D)}{\pi^*(a^D \mid W)}\int_0^\tau \frac{h_{13}^*(s,a^D,a^I,W)dM_{13}^{F*}(s,a^D,W)}{S_1^*(s \mid a^D,W)}\right] \\
&\quad = E\left[\frac{\pi(a^D \mid W)}{\pi^*(a^D \mid W)}\right. \\
&\quad \left.\int_0^\tau \frac{h_{13,}^*(s,a^D,a^I,W)S_1(s \mid a^D,W)}{S_1^*(s \mid a^D,W)}\left\{\mathrm{d}\Lambda_{13}(s \mid a^D,W)-\mathrm{d}\Lambda_{13}^*(s \mid a^D,W)\right\}\right],
\end{aligned}
\tag{24}
$$

and

$$
\begin{aligned}
&E\left[\frac{I(A=a^D)}{\pi^*(a^D \mid W)}\frac{\eta h_{23}^*(T_1,a^D,a^I,W)}{S_1^*(T_1 \mid a^D,W)}\int_{T_1}^\tau \frac{dM_{23}^{F*}(s,a^D,T_1,W)}{S_2^*(s \mid T_1,a^D,W)}\right] \\
&\quad = E\left[\frac{\pi(a^D \mid W)}{\pi^*(a^D \mid W)}E\left\{\frac{h_{23}^*(T_1,a^D,a^I,W)}{S_1^*(T_1 \mid a^D,W)}\int_{T_1}^\tau \frac{S_2(s \mid T_1,a^D,W)}{S_2^*(s \mid T_1,a^D,W)}\left\{\mathrm{d}\Lambda_{23}(s \mid T_1,a^D,W)\right.\right.\right. \\
&\quad \left.\left.\left.-\mathrm{d}\Lambda_{23}^*(s \mid T_1,a^D,W)\right\}\middle| a^D,W\right\}\right] \\
&\quad = E\left[\frac{\pi(a^D \mid W)}{\pi^*(a^D \mid W)}\int_0^\tau \frac{h_{23}^*(s,a^D,a^I,W)}{S_1^*(s \mid a^D,W)}\left\{1-\frac{S_2(\tau \mid s,a^D,W)}{S_2^*(\tau \mid s,a^D,W)}\right\}\right. \\
&\quad \left. S_1(s \mid a^D,W)\,\mathrm{d}\Lambda_{12}(s \mid a^D,W)\right]
\end{aligned}
\tag{25}
$$

Suppose π, Λ_{13} and Λ_{23} are correctly specified, but Λ_{12} is not. Then the terms (24) and (25) are zero, and we have

$$
\begin{aligned}
&E\left[\tilde{\psi}(Q^*)(Z,a^D,a^I,W)\right]\\
&= E\left[e^{-\Lambda^*_{12}(\tau|a^I,W)-\Lambda_{13}(\tau|a^D,W)}\int_0^\tau e^{-\Lambda_{12}(s|a^I,W)+\Lambda^*_{12}(s|a^I,W)}\left\{\mathrm{d}\Lambda_{12}(s\mid a^I,W)\right.\right.\\
&\quad\left.\left.-\,\mathrm{d}\Lambda^*_{12}(s\mid a^I,W)\right\}\right]\\
&-E\left[\int_0^\tau S_2(\tau\mid s,a^D,W)\Omega_s(a^D,a^I,W)\left\{\mathrm{d}\Lambda_{12}(s\mid a^I,W)-\mathrm{d}\Lambda^*_{12}(s\mid a^I,W)\right\}\right]\\
&+E\left[\int_0^\tau\left\{\int_0^s e^{-\Lambda_{12}(s|a^I,W)+\Lambda^*_{12}(s|a^I,W)}\left\{\mathrm{d}\Lambda_{12}(s\mid a^I,W)-\mathrm{d}\Lambda^*_{12}(s\mid a^I,W)\right\}\right\}\right.\\
&\quad S_2(\tau\mid sa^D,W)\\
&\quad\left.\times e^{-\Lambda^*_{12}(s|a^I,W)-\Lambda_{13}(s|a^D,W)}\,\mathrm{d}\Lambda^*_{12}(s\mid a^I,W)\right]\\
&\quad+E\left[1-e^{-\Lambda^*_{12}(\tau|a^I,W)-\Lambda_{13}(\tau|a^D,W)}\right.\\
&\quad\left.-\int_0^\tau S_2(\tau\mid s,a^D,W)e^{-\Lambda^*_{12}(s|a^I,W)-\Lambda_{13}(s|a^D,W)}\,\mathrm{d}\Lambda^*_{12}(s\mid a^I,W)\right]\\
&\quad-\psi(Q^*;\tau,a^D,a^I)\\
&\quad= E\left[e^{-\Lambda^*_{12}(\tau|a^I,W)-\Lambda_{13}(\tau|a^D,W)}\left\{1-e^{-\Lambda_{12}(\tau|a^I,W)+\Lambda^*_{12}(\tau|a^I,W)}\right\}\right]\\
&\quad-E\left[\int_0^\tau S_2(\tau\mid s,a^D,W)\Omega_s(a^D,a^I,W)\left\{\mathrm{d}\Lambda_{12}(s\mid a^I,W)-\mathrm{d}\Lambda^*_{12}(s\mid a^I,W)\right\}\right]\\
&\quad+E\left[\int_0^\tau\left\{1-e^{-\Lambda_{12}(s|a^I,W)+\Lambda^*_{12}(s|a^I,W)}\right\}\right.\\
&\quad\left.S_2(\tau\mid s,a^D,W)e^{-\Lambda^*_{12}(s|a^I,W)-\Lambda_{13}(s|a^D,W)}\,\mathrm{d}\Lambda^*_{12}(s\mid a^I,W)\right]\\
&\quad+E\left[1-e^{-\Lambda^*_{12}(\tau|a^I,W)-\Lambda_{13}(\tau|a^D,W)}\right.\\
&\quad\left.-\int_0^\tau S_2(\tau\mid s,a^D,W)e^{-\Lambda^*_{12}(s|a^I,W)-\Lambda_{13}(s|a^D,W)}\,\mathrm{d}\Lambda^*_{12}(s\mid a^I,W)\right]\\
&\quad-\psi(Q^*;\tau,a^D,a^I)\\
&\quad= E\left[1-\Omega_t(a^D,a^I,W)+\int_0^\tau S_2(\tau\mid s,a^D,W)\Omega_s(a^D,a^I,W)\,\mathrm{d}\Lambda_{12}(s\mid a^I,W)\right]\\
&\quad-\psi(Q^*;\tau,a^D,a^I)\\
&\quad=\psi(Q;\tau,a^D,a^I)-\psi(Q^*;\tau,a^D,a^I)
\end{aligned}
$$

Similarly, suppose π, Λ_{12} and Λ_{23} are correctly specified, but Λ_{13} is not. Then (23) and (25) are 0, and

$$
\begin{aligned}
&E\big[\tilde{\psi}(Q^*)(Z, a^D, a^I, W)\big] \\
&\quad = E\left[e^{-\Lambda_{12}(\tau|a^I,W)-\Lambda^*_{13}(\tau|a^D,W)}\left\{1 - e^{-\Lambda_{13}(\tau|a^I,W)+\Lambda^*_{13}(\tau|a^I,W)}\right\}\right] \\
&\quad + E\left[\int_0^\tau \left\{1 - e^{-\Lambda_{13}(s|a^I,W)+\Lambda^*_{13}(s|a^I,W)}\right\}\right. \\
&\quad \left. S_2(\tau \mid s, a^D, W) e^{-\Lambda_{12}(s|a^I,W)-\Lambda^*_{13}(s|a^D,W)}\, \mathrm{d}\Lambda_{12}(s \mid a^I, W)\right] \\
&\quad + E\left[1 - e^{-\Lambda_{12}(\tau|a^I,W)-\Lambda^*_{13}(\tau|a^D,W)}\right. \\
&\quad \left. - \int_0^\tau S_2(\tau \mid s, a^D, W) e^{-\Lambda_{12}(s|a^I,W)-\Lambda^*_{13}(s|a^D,W)}\, \mathrm{d}\Lambda_{12}(s \mid a^I, W)\right] \\
&\quad - \psi(Q^*;\tau, a^D, a^I) \\
&\quad = E\left[1 - \Omega_t(a^D, a^I, W) + \int_0^\tau S_2(\tau \mid s, a^D, W)\Omega_s(a^D, a^I, W)\, \mathrm{d}\Lambda_{12}(s \mid a^I, W)\right] - \psi(Q^*;\tau, a^D, a^I) \\
&\quad = \psi(Q;\tau, a^D, a^I) - \psi(Q^*;\tau, a^D, a^I)
\end{aligned}
$$

Finally, suppose π, Λ_{12} and Λ_{13} are correctly specified, but Λ_{23} is not. Then (23) and (24) are 0, and

$$
\begin{aligned}
&E\big[\tilde{\psi}(Q^*)(Z, a^D, a^I, W)\big] \\
&\quad = E\left[\int_0^\tau S^*_2(\tau \mid s, a^D, W)\left\{1 - \frac{S_2(\tau \mid s, a^D, W)}{S^*_2(\tau \mid s, a^D, W)}\right\}\Omega_s(a^D, a^I, W)\, \mathrm{d}\Lambda_{12}(s \mid a^I, W)\right] \\
&\quad + E\left[1 - \Omega_t(a^D, a^I, W) - \int_0^\tau S^*_2(\tau \mid s, a^D, W)\Omega_s(a^D, a^I, W)\, \mathrm{d}\Lambda_{12}(s \mid a^I, W)\right] \\
&\quad - \psi(Q^*;\tau, a^D, a^I) \\
&\quad = \psi(Q;\tau, a^D, a^I) - \psi(Q^*;\tau, a^D, a^I)
\end{aligned}
$$

F Proof of Lemma 3

Using the representation of the full-data influence function in (19) we need to show

$$
E\left[\int \frac{\left\{\tilde{\psi}(Z;\tau, a^D, a^I) - E^*\{\tilde{\psi}(Z;\tau, a^D, a^I) \mid G_s(Z)\}\right\}}{K^*_{\tilde{C}}(s \mid G_s(Z))}\, \mathrm{d}M^*_{\tilde{C}}(s, G_s(Z))\right] = 0 \tag{26}
$$

when either $\Lambda_{\tilde{C}}$ is correctly specified or the entire outcome distribution, that is Λ_{12}, Λ_{13} and Λ_{23}, are correctly specified.

By iterated expectations (26) holds if we show

(a) $E\Big[\mathrm{d}M^*_{\tilde{C}}(s, G_s(Z)) \mid Z\Big] = 0$ when $\Lambda_{\tilde{C}}$ is correctly specified

(b) $E\{\tilde{\psi}(Z;\tau,a^D,a^I) \mid G_u(Z)\} - E^*\{\tilde{\psi}(Z;\tau,a^D,a^I) \mid G_u(Z)\} = 0$ if Λ_{12}, Λ_{13} and Λ_{23} are correctly specified

(a)

Note that, under CAR, we have

$$\begin{aligned}
&E\{\mathrm{d}N_{\tilde{C}}(u) \mid T_1, T_2, \eta, A, W\} \\
&\quad = E\{I(\tilde{C} = u, T_1 \geq u) \mid T_1, T_2, \eta, A, W\} + E\{\eta I(\tilde{C} = u, T_1 \leq u < T_2) \mid T_1, T_2, \eta, A, W\} \\
&\quad = I(u \leq T_1) K_{\tilde{C},1}(u \mid A, W) \alpha_{\tilde{C},1}(u \mid A, W) \\
&\quad + I(T_1 \leq u < T_2) K_{\tilde{C},2}(u \mid T_1, A, W) \alpha_{\tilde{C},2}(u \mid T_1, A, W),
\end{aligned}$$

and

$$\begin{aligned}
E\{I(u \leq \tilde{T}_2) \mid Z\} &= I(u \leq T_1) K_{\tilde{C},1}(u \mid A, W), \\
E\{I(T_1 \leq u < \tilde{T}_2) \mid Z\} &= I(T_1 \leq u < T_2) K_{\tilde{C},2}(u \mid T_1, A, W).
\end{aligned}$$

Then

$$\begin{aligned}
E\Big\{ \mathrm{d}M^*_{\tilde{C}}(u, G_u(Z)) \mid Z \Big\} &= I(u \leq T_1) K_{\tilde{C},1}(u \mid A, W) \Big\{ \alpha_{\tilde{C},1}(u \mid A, W) - \alpha^*_{\tilde{C},1}(u \mid A, W) \Big\} \\
&+ I(T_1 \leq u < T_2) K_{\tilde{C},2}(u \mid T_1, A, W) \big\{ \alpha_{\tilde{C},2}(u \mid T_1, A, W) \\
&- \alpha^*_{\tilde{C},2}(u \mid T_1, A, W) \Big\},
\end{aligned}$$

which is zero when the censoring distribution is correctly specified.

(b)

Note that for any $\int h(u;Z)\,\mathrm{d}M^F(u;Z) \in \mathcal{T}_i$ for $i = 1, 2, 3$

$$\begin{aligned}
&E\left\{ \int h(u;Z)\,\mathrm{d}M^F(u;Z) \mid G_r(Z) \right\} - E^*\left\{ \int h(u;Z)\,\mathrm{d}M^F(u;Z) \mid G_r(Z) \right\} \\
&= \int h(u;Z) \big[E\{\mathrm{d}M^F(u;Z) \mid G_r(Z)\} - E^*\{\mathrm{d}M^F(u;Z) \mid G_r(Z)\} \big]
\end{aligned}$$

Consider first the $1 \to 2$ terms. Note that

$$\begin{aligned}
&E^*\{\mathrm{d}M_{12}(u \mid A, W) \mid G_s(Z)\} \\
&\quad = I(T_1 > s) \left[I(u \geq s) \frac{S^*_1(u \mid A, W)}{S^*_1(s \mid A, W)} \{\mathrm{d}\Lambda^*_{12}(u \mid A, W) \right. \\
&\quad \left. - \mathrm{d}\Lambda_{12}(u \mid A, W)\} - I(u < s)\,\mathrm{d}\Lambda_{12}(u \mid A, W) \right] \\
&\quad + I(T_1 \leq s < T_2)\,\mathrm{d}M_{12}(u \mid A, W) \\
&\quad + I(s > T_2)\,\mathrm{d}M_{12}(u \mid A, W),
\end{aligned}$$

and

$$
\begin{aligned}
&E\big\{\mathrm{d}M_{12}(u \mid A, W) \mid G_s(Z)\big\} \\
&\quad = I(s < T_1)\Big[- I(u < s)\,\mathrm{d}\Lambda_{12}(u \mid A, W)\Big] + I(T_1 \le s < T_2)\,\mathrm{d}M_{12}(u \mid A, W) \\
&\quad + I(s > T_2)\,\mathrm{d}M_{12}(u \mid A, W),
\end{aligned}
$$

so,

$$
\begin{aligned}
&E\big\{\mathrm{d}M_{12}(u \mid A, W) \mid G_s(Z)\big\} - E^*\big\{\mathrm{d}M_{12}(u \mid A, W) \mid G_s(Z)\big\} \\
&\quad = I(s < T_1)\left[I(u \ge s)\frac{S_1^*(u \mid A, W)}{S_1^*(s \mid A, W)}\big\{\mathrm{d}\Lambda_{12}(u \mid A, W) - \mathrm{d}\Lambda_{12}^*(u \mid A, W)\big\}\right].
\end{aligned}
$$

By similar calculations it holds for the $1 \to 3$ term that

$$
\begin{aligned}
&E\big\{\mathrm{d}M_{13}(u \mid A, W) \mid G_s(Z)\big\} - E^*\big\{\mathrm{d}M_{13}(u \mid A, W) \mid G_s(Z)\big\} \\
&\quad = I(s < T_1 \wedge T_2)\left[I(u \ge s)\frac{S_1^*(u \mid A, W)}{S_1^*(s \mid A, W)}\big\{\mathrm{d}\Lambda_{13}(u \mid A, W) - \mathrm{d}\Lambda_{13}^*(u \mid A, W)\big\}\right].
\end{aligned}
$$

and for the $2 \to 3$ term that

$$
\begin{aligned}
&E\big\{\mathrm{d}M_{23}(u \mid T_1, A, W) \mid G_s(Z)\big\} - E^*\big\{\mathrm{d}M_{23}(u \mid T_1, A, W) \mid G_s(Z)\big\} \\
&\quad = I(T_1 \le s < T_2)\left[I(u \ge s)\frac{S_2^*(u \mid T_1, A, W)}{S_2^*(s \mid T_1, A, W)}\big\{\mathrm{d}\Lambda_{23}(u, u - T_1 \mid A, W)\right. \\
&\quad \left. - \mathrm{d}\Lambda_{23}^*(u, u - T_1 \mid A, W)\big\}\right].
\end{aligned}
$$

G Simulation details

The following simulation scenarios were considered:

(i) all models are correctly specified: $\zeta = \gamma_{AW}^{12} = \gamma_{AW}^{13} = \gamma_{AW}^{23} = \theta = 0$
(ii) censoring and propensity score mis-specified: $\zeta = 2$, $\theta = 0.8$ and $\gamma_{AW}^{12} = \gamma_{AW}^{13} = \gamma_{AW}^{23} = 0$
(iii) Λ_{12} misspecified: $\gamma_{AW}^{12} = 4$ and $\zeta = \gamma_{AW}^{13} = \gamma_{AW}^{23} = \theta = 0$
(iv) Λ_{13} misspecified: $\gamma_{AW}^{13} = 1$ and $\zeta = \gamma_{AW}^{12} = \gamma_{AW}^{23} = \theta = 0$
(v) Λ_{23} misspecified: $\gamma_{AW}^{23} = 1$ and $\zeta = \gamma_{AW}^{12} = \gamma_{AW}^{13} = \theta = 0$
(vi) Λ_{12} and Λ_{13} misspecified: $\gamma_{AW}^{12} = 4\,\gamma_{AW}^{13} = 1$ and $\zeta = \gamma_{AW}^{23} = \theta = 0$
(vii) $\Lambda_{\tilde{C}}$ and Λ_{23} misspecified: $\theta = 0.8$, $\gamma_{AW}^{23} = 1$ and $\zeta = \gamma_{AW}^{12} = \gamma_{AW}^{12} = \gamma_{AW}^{13} = 0$
(viii) π and Λ_{23} misspecified: $\zeta = 2$, $\gamma_{AW}^{23} = 1$ and $\gamma_{AW}^{12} = \gamma_{AW}^{12} = \gamma_{AW}^{13} = \theta = 0$

Acknowledgements The authors thank Thomas Sehested and Christian Torp-Pedersen for sharing the data used in Sect. 6.

Funding Open access funding provided by Royal Danish Library.

