Asbjrn Thomas Maltesen

Analysis of multiple domains in longitudinal epidemiological studies using modern causal inference methodology

UNIVERSITY OF COPENHAGEN
FACULTY OF HEALTH AND MEDICAL SCIENCES

PhD thesis

Asbjørn Thomas Maltesen

H. Lundbeck A/S &
Section of Biostatistics, University of Copenhagen

Analysis of multiple domains in longitudinal epidemiological studies using modern causal inference methodology

A longitudinal study consisting of patients with major depression disorder

Academic supervisor: Torben Martinussen
Industrial supervisors: Klaus Groes Larsen and Lene Hammer-Helmich

This thesis has been submitted to the Graduate School of Health and Medical Sciences, University of Copenhagen: 31/7/2020

Analysis of multiple domains in longitudinal epidemiological studies using modern causal inference methodology - A longitudinal study consisting of patients with major depression disorder

PhD thesis

Asbjørn Thomas Maltesen

H. Lundbeck A/S
Business Unit: Research & Development
Division: Experimental Medicine & Clinical Development
Department: Health Economics & Epidemiology Statistics
Ottiliavej 9
2500 Valby
Denmark
thomas.maltesen@protonmail.com

Section of Biostatistics
University of Copenhagen
Øster Farimagsgade 5
Entrance B, 2nd floor
P.O.B. 2099
DK-1014 Copenhagen K

Academic supervisor: Torben Martinussen
Industrial supervisors: Klaus Groes Larsen and Lene Hammer-Helmich
Submitted: 31/7/2020

Preface

This thesis has been submitted to the Graduate School of the Faculty of Health and Medical Sciences, University of Copenhagen, Denmark. The work was carried out in collaboration between H. Lundbeck A/S, Denmark and Section of Biostatistics, University of Copenhagen, Denmark. This industrial PhD was funded by the Innovation Fund Denmark and H. Lundbeck A/S. I would like to thank my supervisors Torben Martinussen, Section of Biostatistics, University of Copenhagen and Klaus Groes Larsen and Lene Hammer-Helmich, H. Lundbeck A/S for advice and support. I would like to thank Ingrid Sofie Harbo for allowing me the opportunity to do an industrial PhD at H. Lundbeck A/S. I would like to thank all my colleagues at HEEstat (Health Economics and Epidemiology Statistics) for the time at H. Lundbeck A/S. I also want to thank for all the years at Section of Biostatistics. A special thanks to Julie L. Forman and Susanne K. Laupstad for the time at Section of Biostatistics. I am grateful to Rhian M. Daniel for giving me the opportunity to visit her and her department. I would like to thank Rhian for her advice, discussions and improvements to the manuscripts. I would like to thank Andrea Gärtner and Laszlo Trefan for the coffee breaks in the afternoons and for making my visit in Wales a great experience. Finally, I want to thank all the people I met at Division of Population Medicine for making my visit a pleasant experience. I would like to thank Marged Hall for helping me improve my English. I would like to thank my parents and all my friends for their support and help.

Abstract

The motivation of this thesis is the observational cohort study called *Prospective Epidemiological Research on Functioning Outcomes Related to Major depressive disorder* (PERFORM) since we are interested in the causal effect of cognitive symptoms (exposure) on functional impairment at a later time (outcome). This PhD thesis consists of three manuscripts and each manuscript contains an analysis of the PERFORM study. The PERFORM study is introduced in Chapter 1.

In Chapter 2, we introduce to causal inference. We consider the g-formula and the inverse probability weights to estimate the causal effect in the presence of time-dependent confounding for longitudinal data. We list the assumptions that we need to assume to identify the causal effect of cognitive symptoms on functional impairment at a later time. We consider two estimators for the g-formula and the inverse probability weighted estimator. Chapter 2 consists only of existing methods from the literature.

We consider mediation analysis in Chapter 3 because we want to estimate the direct effect of cognitive symptoms on functional impairment at a later time. We propose with Manuscript II a new definition of sequential mediation for the interventional direct effect and the interventional indirect effects for multiple mediators. We obtain the overall effect to be equal to the total causal effect using our new definition. The new definition is shown in Chapter 3.

We consider data containing missing observations in Chapter 4. Patients tend to drop-out of studies. It applies for observational studies as well as interventional studies. This will cause data to contain missing observations. We could reduce the data to a subset of fully observed patients but this may result in biased estimates. We propose a doubly robust estimator in Manuscript I for the g-formula when the data contains missing observations. The estimator is unbiased even if the models relating to the missingness mechanism in the data are misspecified. The models relating to the missingness mechanism may be misspecified since the knowledge of the models may be unknown. We also propose a doubly robust estimator in Manuscript III for sequential mediation for multiple mediators when the data contains missing observations. The two estimators are shown in Chapter 4.

In Chapter 5, we analyse the data of the PERFORM study that were not shown in the three manuscripts. The reason for the additional analysis of the PERFORM study is to connect the dots between the two time points that are shown in Manuscript I and Manuscript III. We

compare our new estimators from Chapter 4 to other existing estimators from the literature. In Chapter 6, we finalize the thesis with a discussion of our findings.

The PERFORM study was used as an example in this thesis. The estimators that have been developed in this thesis can be applied to a longitudinal data with repeated measurements and monotone missingness. We have shown a list at page 51 of the abbreviations that we have used in this thesis.

Resumé

Motivationen for denne afhandling kommer fra observationsstudiet *Prospective Epidemiological Research on Functioning Outcomes Related to Major depressive disorder* (PERFORM), da vi er interesseret i at estimere den kausale effekt af kognitive symptomer (exposure) på funktionsnedsættelse til et senere tidspunkt (outcome). Denne afhandling indeholder tre manuskripter og hvert manuskript indeholder en analyse af PERFORM studiet. Studiet PERFORM er introduceret i kapitel 1.