References

Aalen OO, Stensrud MJ, Didelez V, Daniel R, Roysland K, Strohmaier S (2020) Time-dependent mediators in survival analysis: modeling direct and indirect effects with the additive hazards model. Biom J 62(3):532–549

Andersen PK, Borgan O, Gill RD, Keiding N (2012) Statistical models based on counting processes. Springer

Bickel PJ, Klaassen CA, Ritov Y, Wellner JA (1993) Efficient and adaptive estimation for semiparametric models. Johns Hopkins University Press Baltimore

Chan CGC, Gao F, Xia F (2021) Discussion on "causal mediation of semicompeting risk" by yen-tsung huang. Biometrics 77(4):1155–1159

Comment L, Mealli F, Haneuse S, Zigler C (2019) Survivor average causal effects for continuous time: a principal stratification approach to causal inference with semicompeting risks. arXiv preprint arXiv:1902.09304

Didelez V (2019) Defining causal mediation with a longitudinal mediator and a survival outcome. Lifetime Data Anal 25(4):593–610

Dinicolantonio JJ, D'Ascenzo F, Tomek A, Chatterjee S, Niazi AK, Biondi-Zoccai G (2013) Clopidogrel is safer than ticagrelor in regard to bleeds: a closer look at the plato trial. Int J Cardiol 168(3):1739–1744

Fine JP, Jiang H, Chappell R (2001) On semi-competing risks data. Biometrika 88(4):907–919

Fleming T, Harrington D (1991) Counting processes and survival analysis. Wiley, New York

Fulcher IR, Shpitser I, Didelez V, Zhou K, Scharfstein DO (2021) Discussion on causal mediation of semicompeting risks by yen-tsung huang. Biometrics 77(4):1165–1169

Gao F, Xia F, Chan KCG (2021) Defining and estimating subgroup mediation effects with semi-competing risks data. arXiv preprint arXiv:2010.00061

Gill RD, van der Laan MJ, Robins JM (1997) Coarsening at random: characterizations, conjectures, counter-examples. In: Proceedings of the first Seattle symposium in biostatistics, Springer, pp 255–294

Huang YT (2021) Causal mediation of semicompeting risks. Biometrics 77(4):1143–1154

Ibragimov IA, Has'minskii RZ (1981) Statistical estimation: Asymptotic theory. Springer

Kazi DS, Leong TK, Chang TI, Solomon MD, Hlatky MA, Go AS (2015) Association of spontaneous bleeding and myocardial infarction with long-term mortality after percutaneous coronary intervention. J Am Coll Cardiol 65(14):1411–1420

Lin SH, Young JG, Logan R, VanderWheele TJ (2017) Mediation analysis for a survival outcome with time-varying exposures, mediators, and confouncers. Stat Med 36(26):4153–4166

Lu X, Tsiatis AA (2008) Improving the efficiency of the log-rank test using auxiliary covariates. Biometrika 95(3):679–694

Martinussen T, Stensrud MJ (2023) Estimation of separable direct and indirect effects in continuous time. Biometrics 79(1):127–139

Newey WK, McFadden D (1994) Large sample estimation and hypothesis testing. Handb Econ 4:2111–2245

Pearl J (2001) Direct and indirect effects. In: Proceedings of the seventeenth conference on uncertainty in artificial intelligence, Morgan Kaufman, pp 411–420

Pfanzagel J, Wefelmeyer W (1985) Contributions to a general asymptotic statistical theory. Stat Risk Model 3(3–4):379–388

Putter H, Fiocco M, Geskus RB (2007) Tutorial in biostatistics: competing risks and multi-state models. Stat Med 26(11):2389–2430

Robins JM (1986) A new approach to causal inference in mortality studies with sustained exposure periods—application to control of the healthy worker survivor effect. Math Model 7(9):1393–1512
Robins JM, Greenland S (1992) Identifiability and exchangeability for direct and indirect effects. Epidemiology 3(2):143–155
Robins JM, Richardson TS (2011) Alternative graphical causal models and the identification of direct effects. In: Shrout PE, Keyes KM, Ornstein K (eds) Causality and psychopathology: Finding the determinants of disorders and their cures, Oxford University Press, chap 6, pp 103–158
Robins JM, Richardson TS, Shpitser I (2021) An interventionist approach to mediation analysis. arXiv preprint arXiv:2008.06019
Rytgaard HC, Gerds TA, van der Laan MJ (2022) Continuous-time targeted minimum loss-based estimation of intervention-specific mean outcomes. Ann Stat 50:2469–2491
Sehested TSG, Carlson N, Hansen PW, Gerds TA, Charlot MG, Torp-Pedersen C, Køber L, Gislason GH, Hlatsky MA, Fosbøl EL (2019) Reduced risk of gastrointestinal bleeding associated with proton pump inhibitor therapy in patients treated with dual antiplatelet therapy after myocardial infarction. Eur Heart J 40(24):1963–1970
Stefanski LA, Boos DD (2002) The calculus of m-estimation. Am Stat 56(1):29–38
Stensrud MJ, Hernan M, Tchetgen EJT, Robins JM, Didelez V (2021) A generalized theory of separable effects in competing event settings. Lifetime Data Anal 27(4):588–631
Stensrud MJ, Young JG, Martinussen T (2021) Discussion on "causal mediation of semicompeting risks by yen-tsung huang. Biometrics 77(4):1160–1164
Stensrud MJ, Young JG, Didelez V, Robins JM, Hernán MA (2022) Separable effects for causal inference in the presence of competing events. J Am Stat Assoc 117(537):175–183
Tsiatis A (2006) Semiparametric theory and missing data. Springer
van der Laan MJ, Robins JM (2003) Unified methods for censored longitudinal data and causality. Springer
van der Vaart AW (2000) Asymptotic statistics, 3rd edn. Cambridge University Press
Valeri L, Proust-Lima C, Fan W, Chen JT, Jacqmin-Gadda H (2021) A multistate approach for mediation analysis in the presence of semi-competing risks with application in cancer survival disparities. arXiv preprint arXiv:2102.13252
VanderWeele TJ (2015) Explanation in causal inference: Methods for mediation and interaction. Oxford University Press
Vansteelandt S, Linder M, Vandenberghe S, Steen J, Madsen J (2019) Mediation analysis of time-to-event endpoints accounting for repeatedly measured mediators subject to time-varying confounding. Stat Med 38(24):4828–4840
Wallentin L, Becker RC, Budaj A, Cannon CP, Emanuelsson H, Held C, Horrow J, Husted S, James S, Katus H, Mahaffey KW, Scirica BM, Skene A, Steg PG, Storey RF, Harrington RA (2009) Ticagrelor versus clopidogrel in patients with acute coronary syndromes. N Engl J Med 361(11):1045–1057
Zhang JL, Rubin DB (2003) Estimation of causal effects via principal stratification when some outcomes are truncated by death. J Educ Behav Stat 28(4):353–368
Zheng W, van der Laan MJ (2017) Longitudinal mediation analysis with time-varying mediators and exposures, with application to survival outcomes. J Causal Inference 5(2)

Authors and Affiliations

Marie Skov Breum[1] · Anders Munch[1] · Thomas A. Gerds[1] · Torben Martinussen[1]

✉ Marie Skov Breum
masb@sund.ku.dk

[1] Section of Biostatistics, Department of Public Health, University of Copenhagen, Copenhagen, Denmark

Manuscript II

Estimation of data-dependent (in)direct effect with a repeatedly measured mediator and missing outcome data

Marie Skov Breum and Mads Sundby Palle

Details:
To be submitted to Statistics in Medicine

Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Estimation of data-dependent (in)direct effects with a repeatedly measured continuous mediator and missing outcome data

Marie Skov Breum*[1] | Mads Sundby Palle[2]

[1]Section of Biostatistics, University of Copenhagen, Copenhagen, Denmark
[2]Novo Nordisk, Bagsværd, Denmark

Correspondence
*Marie Skov Breum, Email: masb@sund.ku.dk

Present Address
Øster Farimagsgade 5, 1014 Copenhagen, Denmark.

Abstract

In this paper we present a method for estimating the extent to which the effect of a (randomized) baseline treatment on an outcome of interest is mediated through a repeatedly measured continuous covariate. The causal estimand that we will consider is an interventional (in)direct effect which intervenes stochastically on the potential mediator using a known distribution which is estimated from the data. For estimation we propose a longitudinal targeted minimum loss-based estimation (LTMLE) method based on the sequential regression technique. We verify the theoretical properties of the estimator in a simulation study, and we illustrate the method by an application to data from the NASH clinical trial.

KEYWORDS:
mediation analysis; causal inference; longitudinal data; targeted maximum likelihood estimator

1 | INTRODUCTION

Causal mediation analysis is an increasingly important tool in epidemiological and medical research, enabling researchers to examine and quantify the mechanisms through which a treatment or exposure exerts it's effects on an outcome of interest[1]. When the potential mediator is assessed at a single time point the causal estimands of choice are often the natural direct and indirect effects[2,3] which are defined in terms of so-called nested or 'cross-world' counterfactual outcomes, i.e. counterfactual outcomes in a world where the exposure is assigned to a particular value and the mediator is assigned to its counterfactual value under a possibly different treatment assignment. Natural (in)direct effects have many appealing properties including adding up to the total treatment effect. However a major limitation is that they are generally not identified in the presence of mediator-outcome confounders that are affected by the exposure[4,5]. In addition to limiting the practical applicability of natural direct and indirect effects in the non-longitudinal setting, this means that they do not immediately generalize to the longitudinal setting that we are considering in this paper where such confounders are inherent.

Much recent work on causal mediation analysis has addressed this limitation of natural (in)direct effects either by proposing causal bounds[6,7,8,9,10] or by proposing alternative causal mediation estimands like randomized interventional effects[11,12,13,5,14] or separable (interventionist) effects[6,15,16] that avoid cross-world notions. In this paper we will focus on the concept of randomized interventional (in)direct effects. Randomized interventional (in)direct effects, also referred to simply as 'interventional' (in)direct effects or 'stochastic' (in)direct effects, are defined in terms of hypothetical outcomes in a world where the exposure is assigned to a particular value and the mediator is drawn from a distribution which is either known (e.g. based on the observed data) or is defined based the potential mediator distribution. Interventional (in)direct effects are identifiable in the presence of post-treatment confounder[5,17] and can therefore be generalized to the longitudinal setting where the mediator is measured repeatedly over time and we expect feedback between certain time-varying covariates and the mediator[18,19]. The presence of time-varying confounders however means that different total effect decompositions are possible with different interpretations

and identifiability assumptions. Two distinct extensions of interventionist effects to the longitudinal setting were proposed by Zheng & van der Laan (2017)[18] and VanderWeele and Tchetgen Tchetgen (2017)[19].

Since in our motivating application we are interested in an indirect effect that includes only paths directly from treatment to the mediator, we focus on the effect decomposition of Zheng & van der Laan (2017)[18] where the stochastic mediator distribution conditions on the entire past. The contribution of this paper is that we will consider data-dependent versions of the mediation target parameters where the stochastic conditional mediator distributions are assumed to be known and estimated from the data. We will show that the data-adaptive interventionist (in)direct effects can be identified under weaker identification assumptions than the interventionist (in)direct effects which assume that the stochastic mediator distribution is the true unknown distribution. This comes at the cost of the direct and indirect effects not necessarily providing a decomposition of the total treatment effect.

We derive the efficient influence function and propose a longitudinal targeted minimum loss-based estimation (LTMLE)[20,21] method based on the sequential regression technique of Bang & Robins (2005)[22]. The TMLE is a multiply robust and locally efficient estimator which allows for the use of data-adaptive nuisance parameter estimation provided certain rate conditions hold[23,24].

1.1 | NASH trial

The motivating application in this paper is the NASH clinical trial conducted by Novo Nordisk. Non-alcoholic steatohepatitis (NASH) is an advanced form of nonalcoholic fatty liver disease which is defined by presence of steatosis, ballooning, inflammation and varying degrees of fibrosis in the liver[25]. NASH is very prevalent in patients with obesity and type II diabetes[26]. If allowed to progress it can lead to cirrhosis and liver failure, in which case liver transplantation is the only treatment option[27,28,29,30]. There are currently no approved drugs for NASH, and first line of treatment is weight management and treatment of comorbidities[31]. NASH is typically asymptomatic and is diagnosed by liver biopsy[32].

The NASH phase II clinical trial is a double-blind randomized six-arm trial which compared three different doses of semaglutide (0.1, 0.2 and 0.4 mg) with placebo in subjects with NASH and obesity (BMI>25). A total of 320 subjects were randomized, stratified on region (Japanese/non-japanese), diabetes status (type II/non-type II) and fibrosis stage (1, 2 or 3) at an initial screening 6 weeks before baseline. The subjects were followed for a maximum of 72 weeks and attended a number of scheduled post-baseline visits. A liver biopsy was performed at the final assessment 72 weeks after baseline. The primary endpoint was histological resolution of NASH after 72 weeks (yes/no). Due to the invasive nature of the procedure a some patients refused to get a biopsy at the final assessment resulting in missing outcome data.

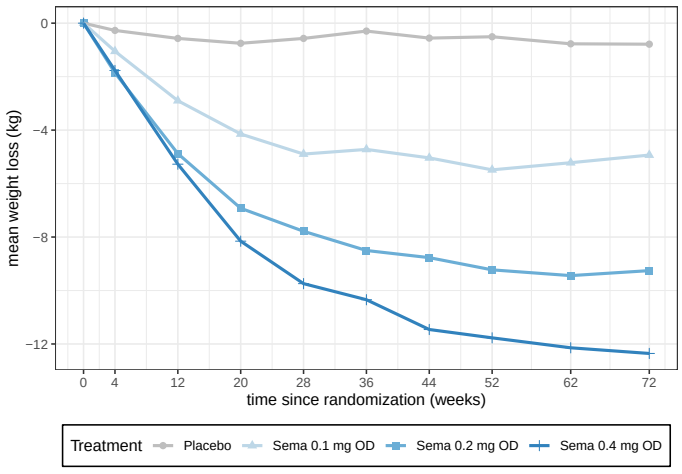


Figure 1 Average weight loss over time in the NASH trial by treatment arm

Semaglutide, which is sold under the brand names Ozempic, Wegovy and Rybelsus, is a once-daily injected glucagon-like peptide-1 (GLP-1) which is used for the treatment of type II diabetes and as anti-obesity medication. As illustrated in Figure 1 patients in the treatment arms loose a lot of weights while patients in the placebo arm do not. The question that motivated this

research was whether the causal pathways through which Semaglutide exerts it's effect on the primary endpoint are different from the pathways through which weight management, which is current first-line treatment, exerts it's effect.

1.2 | Organization of paper

This paper is organized as follows. In the following section we describe the setting and notation that we will use throughout the paper. In Section 3 we introduce the the data-dependent causal mediation estimand and provide the necessary identification assumptions. In Section 4 we propose a longitudinal targeted minimum loss-based estimation (LTMLE)[23,21] method based on the sequential regression technique[22] and describe it's implementation in detail. In Section 5 we conduct a simulation study to demonstrate the estimator's finite sample performance and robustness properties. Section 6 illustrates the method by an application to data from the NASH clinical trial. Some final remarks and further discussion is provided in Section 7. Proofs and technical details are given in the Appendix.

2 | SETTING AND NOTATION

We assume that the trial data is a sample of independent observations $(O_1, ..., O_n)$ which are identically distributed according to some unknown probability distribution denoted P_0, which is assumed to lie in a non-parametric model $\mathcal{P}$.

Let $k = 1, ..., K$ be discrete time-points representing the K follow-up visits. We can write the observed data vector O as $O = (L_0, A, C_1, L_1, M_1, ..., C_k, L_k, M_k, ..., C_K, L_K, M_K, R_Y, R_Y Y)$, where $L_0 \in \mathbb{R}^q$ represents the vector of baseline covariates, $A \in \mathcal{A}$ is the baseline treatment randomization, $C_t \in \{0, 1\}$ is the censoring indicator representing whether a subject has discontinued treatment or has dropped out by time t, $L_t \in \mathbb{R}^d$ is a vector of time-varying covariates collected at each follow-up visit and $M_t \in \mathbb{R}$ is the potential mediator. Note that we can also define $C_t = (C_t^D, C_t^A)$ to distinguish between censoring due to drop-out and censoring due to non-adherence. Finally $Y \in \{0, 1\}$ represents the outcome and $R_Y \in \{0, 1\}$ is the outcome missingness indicator. If a subject is censored then subsequent C_t, L_t, M_t and $(R_Y, R_Y Y)$ are encoded with default values.

Let $X_{m:k} = (X_m, ..., X_k)$ and let $\bar{X}_k$ denote the history of a random variable up to time k. We assume that the outcome is coarsened at random (CAR) i.e. $Y \perp\!\!\!\perp R_Y \mid A, \bar{C}_K, \bar{L}_K, \bar{M}_K$. We further assume that the data can be represented using the following Structural Causal Model[33]

$$
\begin{aligned}
L_0 &= f_{L_0}(U_{L_0}), \\
A &= f_A(L_0, U_A), \\
C_k &= f_{C_k}(A, \bar{C}_{k-1}, \bar{L}_{k-1}, \bar{M}_{k-1}, U_{C_k}), k = 1, ..., K, \\
L_k &= f_{L_k}(A, \bar{C}_k, \bar{L}_{k-1}, \bar{M}_{k-1}, U_{L_k}), k = 1, ..., K, \\
M_k &= f_{M_k}(A, \bar{C}_k, \bar{L}_k, \bar{M}_{k-1}, U_{M_k}), k = 1, ..., K, \\
R_Y &= f_{R_Y}(A, \bar{C}_K, \bar{L}_K, \bar{M}_K, U_{R_Y}), \\
R_Y Y &= R_Y f_Y(A, \bar{C}_K, \bar{L}_K, \bar{M}_K, U_Y),
\end{aligned}
\tag{1}
$$

where $U = (U_{L_0}, U_A, (U_{C_k} : k = 1, ...K), (U_{L_k} : k = 1, ...K), (U_{M_k} : k = 1, ...K), U_{R_Y}, U_Y)$ are exogenous random variables, and $f_{L_0}, f_A, (f_{C_k} : k = 1, ...K)$, $(f_{L_k} : k = 1, ...K)$, $(f_{M_k} : k = 1, ...K)$, f_{R_Y} and f_Y are deterministic mappings. Note that we do not require for the baseline treatment to be randomized.

3 | CAUSAL ESTIMAND(S)

The causal estimand that we consider is an randomized interventional (in)direct effect[11,12,13,5]. Let

$$
g_t^a(M_t \mid \bar{L}_t, \bar{M}_{t-1}) = P(M_t \mid A = a, C_t = 0, \bar{L}_t, \bar{M}_{t-1}),
$$

be the stochastic distribution of M_t under an intervention that sets $A = a$ and $C_t = 0$. We assume that this distribution is known and estimated from data, and we denote it by $\hat{g}_t^a$.

Consider an intervention on the Structural Causal Model (SCM) in (1) to set $A = a$ for $a \in \mathcal{A}$, and set $C_t = 0$ and randomly draw $M_t \sim \hat{g}_t^a$ for $t = 1, ..., K$. Let $\hat{\mathbf{g}}^a = (\hat{g}_t^a : t = 1, ..., K)$ and let $Y(a, \hat{\mathbf{g}}^a)$ denote the resulting counterfactual outcome.

We can then define the data-dependent stochastic interventional direct and indirect effects as

$$SDE(a, a') = E\left\{Y(a, \hat{\mathbf{g}}^a) - Y(a', \hat{\mathbf{g}}^a)\right\}, \tag{2}$$

and

$$SIE(a, a') = E\left\{Y(a', \hat{\mathbf{g}}^a) - Y(a', \hat{\mathbf{g}}^{a'})\right\}. \tag{3}$$

That is, the stochastic interventional indirect effect (SIE) is the effect of fixing the mediator to a random draw from the observed distribution of the mediator in the population when given treatment $A = a$, versus a random draw from the distribution of the population when given treatment $A = a'$ while assigning treatment to the value a. The stochastic interventional direct (SDE) effect is the effect of setting treatment $A = a$ versus $A = a'$ under an intervention that fixes the mediator to a random draw from the observed distribution of the mediator in the populationn when given treatment $A = a$.

The stochastic direct and indirect effect defined in (2) and (3) do not necessarily provide a decomposition of the total treatment effect. Instead they provide a decomposition of the so-called overall effect

$$OE(a, a') = E\left\{Y(a, \mathbf{g}^a) - Y(a', \mathbf{g}^{a'})\right\} = SIE(a', a) + SDE(a', a). \tag{4}$$

This overall effect (OE) can be interpreted as the difference in expected outcome between being in treatment arm $A = a$ with the mediator randomly drawn from the distribution of the population when given treatment $A = a$, and the expected outcome when being in treatment arm $A = a'$ with the mediator randomly drawn from the distribution of the population when given treatment $A = a'$.

Note that

$$OE(a', a) = E\left\{Y(a) - Y(a')\right\} + E\left\{Y(a, \hat{\mathbf{g}}^a) - Y(a)\right\} + E\left\{Y(a') - Y(a', \hat{\mathbf{g}}^{a'})\right\},$$

where the first term is the total effect and the two last terms are related to the difference between drawing the mediator from the observed distribution and setting the mediator to it's natural level.

3.1 | Identifiability

Suppose the following assumptions hold for all $t = 1, ..., K$

A.0 $L_{1:K}(a), L_{1:K}(a, \bar{m}), Y(a, \bar{m}) \perp\!\!\!\perp A \mid L_0$,

A.1 $L_{t:K}(a), L_{t:K}(a, \bar{m}), Y(a, \bar{m}) \perp\!\!\!\perp C_t \mid A = a, C_{t-1} = 0, \bar{M}_{t-1}, L_{t-1}$,

A.2 $L_{(t+1):K}(a, \bar{m}), Y(a, \bar{m}) \perp\!\!\!\perp M_t \mid A = a, C_t = 0, \bar{L}_t, \bar{M}_{t-1}$,

A.3 Positivity/overlap:

(i) $p_0(l_0) > 0 \Rightarrow p_0(a \mid l_0) > 0$,

(ii) $p_0(a, \bar{0}, \bar{m}_{t-1}, \bar{l}_{t-1}) > 0 \Rightarrow p_0(c_t = 1 \mid a, \bar{0}, \bar{m}_{t-1}, \bar{l}_{t-1}) < 1$,

(iii) $\sup_{m_t} \frac{\hat{g}_t^{a'}(m_t|\bar{l}_t, \bar{m}_{t-1})}{p_0(m_t|a, \bar{0}, \bar{l}_t, \bar{m}_{t-1})} < \infty$.

Assumption A.0 is the treatment randomization assumption, which holds by construction in a randomized trial such as the NASH trial. Assumptions A.1 is a sequential exchangeability assumption for censoring. In the NASH trial assumption A.1 requires that treatment randomization and histories of weight loss and covariates are sufficient to adjust for confounding between current censoring, and current and future covariate values and the outcome. Assumptions A.2 is a sequential exchangeability assumption for the mediator. Assumption A.2 requires that treatment randomization and histories of weight loss and covariates are sufficient to adjust for confounding between current weight loss, and current and future covariate values and the outcome.

The conditions in assumption A.3 are the positivity assumptions which ensure that the identifying formula below is well-defined. Condition (i) states that there should be a positive probability of being assigned to all treatment arms. Condition (ii) requires that within all strata in the data the probability of being censored should be less than one. Condition (iii) states that the stochastic distribution from which the mediator values are drawn should be supported in the data. Conditions A.3 (i) and A.3 (ii) are both unproblematic in the NASH trial. Condition A.3 (iii) requires further scrutiny as subjects in the placebo-arm do not loose weight while most subjects in the treatment arms loose weight. We discuss this further in Section 6.

Under assumptions (A.0-A.3) we have

$$\Psi(P_0)(a,\hat{\mathbf{g}}^{a'}) \equiv E\left\{Y(a,\hat{\mathbf{g}}^{a'})\right\} = \int_{\mathcal{L}_0} \prod_{k=1}^{K} \int_{\mathcal{L}_k\times\mathcal{M}_k} \sum_{a,y} \Big\{ y p_{0,Y}(y \mid a,\bar{0},\bar{m}_K,\bar{l}_K) p_{0,L_0}(l_0) \times p_{0,L_k}(l_k \mid a,\bar{0},\bar{m}_{k-1},\bar{l}_{k-1}) \hat{g}_k^{a'}(m_k \mid \bar{l}_k,\bar{m}_{k-1})\, \mathrm{d}\mu_{L_k}(l_k)\, \mathrm{d}\mu_{M_k}(m_k) \Big\}\, \mathrm{d}\mu_{L_0}(l_0). \quad (5)$$

where $\mathcal{L}_k$ is the support of L_k, $\mathcal{M}_k$ is the support of M_k and $\mu_{L_k}(l_k)$, $\mu_{M_k}(m_k)$ are some dominating measures. The proof is given in Appendix A. Note that $\Psi(P)(a,\hat{\mathbf{g}}^{a'})$ is data-dependent target parameter mapping because it depends on the data through $\hat{g}^a(P_n)$.

Then the SDE and the SIE respectively identifies to

$$SDE(a,a') = \Psi(P)(a,\hat{\mathbf{g}}^a) - \Psi(P)(a',\hat{\mathbf{g}}^a),$$

and

$$SDE(a,a') = \Psi(P)(a',\hat{\mathbf{g}}^a) - \Psi(P)(a',\hat{\mathbf{g}}^{a'}).$$

As mentioned the identification assumptions above are weaker than the assumptions required for identifying stochastic (in)direct effect when assuming that the stochastic mediator distribution is the true unknown distribution. In particular this would require the additional sequential exchangeability assumptions (A.0*) $\bar{M}_K(a) \perp\!\!\!\perp A \mid L_0$ and (A.1*) $M_{t:K}(a) \perp\!\!\!\perp C_t \mid A = a, C_{t-1} = 0, \bar{M}_{t-1}, \bar{L}_{t-1}$, and the additional positivity assumption that covariate values supported under one treatment arm is also supported under the other treatment arm.

4 | ESTIMATION

In this section we propose a longitudinal targeted minimum loss-based estimation (LTMLE)[23,24] method based on the sequential regression technique of Bang & Robins (2005)[22].

In particular we note that the target parameter in (5) can also be represented as a nested expectation

$$\Psi(P_0)(a,\hat{\mathbf{g}}^{a'}) = E\left\{ Q_{L_1}^{a,\hat{\mathbf{g}}^{a'}}(P_0)(L_0) \right\}, \quad (6)$$

where for $t = 1,...K$ and $P \in \mathcal{P}$

$$\begin{aligned}
Q_Y^{a,\hat{\mathbf{g}}^{a'}}(P)(\bar{L}_K,\bar{M}_K) &= Q_{L_{K+1}}^{a,\hat{\mathbf{g}}^{a'}}(P)(\bar{L}_K,\bar{M}_K) = E_P\left\{ Y \mid A = a, C_K = 0, \bar{L}_K, \bar{M}_K, R_Y = 1 \right\}, \\
Q_{M_t}^{a,\hat{\mathbf{g}}^{a'}}(P)(\bar{L}_t,\bar{M}_{t-1}) &= \int_{\mathcal{M}_t} Q_{L_{t+1}}^{a,\hat{\mathbf{g}}^{a'}}(P)(\bar{L}_t,m_t,\bar{M}_{t-1}) \hat{g}_t^{a'}(m_t \mid A,\bar{L}_t,\bar{M}_{t-1})\, \mathrm{d}\mu_{M_t}(m_t), \\
Q_{L_t}^{a,\hat{\mathbf{g}}^{a'}}(P)(\bar{L}_{t-1},\bar{M}_{t-1}) &= E_P\left\{ Q_{M_t}^{a,\hat{\mathbf{g}}^{a'}}(P)(\bar{L}_t,\bar{M}_{t-1}) \mid A = a, C_t = 0, \bar{L}_{t-1},\bar{M}_{t-1} \right\}.
\end{aligned}$$

We show in Appendix B that the efficient influence function for the target in (6) is given as follows

$$\begin{aligned}
D^*_{a,\hat{\mathbf{g}}^{a'}}(P)(O) =& H_{K+1}^{a,\hat{\mathbf{g}}^{a'}}(\bar{L}_K,\bar{M}_K)\left\{ Y - Q_Y^{a,\hat{\mathbf{g}}^{a'}}(\bar{L}_K,\bar{M}_K) \right\} \\
&+ \sum_{k=1}^{K} H_k^{a,\hat{\mathbf{g}}^{a'}}(\bar{L}_{k-1},\bar{M}_{k-1}) \left\{ Q_{M_k}^{a,\hat{\mathbf{g}}^{a'}}(\bar{L}_k,\bar{M}_{k-1}) - Q_{L_k}^{a,\hat{\mathbf{g}}^{a'}}(\bar{L}_{k-1},\bar{M}_{k-1}) \right\} \\
&+ Q_{L_k}^{a,\hat{\mathbf{g}}^{a'}}(L_0) - \Psi(P)(a,\hat{\mathbf{g}}^{a'}),
\end{aligned} \quad (7)$$

where

$$H_{K+1}^{a,\hat{\mathbf{g}}^{a'}}(\bar{L}_K,\bar{M}_K) = \frac{I(A=a)}{p_A(a \mid L_0)} \frac{I(C_K=0)}{\delta_K(a,\bar{L}_{K-1},\bar{M}_{K-1})} \frac{I(R_Y=1)}{p_{R_Y}(R_Y=1 \mid a,\bar{0},\bar{L}_K,\bar{M}_K)} \prod_{j=1}^{K} \frac{\hat{g}_j^{a'}(M_j \mid \bar{L}_j,\bar{M}_{j-1})}{p_{M_j}(M_j \mid a,\bar{0},\bar{L}_j,\bar{M}_{j-1})}, \quad (8)$$

and

$$H_k^{a,\hat{\mathbf{g}}^{a'}}(\bar{L}_{k-1},\bar{M}_{k-1}) = \frac{I(A=a)}{p_A(a \mid L_0)} \frac{I(C_k=0)}{\delta_k(a,\bar{L}_{k-1},\bar{M}_{k-1})} \prod_{j=1}^{k-1} \frac{\hat{g}_j^{a'}(M_j \mid \bar{L}_j,\bar{M}_{j-1})}{p_{M_j}(M_j \mid a,\bar{0},\bar{L}_j,\bar{M}_{j-1})},\ k = 1,...,K \quad (9)$$

for

$$\delta_k(A, \bar{L}_{k-1}, \bar{M}_{k-1}) = \prod_{j=1}^{k} p_{C_j}(C_j = 0 \mid A, C_{j-1} = 0, \bar{L}_{j-1}, \bar{M}_{j-1}),\ k = 1, ..., K.$$

4.1 | LTMLE algorithm

We will use the loss functions

$$\mathcal{L}(Q^{a,\hat{\mathbf{g}}^{a'}}_{L_{K+1}}) = -\left\{ Y \log\left(Q^{a,\hat{\mathbf{g}}^{a'}}_{L_{K+1}}\right) + (1-Y)\log\left(1 - Q^{a,\hat{\mathbf{g}}^{a'}}_{L_{K+1}}\right)\right\},$$
$$\mathcal{L}(Q^{a,\hat{\mathbf{g}}^{a'}}_{L_t}) = -\left\{ Q^{a,\hat{\mathbf{g}}^{a'}}_{M_t} \log\left(Q^{a,\hat{\mathbf{g}}^{a'}}_{L_t}\right) + (1 - Q^{a,\hat{\mathbf{g}}^{a'}}_{M_t})\log\left(1 - Q^{a,\hat{\mathbf{g}}^{a'}}_{L_t}\right)\right\},$$

and the least favorable submodels

$$Q^{a,\hat{\mathbf{g}}^{a'}}_{L_{K+1}}(\varepsilon) = \text{expit}\left(\text{logit}\left(Q^{a,\hat{\mathbf{g}}^{a'}}_{L_{K+1}} + \varepsilon\right)\right),$$
$$Q^{a,\hat{\mathbf{g}}^{a'}}_{L_t}(\varepsilon) = \text{expit}\left(\text{logit}\left(Q^{a,\hat{\mathbf{g}}^{a'}}_{L_t} + \varepsilon\right)\right).$$

Then the implementation of the LTMLE algorithm can be described as follows

1. Obtain initial estimates $\hat{H}^{a,\hat{\mathbf{g}}^{a'}}_{t,n}$ of $H^{a,\hat{\mathbf{g}}^{a'}}_{t}$ for $t = 1, ..., K+1$.
2. Regress Y on $(A, \bar{L}_K, \bar{M}_K)$ among those who are uncensored at time K with $R_Y = 1$. Evaluate the fitted function at $A = a$ and the observed covariates $(\bar{L}_K, \bar{M}_K)$ to obtain an estimate $\hat{Q}^{a,\hat{\mathbf{g}}^{a'}}_{n,L_{K+1}}(\bar{L}_K, \bar{M}_K)$ of $Q^{a,\hat{\mathbf{g}}^{a'}}_{L_{K+1}}(\bar{L}_K, \bar{M}_K)$. Update the estimate by setting $Q^{*,a,g}_{n,L_{K+1}} = \hat{Q}^{a,\hat{\mathbf{g}}^{a'}}_{n,L_{K+1}}(\varepsilon_{n,L_{K+1}})$, where

   $$\varepsilon_{n,L_{K+1}} = \arg\min_{\varepsilon} P_n \hat{H}^{a,\hat{\mathbf{g}}^{a'}}_{K+1,n} \mathcal{L}\left(\hat{Q}^{a,\hat{\mathbf{g}}^{a'}}_{n,L_{K+1}}(\varepsilon)\right),$$

   is the coefficient of a weighted logistic regression of Y onto the intercept model with an offset $\text{logit}\left(\hat{Q}^{a,\hat{\mathbf{g}}^{a'}}_{n,L_{K+1}}(\bar{L}_K, \bar{M}_K)\right)$ and weights $\hat{H}^{a,\hat{\mathbf{g}}^{a'}}_{K+1,n}(\bar{L}_K, \bar{M}_K)$.
3. For t=K,...,1
   (a) Compute an estimate $\hat{Q}^{a,\hat{\mathbf{g}}^{a'}}_{n,M_t}(\bar{L}_t, \bar{M}_{t-1}) = \sum_j \hat{Q}^{*,a,\hat{\mathbf{g}}^{a'}}_{n,L_{t+1}}(\bar{L}_t, m_{t,j}, \bar{M}_{t-1}) g(m_{t,j} \mid \bar{L}_t, \bar{M}_{t-1}) \Delta m_{t,j}$ of $Q^{a,\hat{\mathbf{g}}^{a'}}_{M_t}(\bar{L}_t, \bar{M}_{t-1})$, where $\Delta m_{t,j} = m_{t,j} - m_{t,j-1}$ and the $m_{t,j}$'s are some appropriately chosen discretization of the support of M_t. Alternatively the integral can be computed using Monte Carlo integration.
   (b) Regress $\hat{Q}^{a,\hat{\mathbf{g}}^{a'}}_{n,M_t}(\bar{L}_t, \bar{M}_{t-1})$ on $(A, \bar{L}_{t-1}, \bar{M}_{t-1})$ among those who are uncensored at time t. Evaluate the fitted function at $A = a$ the observed covariates $(\bar{L}_{t-1}, \bar{M}_{t-1})$ to obtain an estimate $\hat{Q}^{a,\hat{\mathbf{g}}^{a'}}_{n,L_t}(\bar{L}_{t-1}, \bar{M}_{t-1})$ of $Q^{a,\hat{\mathbf{g}}^{a'}}_{L_t}(\bar{L}_{t-1}, \bar{M}_{t-1})$. Update the estimate by setting $Q^{*,a,g}_{n,L_t} = \hat{Q}^{a,\hat{\mathbf{g}}^{a'}}_{n,L_t}(\varepsilon_{n,L_t})$ where

   $$\varepsilon_{n,L_t} = \arg\min_{\varepsilon} P_n \hat{H}^{a,\hat{\mathbf{g}}^{a'}}_{t,n} \mathcal{L}\left(\hat{Q}^{a,\hat{\mathbf{g}}^{a'}}_{n,L_t}(\varepsilon)\right)$$

   is the coefficient of a weighted logistic regression of $\hat{Q}^{a,\hat{\mathbf{g}}^{a'}}_{n,M_t}(\bar{L}_t, \bar{M}_{t-1})$ onto the intercept model with an offset $\text{logit}\left(\hat{Q}^{a,\hat{\mathbf{g}}^{a'}}_{n,L_t}(\bar{L}_{t-1}, \bar{M}_{t-1})\right)$ and weights $\hat{H}^{a,\hat{\mathbf{g}}^{a'}}_{t,n}(\bar{L}_{t-1}, \bar{M}_{t-1})$.
4. Then the TMLE is

$$\hat{\psi}^{\text{tmle}}_n = \frac{1}{n}\sum_{i=1}^{n}\left\{ Q^{*,a,\hat{\mathbf{g}}^{a'}}_{n,L_1}(L_{0,i})\right\}. \tag{10}$$

4.2 | Inference

Let $\delta = (\delta_k, k = 1, ..., K)$, $p_M = (p_{M_k} : k = 1, ..., K)$ and $Q^{a,}_{L} = (Q^{a,\hat{\mathbf{g}}^{a'}}_{L_k} : k = 1, ..., K+1)$. We show in Appendix C that the TMLE in (10) is a consistent estimator of $\Psi(P_0)(a, \hat{\mathbf{g}}^{a'})$ if either of the following conditions hold

(i) p_L and p_Y are correctly specified,

(ii) p_{R_Y}, δ and p_M are correctly specified,

(iii) p_Y, δ_n and p_M are correctly specified.