Vi introducerer i kapitel 2 til kausal inference, hvor vi betragter g-formlen og de inverse sandsynlighed vægte justeret for tidsafhængig confounding for longitudinelt data. Vi opskriver antagelserne, som vi er nødt til at antage for at kunne identificere den kausale effekt af kognitive symptomer på funktionsnedsættelse til et senere tidspunkt. Vi betragter to estimatorer for g-formlen og estimatoren for de inverse sandsynlighed vægte i kapitel 2. Kapitlet indeholder kun eksisterende metoder fra litteraturen.

Vi betragter mediation analyse i kapitel 3, da vi er interesseret i den direkte effekt af kognitive symptomer på funktionsnedsættelse til et senere tidspunkt. I Manuskript II foreslår vi en ny definition af sekventiel mediation for den interventionale direkte effekt og de interventionale indirekte effekter for multiple mediatorer. Vi kan med den nye definition opnå at den samlet (overall) effekt er lig med den totale kausale effekt. Den nye definition er vist i kapitel 3.

I kapitel 4 betragter vi data, som ikke er fuldt observeret. Patienter har en tendens til at droppe ud af studier. Det gælder både for observationsstudier og interventionsstudier og det bidrager til manglende observationer i data. Vi kunne reducere data til kun fuldt observeret patienter, men det kan måske resultere i ikke-centrale (biased) estimater. Vi foreslår en dobbelt robust estimator i Manuskript I for g-formlen når data mangler observationer. Den nye estimator er robust i tilfælde af at man skulle vælge de forkerte modeller for mekanismen, som skaber de manglende observationer. Man kan komme til at vælge de forkerte modeller for mekanismen, som skaber de manglende observationer, da man måske ikke kender dem på forhånd. Vi foreslår også en dobbelt robust estimator i Manuskript III med den nye definition af sekventiel mediation når data ikke er fuldt observeret. De to nye estimatorer er vist i kapitel 4.

I kapitel 5 analyser vi resten af PERFORM studiet, som ikke blev vist i de tre manuskripter. De ekstra analyser af PERFORM studiet er lavet, fordi vi vil forbinde de to tidspunkter, som er vist i Manuskript I og Manuskript III. Vores nye estimatorer fra kapitel 4 sammenlignes med

eksisterende estimatorer fra litteraturen. Kapitel 6 afslutter afhandlingen med en diskussion af vores fund.

Vi har brugt PERFORM studiet, som et eksempel. Estimatorne, som er blevet udviklet i denne afhandling, vil også kunne bruges til at analysere andre studier. Data kan være longitudinelt med gentagende målinger og manglende observationer. På side 51 finder man en liste over de forkortelser, som vi har brugt i denne afhandling.

Contents

Preface **i**

Abstract **iv**

Resumé **vi**

1 Introduction **1**

1.1 The aim 1
1.2 The PERFORM study 2

2 The causal effect **5**

2.1 One exposure 5
2.2 Time-varying exposure 6
2.3 Three estimators for estimating the causal effect 8
2.4 Simulation study 10
2.5 Simulation study based on the PERFORM study 12

3 Mediation **15**

3.1 Direct and indirect effects 15
3.2 Multiple mediators 17
3.3 Simulation study 18
3.4 Simulation study based on the PERFORM study 20

4 Missing observations **21**

4.1 Missing observations in data 21
4.2 Adaptive estimator 23
4.3 Simulation study 26
4.4 Simulation study based on the PERFORM study 27
4.5 The choice of the U-function 28
4.6 Nonmonotone missingness 29

5 Analysing the PERFORM study **31**

5.1 Statistical analysis 31
5.2 Result 32

6 Discussion **37**

6.1 Perspectives and limitations 39

7 Summary of manuscripts **41**

7.1 Summary of Manuscript I 41
7.2 Summary of Manuscript II 41
7.3 Summary of Manuscript III 42

A The doubly robust estimator **43**

B Tables from the analysis of the PERFORM study **47**

List of Abbreviations **51**

Bibliography **53**

Manuscript I **59**

Manuscript II **81**

Manuscript III **93**

1 Introduction

Major depressive disorder (MDD) is a multidimensional disease characterised by emotional, physical and cognitive symptoms. Treatment of cognitive symptoms may hold the key to achieving functional recovery in patients with MDD and the relationship between cognitive symptoms and functional impairment is not well understood (Chokka et al. (2019)). The Prospective Epidemiological Research on Functioning Outcomes Related to Major depressive disorder (PERFORM) study was conducted to better understand the course of a depressive episode and its impact on patient functioning in outpatients with MDD. The PERFORM study describes the course of functional impairment, perceived cognitive symptoms and depression symptoms over two years in outpatients with MDD (Hammer-Helmich et al. (2018)). The work in this thesis is motivated by the PERFORM study. We are interested in the effect of cognitive symptoms on functional impairment at a later time. The g-formula is needed to estimate the causal effect of the time-varying exposure (cognitive symptoms) in the presence of time-dependent confounding (Robins (1986); Daniel et al. (2013)). Mediation analysis is also needed because we have assumed that the causal effect of cognitive symptoms on functional impairment at a later time contains one direct path and three indirect paths. We are interested in estimating the direct effect of cognitive symptoms on functional impairment at a later time. Methods for adjusting time-dependent confounding and estimating the direct and indirect effects already exists in the literature. Methods for working with data containing missing observations exist in the literature too. However, none of the existing methods can be used for our data. See the three manuscripts (Manuscript I, Manuscript II and Manuscript III) for the different methods and the reason for why the existing methods do not work for our data.

1.1 The aim

The aim of this thesis is to develop an estimator to analyse longitudinal data with time-dependent confounding and missing observations that follow a monotone pattern. We also want to develop an estimator for mediation analysis for multiple mediators so that the overall effect is equal to the total causal effect while data contains missing observations that follow a monotone pattern. To the best of our knowledge the two estimators do not exist in the literature. The importance of the two estimators is that they utilize data better and reducing bias of the estimates compared to estimators using only complete cases. The assumption about

the missing observations for our two estimators is less strict compared to estimators using only complete cases.

1.2 The PERFORM study

The patients in the PERFORM study were either starting their first course of antidepressant monotherapy or undergoing their first switch of antidepressant. The patients were enrolled by a general practitioner or a psychiatrist. All the patients have been measured on three self-reported scales (Sheehan Disability Scale, Perceived Deficit Questionnaire and Patient Health Questionnaire) at six time points. The patients have a baseline and they have been measured again after 2, 6, 12, 18 and 24 months since baseline. Data were collected in five European countries: France, Germany, Spain, Sweden and United Kingdom. Eligible patients had a current diagnosis of MDD. Participation in the study was independent of the choice of antidepressant prescribed to the patient.

The Sheehan Disability Scale (SDS). The scale was used to measure the patient's functional impairment. The Sheehan Disability Scale assesses the functional impairment over the previous seven days. The scale consists of three items and the scale covers: work/school, social life/leisure activities and family life/home duties. Each item ranges from 0 to 10 with a global score ranging from 0 to 30. The global score of functional impairment at 0 corresponds to be unimpaired and the global score at 30 corresponds to be highly impaired. The score of SDS is categorised as follows: $0-5$ corresponds to minimal functional impairment, $6-11$ corresponds to mild functional impairment, $12-20$ corresponds to moderate functional impairment and $21-30$ corresponds to moderately functional impairment (The categories were introduced at the 2019 ECNP Congress (Llora et al. (2019))).

The Perceived Deficit Questionnaire (PDQ-5). The scale was used to measure the patient's cognitive symptoms: memory, concentration and executive function over the past four weeks (we suppress the "-5" in the name of the scale PDQ-5 to simplify the notation). The scale consists of five items with each item ranging from 0 to 4 with a global score ranging from 0 to 20. A higher score of PDQ corresponds to the patient suffers greater severity of their cognitive symptoms.

The Patient Health Questionnaire (PHQ-9). The scale was used to measure depression severity of the patient (we suppress the "-9" in the name of the scale PHQ-9 to simplify the notation). The scale consists of nine items with each item ranging from 0 to 3 with a global score ranging from 0 to 27. The global score at 0 corresponds to absence of depression and the global score at 27 corresponds to severe depression. The score of PHQ is categorised as follows: $0-4$ corresponds to none or minimal depression, $5-9$ corresponds to mild depression, $10-14$

corresponds to moderate depression, $15-19$ corresponds to moderately severe depression and $20-27$ corresponds to severe depression (Kroenke and Spitzer (2002)). See Hammer-Helmich et al. (2018) for further information about the PERFORM study.

We assume in Manuscript I that depression severity (PHQ) affects both cognitive symptoms (PDQ) and functional impairment (SDS) and that cognitive symptoms affect functional impairment. We also assume in Manuscript I that all the present measurements affect all the future measurements at the next time point and all the present measurements do not affect the past measurements (Haro et al. (2019)). Let t denote the time point. Let $t = b$ denote the baseline, and let t be equal to 2, 6, 12, 18 and 24 (months) which denotes the measurement time points since baseline. Let SDS_t denote the functional impairment at time t. Let PDQ_t denote the cognitive symptoms at time t. Let PHQ_t denote the depression severity at time t. Let W_t denote the vector of all three measurements at time $t \in \{b, 2, 6, 12, 18, 24\}$, $W_t = (PHQ_t, PDQ_t, SDS_t)$. The process is indicated by the Directed Acyclic Graph (DAG) in Figure 1.1 with all six time points over the two years.

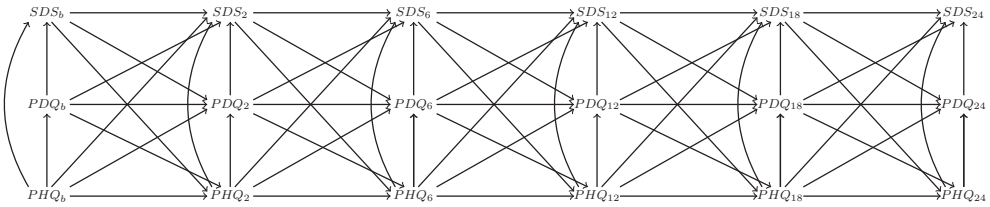


Figure 1.1: Let SDS_t denote the functional impairment at time $t \in \{b, 2, 6, 12, 18, 24\}$. Let PDQ_t denote the cognitive symptoms at time $t \in \{b, 2, 6, 12, 18, 24\}$. Let PHQ_t denote the depression severity at time $t \in \{b, 2, 6, 12, 18, 24\}$. Let $t = b$ denote the baseline, and let t be equal to 2, 6, 12, 18 and 24 (months) which denotes the measurement time points since baseline.

We let *pt* denote the *prior time* point before time t and we let *st* denote the *subsequent time* point after time t. An example: if t is equal to b (the time point is baseline) then *pt* does not exist and *st* is equal to 2. If t is equal to 18 then *pt* is equal to 12 and *st* is equal to 24. See Manuscript I for further information.

Patients in both observational studies and interventional studies tend to drop-out which means that the data contains missing observations. The analysis of a longitudinal study with repeated measurements and time-dependent confounding may be complicated when data contains missing observations. The PERFORM study has substantial many missing observations besides time-dependent confounding and mediated effects. Table 1.1 shows the numbers of the observed patients for each scale at each time point.