Provided certain rate conditions hold[21,34,24] the TMLE is an asymptotically efficient estimator of $\psi_0 = \Psi(P_0)(a, \hat{\mathbf{g}}^{a'})$. In particular, $\sqrt{n}(\hat{\psi}_n^{\text{tmle}} - \psi_0) \to N(0, \sigma_0^2)$, where $\sigma_0^2 = ED^*_{a,\hat{\mathbf{g}}^{a'}}(P_0)(O)^2$ is the variance of the efficient influence curve. We can estimate σ_0^2 with the empirical sample variance of the estimated efficient influence curve.

5 | SIMULATION STUDY

In this section we conduct a simulation study to demonstrate the estimator's finite sample performance and robustness properties.

5.1 | Data generating distribution

We consider the following data-generating mechanism

$$
\begin{aligned}
L_0 &\sim N(4, 1),\\
A &\sim Bern(0.5),\\
C_1 &\sim Bern\left(\text{expit}(-2 - 0.5A\right),\\
L_1 &\sim N(0.5 + 0.85L_0 + \beta^A_{L_1} A, 1),\\
M_1 &\sim N(2.3 + \beta^A_{M_1} A - 0.2L_1, 1),\\
C_2 &\sim Bern\left(\text{expit}(-2 - 0.5A - 0.05M_1\right),\\
L_2 &\sim N(0.5 + 0.1L_0 + 0.75L_1 + \beta^A_{L_2} A + 0.2M_2, 1),\\
M_2 &\sim N(0.5 + 0.9M_1 + \beta^A_{M_2} A - 0.2L_2, 1),\\
R_Y &\sim Bern\left(\text{expit}(2.5 + 0.2A + 0.1M_2 - 0.1L_2)\right),\\
Y &\sim Bern\left(\text{expit}(-1 + \beta^A_Y A + \beta^M_Y M_2 + \beta^L_Y L_2)\right),
\end{aligned}
$$

where $\text{expit}(x) = 1/(1 + \exp(-x))$.

5.2 | Results

In Table 1 we consider four simulation scenarios. In the first scenario (i) we consider the case where there is no direct effect because exposure has no effect on the covariates or on the outcome, and we set $(\beta^A_{L_1}, \beta^A_{L_2})$ equal to $(0.00, 0.00)$, $(\beta^A_{M_1}, \beta^A_{M_2})$ equal to $(0.75, 1.00)$ and $(\beta^A_Y, \beta^M_Y, \beta^L_Y)$ equal to $(0.00, 0.20, -0.15)$. In scenario (ii) we consider the case where both a direct and an indirect effect is present and we set $(\beta^A_{L_1}, \beta^A_{L_2})$ equal to $(0.75, 1.00)$,$(\beta^A_{M_1}, \beta^A_{M_2})$ equal to $(0.75, 1.00)$ and $(\beta^A_Y, \beta^M_Y, \beta^L_Y)$ equal to $(0.75, 0.20, -0.15)$. In the third scenario (iii) we consider the case where there is no indirect effect and we set $(\beta^A_{L_1}, \beta^A_{L_2})$ equal to $(0.75, 1.00)$ $(\beta^A_{M_1}, \beta^A_{M_2})$ equal to $(0.75, 1.00)$ and $(\beta^A_Y, \beta^M_Y, \beta^L_Y)$ equal to $(0.75, 0.00, -0.15)$.

	n	400					4000				
setting		mean	bias	sd	se	cov	mean	bias	sd	se	cov
	SDE	0.44	0.437	7.02	6.55	94.0	-0.09	-0.088	2.27	2.20	94.6
(i)	SIE	4.67	-0.595	5.38	4.74	92.2	5.23	-0.032	1.76	1.68	93.6
	OE	5.10	-0.158	4.67	4.78	95.3	5.14	-0.119	1.50	1.52	96.2
	SDE	7.52	-0.112	8.25	7.25	90.1	7.54	-0.104	2.50	2.47	94.2
(ii)	SIE	5.56	-0.769	6.45	5.42	89.7	6.26	-0.093	2.04	1.95	92.4
	OE	13.1	-0.881	5.11	5.02	93.3	13.8	-0.197	1.61	1.59	94.3
	SDE	8.28	-0.184	8.13	7.38	90.6	8.44	-0.016	2.52	2.56	95.1
(iii)	SIE	0.39	0.035	6.67	5.64	87.4	-0.45	-0.024	2.16	2.07	94.5
	OE	7.89	-0.149	4.55	4.65	95.6	8.00	-0.040	1.45	1.47	94.9

Table 1 setting refers to the simulation scenarios described in the body of the text; mean is the average estimate across simulations (×100); bias is the average bias across simulations (×100); sd is the standard deviation (×100); se is the average of the estimated influence function based standard error across simulations (×100); cov is the coverage probability of a 95% Wald type confidence interval (×100). Each entry is based on 1000 replicates.

In all four scenarios models for Y, R_Y and C_t were fitted using correctly specified logistic regression models. The regressions in the targeting step in 3b) of the algorithm were fitted with a quasibinomial regression adjusted for all relevant covariates. The conditional density of M_t was estimated from a normal density with homoscedastic variance. The results in Table 1 show that as expected the estimator is unbiased when all models are correctly specified. We see that the coverage is close to 95% for the larger sample size $n = 4000$ and close to 90% for the smaller sample size $n = 400$.

	n	400					4000				
		mean	bias	sd	se	cov	mean	bias	sd	se	cov
	SDE	7.52	-0.113	8.26	7.43	90.7	7.53	-0.117	2.54	2.53	94.3
(a) $Q_L^{a,\hat{\mathbf{g}}^{a'}}$ and p_Y misspec	SIE	5.86	-0.480	6.55	5.62	91.7	6.31	-0.035	2.09	2.02	93.1
	OE	13.4	-0.592	5.10	5.08	94.1	13.8	-0.151	1.61	1.61	94.7
	SDE	7.48	-0.150	8.14	7.18	89.6	7.53	-0.113	2.48	2.45	93.9
(b) $Q_L^{a,\hat{\mathbf{g}}^{a'}}$ and p_{R_Y} misspec	SIE	5.67	-0.670	6.36	5.35	89.6	6.28	-0.073	2.02	1.92	92.2
	OE	13.1	-0.821	5.10	5.06	93.8	13.8	-0.186	1.61	1.60	94.5
	SDE	6.11	-1.53	8.40	9.89	95.0	5.85	-1.80	2.59	3.42	94.9
(c) $Q_L^{a,\hat{\mathbf{g}}^{a'}}$, p_Y and p_{R_Y} misspec	SIE	7.28	0.94	6.88	8.04	94.5	8.02	1.67	2.17	2.91	94.3
	OE	13.4	-0.58	5.11	5.10	94.2	13.9	-0.13	1.62	1.61	95.0

Table 2 mean is the average estimate across simulations (×100); bias is the average bias across simulations (×100); sd is the standard deviation (×100); se is the average of the estimated influence function based standard error across simulations (×100); cov is the coverage probability of a 95% Wald type confidence interval (×100). Each entry is based on 1000 replicates.

In Table 2 we consider the simulation setting as in (iii) where both a direct and indirect effect is present, and we demonstrate the robustness properties of the estimator. In (a) we consider misspecification of the models for Q_L and p_Y, in (b) we consider misspecification of the models for Q_L and p_{R_Y} and in (c) we consider misspecification of the models for Q_L, p_Y and p_{R_Y}. In all cases we misspecify the models by only including a term for A. We see that as expected from the robustness properties in Section 4.2 the estimator remains unbiased in (a)-(b) but is biased in (c).

6 | ANALYSIS OF THE NASH TRIAL

We now apply our proposed method to the NASH clinical trial described in Section 1.1. The potential mediator is weight loss (kg) which was measured at the follow-up visits at weeks 4, 12, 20, 28, 36, 44, 52, 63 and 72. To fulfill the sequential exchangeability assumptions described in Section 3.1 we include the time-varying covariates Alanine Aminotransferase (ALT) and Aspartate Aminotransferase (AST) measured at the same visits as weight loss, Serum Enhanced Liver Fibrosis (ELF) measured at weeks 28, 52 and 72, FIB4 score measured at weeks 4, 12, 28, 36, 52 and 72 and triglycerides (TG) measured at weeks 4, 12, 28, 52 and 72 AST, ALT and TG values were all log transformed. Missing covariate values were imputed by carrying forward last recorded value. Note that there are very few missing values and this is unlikely to affect the results. We included sex, age, baseline weight, diabetes type and baseline fibrosis stage as baseline covariates, as well as baseline values of ALT, AST, ELF, FIB4 and TG. Table 3 shows the number of subjects who are in the study and on treatment across treatment arms at the follow-up visit where weight loss was measured, along with the number of biopsies performed at the end of follow-up.

week	0	4	12	20	28	36	44	52	62	72	biopsy
Placebo	80	80	78	78	75	75	75	73	73	72	67
Sema 0.1 mg OD	80	80	77	76	76	76	76	76	75	74	72
Sema 0.2 mg OD	78	77	75	70	69	68	68	67	67	66	61
Sema 0.3 mg OD	82	81	78	76	75	74	73	73	73	73	66
total	320	318	308	300	295	293	292	289	288	285	266

Table 3 Number of subjects who are in the study and on treatment across treatment arms, along with the number of non-missing biopsies among those that are still in the study and on treatment at the final visit.

Models for p_Y and p_{R_Y} as well as the regression in step 3b) were estimated using an ensemble Super Learner[35,36] which included glm with and without AIC based stepwise covariate selection, bayes glm, generalized additive model, penalized regression and tree based methods. Each algorithm was coupled with both a variable importance-based covariate screener, a correlation-based covariate screener and a coefficient threshold-based covariate screener. Finally p_C was estimated via logistic regression and the conditional density p_M was estimated from a normal density.

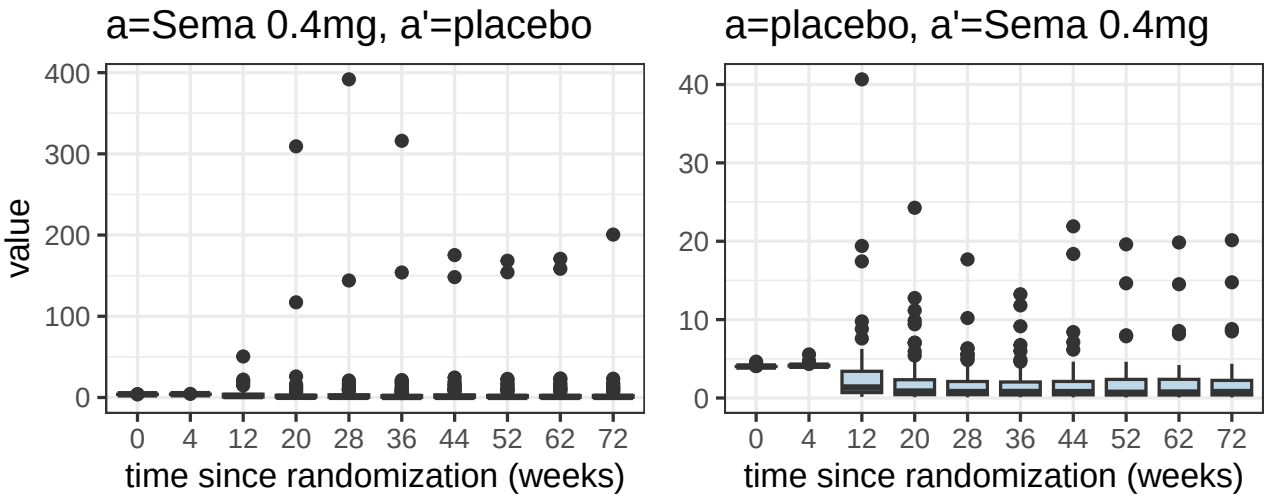


Figure 2 Boxplots of the clever covariates/weights.

The baseline treatment variable has four levels corresponding to placebo, Semaglutide 0.1mg, Semaglutide 0.2mg and Semaglutide 0.4mg respectively. We will focus on the comparison Semaglutide 0.4mg versus placebo. To investigate the positivity assumption in A.3 (iii) we compute the weights/clever covariates in (9) both with a=placebo and a'=Sema 0.4 mg, and with a=Sema 04.mg and a'=placebo. A boxplot of the weights is presented in Figure 2 below. We see that letting a=Sema 04.mg and a'=placebo results in some very extreme outliers in the weights. This was expected since there are no subjects in the placebo group who loose weight resulting in severe violations of assumption in A.3 (iii). Letting a=placebo and a'=Sema 0.4 mg still results in some outliers but not as extreme. This is because there are some subjects in the treatment arm who do not loose weight.

Based on the distribution of the weights we let a=placebo and a'=Sema 0.4 mg when computing the direct and indirect effects. This means that the direct effect can be interpreted as the difference in expected outcomes of assigning treatment to placebo versus Sema 0.4mg while fixing the mediator to a random draw from the observed distribution of weight loss when assigned to placebo. The indirect effect can be interpreted as the difference in expected outcome of assigning the mediator to a random draw from the observed distribution of the weight loss when given placebo versus assigning the mediator to a random draw from the observed distribution of the weight loss when given Sema 0.4mg, while fixing the treatment to placebo.

		est	se	CI lower	CI upper
a=placebo, a'=Sema 0.4 mg OD	SDE	-0.0447	0.0774	-0.196	0.107
	SIE	-0.121	0.0538	-0.226	-0.0152
	OE	-0.165	0.0794	-0.321	-0.00958

Table 4 NASH analysis results. est is the targeted maximim likelihood estimate; se is the influence function based standard error estimate; CI lower and CI upper are the upper and lower bounds respectively of a Wald-type confidence interval.

The results of the analysis are summarized in Table 4. The indirect effect is estimated to be -0.121 (95% CI [-0.226; -0.015]) and the direct effect is estimated to be -0.045 with a confidence interval that includes zero. That is, we find no evidence of a direct effect of Semaglutide on the primary endpoint not mediated through weight loss. We note that the results should only be considered hypothesis generating. Further research is needed to asses the effect of near-violations of the positivity assumption A.3 (iii).

7 | DISCUSSION

In this paper we have proposed class of data-dependent interventional (in)direct effects for estimating the extent to which the effect of a (randomized) baseline treatment on an outcome of interest is mediated through a repeatedly measured continuous covariate. For estimation we proposed a longitudinal targeted minimum loss-based estimation (LTMLE) method based on the sequential regression technique.

We argued that the data-dependent interventional (in)direct effects may sometimes be preferred by researchers because they are identified under weaker assumptions than the interventionist (in)direct effects which assume that the stochastic mediator distribution is the true unknown distribution. Moreover the estimation procedure is simpler than the TMLE algorithm of Zheng & van der Laan (2017)[18] as there is only one targeting step in each iteration, and the clever covariates do not depend on the conditional density of the covariates. This makes the implementation barriers lower.

We applied the method to the NASH clinical trial where we found no evidence of a direct effect of Semaglutide on NASH resolution not mediated through weight loss.

A weakness of the method is that when there no or limited overlap between the conditional mediator distributions in the control and treatment arms the weights for the TMLE algorithm may become very extreme. To address this a possible extension of the method is to consider interventional direct effects where the stochastic mediator distribution g^* marginalizes over the treatment variable. That is

$$g_t^*(M_t \mid \bar{L}_t, \bar{M}_{t-1}) = \sum_a P(M_t \mid A = a, \bar{C}_t = 0, \bar{L}_t, \bar{M}_{t-1}) P(A = a \mid L_0),$$

A further extension of this is to identify the target with most support in the data by choosing g^* to minimize the dissimilarity

$$\text{var}\left\{D^*(P)(a,g^*)(P)(O) - D^*(P)(a,g^{obs})(P)(O)\right\},$$

where g^{obs} is the observed mediator distribution. This will be the topic of future research.

ACKNOWLEDGMENTS

This research was funded by a research gift from Novo Nordisk to the Section of Biostatistics, University of Copenhagen.

How to cite this article: Breum, M. S. and Palle, M. S.,(2023) Estimation of data-dependent stochastic (in)direct effects with a repeatedly measured continuous mediator and missing outcome data, *Statistics in Medicine*, *00:1-15*.

APPENDIX

A IDENTIFICATION

Let $G_k^{a'}$ denote a variable drawn from the known distribution $g_k^{a'}(m_k \mid \bar{l}_k, \bar{m}_{k-1})$. By definition we have

$$E\left\{Y(a,\mathbf{g}^{a'})\right\} = \int_{\mathcal{L}_0} \prod_{k=1}^{K} \int_{\mathcal{L}_k \times \mathcal{M}_k} \Big[E\left\{Y(a,\mathbf{g}^{a'}) \mid \bar{G}_K^{a'} = \bar{m}_K, \bar{L}_K(a,\mathbf{g}^{a'}) = \bar{l}_K\right\} g_k^{a'}(m_k \mid \bar{l}_k, \bar{m}_{k-1})$$
$$\times P(L_k(a,\mathbf{g}^{a'}) = l_k \mid \bar{G}_{k-1}^{a'} = \bar{m}_{k-1}, \bar{L}_{k-1}(a,\mathbf{g}^{a'}) = \bar{l}_{k-1}) p(L_0 = l_0)\Big] \, \mathrm{d}\mu_{M_k}(m_k) \, \mathrm{d}\mu_{L_k}(l_k) \, \mathrm{d}\mu_{L_0}(l_0).$$

We first demonstrate how to identify the conditional probability of $L_k(a,\mathbf{g}^{a'})$. We have

$$\begin{aligned}
&P(L_k(a,\mathbf{g}^{a'}) = l_k \mid L_0 = l_0, G_{k-1}^{a'} = \bar{m}_{k-1}, \bar{L}_{k-1}(a,\mathbf{g}^{a'}) = \bar{l}_{k-1}) \\
&\overset{(i)}{=} P(L_k(a,m_k) = l_k \mid L_0 = l_0, G_{k-1}^{a'} = \bar{m}_{k-1}, \bar{L}_{k-1}(a,\mathbf{g}^{a'}) = \bar{l}_{k-1}) \\
&\overset{(ii)}{=} P(L_k(a,m_k) = l_k \mid L_0 = l_0, A = a, G_{k-1}^{a'} = \bar{m}_{k-1}, \bar{L}_{k-1}(a,\mathbf{g}^{a'}) = \bar{l}_{k-1}) \\
&\overset{(iii)}{=} P(L_k(a,m_k) = l_k \mid L_0 = l_0, A = a, C_1 = 0, G_{k-1}^{a'} = \bar{m}_{k-1}, \bar{L}_{k-1}(a,\mathbf{g}^{a'}) = \bar{l}_{k-1}) \\
&\overset{(iv)}{=} P(L_k(a,m_k) = l_k \mid L_0 = l_0, A = a, C_1 = 0, G_{2:k-1}^{a'} = \bar{m}_{2:k-1}, \bar{L}_{k-1}(a,\mathbf{g}^{a'}) = \bar{l}_{k-1}) \\
&\overset{(v)}{=} P(L_k(a,m_k) = l_k \mid L_0 = l_0, A = a, C_1 = 0, M_1 = m_1, G_{2:k-1}^{a'} = \bar{m}_{2:k-1}, L_1 = l_1, \bar{L}_{2:k-1}(a,\mathbf{g}^{a'}) = \bar{l}_{2:k-1}) \\
&= P(L_k = l_k \mid L_0 = l_0, A = a, \bar{C}_k = \bar{0}, \bar{M}_{k-1} = \bar{m}_{k-1}, \bar{L}_{k-1} = \bar{l}_{k-1}),
\end{aligned}$$

where the first equality follows from the definition of $L(a,\mathbf{g}^{a'})$, the second equality follows from (A.0) and the third equality from (A.1). The fourth equality follows because by definition $L_k(a,m_k) \perp\!\!\!\perp G_1^{a'} \mid L_0 = l_0, L_1(a) = l_1$ and the fifth equality follows from (A.2). The last equality follows from iteratively repeating the arguments in steps (iii)-(v).

Analogously assumptions (A.0)-(A.2) imply that the conditional expectation of $Y(a,\mathbf{g}^{a'})$ identifies to

$$E\left\{Y(a,\mathbf{g}^{a'}) \mid \bar{G}_K^{a'} = \bar{m}_K, \bar{L}_K(a,\mathbf{g}^{a'}) = \bar{l}_K\right\} = E\left\{Y \mid A = a, \bar{C}_K = \bar{0}, \bar{M}_K = \bar{m}_K, \bar{L}_K = \bar{l}_K\right\}.$$

B DERIVATION OF THE EFFICIENT INFLUENCE FUNCTION

There are several ways to derive efficient influence functions. We here use the Gâteaux derivative approach[37,38]. Specifically we will compute the so-called Gâteaux derivative[39] of the target parameter in the direction of a point mass at single observation $\tilde{o}$, defined as

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} \psi\left(P_\varepsilon\right),$$

where $P_\varepsilon = (1-\varepsilon)P + \varepsilon\delta_{\tilde{o}}$ and $\delta_{\tilde{o}}$ is the Dirac measure at $O = \tilde{o}$.

For simplicity we derive the efficient influence function for the special case of $K = 2$ and we do the calcultions assuming all variables are continuous. It follows from the chain rule that

$$
\begin{aligned}
\frac{d}{d\varepsilon}\Big|_{\varepsilon=0}\Psi(P_\varepsilon)(a,\mathbf{g}^{a'}) =& E\left(\int_{\mathcal{M}_{\bar{1}}} E\left[\int_{\mathcal{M}_2}\left\{\int_{\mathcal{Y}} y\frac{d}{d\varepsilon}\Big|_{\varepsilon=0} p_{Y,\varepsilon}(y \mid a,\bar{0},\bar{m}_2,\bar{L}_2,R_Y=1)\,\mathrm{d}\mu_Y(y)\right\} g_2^{a'}(m_2 \mid \bar{L}_2,m_1)\,\mathrm{d}\mu_{M_2}(m_2)\,\middle|\, A=a,\bar{L}_1,m_1\right] g_1^{a'}(m_1 \mid \bar{L}_1)\,\mathrm{d}\mu_{M_1}(m_1)\right) \\
&+ E\left[\int_{\mathcal{M}_1}\left\{\int_{\mathcal{L}_2} Q_{M_2}^{a,\mathbf{g}^{a'}}(l_2,\bar{L}_1,M_1)\frac{d}{d\varepsilon}\Big|_{\varepsilon=0} p_{L_2,\varepsilon}(l_2 \mid a,\bar{L}_1,M_1)\,\mathrm{d}\mu_{L_2}(l_2)\right\} g_1^{a'}(m_1 \mid \bar{L}_1)\,\mathrm{d}\mu_{M_1}(m_1)\right] \\
&+ E\left\{\int_{\mathcal{L}_1} Q_{M_1}^{a,\mathbf{g}^{a'}}(l_1,L_0)\frac{d}{d\varepsilon}\Big|_{\varepsilon=0} p_{L_1,\varepsilon}(l_1 \mid a,\bar{L}_0)\,\mathrm{d}\mu_{L_1}(l_1)\right\} \\
&+ \int_{\mathcal{L}_0} Q_{L_1}(l_0)\frac{d}{d\varepsilon}\Big|_{\varepsilon=0} p_{L_0,\varepsilon}(l_0)\,\mathrm{d}\mu_{L_0}(l_0) \\
=& \frac{\delta_{\tilde{a}}(a)}{p_A(a \mid \tilde{l}_0)}\frac{\delta_{\tilde{r}_y}(1)}{p_{R_Y}(1 \mid \tilde{l}_{0:2},\tilde{m}_{1:2})}\prod_{k=1}^{2}\frac{\delta_{\tilde{c}_k}(0)}{p_{C_k}(0 \mid \tilde{l}_{0:k-1},\tilde{m}_{1:k-1})}\frac{g_k^{a'}(\tilde{m}_k \mid \tilde{l}_{0:k},\tilde{m}_{1:k-1})}{p_{M_k}(\tilde{m}_k \mid \tilde{a},\bar{0},\tilde{l}_{0:k},\tilde{m}_{1:k-1})}\left\{\tilde{y} - Q_Y^{a,g^{a'}}(\tilde{m}_{1:2},\tilde{l}_{0:2})\right\} \\
&+ \frac{\delta_{\tilde{a}}(a)}{p_A(a \mid \tilde{l}_0)}\frac{g_1^{a'}(\tilde{m}_1 \mid \tilde{l}_{0:1})}{p_{M_1}(\tilde{m}_1 \mid \tilde{a},0,\tilde{l}_{0:1})}\prod_{k=1}^{2}\frac{\delta_{\tilde{c}_k}(0)}{p_{C_k}(0 \mid \tilde{l}_{0:k-1},\tilde{m}_{1:k-1})}\left\{Q_{M_2}^{a,g^{a'}}(\tilde{m}_1,\tilde{l}_{0:2}) - Q_{L_2}^{a,g^{a'}}(\tilde{m}_1,\tilde{l}_{0:1})\right\} \\
&+ \frac{\delta_{\tilde{a}}(a)}{p_A(a \mid \tilde{l}_0)}\frac{\delta_{\tilde{c}_1}(0)}{p_{C_1}(0 \mid \tilde{l}_0)}\left\{Q_{M_1}^{a,g^{a'}}(\tilde{l}_{0:1}) - Q_{L_1}^{a,g^{a'}}(\tilde{l}_0)\right\} \\
&+ Q_{L_1}^{a,g^{a'}}(\tilde{l}_0) - \int Q_{L_1}^{a,g^{a'}}(l_0)p_{L_0}(l_0)\,\mathrm{d}\mu_{L_0}(l_0)
\end{aligned}
$$

Since A, R_Y and C_k are discrete we can replace the dirac delta functions with the indicator functions to obtain the efficient influence function

$$
\begin{aligned}
D^*(P)(O) =& \frac{I(A=a)}{p_A(a \mid L_0)}\frac{I(R_Y=1)}{p_{R_Y}(1 \mid a,0,\bar{L}_2,\bar{M}_2)}\prod_{k=1}^{2}\frac{I(C_k=0)}{\delta_k(a,\bar{L}_{k-1},\bar{M}_{k-1})}\frac{g_k^{a'}(\tilde{M}_k \mid \bar{L}_k,\bar{M}_{k-1})}{p_{M_k}(M_k \mid a,\bar{0},\bar{L}_k,\bar{M}_{k-1})}\left\{Y - Q_Y^{a,g^{a'}}(\bar{M}_2,\bar{L}_2)\right\} \\
=& \frac{I(A=a)}{p_A(a \mid L_0)}\frac{g_1^{a'}(M_1 \mid \bar{L}_1)}{p_{M_1}(M_1 \mid L_0,a,0,L_1)}\prod_{k=1}^{2}\frac{I(C_k=0)}{\delta_k(a,\bar{L}_{k-1},\bar{M}_{k-1})}\left\{Q_{M_2}^{a,g^{a'}}(M_1,\bar{L}_2) - Q_{L_2}^{a,g^{a'}}(M_1,\bar{L}_1)\right\} \\
&+ \frac{I(A=a)}{p_A(a \mid L_0)}\frac{I(C_1=0)}{\delta_1(a,L_0)}\left\{Q_{M_1}^{a,g^{a'}}(\bar{L}_1) - Q_{L_1}^{a,g^{a'}}(L_0)\right\} \\
&+ Q_{L_1}^{a,g^{a'}}(L_0) - \Psi(P)(a,\hat{\mathbf{g}}^{a'}).
\end{aligned} \tag{B1}
$$

The generalization to arbitrary K follows immediately.

C MULTIPLE ROBUSTNESS

We here prove the robustness conditions stated in Section 4.2. We again consider the case of $K = 2$ where the influence function can be written as in (B1).

(i)

Suppose p_L and p_Y are correctly specified. Then it follows by a simple application of iterated expectations that

$$E\left\{D^*(P)(O)\right\} = \Psi(P_0)(a,\hat{\mathbf{g}}^{a'}) - \Psi(P)(a,\hat{\mathbf{g}}^{a'})$$

(ii)

Suppose p_Y δ and p_M are correctly specified. Then

$$\begin{aligned}
E\{D^*(P)(O)\} =& E\left(E\left[\frac{g_1^{a'}(M_1 \mid \bar{L}_1)}{p_{0,M_1}(M_1 \mid a, \bar{0}, \bar{L}_1)} E\left\{Q_{M_2}^{a,g^{a'}}(M_1, \bar{L}_2) \,\middle|\, A = a, \bar{C}_2 = 0, \bar{L}_1, \bar{M}_1\right\} \mid A = a, C_1 = 0, \bar{L}_1\right]\right) \\
&- E\left[E\left\{\frac{g_1^{a'}(M_1 \mid \bar{L}_1)}{p_{0,M_1}(M_1 \mid a, \bar{0}, \bar{L}_1)} Q_{L_2}^{a,g^{a'}}(M_1, \bar{L}_1) \,\middle|\, A = a, C_1 = 0, \bar{L}_1\right\}\right] \\
&+ E\left[E\left\{Q_{M_1}^{a,g^{a'}}(\bar{L}_1) \,\middle|\, A = a, C_1 = 0, L_0\right\}\right] - E\left\{Q_{L_1}^{a,g^{a'}}(L_0)\right\} \\
&+ E\left\{Q_{L_1}^{a,g^{a'}}(L_0)\right\} - \Psi(P)(a, \hat{\mathbf{g}}^{a'}) \\
=& \Psi(P_0)(a, \hat{\mathbf{g}}^{a'}) - E\left[E\left\{\frac{g_1^{a'}(M_1 \mid \bar{L}_1)}{p_{0,M_1}(M_1 \mid a, \bar{0}, \bar{L}_1)} Q_{L_2}^{a,g^{a'}}(M_1, \bar{L}_1) \,\middle|\, A = a, C_1 = 0, \bar{L}_1\right\}\right] \\
&+ E\left[E\left\{Q_{M_1}^{a,g^{a'}}(\bar{L}_1) \,\middle|\, A = a, C_1 = 0, L_0\right\}\right] - E\left\{Q_{L_1}^{a,g^{a'}}(L_0)\right\} \\
&+ E\left\{Q_{L_1}^{a,g^{a'}}(L_0)\right\} - \Psi(P)(a, \hat{\mathbf{g}}^{a'}) \\
=& \Psi(P_0)(a, \hat{\mathbf{g}}^{a'}) - \Psi(P)(a, \hat{\mathbf{g}}^{a'}).
\end{aligned}$$

(iii)

Suppose p_{R_Y} δ and p_M are correctly specified. Then we again get a telescoping sum

$$\begin{aligned}
E\{D^*(P)(O)\} =& \Psi(P_0)(a, \hat{\mathbf{g}}^{a'}) \\
&- E\left(E\left[\frac{g_1^{a'}(M_1 \mid \bar{L}_1)}{p_{0,M_1}(M_1 \mid a, \bar{0}, \bar{L}_1)} E\left\{\frac{g_2^{a'}(M_2 \mid \bar{L}_2, M_1)}{p_{0,M_2}(M_2 \mid a, \bar{0}, \bar{L}_2, M_1)} Q_Y^{a,g^{a'}}(\bar{M}_2, \bar{L}_2) \,\middle|\, A = a, \bar{C}_2 = 0, M_1, \bar{L}_2\right\} \,\middle|\, A = a, C_1 = 0, \bar{L}_1\right]\right) \\
&+ E\left(E\left[\frac{g_1^{a'}(M_1 \mid \bar{L}_1)}{p_{0,M_1}(M_1 \mid a, \bar{0}, \bar{L}_1)} E\left\{Q_{M_2}^{a,g^{a'}}(M_1, \bar{L}_2) \,\middle|\, A = a, \bar{C}_2 = 0, \bar{L}_1, \bar{M}_1\right\} \mid A = a, C_1 = 0, \bar{L}_1\right]\right) \\
&- E\left[E\left\{\frac{g_1^{a'}(M_1 \mid \bar{L}_1)}{p_{0,M_1}(M_1 \mid a, \bar{0}, \bar{L}_1)} Q_{L_2}^{a,g^{a'}}(M_1, \bar{L}_1) \,\middle|\, A = a, C_1 = 0, \bar{L}_1\right\}\right] \\
&+ E\left[E\left\{Q_{M_1}^{a,g^{a'}}(\bar{L}_1) \,\middle|\, A = a, C_1 = 0, L_0\right\}\right] - E\left\{Q_{L_1}^{a,g^{a'}}(L_0)\right\} \\
&+ E\left\{Q_{L_1}^{a,g^{a'}}(L_0)\right\} - \Psi(P)(a, \hat{\mathbf{g}}^{a'}) \\
=& \Psi(P_0)(a, \hat{\mathbf{g}}^{a'}) - \Psi(P)(a, \hat{\mathbf{g}}^{a'}).
\end{aligned}$$

D ADDITIONAL SIMULATIONS

In Table D1 below we report the results from a simulation study investigating the effect of the number of cutpoints used for the numerical integration in step 3a) of the algorithm in Section 4. The simulation setting is identical to setting (iii) of Section 5 and we report only estimates of $\Psi(1, \hat{g}^0)(P)$.

n	400				4000			
bins	bias	sd	se	cov	bias	sd	se	cov
10	-0.775	6.91	6.18	88.2	-0.194	2.27	2.22	92.4
20	-0.619	6.91	6.18	88.6	-0.144	2.27	2.22	92.7
40	-0.620	6.89	6.18	88.7	-0.144	2.27	2.22	92.8
80	-0.634	6.89	6.18	88.6	-0.146	2.27	2.22	92.8
160	-0.642	6.88	6.18	88.6	-0.148	2.27	2.22	92.8

Table D1 Estimates of $\Psi(1, \hat{g}^0)(P)$. bins is the number of cutpoints used for the numerical integration; bias is the average bias across simulations (×100); sd is the standard deviation (×100); se is the average of the estimated standard error across simulations (×100); cov is the coverage propbability of a 95% Wald type confidence interval (×100). Each entry is based on 1000 replicates.

References

1. VanderWeele TJ. *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press . 2015.
2. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 1992; 3: 143-155.
3. Pearl J. Direct and Indirect Effects. In: Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers Inc; 2001: 411-420.
4. Avin C, Shpitser I, Pearl J. Identifiability of path-specific effects. In: Proceedings of the International Joint Conference on Artificial Intelligence. Morgan Kaufmann Publishers Inc; 2005: 357-363.
5. VanderWeele TJ, Vansteelandt S, Robins JM. Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology* 2014; 25(2): 300-6.
6. Robins JM, Richardson TS. Alternative graphical causal models and the identification of direct effects. In: Shrout PE, Keyes KM, Ornstein K., eds. *Causality and psychopathology: Finding the determinants of disorders and their cures*Oxford University Press. 2011 (pp. 103–158).
7. Cai Z, Kuroki M, Pearl J, Tian J. Bounds on direct effects in the presence of confounded intermediate variables. *Biometrics* 2008; 64(3): 695-701.
8. Kaufman S, Kaufman J, R.F. M, Greenland S, Poole C. Improved estimation of controlled direct effects in the presence of unmeasured confounding of intermediate variables. *Statistics in Medicine* 2005; 24(11): 1683-1702.
9. Tchetgen Tchetgen E, Phiri K. Bounds for pure direct effect. *Epidemiology* 2014; 25(5): 775-776.
10. Miles C, Kanki P, Meloni E. On partial identification of the pure direct effect. arXiv preprint arXiv:1509.01652; 2021.
11. van der Laan MJ, Petersen ML. Direct effect models. *The International Journal of Biostatistics* 2008; 4(1): Article 23.
12. Didelez V, Dawid AP, Geneletti S. Direct and indirect effects of sequential treatments. In: AUAI Press; 2006: 138-146.
13. Geneletti S. Identifying direct and indirect effects in a non-counterfactual framework. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 2007; 69(2): 199-215.
14. Rudolph KE, Sofrygin O, Zheng W, van der Laan MJ. Robust and Flexible Estimation of Stochastic Mediation Effects: A Proposed Method and Example in a Randomized Trial Setting. *Epidemiologic Methods* 2018; 1(7).
15. Robins JM, Richardson TS, Shpitser I. An Interventionist Approach to Mediation Analysis. In: Geffner G, Dechter R, Halpern JY., eds. *Probabilistic and Causal Inference: The Works of Judea Pearl*New York: Association for Computing Machinery. 1 ed. 2022 (pp. 713–764).
16. Stensrud MJ, Young JG, Didelez V, Robins JM, Hernán. M. Separable Effects for Causal Inference in the Presence of Competing Events. *Journal of the American Statistical Association* 2022; 117(537): 175–183.

17. Vansteelandt S, Daniel RM. Interventional effects for mediation analysis with multiple mediators. *Epidemiology* 2017; 28(2): 258-265.

18. Zheng W, van der Laan MJ. Longitudinal Mediation Analysis with Time-varying Mediators and Exposures, with Application to Survival Outcomes. *Journal of Causal Inference* 2017; 5(2).

19. VanderWeele TJ, Tchetgen Tchetgen EJ. Mediation analysis with time varying exposures and mediators. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 2017; 79(3): 917-938.

20. van der Laan MJ, Rubin D. Targeted Maximum Likelihood Learning. *International Journal of Biostatistics* 2006; 2(1).

21. van der Laan MJ, Gruber S. Targeted minimum loss based estimation of causal effects of multiple time point interventions. *International Journal of Biostatistics* 2012; 8(1).

22. Bang H, Robins JM. Doubly Robust Estimation in Missing Data and Causal Inference Models. *Biometrics* 2005; 61(4): 962-973.

23. van der Laan MJ, Rose S. *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer Science & Business Media . 2011.

24. van der Laan MJ, Rose S. *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies*. Springer Science & Business Media . 2018.

25. Than NN, Newsome PN. A concise review of non-alcoholic fatty liver disease. *Atherosclerosis* 2015; 239(1): 192-202.

26. Williams C, Stengel J, Asike M, Torres D, Shaw J, Contreras M. Prevalence of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis among a largely middleaged population utilizing ultrasound and liver biopsy: a prospective study. . *Gastroenterology* 2011; 140(1): 124-131.