Table 1.1 shows that the number of patients with fully observed vectors will decrease when we combine the three different scales across the same time point. For example, the three scales at baseline ($t = b$) have at least/minimum 750 patients who have answered on each scale but

	Time points					
Scale	b	2	6	12	18	24
Sheehan Disability Scale (SDS)	750	607	586	554	486	458
Perceived Deficit Questionnaire (PDQ)	770	714	644	654	580	564
Patient Health Questionnaire (PHQ)	940	805	740	701	638	604
Fully observed vectors across the time point (W)	564	474	458	450	399	379

Table 1.1: The *Scale* column shows the three different scales. The different numbers in the six columns: b, 2, 6, 12, 18 and 24 represent the number of patients who have an observation for a specific time point and a scale. The *Fully observed vectors across the time point (W)* row shows the number of patients who have all three scales observed at the same time point.

only 564 patients have answered all three scales at the same time point. The Table also shows that the number of patients with fully observed vectors will decrease when we combine different time points since the numbers of W_t decrease over the six time points. See a similar Table in Haro et al. (2019).

2 | The causal effect

Section 2.1 introduces causal inference and the assumptions for identifying the causal effect for one binary exposure. Section 2.2 considers the g-formula with a time-varying exposure in the presence of time-dependent confounding. Section 2.3 shows three estimators for estimating a time-varying exposure (binary) in the presence of time-dependent confounding. Section 2.4 and Section 2.5 (based on the PERFORM study) show two different simulation studies. The two simulation studies are used to compare the three different estimators to each other.

2.1 One exposure

Let L, A and Y be observed. Let Y denote the continuous outcome variable. Let A denote a binary exposure ($A = 1$ is exposed and $A = 0$ is un-exposed). Let Y^a be the potential outcome that would have been observed if A is set to a (Rubin (1978)). The variable $Y^{a=1}$ defines the outcome Y that would have been observed if the subject had been exposed. The variable $Y^{a=0}$ defines the outcome Y that would have been observed if the subject had been un-exposed. The exposure A has a causal effect on the subject if $Y^{a=1} \neq Y^{a=0}$. The average causal effect for the exposed is not null if $E(Y^{a=1}) \neq E(Y^{a=0})$. The assumptions to identifying the causal effect are: exchangeability, positivity and consistency. The assumption *exchangeability* states that: $Y^a \perp\!\!\!\perp A$ for all $a \in A$. We assume that some will be exposed and some will be un-exposed with the *positivity* assumption given by $P(A = a) > 0$ for all $a \in A$. The *consistency* assumption is given by

$$\text{if } A = a \text{ then } Y^a = Y^A = Y. \tag{2.1}$$

We assume with the assumption that the observed outcome Y is equal to the potential outcome Y^a when the observed exposure A is equal to a. Let L be a potential confounder. Let the exposure A and the outcome Y share a cause L. The conditional exchangeability is given by

$$Y^a \perp\!\!\!\perp A \mid L \tag{2.2}$$

for all $a \in A$. The assumption states that the potential outcome is independent of the observed exposure given the measured confounder. The assumption is *not* equal to $Y \perp\!\!\!\perp A \mid L$. Unmeasured confounding between the exposure and the outcome may bias the estimation of the true

causal effect (Ding and VanderWeele (2015)). The assumption for positivity is given by

$$0 < P(A = a \mid L = l) < 1. \tag{2.3}$$

for all $l \in L$ with $P(L = l) > 0$. Unfortunately, the two assumptions (2.1) and (2.2) are untestable on the observed data (Robins et al. (2000); Cole and Hernán (2008); Cole and Frangakis (2009); VanderWeele (2009c); Pearl (2009, 2010)).

The marginal structural model (MSM) with one binary exposure A may be given by $E(Y^a) = \beta_I + \beta_1 a$. The coefficient β_I denotes the average causal effect for the un-exposed and the sum of the two coefficients β_I and β_1 denotes the average causal effect for the exposed (Hernán and Robins (2017)). The g-formula (Robins (1986)) with one exposure is given by

$$E\left(Y^a\right) = \int_{\mathcal{L}} E(Y^a \mid L = l) f_L(l) dl \tag{2.4}$$

$$= \int_{\mathcal{L}} E(Y^a \mid A = a, L = l) f_L(l) dl \tag{2.5}$$

$$= \int_{\mathcal{L}} E(Y \mid A = a, L = l) f_L(l) dl. \tag{2.6}$$

We use the conditional exchangeability assumption (2.2) to obtain the equation at (2.5) from (2.4) and we use the consistency assumption (2.1) to obtain the equation at (2.6) from (2.5). The inverse probability weight (IPW) with one binary exposure is given by

$$\begin{aligned} E\left(Y^a\right) &= \int_{\mathcal{L}} E(Y \mid A = a, L = l) f_L(l) dl \\ &= \int_{\mathcal{L}} \int_{\mathcal{Y}} y f_{Y|L,A}(y \mid l, a) dy f_L(l) dl \\ &= \int_{\mathcal{Y} \times \mathcal{L}} y \frac{I(A = a)}{f_{A|L}(a \mid l)} f_{Y,L,A}(y, l, a) d(y, l) \end{aligned}$$

where $f_{A|L}(a \mid l)$ is the probability for receiving the exposure A given L (Robins et al. (2000)). The last equality is only true if assumption (2.3) holds.

Studies may have repeated measurements. It applies for both observational studies and interventional studies. The next Section considers the g-formula for estimating the causal effect of a time-varying exposure in the presence of time-dependent confounding in longitudinal studies with repeated measurements.

2.2 Time-varying exposure

Suppose that our data comprises of n independent and identically distributed (iid) realization, $Z_1, \ldots, Z_n$. Let Z_i denote an ordered sequence $(L_{0,i}, A_{0,i}, \ldots, L_{T,i}, A_{T,i}, Y_i)$. We suppress the

index i to simplify the notation. Let Y denote the continuous outcome variable and the variable is measured at time $T+1$. Let A_t denote the exposure at time $t \in \{0, \ldots, T\}$. Let L_t denote the measured potential confounders at time $t \in \{0, \ldots, T\}$. Let $\overline{A}_T$ denote the vector of all exposures up to time T, $\overline{A}_T = (A_0, \ldots, A_T)$. Let $\overline{L}_T$ denote the vector of all measured potential confounders up to time T, $\overline{L}_T = (L_0, \ldots, L_T)$. Let $\overline{a}_T$ denote the vector $(a_0, \ldots, a_T)$ and let $\overline{l}_T$ denote the vector $(l_0, \ldots, l_T)$. See Manuscript I for further information. Let the potential outcome $Y^{\overline{a}_T}$ be the outcome that would have been observed if the vector $\overline{A}_T$ is set to $\overline{a}_T$. The outcome Y may be causally influenced by the whole history of $\overline{A}_T$ and $\overline{L}_T$. See the DAG in Figure 2.1. See a similar DAG in Daniel et al. (2013). If the potential confounders $\overline{L}_T$ are ignored in the analysis then may the effect of $\overline{A}_T$ on Y be confounded (Robins et al. (2007); Lok and DeGruttola (2012); Robins and Wasserman (2013); Daniel et al. (2013); Vansteelandt and Sjolander (2016); Keogh et al. (2018)).

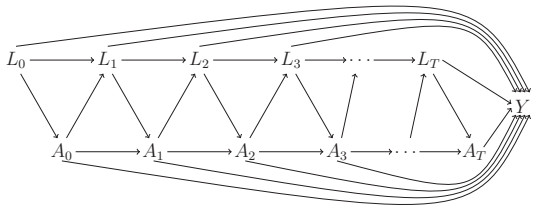


Figure 2.1: Let A_t denote the exposure at time t, let L_t denote the measured potential confounders at time t and let Y denote the continuous outcome variable.

The consistency assumption for $T+1$ exposures is given by

$$Y^{\overline{a}_T} = Y \text{ if } \overline{A}_T = \overline{a}_T; L_t^{\overline{a}_{t-1}} = L_t \text{ if } \overline{A}_{t-1} = \overline{a}_{t-1} \tag{2.7}$$

for time $t \in \{0, \ldots, T\}$. The conditional exchangeability for a time-varying exposure in the presence of time-dependent confounding is given by

$$Y^{\overline{a}_T} \perp\!\!\!\perp A_t \mid \overline{L}_t, \overline{A}_{t-1} \tag{2.8}$$

for all $\overline{a}_T \in \overline{A}_T$ for all $t \in \{0, \ldots, T\}$. The positivity assumption for a time-varying exposure in the presence of time-dependent confounding is given by

$$0 < f_{A_t|\overline{L}_t, \overline{A}_{t-1}}(a_t \mid \overline{l}_t, \overline{a}_{t-1}) < 1 \tag{2.9}$$

for all $\overline{l}_t, \overline{a}_{t-1} \in \overline{L}_t, \overline{A}_{t-1}$ with $0 < f_{\overline{L}_t, \overline{A}_{t-1}}(\overline{l}_t, \overline{a}_{t-1})$ for $t = 0, \ldots, T$ with probability one. The g-formula (Robins (1986)) for a time-varying exposure in the presence of time-dependent confounding is given by

$$E\left(Y^{\overline{a}_T}\right) = \int_{\mathcal{L}} E(Y \mid \overline{A}_T = \overline{a}_T, \overline{L}_T = \overline{l}_T) \prod_{t=0}^{T} f_{L_t|\overline{L}_{t-1}, \overline{A}_{t-1}}(l_t \mid \overline{l}_{t-1}, \overline{a}_{t-1}) dl_t \tag{2.10}$$

with the set $(\overline{L}_{-1}, \overline{A}_{-1})$ as the empty set. The inverse probability weights for a time-varying exposure in the presence of time-dependent confounding is obtained by rewriting (2.10) to

$$E\left(Y^{\overline{a}_T}\right) = \int_{\mathcal{Y}\times\mathcal{L}} y \prod_{t=0}^{T} \left\{ \frac{I(A_t = a_t)}{f_{A_t|\overline{L}_{t-1},\overline{A}_{t-1}}(a_t \mid \overline{l}_{t-1}, \overline{a}_{t-1})} \right\} f_{Y,\overline{L}_T,\overline{A}_T}(y, \overline{l}_T, \overline{a}_T) d(y, \overline{l}_T). \tag{2.11}$$

The weight for receiving the exposure A_t given $\overline{L}_{t-1}$ and $\overline{A}_{t-1}$ will become large if the probability $f_{A_t|\overline{L}_{t-1},\overline{A}_{t-1}}(a_t \mid \overline{l}_{t-1}, \overline{a}_{t-1})$ in the denominator is close to zero while the assumption (2.9) holds. This can lead to a biased estimate. Stabilized inverse probability weights are sometimes used instead to avoid the possibility that the weights explode. See Daniel et al. (2013) for further information. The MSM may be given by $E\left(Y^{\overline{a}_T}\right) = \boldsymbol{\beta}\overline{a}_T^{\star}$ with $\boldsymbol{\beta}$ as the row vector of causal parameter values and the vector has the same length as the column vector $\overline{a}_T^{\star}$. The column vector $\overline{a}_T^{\star}$ contains the value 1 for the intercept β_I, all the exposures and all the possible interactions between the different exposures. It is exemplified in Section 2.4.