27. Adams L, Lymp J, St Sauver J, Sanderson S, Lindor K, Feldstein A. The natural history of nonalcoholic fatty liver disease: a population-based cohort study. *Gastroenterology* 2005; 129(1): 113-21.

28. Shaker M, Tabbaa A, Albeldawi M, Alkhouri N. Liver transplantation for nonalcoholic fatty liver disease: new challenges and new opportunities. *World J Gastroenterol.* 2014; 20(18): 5320-30.

29. Charlton M, Burns PR, Watt K, Heimbach J, Dierkhising R. Frequency and outcomes of liver transplantation for nonalcoholic steatohepatitis in the United States. *Gastroenterology* 2011; 141(4): 1249-53.

30. Merola J, Liapakis A, Mulligan D, Yoo P. Non-alcoholic fatty liver disease following liver transplantation: a clinical review. *Clin Transplant.* 2015; 29(9): 728-37.

31. Chalasan N, Younossi Z, Lavine J, Diehl A, Brunt E, Cusi K. The diagnosis and management of non-alcoholic fatty liver disease: practice Guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association. *Hepatology* 2012; 55(6): 2005-23.

32. Dowman TJNP. Systematic review: the diagnosis and staging of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis. *Aliment Pharmacol Ther.* 2011; 33(5): 525-540.

33. Pearl J. *Causality*. Cambridge University Press . 2000.

34. Petersen M, Schwab J, Gruber S, Blaser N, Schomaker M, van der Laan M. Targeted maximum likelihood estimation for dynamic and static longitudinal marginal structural working models. *Journal of Causal Inference* 2014; 2(2): 147-185.

35. van der Laan MJ, Dudoit S. Unified cross-validation methodology for selection among estimators and a general cross-validated adaptive epsilon-net estimator: Finite sample oracle inequalities and examples.. 2003.

36. van der Laan MJ, Polley EC, Hubbard AE. Super Learner. *Stat Appl Genet Mol Biol.* 2007; 6: Article 25.

37. Hines O, Dukes O, Diaz-Ordaz K, Vansteelandt S. Demystifying Statistical Learning Based on Efficient Influence Functions. *The American Statistician* 2022; 76(3): 292–304.

38. Kennedy EH. Semiparametric doubly robust targeted double machine learning: a review. arXiv preprint arXiv:2203.06469; 2023.

39. van der Vaart AW. *Asymptotic Statistics*. Cambridge University Press. 3 ed. 2000.

Manuscript III

Efficient nonparametric estimators of discrimination measures with censored survival data

Marie Skov Breum and Torben Martinussen

Details:
Under preparation

Efficient nonparametric estimators of discrimination measures with censored survival data

Marie Breum[1] and Torben Martinussen[1]

[1] Section of Biostatistics, Department of Public Health, University of Copenhagen

Abstract

Both the concordance index (C-index) and the concordance probability of Gönen and Heller are widely used in the medical literature for evaluating predictive accuracy of survival regression models. In this paper we propose novel estimators of these two discrimination measures based on the efficient influence functions. The proposed estimators are non-parametric and locally efficient, and allow for covariate-dependent censoring. Another much used discrimination measure is the t-year area under the receiver operating characteristic (ROC) curve (AUC_t). Since the AUC_t is proper, as opposed to the C-index, it has been recommended as a better discrimination measure than the C-index if t-years risk prediction is the main goal rather than a global assessment of the fitted model. We also develop a novel estimator for the AUC_t with similar desirable properties. We verify the theoretical properties of the proposed estimators in a simulation study, and illustrate the method by an application to data from the German Breast Cancer Study Group.

Key words: concordance probability, efficient influence function, predictive accuracy, right-censoring, survival data.

1 Introduction

A popular tool in the medical literature for evaluating the predictive accuracy of a time-to-event model is to look at concordance measures such as the C-index [Harrell et al., 1982, Harrell Jr et al., 1996, Heagerty and Zheng, 2005, Uno et al., 2011] or the concordance probability of Gönen and Heller [2005]. These measures are used to quantify how well an estimated risk score discriminates between subjects with different event times. Considering two random subjects (i, j) from the population of interest the C-index is informally defined as

$$\mathcal{C} = Pr(risk(i) \geq risk(j) \mid i \text{ has event before } j), \tag{1}$$

while the concordance probability of Gönen and Heller [2005] is informally defined as

$$K = Pr(i \text{ has event before } j \mid risk(i) \geq risk(j)). \tag{2}$$

The C-index and the concordance probability both range from 0.5 to 1 and share the interpretation that a value of 1 represents a model that perfectly discriminates between subjects with different event times while a value of 0.5 represents a model with no discriminatory power. Note than when the event time and risk score are both continuous and we do not truncate the event-time then the measures are identical. If, as is commonly done, we restrict to a specific time horizon (i.e we only observe up to a given point in time) then two measures are proportional.

An early estimator of the C-index for survival data was proposed by Harrell Jr et al. [1996]. However as noted by Gönen and Heller [2005] and Uno et al. [2011] the limiting value of Harell's C-index depends on the censoring distribution. To address this limitation Uno et al. [2011] develop an ICPW estimator assuming independent censoring which was extended by Gerds et al. [2013] to allow for covariate-dependent censoring. Hartman et al. [2023b] similarly note that the limiting value of the c-index depends on the distribution of the truncation times and propose an IPW estimator for left-truncated data. Gönen and Heller [2005] propose a plug-in estimator of the concordance probability assuming that data are generated by is a proportional hazards model. If this is not the case then their

estimator will be biased, and moreover it depends on the censoring distribution in the specific study which clearly is of no scientific interest.

The contribution of this paper is that we propose an estimator of the concordance probability that does not require data to be generated from a Cox model. This is done by viewing the concordance probability as an estimand and then deriving the corresponding efficient influence function. The so-called one-step estimator based on this efficient influence function has a term identically to the estimator proposed by Gönen and Heller [2005] but also a debiasing term reflecting the fact that data might not be generated from a Cox model. The estimator we propose makes efficient use of data and at the same time possess some desirable robustness properties. In the same vein we also propose a novel estimator of the C-index with the same desirable properties.

Recently, it was pointed out by [Blanche et al., 2019] that the C-index is not proper for the evaluation of t-year predicted risks. They recommended instead to look at the so-called cumulative-dynamic time-dependent area under the ROC-curve (AUC_t) [Heagerty et al., 2000, Heagerty and Zheng, 2005, Blanche et al., 2013] that is proper in this scenario. The AUC_t is informally defined as

$$\text{AUC}_t = Pr(risk(i) \geq risk(j) \mid i \text{ has event before } t \text{ and } j \text{ has event after } t). \tag{3}$$

For this estimand, we derive similarly its corresponding efficient influence function and exploit this to propose a novel estimator of the AUC_t.

The three discrimination measures mentioned above all rely on a given scoring rule, which we denote by $\beta^T X$ with X being the available markers that we wish to use to form the scoring rule. It turns out to be important how the coefficient vector β is chosen. We use a well defined estimand that does not rely on data being generated from any specific model such as the Cox-model. To estimate β we first calculate its efficient influence function as use the corresponding one-step estimator. This turns out to be crucial in order to obtain simple estimators for the three discrimination measures of interest.

This paper is organized as follows. In Section 2 we define the concordance measures more formally. In Section 3 we propose locally efficient non-parametric estimators of the truncated versions of K and C based on the efficient influence function and we discuss estimation of $\beta(P)$. In Section 4 we define AUC_t more formally and we propose an estimator based on the efficient influence function. In Section 5 we conduct a simulation study to illustrate the large sample properties of the estimators and in Section 6 we apply the proposed novel estimators to data from the German Breast Cancer Study Group (GBCSG). Some final remarks are provided in Section 7.

2 Concordance measures

Let T denote the continuous failure time and X the d-dimensional vector of markers. Let C denote the censoring time so that we observe $\tilde{T} = T \wedge C$ and $\Delta = I(T \leq C)$, and assume that T and C are conditionally independent given X. Let $O = (\tilde{T}, \Delta, X)$ and $Z = (T, X)$ with the latter corresponding to the case where there is no censoring which we informally refer to as full data. We observe data in the time interval $[0, \tau]$ with $\tau < \infty$. We let $N^o(t) = I(\tilde{T} \leq t, \Delta = 1)$ denote the counting process that jumps when an event time of interest is observed, and also define $N(t) = I(T \leq t)$ corresponding to the counting process in the case without censoring. Let $\beta(P)$ be some estimand and define the scoring rule

$$Y = \beta^T X,$$

that we wish to evaluate in terms of predictive power. Here β is short for $\beta(P)$. We return to how $\beta(P)$ can be chosen. We further assume that the covariate vector X contains at least one continuous covariate that we call W, and we write $X = (V, W)$ with V denoting the remaining covariates. Also,

$$\beta^T X = \beta_V^T V + \beta_W W$$

and for now assume $\beta_W > 0$. The concordance probability of Gönen and Heller [2005] is

$$K_\tau(P) = P(T_2 > T_1, T_1 \leq \tau \mid Y_1 \geq Y_2), \tag{4}$$

where T_j is shorthand notation for a draw of the event time from the population where we restrict to Y_j. In Gönen and Heller [2005], the truncation by τ is left out and the resulting concordance probability is

calculated assuming the data are generated by a Cox-model which further makes estimation possible. However, without the finite maximum follow-up time τ, the concordance probability is not in general identifiable. The concordance index, also known as the C-index [Harrell et al., 1982, Harrell Jr et al., 1996, Pencina and D'agostino, 2004], for survival data is in the truncated form given by

$$C_\tau(P) = P(Y_1 \geq Y_2 \mid T_2 > T_1, T_1 \leq \tau), \tag{5}$$

see [Uno et al., 2011, Heagerty and Zheng, 2005]. Clearly, these two concordance measures are proportional to

$$\Psi_\tau(P) = P(T_2 > T_1, T_1 \leq \tau, Y_1 \geq Y_2), \tag{6}$$

in the considered setting, where both Y and T are continuous. Specifically,

$$K_\tau(P) = 2\Psi_\tau(P) \quad \text{and} \quad C_\tau(P) = 2\Psi_\tau(P)/\left\{1 - S^2(\tau)\right\}.$$

In the following we develop doubly robust nonparametric estimators of both $K_\tau(P)$ and $C_\tau(P)$ mitigated by calculating the efficient influence function of $\Psi_\tau(P)$. We will write K_τ for $K_\tau(P)$ and likewise for the other esitmands unless we wish to stress the dependence on P.

3 Doubly robust nonparametric estimators of K_τ and C_τ

The key to develop the novel estimators of the two discrimination measures is to calculate the efficient influence function (EIF) of $\Psi_\tau(P)$ without any reference to a specific statistical model such as the Cox model. For $y_j = \beta^T x_j$, $j = 1, 2$, define

$$h_\tau(y_1, y_2) = \int_0^\tau S(t|y_2) S(t|y_1) \lambda(t|y_1) dt, \tag{7}$$

where $S(t|y) = P(T > t|Y = y)$ and $\lambda(t|y)$ is the corresponding conditional hazard function. Thus,

$$\Psi_\tau = \int\int_{y_1 > y_2} h_\tau(y_1, y_2) dF_Y(y_1) dF_Y(y_2),$$

where F_Y denotes the distribution function for Y. We show in the Appendix that the efficient influence function corresponding to Ψ_τ and based on full data is

$$D^*_{\Psi_\tau}(Z, P) = g_\tau(X; P) + G_\tau(Z, \dot\beta, P), \tag{8}$$

where

$$g_\tau(X; P) = \int I(y > Y) h_\tau(Y, y) dF_Y(y) + \int I(y < Y) h_\tau(y, Y) dF_Y(y) - 2\Psi_\tau$$

and

$$\begin{aligned} G_\tau(Z, \dot\beta, P) = &\int_Y^\infty \int_0^\tau S(t|Y) \int_0^t \frac{dM(u|Y)}{S(u|Y)} S(t|y) d\Lambda(t|y) dF_Y(y) \\ &+ \int_{-\infty}^Y \left[\int_0^\tau S(t|y) dM(t|Y) - \int_0^\tau S(t|y) S(t|Y) \int_0^t \frac{dM(u|Y)}{S(u|Y)} d\Lambda(t|Y) \right] dF_Y(y) \\ &+ \tilde H(\beta) \dot\beta(P), \end{aligned}$$

with $\tilde H(\beta)$ a constant (given in the Appendix) and $\dot\beta$ is the efficient influence function of β. Further, $dM(t|Y)$ denotes the martingale (increment) corresponding to the full data counting process $N(t)$ and conditioning on Y, and $\Lambda(t|y) = \int_0^t \lambda(u|y) du$. The only term in the full data EIF that is affected when moving to the observed data case is the term $G_\tau(Z, \dot\beta, P)$. If we were willing to assume conditionally independent censoring given Y then we get the wanted observed data $G_\tau(O, \dot\beta, P)$ by replacing $dM(\cdot|Y)$ in $G_\tau(Z, \dot\beta, P)$ with $dM^o(\cdot|Y)/K_C(\cdot|Y)$ where $dM^o(t|X)$ denotes the martingale (increment) corresponding to the observed counting process $N^o(t)$ and conditioning on Y, and with

$K_C(u \mid Y) = P(C > u|Y)$. Under the earlier stated conditionally independent censoring assumption given X we need first to rewrite

$$dM(v|Y) = dM(v|X) + S(v|X)d\{\Lambda(v|X) - \Lambda(v|Y)\} - S(v|X)\int_0^v \frac{dM(r|X)}{S(r|X)}d\{\Lambda(v|X) - \Lambda(v|Y)\} \quad (9)$$

We then obtain $G_\tau(O, \dot{\beta}, P)$ by replacing $dM(\cdot|X)$ in $G_\tau(Z, \dot{\beta}, P)$ with $dM^o(\cdot|X)/K_C(\cdot|X)$. The explicit expression for $G_\tau(O, \dot{\beta}, P)$ is given in the Appendix. Finally, this gives us the wanted efficient influence function based on the observed data

$$D^*_{\Psi_\tau}(O, P) = g_\tau(X) + G_\tau(O, \dot{\beta}, P). \quad (10)$$

3.1 One-step estimator

The one-step estimator of Ψ_τ is obtained by setting the empirical mean of $D^*_{\Psi_\tau}(O, P)$ equal to zero and solve for Ψ_τ. This will obviously depend on unknown quantities, however. To this end we replace β with some estimator $\hat{\beta}$, which we give in a moment, and F_Y is replaced with its empirical counterpart. Since $K_\tau = 2\Psi_\tau$, this results in the following one-step estimator of the concordance probability K_τ

$$\hat{K}_\tau = \frac{2}{n(n-1)}\sum_{i\neq j} I\left(\hat{\beta}^T X_i > \hat{\beta}^T X_j\right) h_\tau\left(\hat{\beta}^T X_j, \hat{\beta}^T X_i\right) + \frac{1}{n}\sum_{i=1}^n G_\tau(O_i, \dot{\beta}, \hat{P}). \quad (11)$$

It is interesting to note that the first term in (11) corresponds to the estimator of Gönen and Heller [2005] (GH-estimator) had we used the Cox partial likelihood estimator of β. The remaining part of (11) is a debiasing term in case the data are not generated from a Cox model (given X). Note also that we have used the notation $\hat{P}$ to indicate that we further need to replace all other unknown quantities, such as $S(\cdot|X)$, by working estimates. We will be more specific about this later but leave a bit vague for now. If data were in fact generated by a Cox model and we use the Cox partial likelihood estimator of β then the debiasing term is negligble. In the more likely case where data are not generated by the Cox model the above one-step estimator can in principle be used to correct the bias of the GH-estimator. However, we do not recommend to use this procedure as it would require estimation of the constant $\tilde{H}(\beta)$, which is not attractive due to the complicated structure of the constant. However, this can be avoided all together if we use the one-step estimator that solves the (empirical) version of the eif corresponding to the estimand $\beta(P)$. Put in other words, by doing so, the term $\tilde{H}(\beta)\dot{\beta}(P)$ can be dropped from the above eif. Since $C_\tau = K_\tau/\left\{1 - S^2(\tau)\right\}$ this suggest the estimator

$$\hat{C}_\tau = \hat{K}_\tau/\left[1 - \{\hat{S}(\tau)\}^2\right], \quad (12)$$

where $\hat{S}(\tau)$ is its corresponding one-step estimator based on the observed data:

$$\hat{S}(t) = n^{-1}\sum_i S_n(t|X_i)\left\{1 - \int_0^t \frac{dM_n^o(u|X_i)}{S_n(u|X_i)K_C^n(u|X_i)}\right\}, \quad (13)$$

evaluated at $t = \tau$, and where $S_n(u|x)$ means an estimator of $S(u|x)$, and similarly with $dM_n^o(u|X_i) = dN_i^o(u) - I(u \leq \tilde{T})d\Lambda_n(u|X_i)$ and K_C^n.

3.2 The estimand $\beta(P)$ and its corresponding EIF

As mentioned earlier, the GH-estimator of K_τ relies on the scoring rule $\check{\beta}^T X$, where $\check{\beta}$ is the Cox partial likelihood estimator that converges to a well defined coefficient vector $\tilde{\beta}$ even if the Cox model is not correctly specified, but $\tilde{\beta}$ has the undesirable feature that it depends on the censoring distribution in the actual study in case the Cox model is misspecified. Their proposed scoring rule may thus reflect properties of the specific censoring distribution, which is of no scientific interest. We take a different approach defining the scoring rule in the setting where there is no censoring so it only reflects the association between survival and the markers. Inspired by the assumption-lean approach by Vansteelandt and Dukes [2022], we propose to use the coefficient $\beta_\tau(P)$ defined as

$$\beta_t(P) = \{\text{var}(X)\}^{-1}\text{cov}[g\{S(t|X)\}, X], \quad (14)$$

where g is some pre-specified link function. If we take $g(x) = \log\{-\log(x)\}$ and if in fact $S(t|X) = \exp\{-\Lambda_0(t)e^{\theta^T X}\}$ (ie the Cox model is correctly specified) then $\beta_t(P) = \theta$, but $\beta_t(P)$ remains well defined otherwise. The EIF of $\beta_t(P)$ is conveniently calculated by calculating the EIF of $\text{var}(X)$ and of $\text{cov}[g\{S(t|X)\}, X]$ separately. One can show that

$$\begin{aligned}\text{EIF}[\text{cov}[g\{S(t|X)\}, X]] =& \{g(S(t|X)) - Eg\}\{X - EX\} - E\{g(S(t|X)) - Eg\}\{X - EX\} \\ & - \{X - EX\}g'\{S(t|X)\}S(t|X)\int_0^t \frac{dM^o(u \mid X)}{K_C(u|X)S(u|X)}.\end{aligned}$$

Similarly one can develop the EIF for $\text{var}(X)$ giving the estimator $\hat{\text{var}}(X) = \mathbb{P}_n\{(X - \mathbb{P}_n X)(X - \mathbb{P}_n X)^T\}$. This results in the desired estimator of $\beta(P)$:

$$\begin{aligned}\hat{\beta}_t(P) = \{\hat{\text{var}}(X)\}^{-1}\mathbb{P}_n\Big[&\{g(S_n(t|X)) - E_n g\}\{X - E_n X\} \\ & - \{X - E_n X\}g'\{S_n(t|X)\}S_n(t|X)\int_0^t \frac{dM_n^o(u \mid X)}{K_n(u|X)S_n(u|X)}\Big] \qquad (15)\end{aligned}$$

that solves the (empirical version) of the EIF of $\beta_t(P)$ equal to zero.