2.3 Three estimators for estimating the causal effect

We show three estimators for estimating $E(Y^{\overline{a}_T})$. Let V_t denote the vector $(\overline{L}_t, \overline{A}_t)$ and $v_t = (\overline{l}_t, \overline{a}_t)$. Manuscript I defines an estimator $\hat{E}(Y^{\overline{a}_T})$ for $E(Y^{\overline{a}_T})$ (at (2.10)) to be given by

$$\hat{E}\left(Y^{\overline{a}_T}\right) = \frac{1}{n}\sum_{i=1}^{n} \mu\{V_{0,i}, \hat{\boldsymbol{\gamma}}\} \tag{2.12}$$

with $m\{v_T, \boldsymbol{\xi}\} = E(Y \mid \overline{L}_T = \overline{l}_T, \overline{A}_T = \overline{a}_T)$, $\mu\{v_{T-1}, \boldsymbol{\gamma}\} = E(m\{V_T, \boldsymbol{\xi}\} \mid \overline{L}_{T-1} = \overline{l}_{T-1}, \overline{A}_{T-1} = \overline{a}_{T-1})$ and $\mu\{v_t, \boldsymbol{\gamma}\} = E(\mu\{V_{t+1}, \boldsymbol{\gamma}\} \mid \overline{L}_t = \overline{l}_t, \overline{A}_t = \overline{a}_t)$. The last model is given by $\mu\{v_0, \boldsymbol{\gamma}\} = E(\mu\{V_1, \boldsymbol{\gamma}\} \mid L_0 = l_0, A_0 = a_0)$. We refer to the $m\{v_T, \boldsymbol{\xi}\}$-model and all the $\mu\{v_t, \boldsymbol{\gamma}\}$-models for $t \in \{0, \ldots, T\}$ as the μ-models. All the μ-models in (2.12) have hats to indicate predicted values from the specified μ-models that have been used for the estimation and the predicted values are plugged into the estimator. Let $\hat{m}(V_T)$ denote the $m\{V_T, \hat{\boldsymbol{\xi}}\}$-model and let $\hat{\mu}(V_t)$ denote the $\mu\{V_t, \hat{\boldsymbol{\gamma}}\}$-model for $t \in \{0, \ldots, T\}$ to simplify the notation. Let $m(v_T, \boldsymbol{\xi}_0)$ denote the true model with the vector of true parameter values $\boldsymbol{\xi}_0$ and let $\mu\{v_t, \boldsymbol{\gamma}_0\}$ denote the true model with the vector of true parameter values $\boldsymbol{\gamma}_0$ for $t \in \{0, \ldots, T\}$. We show in the Appendix in Manuscript I that the estimator (2.12) is asymptotically normally distributed in the situation when T is equal to 1. The estimator $\hat{E}(Y^{\overline{a}_T})$ solves the estimating equation $0 = \sum_{i=1}^{n} U(Z_i)$ with

$$U(Z_i) = \mu\{V_{0,i}, \boldsymbol{\gamma}_0\} - E\left(Y^{\overline{a}_T}\right). \tag{2.13}$$

The $\mu\{V_t, \boldsymbol{\gamma}\}$-model can be extended if the confounder L_t is multivariate. Section 2.4 shows an example where the confounder is multivariate consisting of two variables. The estimator (2.12) is unbiased if the μ-models are correctly specified. The estimator (2.12) is obtained by a series of iterated conditional expectations (Kreif et al. (2017)). See Manuscript I for further information.

The IPW estimator: The inverse probability weighted (IPW) estimator $\hat{E}(Y^{\overline{a}_T})$ for $E(Y^{\overline{a}_T})$ (at (2.11)) is given by

$$\hat{E}\left(Y^{\overline{a}_T}\right) = \frac{1}{n}\sum_{i=1}^{n} Y_i \prod_{t=0}^{T} \frac{I(a_t)}{\pi\{V_{t,i}, \hat{\boldsymbol{\alpha}}\}} \tag{2.14}$$

with $\pi\{v_t, \boldsymbol{\alpha}\} = \pi(A_t = a_t \mid \overline{L}_t = \overline{l}_t, \overline{A}_{t-1} = \overline{a}_{t-1})$ which denotes the probability for the exposure A_t is equal to a_t given $\overline{L}_t$ and $\overline{A}_{t-1}$ for $t \in \{0, \ldots, T\}$. The indicator $I(A_t = a_t)$ is denoted by $I(a_t)$ for $t \in \{0, \ldots, T\}$. We refer to all the $\pi\{v_t, \boldsymbol{\alpha}\}$-models for $t \in \{0, \ldots, T\}$ as the π-models. All the hats in (2.14) indicate predicted values from the specified π-models that have been used for the estimation and the predicted values are plugged into the estimator. Let $\pi\{v_t, \boldsymbol{\alpha}_0\}$ denote the true model with the vector of true parameter values $\boldsymbol{\alpha}_0$ for $t \in \{0, \ldots, T\}$. The estimator (2.14) is unbiased if the π-models are correctly specified. The estimator (2.14) solves the estimating equation $0 = \sum_{i=1}^{n} U_{IPW}(Z_i)$ with

$$U_{IPW}(Z_i) = Y_i \prod_{t=0}^{T} \frac{I(a_t)}{\pi\{V_{t,i}, \boldsymbol{\alpha}_0\}} - E\left(Y^{\overline{a}_T}\right). \tag{2.15}$$