4 Estimation of AUC_t

The C-index is much used in practise, but it was recently argued by [Blanche et al., 2019] that the cumulative-dynamic time-dependent area under the ROC-curve (AUC_t) [Heagerty et al., 2000, Heagerty and Zheng, 2005, Blanche et al., 2013] defined for all $t \leq \tau$ by

$$\text{AUC}_t = P(Y_1 \geq Y_2 \mid T_1 \leq t, T_2 > t) \qquad (16)$$

should be preferred if the aim is to predict risk of an event for a specific time horizon, $[0, t]$, say. For the same reason, we redefine $\beta(P)$ to be $\beta_t(P)$. Specifically, [Blanche et al., 2019] showed that the C-index is not proper while the AUC_t is proper. Define

$$v_t(y_1, y_2) = \{1 - S(t|y_1)\}S(t|y_2)$$

and

$$\Theta_t = P(Y_1 \geq Y_2, T_1 \leq t, T_2 > t) \qquad (17)$$

so that $\text{AUC}_t = \Theta_t/[\{1 - S(t)\}S(t)]$. As we know how to estimate $S(t)$ using (13), we can concentrate on the estimand Θ_t. By similar calculations as in Section 3, we get

$$D^*_{\Theta_t}(Z, P) = \tilde{g}_t(X; P) + \tilde{G}_t(Z, \dot{\beta}, P),$$

where

$$\tilde{g}_t(X; P) = \int I(y < Y)v_t(Y, y)dF_Y(y) + \int I(y > Y)v_t(y, Y)dF_Y(y) - 2\Theta_t$$

and

$$\tilde{G}_t(Z, \dot{\beta}, P) = \{F_Y(Y) - F_T(t)\}\, S(t|Y)\int_0^t \frac{dM(u|Y)}{S(u|Y)} + \check{H}(\beta)\dot{\beta}$$

with $\check{H}(\beta)$ a constant (given in the Appendix) and $\dot{\beta}$ is the efficient influence function of β. This leads to the following one-step estimator of Θ_t

$$\hat{\Theta}_t = \frac{1}{n(n-1)}\sum_{i \neq j} I\big(\hat{\beta}^T X_i > \hat{\beta}^T X_j\big)\hat{v}_t\big(\hat{\beta}^T X_j, \hat{\beta}^T X_i\big) + \frac{1}{2n}\sum_{i=1}^n \tilde{G}_t(O_i, \dot{\beta}, \hat{P}) \qquad (18)$$

where $\tilde{G}_t(O, \dot{\beta}, \hat{P})$ is obtained as in Section 3 using (9) and where the part involving the complicated constant $\check{H}(\beta)$ can be dropped as long as we use the estimator $\hat{\beta}_t$ given in the previous subsection. We thus propose the following estimator

$$\widehat{\text{AUC}}_t = \hat{\Theta}_t/[\{1 - \hat{S}(t)\}\hat{S}(t)]. \qquad (19)$$

5 Simulation study

5.1 Set-up

In this section we illustrate the proposed method in a simulation study based on the German Breast Cancer Study (GBCS) data from Section 6. We simulate a covariate vector $X = (X_1, X_2, X_3)$ mimicking the covariates log(prog_recp), tumor grade and nodes from the GBCS data. Event times are generated from a Weibull regression model $\lambda(t \mid X) = k\lambda^{-1}(t/\lambda)^{k-1}\exp(\beta^T X)$ where the parameters k, λ and β are estimated from the GBCS data. In the real data the censoring times do not appear to depend on the covariates. However to illustrate our method for covariate dependent censoring we generate censoring time from a Weibull regression model $\lambda_C(t \mid X_1) = k_C\lambda_C^{-1}(t/\lambda_C)^{k_C-1}\exp(\gamma X_1)$ where X_1 is the covariate log(prog_recp) and k_C and λ_C are estimated from the data. We choose γ so that the effect of log(prog_recp) on the censoring times is similar to the effect of log(prog_recp) on the event times (i.e. $\gamma \approx \beta_1 k k_C^{-1}$). To illustrate the performance of our method in settings where the proportional hazards assumption does not hold we also consider event times generated from a stratified Weibull regression model $\lambda_j(t \mid X_1, X_3) = k_j\lambda^{-1}(t/\lambda)^{k_j-1}\exp((X_1, X_3)^T\phi)$ where we stratify on the covariate tumor grade. We set end of follow-up τ to be equal to 2192 days (approx 6 years) across simulations.

5.2 Results

In Table 1 we consider the setting where the proportional hazards assumption holds and we simulate data with a sample size of $n \in \{300, 600, 1200, 2400\}$. For each simulated data set we compute the one-step estimator of the concordance propbability $\hat{K}_\tau$ defined in Equation (11) as well as the plug-in estimator of the concordance probability $\hat{K}_\tau^{\text{plug-in}}$ which is equal to (11) without the debiasing term. We also compute the one-step estimator of the c-index $\hat{C}_\tau$ defined in Equation (12) and the one-step estimator of AUC_t defined in (19) with t equal to 1096 days. To estimate the working models we considered both a penalized Poisson regression approach which was implemented using the R-package `glmnet` [Friedman et al., 2010], as well as correctly specified Cox proportional hazards regression using the `coxph` function from the `survival` package [Therneau, 2023]. This was repeated 500 times and we computed the bias, standard deviation (sd), and mean squared error (mse) across simulations.

		Penalized Poisson				Cox PH			
	n	300	600	1200	2400	300	600	1200	2400
$\hat{K}_\tau$	bias	4.56	2.89	1.93	1.20	1.03	0.42	0.20	0.07
	sd	2.56	1.83	1.37	1.02	2.92	1.83	1.36	1.04
	mse	2.73	1.17	0.56	0.25	0.96	0.35	0.19	0.11
$\hat{K}_\tau^{\text{plug-in}}$	bias	7.17	4.52	3.09	2.02	1.31	0.57	0.28	0.11
	sd	3.01	2.07	1.51	1.12	2.96	1.83	1.24	1.01
	mse	6.04	2.47	1.18	0.53	1.05	0.37	0.19	0.10
$\hat{C}_\tau$	bias	2.83	1.92	1.64	1.14	1.00	0.43	0.33	0.15
	sd	2.70	1.87	1.14	0.89	2.76	1.73	1.24	0.81
	mse	1.53	0.72	0.46	0.21	0.86	0.32	0.17	0.07
$\widehat{\text{AUC}}_t$	bias	2.46	1.97	1.17	1.08	−0.46	−0.34	−0.02	−0.05
	sd	3.68	2.48	2.00	1.13	2.85	2.06	1.63	1.10
	mse	1.95	1.00	0.70	0.29	0.83	0.43	0.27	0.12

Table 1: Simulation results under proportional hazards. bias is the average bias across simulations (×100); sd is the empirical standard deviation (×100); mse is the mean squared error (×1000). Each entry is based on 500 replicates.

The results summarized in Table 1 show that for all estimators the bias and the mse decrease with sample size. The bias and mse are generally smaller when the nuisance models are estimated using a correctly specified Cox proportional hazards model than when using penalized Poisson regression, while the sd is similar. When using the penalized Poisson regression approach the one-step estimator

$\hat{K}_\tau$ provides a bias reduction compared to the plug-in estimator $\hat{K}_\tau^{\text{plug-in}}$. When the working models are estimated using a correctly specified Cox model the plug-in estimator $\hat{K}_\tau^{\text{plug-in}}$, which then is equal to a stratified version of the Gönen and Heller estimator, is very similar to $\hat{K}_\tau$ in terms of bias and mse. This was expected as the debiasing term should be close to zero in this setting.

In Table 2 we consider the setting where the proportional hazards assumption does not hold and we compute $\hat{K}_\tau$ and $\hat{K}_\tau^{\text{plug-in}}$ for each simulated data set. We estimate the working models using both a penalized Poisson regression approach and using a misspecified Cox proportional hazards regression.

		Penalized Poisson				Cox PH			
	n	300	600	1200	2400	300	600	1200	2400
$\hat{K}_\tau$	bias	2.79	1.37	0.33	-0.66	-0.50	-1.03	-1.43	-1.58
	sd	2.85	2.06	1.48	1.04	3.72	2.80	1.59	1.19
	mse	1.59	0.89	0.23	0.15	1.41	0.85	0.46	0.39
$\hat{K}_\tau^{\text{plug-in}}$	bias	5.19	2.87	1.43	0.16	-0.05	-0.71	-1.22	-1.41
	sd	3.20	2.28	1.69	1.10	3.77	2.78	1.57	1.16
	mse	3.72	1.34	0.49	0.12	1.42	0.82	0.39	0.33

Table 2: Simulation results under non-proportional hazards. bias is the average bias across simulations ($\times 100$); sd is the empirical standard deviation ($\times 100$); mse is the mean squared error ($\times 1000$). Each entry is based on 500 replicates.

The results show that, as expected, the bias descreases with sample size when using penalized Poisson regression and increases with sample size when using (misspecified) Cox proportional hazards regression. The one-step estimator $\hat{K}_\tau$ still provides a bias reduction compared to the plug-in estimator $\hat{K}_\tau^{\text{plug-in}}$ when using penalized Poisson regression. This is not the case when using (misspecified) Cox proportional hazards regression.

6 Empirical study: German Breast Cancer Study Data

We now illustrate our method through an analysis of a trial conducted by the German Breast Cancer Study Group (GBSG) [Schumacher et al., 1994, Schmoor et al., 1996, Sauerbrei and Royston, 1999]. The main objective of the trial was to investigate the effect of different adjuvant therapies on recurrence-free survival in node-positive breast cancer patients. In the original study a total of 720 patients were recruited between 1984 and 1989. We will be using the data for the 686 patients who had complete data for the predictors age, tumour size, number of positive lymph nodes, progesterone and oestrogen receptor status, menopausal status and tumour grade. The data is publicly available in the R-package `condSURV` [Meira-Machado and Sestelo, 2023]. Figure 1 shows the Kaplan-Meier curves of the recurrence-free survival probability and the censoring probability, as well as the number of people at risk and the cummulative number of events.

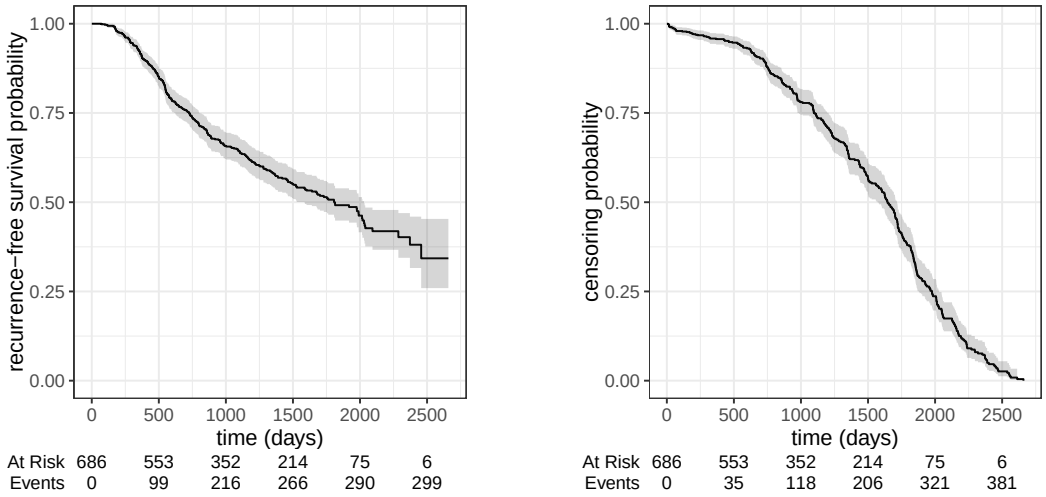
We shall focus on the predictors 'prog_recp' which is the progesterone receptor concentration, 'nodes' which is the number of positive lymph nodes and tumor grade which takes the values I, II and III. We discretize the 'nodes' variable such that it has three levels corresponding to < 3 nodes, $3-5$ nodes and ≥ 6 nodes.

We want to compare the discriminatory power of the following risk scores

$$Y_{t,A} = \hat{\beta}^A_{t,1} \log(\text{prog\_recp} + 1) + \hat{\beta}^A_{t,2} I(\text{nodes} = 2) + \hat{\beta}^A_{t,3} I(\text{nodes} = 3),$$
$$Y_{t,B} = \hat{\beta}^B_{t,1} \log(\text{prog\_recp} + 1) + \hat{\beta}^B_{t,2} I(\text{nodes} = 2) + \hat{\beta}^B_{t,3} I(\text{nodes} = 3) + \hat{\beta}^B_{t,4} I(\text{grade} = 2) + \hat{\beta}^B_{t,5} I(\text{grade} = 3),$$

where $\beta^A_{t,j}$ and $\beta^B_{t,j}$ are the assumption lean Cox regression coefficient described in Section 3.2 for respectively a model "A" which includes only the covariates log(prog_recp) and nodes, and a model "B" which includes the covariates log(prog_recp), nodes and tumor grade. The estimates of the coefficients for t equal to 1 year, 3 years, 5 years and at the maximum follow-up time (2659 days) respectively are given in Table 3 below.

Tumor grade is believed to be an important predictor of recurrence-free survival. This is corroborated by Figure 2 which shows the estimated survival probabilities stratified on tumor grade. Moreover



(a) (b)

Figure 1: Kaplan-Meier estimates of (a) the survival probability and (b) the censoring probability. The numbers at risk at the cummulative number of events are given below.

the estimated assumption lean coefficients in Table 3 are very large. Therefore we would expect model B to have better discriminatory power than model A.

			Penalized Poisson					Cox PH				
t	model		$\hat{\beta}_{t,1}$	$\hat{\beta}_{t,2}$	$\hat{\beta}_{t,3}$	$\hat{\beta}_{t,4}$	$\hat{\beta}_{t,5}$	$\hat{\beta}_{t,1}$	$\hat{\beta}_{t,2}$	$\hat{\beta}_{t,3}$	$\hat{\beta}_{t,4}$	$\hat{\beta}_{t,5}$
365	A	est	−0.23	0.47	1.50	-	-	−0.25	0.39	1.65	-	-
		se	0.098	0.39	0.40	-	-	0.12	0.43	0.44	-	-
	B	est	−0.18	0.42	1.40	0.73	1.26	−0.17	0.43	1.50	1.64	2.12
		se	0.097	0.38	0.41	0.36	0.45	0.11	0.41	0.41	0.40	0.47
1096	A	est	−0.18	0.53	1.29	-	-	−0.16	0.44	1.13	-	-
		se	0.047	0.22	0.19	-	-	0.051	0.22	0.19	-	-
	B	est	−0.14	0.50	1.21	0.48	0.79	−0.12	0.48	1.11	0.44	0.67
		se	0.049	0.22	0.20	0.21	0.24	0.054	0.23	0.19	0.23	0.22
1826	A	est	−0.15	0.55	0.96	-	-	−0.13	0.45	0.81	-	-
		se	0.044	0.21	0.18	-	-	0.049	0.21	0.19	-	-
	B	est	−0.11	0.53	0.88	0.69	0.78	−0.095	0.49	0.80	0.66	0.72
		se	0.046	0.22	0.18	0.19	0.20	0.049	0.21	0.18	0.20	0.19
2659	A	est	−0.12	0.24	0.70	-	-	−0.12	−0.06	0.55	-	-
		se	0.062	0.27	0.23	-	-	0.10	0.47	0.32	-	-
	B	est	−0.12	0.25	0.72	0.49	0.28	−0.14	−0.12	0.56	−0.55	−0.81
		se	0.057	0.25	0.21	0.23	0.22	0.12	0.49	0.36	0.53	0.34

Table 3: Estimates of regression coefficients for models A and B; est is the estimates of the regression coefficients; se is the influence function based standard error estimate.

Table 4 shows the estimated AUC_t for $t \in \{365, 731, 1096, 1461, 1826, 2192, 2555\}$ as well as the estimated C_τ and K_τ, for both models A and B. As in the simulation study we compute the estimators using both the penalized Poisson regression method and Cox proportional hazards regression. The estimates of the overall discrimination measures C_τ and K_τ computed using penalized Poisson regression suggest that the discriminatory abilities of the two models are similar while the estimates computed using Cox proportional hazards regression suggest that model A is slightly better. The

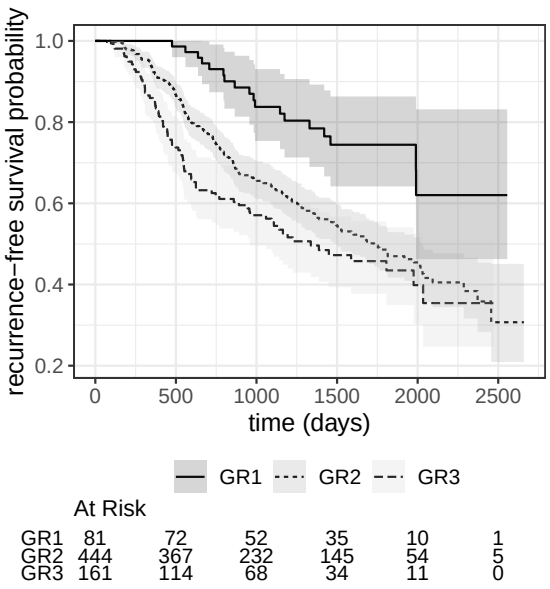


Figure 2: Kaplan-Meier estimates of the survival probability stratified by tumor grade. The numbers at risk in each strata are given below.

	Penalized Poisson		Cox PH	
	model A	model B	model A	model B
AUC_{365}	0.632	0.636	0.630	0.641
AUC_{731}	0.668	0.677	0.662	0.671
AUC_{1096}	0.700	0.713	0.699	0.708
AUC_{1461}	0.716	0.713	0.718	0.726
AUC_{1826}	0.738	0.734	0.755	0.764
AUC_{2192}	0.801	0.791	0.800	0.808
AUC_{2555}	0.891	0.887	0.974	0.939
C_{2659}	0.621	0.625	0.615	0.579
K_{2659}	0.596	0.595	0.597	0.562

Table 4: Estimates of AUC_t for $t \in \{365, 731, 1096, 1461, 1826, 2192, 2555\}$, as well as C_τ and K_τ with τ equal to end of follow-up (2659 days).

estimates of AUC_t suggest that the discriminatory ability of the two models is very similar for t-year predicted risk. The estimates of AUC_t computed using penalized Poisson regression are very similar to those computed using Cox proportional hazards regression.

7 Concluding remarks

In this paper we have proposed novel estimators of the C-index, the concordance probability of Gönen and Heller [2005], and the t-year area under the receiver operating characteristic (ROC) curve (AUC_t) based on the efficient influence function. The estimators are non-parametric and locally efficient, and allow for covariate-dependent censoring. We conducted a simulation study to examine the finite sample performace of the estimators. The simulation study showed that the proposed estimators are unbiased under covariate-dependent censoring, and that the one-step estimator of K_τ provides a bias reduction compared to the plug-in estimator. We illustrated the method by an application the data from the German Breast Cancer Study Group where we compared the discriminatory power of a model (model A) which included the progesterone receptor concentration and the number of positive lymph nodes as predictors to a model (model B) which included the predictors from model A as well as the additional predictor tumor grade. The analysis showed that the discriminatory ability of the two models were

similar both in terms of overall risk and t-year predicted risk.

The clinical utility of concordance measures such as the C-index for survival outcomes is very much up for debate. As discussed in Hartman et al. [2023a] the C-index will be deflated if there are many low risk patients with similar risk scores. This means that a clinically useful model may have a very low concordance. This could be a possible explanation as to why we don't see a difference in discriminatory power between the two models in our data application. A potential solution to this limitation is to construct alternative concordance measures which only make comparisons between subjects with either high or low risk scores. This will be the topic of future research.

Acknowledgements

This research was partly funded by a research gift from Novo Nordisk to the Section of Biostatistics, University of Copenhagen.

A Derivation of the efficient influence function

In this section we derive the efficient influence function (EIF) of the target parameters in (6), (17) and (14). There are many ways of deriving efficient influence functions. As advocated by e.g. Hines et al. [2022] and Kennedy [2023] we will derive the EIF by computing the so-called Gâteaux derivative [van der Vaart, 2000] which is defined as

$$\frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} \Psi(P_\varepsilon),$$

where $P_\varepsilon = (1-\varepsilon)P + \varepsilon\delta_{\tilde{z}}$ and $\delta_{\tilde{z}}$ is the Dirac measure at $Z = \tilde{z}$. We will start by deriving the EIF for the 'full' data and then use Tsiatis [2006] Theorem 10.1 and Theorem 10.4 to map the full-data EIF to the observed-data EIF.

A.1 Efficient influence function of $\Psi_\tau(P)$

We first note that

$$F_Y(y) = P(Y \leq y) = \int\int^{\frac{y-\beta_V^T v}{\beta_W}} f_{W|V}(w)dw f_V(v)dv,$$

so that by the chain rule and the Leibniz integral rule we have

$$\begin{aligned}\frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} F_{Y,\varepsilon}(y) &= F_{W|V}(\frac{y-\beta_V^T V}{\beta_W}) - F_Y(y) + H(\beta;y)\dot{\beta}(P) + I(\beta_V^T V + \beta_W W \leq y) - F_{W|V}(\frac{y-\beta_V^T V}{\beta_W}) \\ &= I(\beta_V^T V + \beta_W W \leq y) - F_Y(y) + H(\beta;y)\dot{\beta}(P), \end{aligned} \tag{20}$$

where

$$H(\beta;y) = -\frac{1}{\beta_W}\int f_{W|V}(\frac{y-\beta_V^T v}{\beta_W})\begin{bmatrix} v^T, \frac{y-\beta_V^T v}{\beta_W}\end{bmatrix} f_V(v)dv,$$

is a non-stochastic constant and $\dot{\beta}(P)$ is the influence function for β.

We further note that

$$\frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} \Lambda(t \mid y) = \frac{\delta_{\tilde{y}}(y)}{f_Y(y)}\int_0^t \frac{dM(u,y)}{S(u \mid y)},$$

and that

$$\frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} S_\varepsilon(t \mid y) = -\frac{\delta_{\tilde{y}}(y)}{f_Y(y)} S(t \mid y)\int_0^t \frac{dM(u,y)}{S(u \mid y)},$$

so that by the chain rule

$$\begin{aligned}\frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} h_{\tau,\varepsilon}(y_1,y_2) &= \int_0^\tau \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} S_\varepsilon(t \mid y_2)S(t \mid y_1)d\Lambda(t \mid y_1) + \int_0^\tau S(t \mid y_2)\frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} S_\varepsilon(t \mid y_1)d\Lambda(t \mid y_1) \\ &\quad + \int_0^\tau S(t \mid y_2)S(t \mid y_1)\frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} d\Lambda_\varepsilon(t \mid y_1) \\ &= -\frac{\delta_{\tilde{y}}(y_1)}{f_Y(y_1)}\int_0^\tau S(t \mid y_2)S(t \mid y_2)\int_0^t \frac{dM(u,y_1)}{S(u \mid y_1)}d\Lambda(t \mid y_1) + \frac{\delta_{\tilde{y}}(y_1)}{f_Y(y_1)}\int_0^\tau S(t \mid y_2)dM(t \mid y_1) \\ &\quad - \frac{\delta_{\tilde{y}}(y_2)}{f_Y(y_2)}\int_0^\tau S(t \mid y_2)\int_0^t \frac{dM(u,y_2)}{S(u \mid y_2)}S(t \mid y_1)d\Lambda(t \mid y_1).\end{aligned}$$

Then

$$\begin{aligned}
\frac{d}{d\varepsilon}\Big|_{\varepsilon=0}\Psi_\tau(P_\varepsilon) =& \int\int_{y_1>y_2} h_\tau(y_1,y_2)dF_Y(y_1)\frac{d}{d\varepsilon}\Big|_{\varepsilon=0}F_{Y,\varepsilon}(y_2) + \int\int_{y_1>y_2} h_\tau(y_1,y_2)\frac{d}{d\varepsilon}\Big|_{\varepsilon=0}F_{Y,\varepsilon}(y_1)dF_Y(y_2)\\
&+\int\int_{y_1>y_2}\frac{d}{d\varepsilon}\Big|_{\varepsilon=0}h_{\tau,\varepsilon}(y_1,y_2)dF_Y(y_1)dF_Y(y_2)\\
=&\int I(y<\tilde{y})h_\tau(\tilde{y},y)dF_Y(y) + \int I(y>Y)h_\tau(y,\tilde{y})dF_Y(y) + \tilde{H}(\beta)\dot{\beta}(P) - 2\Psi_\tau(P)\\
&-\int_{\tilde{y}}^{\infty}\int_0^\tau S(t|\tilde{y})\int_0^t \frac{dM(u|\tilde{y})}{S(u|\tilde{y})}S(t|y)d\Lambda(t|y)dF_Y(y)\\
&+\int_{-\infty}^{\tilde{y}}\left[\int_0^\tau S(t|y)dM(t|\tilde{y}) - \int_0^\tau S(t|y)S(t|\tilde{y})\int_0^t\frac{dM(u|\tilde{y})}{S(u|\tilde{y})}d\Lambda(t|\tilde{y})\right]dF_Y(y),
\end{aligned}$$

where

$$\tilde{H}(\beta) = \int I(y<\tilde{y})h_\tau(\tilde{y},y)H(\beta;y)dF_Y(y) + \int I(y>\tilde{y})h_\tau(y,\tilde{y})H(\beta;y)dF_Y(y).$$

We note that the martingale increment $dM(t,Y)$ can be written as follows

$$\begin{aligned}
dM(t,Y) =& dN(t) - I(T\geq t)d\Lambda(t\mid Y)\\
=& dN(t) - I(T\geq t)d\Lambda(t\mid X) + I(T\geq t)\left\{d\Lambda(t\mid X) - d\Lambda(t\mid Y)\right\}\\
=& dM(t,X) + I(T\geq t)\left\{d\Lambda(t\mid X) - d\Lambda(t\mid Y)\right\}\\
=& dM(t,X) - S(t\mid X)\int_0^t\frac{dM(u,X)}{S(u\mid X)}\left\{d\Lambda(t\mid X) - d\Lambda(t\mid Y)\right\}\\
&+ S(t\mid X)\left\{d\Lambda(t\mid X) - d\Lambda(t\mid Y)\right\},
\end{aligned}\tag{21}$$

where in the last equality we have used that

$$I(T\geq t) = I(T\geq t) - S(t\mid X) + S(t\mid X) = S(t\mid X) - S(t\mid X)\int_0^t\frac{dM(u,X)}{S(u\mid X)}.$$

Then it follows by Tsiatis equation (10.76) that the eif under censoring is.

$$\begin{aligned}
D^*_{\Psi_\tau}(Z)(P) =& \int I(y<Y)h_\tau(Y,y)dF_Y(y) + \int I(y>Y)h_\tau(y,Y)dF_Y(y) + \tilde{H}(\beta)\dot{\beta}(P) - 2\Psi_\tau(P)\\
&-\int_Y^\infty\int_0^\tau S(t\mid Y)\int_0^t\frac{dM^o(u|X)}{K_C(u\mid X)S(u\mid Y)}S(t\mid y)d\Lambda(t\mid y)dF_Y(y)\\
&-\int_Y^\infty\int_0^\tau S(t\mid Y)\int_0^t\frac{S(u\mid X)}{S(u\mid Y)}\left\{d\Lambda(u\mid X) - d\Lambda(u\mid Y)\right\}S(t\mid y)d\Lambda(t\mid y)dF_Y(y)\\
&+\int_Y^\infty\int_0^\tau S(t\mid Y)\int_0^t\frac{S(u\mid X)}{S(u\mid Y)}\int_0^u\frac{dM(v,X)}{K_C(v\mid X)S(v\mid X)}\left\{d\Lambda(v\mid X) - d\Lambda(v\mid Y)\right\}S(t\mid y)d\Lambda(t\mid y)dF_Y(y)\\
&+\int_{-\infty}^Y\int_0^\tau S(t\mid y)\frac{dM^o(u|X)}{K_C(u\mid X)}dF_Y(y)\\
&+\int_{-\infty}^Y\int_0^\tau S(t\mid y)S(t\mid X)\left\{d\Lambda(t\mid X) - d\Lambda(t\mid Y)\right\}dF_Y(y)\\
&-\int_{-\infty}^Y\int_0^\tau S(t\mid y)S(u\mid X)\int_0^u\frac{dM^o(v,X)}{K_C(v\mid X)S(v\mid X)}\left\{d\Lambda(u\mid X) - d\Lambda(u\mid Y)\right\}dF_Y(y)\\
&-\int_{-\infty}^Y\int_0^\tau S(t\mid y)\int_0^t\frac{dM^o(u|X)}{K_C(u\mid X)S(u\mid Y)}S(t\mid Y)d\Lambda(t\mid Y)dF_Y(y)\\
&-\int_{-\infty}^Y\int_0^\tau S(t\mid y)\int_0^t\frac{S(u\mid X)}{S(u\mid Y)}\left\{d\Lambda(u\mid X) - d\Lambda(u\mid Y)\right\}S(t\mid Y)d\Lambda(t\mid Y)dF_Y(y)\\
&+\int_{-\infty}^Y\int_0^\tau S(t\mid y)\int_0^t\frac{S(u\mid X)}{S(u\mid Y)}\int_0^u\frac{dM(v,X)}{K_C(v\mid X)S(v\mid X)}\left\{d\Lambda(v\mid X) - d\Lambda(v\mid Y)\right\}S(t\mid Y)d\Lambda(t\mid Y)dF_Y(y).
\end{aligned}\tag{22}$$

A.2 Efficient influence function of $\Theta_t(P)$

Note that

$$\frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} v_{t,\varepsilon}(y_1,y_2) = \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} S_\varepsilon(t \mid y_2)\left\{1 - S_\varepsilon(t \mid y_1)\right\} = -\frac{\delta_{\tilde{y}}(y_2)}{f_Y(y_2)} S(t \mid y_2) \int_0^t \frac{dM(u,y_2)}{S(u \mid y_2)} \left\{1 - S(t \mid y_1)\right\} + S(t \mid y_2)\frac{\delta_{\tilde{y}}(y_1)}{f_Y(y_1)} S(t \mid y_1) \int_0^t \frac{dM(u,y_1)}{S(u \mid y_1)}.$$

Then by the chain rule and (20) we have

$$\begin{aligned}\frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} \Theta_t(P_\varepsilon) = &\int I(y<\tilde{y}) v_t(\tilde{y},y) dF_Y(y) + \int I(y>\tilde{y}) v_t(y,\tilde{y}) dF_Y(y) \\ &- \int_{\tilde{y}}^{\infty} v_t(y,\tilde{y}) \int_0^t \frac{dM(u,\tilde{y})}{S(u \mid \tilde{y})} dF_Y(y) + \int_{-\infty}^{\tilde{y}} S(t \mid y) S(t \mid \tilde{y}) \int_0^t \frac{dM(u,\tilde{y})}{S(u \mid \tilde{y})} dF_Y(y) \\ &+ \bar{H}(\beta)\dot{\beta}(P) - 2\Theta_t(P),\end{aligned}$$

where

$$\bar{H}(\beta) = \int I(y<\tilde{y}) v_t(\tilde{y},y) H(\beta;y) dF_Y(y) + \int I(y>\tilde{y}) v_t(y,\tilde{y}) H(\beta;y) dF_Y(y).$$

From (21) and Tsiatis equation (10.76) the EIF under censoring is

$$\begin{aligned} D^*_{\Theta_t}(Z)(P) = &\int I(y<Y) v_t(Y,y) dF_Y(y) + \int I(y>Y) v_t(y,Y) dF_Y(y) \\ &- \int_Y^{\infty} v_t(y,Y) \int_0^t \frac{dM^o(u,X)}{K_C(u \mid X) S(u \mid Y)} dF_Y(y) \\ &+ \int_Y^{\infty} v_t(y,Y) \int_0^t \frac{S(u \mid X)}{S(u \mid Y)} \int_0^u \frac{dM^o(v,X)}{K_C(v \mid X) S(v \mid X)} \left\{d\Lambda(u \mid X) - d\Lambda(u \mid Y)\right\} dF_Y(y) \\ &- \int_Y^{\infty} v_t(y,Y) \int_0^t \frac{S(u \mid X)}{S(u \mid Y)} \left\{d\Lambda(u \mid X) - d\Lambda(u \mid Y)\right\} dF_Y(y) \\ &+ \int_{-\infty}^{Y} S(t \mid y) S(t \mid Y) \int_0^t \frac{dM^o(u,X)}{K_C(u \mid X) S(u \mid Y)} dF_Y(y) \\ &- \int_{-\infty}^{Y} S(t \mid y) S(t \mid Y) \int_0^t \frac{S(u \mid X)}{S(u \mid Y)} \int_0^u \frac{dM^o(v,X)}{K_C(v \mid X) S(v \mid X)} \left\{d\Lambda(u \mid X) - d\Lambda(u \mid Y)\right\} dF_Y(y) \\ &+ \int_{-\infty}^{Y} S(t \mid y) S(t \mid Y) \int_0^t \frac{S(u \mid X)}{S(u \mid Y)} \left\{d\Lambda(u \mid X) - d\Lambda(u \mid Y)\right\} dF_Y(y) \\ &+ H(\beta)\dot{\beta}(P) - 2\Theta_t(P). \end{aligned} \tag{23}$$

References

P. Blanche, J.-F. Dartigues, and H. Jacqmin-Gadda. Review and comparison of roc curve estimators for a time-dependent outcome with marker-dependent censoring. *Biometrical Journal*, 55(5):687–704, 2013.

P. Blanche, M. W. Kattan, and T. A. Gerds. The c-index is not proper for the evaluation of-year predicted risks. *Biostatistics*, 20(2):347–357, 2019.

J. H. Friedman, T. Hastie, and R. Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22, 2010.

T. A. Gerds, M. W. Kattan, M. Schumacher, and C. Yu. Estimating a time-dependent concordance index for survival prediction models with covariate dependent censoring. *Statistics in Medicine*, 32 (13):2173–2184, 2013.

M. Gönen and G. Heller. Concordance probability and discriminatory power in proportional hazards regression. *Biometrika*, 92(4):965–970, 2005.

F. E. Harrell, R. M. Califf, D. B. Pryor, K. L. Lee, and R. A. Rosati. Evaluating the yield of medical tests. *Journal of the American Medical Association*, 247(18):2543–2546, 1982.

F. E. Harrell Jr, K. L. Lee, and D. B. Mark. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine*, 15(4):361–387, 1996.

N. Hartman, S. Kim, K. He, and J. D. Kalbfleisch. Pitfalls of the concordance index for survival outcomes. *Statistics in Medicine*, 42(13):2179–2190, 2023a.

N. Hartman, S. Kim, K. He, and J. D. Kalbfleisch. Concordance indices with left-truncated and right-censored data. *Biometrics*, 79(3):1624–1634, 2023b.

P. J. Heagerty and Y. Zheng. Survival model predictive accuracy and roc curves. *Biometrics*, 61(1): 92–105, 2005.

P. J. Heagerty, T. Lumley, and M. S. Pepe. Time-dependent roc curves for censored survival data and a diagnostic marker. *Biometrics*, 56(2):337–344, 2000.

O. Hines, O. Dukes, K. Diaz-Ordaz, and S. Vansteelandt. Demystifying statistical learning based on efficient influence functions. *The American Statistician*, 76(3):292–304, 2022.

E. H. Kennedy. Semiparametric doubly robust targeted double machine learning: a review. arXiv preprint arXiv:2203.06469, 2023.

L. Meira-Machado and M. Sestelo. *Estimation of the Conditional Survival Function for Ordered Multivariate Failure Time Data*, 2023. URL `https://CRAN.R-project.org/package=condSURV`. R package version 2.0.4.

M. J. Pencina and R. B. D'agostino. Overall c as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Statistics in medicine*, 23(13): 2109–2123, 2004.

W. Sauerbrei and P. Royston. Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. *Journal of the Royal Statistical Society A*, 162(1): 71–94, 1999.

C. Schmoor, M. Olschewski, and M. Schumacher. Randomized and non-randomized patients in clinical trials: Experiences with comprehensive cohort studies. *Statistics in Medicine*, 15:253–271, 1996.

M. Schumacher, G. Basters, H. Bojar, K. Hobner, M. Olschewski, W. Sauerbrei, C. Chmoor, and C. Beyerle. Randomized 2 x 2 trial evaluating hormonal treatment and the duration of chemotherapy in node-positive breast cancer patients. *Journal of Clinical Oncology*, 12(10):2086–2093, 1994.

T. M. Therneau. *A Package for Survival Analysis in R*, 2023. URL https://CRAN.R-project.org/package=survival. R package version 3.5-7.

A. Tsiatis. *Semiparametric theory and missing data*. Springer Science & Business Media, 2006.

H. Uno, T. Cai, M. J. Pencina, R. B. D'agostino, and L. J. Wei. On the c-statistic for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in Medicine*, 30(10):1105–1117, 2011.

A. W. van der Vaart. *Asymptotic statistics*. Cambridge University Press, 3 edition, 2000.

S. Vansteelandt and O. Dukes. Assumption-lean inference for generalised linear model parameters. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(3):657–685, 2022.