The doubly robust estimator: Bang and Robins (2005) show an augmented inverse probability weighted estimator of the g-formula with a time-varying exposure in the presence of time-dependent confounding. The doubly robust estimator $\hat{E}(Y^{\overline{a}_T})$ for $E(Y^{\overline{a}_T})$ is given by

$$\hat{E}\left(Y^{\overline{a}_T}\right) = \frac{1}{n}\sum_{i=1}^{n}\left[Y_i \prod_{t=0}^{T} \frac{I(a_t)}{\pi\{V_{t,i}, \hat{\boldsymbol{\alpha}}\}} + \sum_{t=0}^{T}\prod_{k=0}^{t-1} \frac{I(a_k)}{\pi\{V_{k,i}, \hat{\boldsymbol{\alpha}}\}}\left(1 - \frac{I(a_t)}{\pi\{V_{t,i}, \hat{\boldsymbol{\alpha}}\}}\right)\mu\{V_{t,i}, \hat{\boldsymbol{\gamma}}\}\right] \tag{2.16}$$

with $\mu\{v_t, \boldsymbol{\gamma}\} = E(m\{V_{t+1}, \boldsymbol{\xi}\} \mid \overline{L}_t = \overline{l}_t, \overline{A}_t = \overline{a}_t)$ and $\pi\{v_t, \boldsymbol{\alpha}\} = \pi(A_t = a_t \mid \overline{L}_t = \overline{l}_t, \overline{A}_{t-1} = \overline{a}_{t-1})$. See the estimator (2.12) for further information about the $\mu\{v_t, \boldsymbol{\gamma}\}$-models and see the estimator (2.14) for further information about the $\pi\{v_t, \boldsymbol{\alpha}\}$-models. The indicator $I(A_t = a_t)$ is denoted by $I(a_t)$ for $t \in \{0, \ldots, T\}$. Let $I(a_{-1})\pi\{V_{-1}, \boldsymbol{\alpha}\}^{-1} = 1$. All the hats in (2.16) indicate predicted values from the specified μ-models and the specified π-models that have been used for the estimation and the predicted values are plugged into the estimator. The estimator (2.16) is unbiased if either the μ-models or the π-models are correctly specified. Table 2.1 shows the different combinations of the μ-models and the π-models to obtain an unbiased estimator. Let $m(V_T)$ denote the true $m\{V_T, \boldsymbol{\xi}_0\}$-model with the vector of true parameter values $\boldsymbol{\xi}_0$ and let $\mu(V_t)$ denote the true $\mu\{V_t, \boldsymbol{\gamma}_0\}$-model with the vector of true parameter values $\boldsymbol{\gamma}_0$ for $t \in \{0, \ldots, T\}$ to simplify the notation. Let $\pi(V_t)$ denote the true $\pi\{V_t, \boldsymbol{\alpha}_0\}$-model with the vector of true parameter values $\boldsymbol{\alpha}_0$ for $t \in \{0, \ldots, T\}$ to simplify the notation. The estimator (2.16) solves the estimating equation $0 = \sum_{i=1}^{n} U_{AIPW}(Z_i)$ with

$$U_{AIPW}(Z_i) = Y_i \prod_{t=0}^{T} \frac{I(a_t)}{\pi(V_{t,i})} + \sum_{t=0}^{T}\prod_{k=0}^{t-1} \frac{I(a_k)}{\pi(V_{k,i})}\left(1 - \frac{I(a_t)}{\pi(V_{t,i})}\right)\mu(V_{t,i}) - E\left(Y^{\overline{a}_T}\right). \tag{2.17}$$

Let $\Upsilon_T(Z_i)$ be given by

$$\Upsilon_T(Z_i) = \sum_{t=0}^{T} \prod_{k=0}^{t-1} \frac{I(a_k)}{\pi(V_{k,i})} \left(1 - \frac{I(a_t)}{\pi(V_{t,i})}\right) \mu(V_{t,i}).$$

We show in Appendix A that $E(\Upsilon_T(Z))$ is equal to zero. We also show in Appendix A that the estimator (2.16) is unbiased when either the μ-models or the π-models are correctly specified.

μ	Correct		Wrong	
π	Correct	Wrong	Correct	Wrong
Unbiased	✔	✔	✔	✗

Table 2.1: The combinations of the μ-models and the π-models for the estimator (2.16). The *Unbiased* row shows if the estimator is unbiased or biased. The estimator is unbiased with the combination of the models denoted by ✔. The estimator is biased with the combination of the models denoted by ✗. The *Correct* column and the *Wrong* column indicate if the μ-models and the π-models are correctly specified or misspecified.

We notice that the first part of the estimator (2.16) is the inverse probability weighted estimator. The estimator (2.16) has the advantage that some of the included models are allowed to be misspecified compared to the estimator (2.12) and the IPW estimator (2.14). However, the estimator (2.16) may also inherit the unstable weights from the IPW estimator.

2.4 Simulation study

The marginal structural model (MSM) is given by

$$E(SDS_2^{(pdq_b, pdq_2)}) = \beta_I + \beta_1 pdq_b + \beta_2 pdq_2 + \beta_3 pdq_b pdq_2. \tag{2.18}$$

Let Z_b denote the vector (W_b, W_2) where W_t is defined in Chapter 1. The sample size of the data is 2000 and the data are replicated 3000 times. The simulated data correspond to the first two time points in the DAG in Figure 1.1. Data are simulated as follows: $PHQ_b \sim \text{Normal}(0, 1.1^2)$, $PDQ_b \sim \text{Bernoulli}(\varkappa_{pdq_b})$, $SDS_b \sim \text{Normal}(\eta_{sds_b}, 1.1^2)$, $PHQ_2 \sim \text{Normal}(\eta_{phq_2}, 1.1^2)$, $PDQ_2 \sim \text{Bernoulli}(\varkappa_{pdq_2})$ and $SDS_2 \sim \text{Normal}(\eta_{sds_2}, 1.1^2)$ where the means are given by

$$\begin{aligned}
\eta_{sds_b} :=& 0.2PHQ_b - PDQ_b, \\
\eta_{phq_2} :=& 0.4PHQ_b + 0.4PDQ_b - SDS_b \text{ and} \\
\eta_{sds_2} :=& 15 - 1.5PHQ_b + 1.7PDQ_b - 0.5SDS_b - PHQ_2 - 3PDQ_2 + PDQ_2PHQ_2 \\
& - 2PDQ_bPHQ_2 - 3.4PDQ_bPDQ_2 - 2PDQ_bPHQ_b - 0.3PDQ_2PHQ_b
\end{aligned}$$

and the two probabilities are given by

$$\begin{aligned}\text{logit}(\varkappa_{pdq_b}) :=& 1 - 1.6PDQ_b \text{ and}\\ \text{logit}(\varkappa_{pdq_2}) :=& -0.8 + 0.6PHQ_b + 0.9PDQ_b + 0.5SDS_b + 0.4PHQ_2\\ & - 0.8PHQ_bPDQ_b - 0.7PDQ_bSDS_b - 0.5PDQ_bPHQ_2\end{aligned}$$

where $\text{logit}(x) = \log(x) - \log(1-x)$. The true causal effects $\boldsymbol{\beta} = (\beta_I, \beta_1, \beta_2, \beta_3)$ are shown in Table 2.2. We use the seed 3 in R (`set.seed(3)`) to make it possible to replicate all the simulation studies. This applies for all the simulation studies in the three manuscripts and all the simulation studies in this thesis.

We let the μ-models in the estimator (2.12) be correctly specified. We let the π-models in the IPW estimator (2.14) be correctly specified. We specify the μ-models and the π-models in the estimator (2.16) according to Table 2.1. The estimation of the causal effects $\boldsymbol{\beta}$ are evaluated by the mean and the standard error. Table 2.2 shows the mean and the standard error of the 3000 estimates of $\boldsymbol{\beta} = (\beta_I, \beta_1, \beta_2, \beta_3)$. We denote the estimator (2.12) with the letters SG (Simple G-formula) and we denote the estimator (2.16) with the letters DR (Doubly Robust).

		SG	IPW	DR			
μ				Correct		Wrong	
π				Correct	Wrong	Correct	Wrong
Mean	β_I	8.999	8.991	8.997	9.002	8.997	8.844
	β_1	1.003	1.017	1.004	0.998	1.003	1.094
	β_2	2.002	2.004	2.003	1.997	2.004	2.330
	β_3	0.994	0.988	0.994	1.003	0.995	0.871
SE	β_I	0.061	0.350	0.083	0.089	0.188	0.231
	β_1	0.096	0.506	0.126	0.150	0.213	0.268
	β_2	0.106	1.825	0.176	0.143	0.298	0.268
	β_3	0.155	1.880	0.229	0.214	0.399	0.342

Table 2.2: The *Mean* row shows the mean of the 3000 estimates of $\boldsymbol{\beta}$ and the *SE* row shows the standard error of the 3000 estimates of $\boldsymbol{\beta}$. The *SG* column shows the estimation using the estimator (2.12). The *IPW* column shows the estimation using the estimator (2.14). The *DR* column shows the estimation using the estimator (2.16). See Table 2.1 for the description of the *Correct* and *Wrong* columns. The true causal effects are $\boldsymbol{\beta} = (\beta_I, \beta_1, \beta_2, \beta_3) = (9, 1, 2, 1)$.

Table 2.2 shows that the SG estimator (2.12) has the smallest standard errors. The higher standard errors obtained using the DR estimator (2.16) may be caused by the π-models that are included in the estimator since the standard errors obtained using the IPW estimator (2.14) are the highest. It appears that the price for using the doubly robust estimator is larger standard errors when we compare the DR estimator (2.16) to the SG estimator (2.12). The IPW estimator (2.14) may also be unstable and this can lead to biased estimates even though the π-models are correctly specified. This happens in the next Section.

2.5 Simulation study based on the PERFORM study

The MSM is given by (2.18). The simulation study in this Section is from Manuscript I but all the vectors are fully observed. See Chapter 4 for further information about fully observed vectors. The simulation study is based on the two vectors (W_b, W_2) from the PERFORM study. The sample size of the data is 1000 and the data are replicated 5000 times. See Manuscript I for further information about the simulation study. The three estimators: SG (2.12), IPW (2.14) and DR (2.16) are used for estimating the four expected potential outcomes: $E(SDS_2^{(1,1)})$, $E(SDS_2^{(0,1)})$, $E(SDS_2^{(1,0)})$ and $E(SDS_2^{(0,0)})$. The μ-models in the SG estimator (2.12) are correctly specified. The π-models in the IPW estimator (2.14) are correctly specified. We specify the μ-models and the π-models in the DR estimator (2.16) according to Table 2.1. Table 2.3 shows the mean, the standard error and the absolute value of bias (the difference between the mean and the true value) of the 5000 estimates of the expected potential outcomes. We use the function *Bias* because the expected potential outcomes are real numbers and it is hard to determine by the function *Mean* if the estimates are biased or unbiased.

		SG	IPW	DR			
	μ			Correct		Wrong	
	π			Correct	Wrong	Correct	Wrong
Mean	$E(SDS_2^{(1,1)})$	15.394	15.396	15.393	15.394	15.393	15.398
	$E(SDS_2^{(0,1)})$	13.401	11.850	13.446	13.404	13.167	12.690
	$E(SDS_2^{(1,0)})$	13.577	13.554	13.585	13.574	13.592	13.825
	$E(SDS_2^{(0,0)})$	9.823	9.563	9.834	9.783	9.862	9.306
SE	$E(SDS_2^{(1,1)})$	0.253	0.276	0.270	0.267	0.270	0.267
	$E(SDS_2^{(0,1)})$	1.343	17.485	6.096	1.759	6.655	1.573
	$E(SDS_2^{(1,0)})$	0.703	6.266	1.266	1.657	1.275	1.571
	$E(SDS_2^{(0,0)})$	0.909	7.807	2.146	4.278	1.997	4.154
Bias	$E(SDS_2^{(1,1)})$	0.005	0.002	0.005	0.005	0.005	0.000
	$E(SDS_2^{(0,1)})$	0.017	1.568	0.028	0.014	0.251	0.728
	$E(SDS_2^{(1,0)})$	0.008	0.032	0.001	0.012	0.006	0.239
	$E(SDS_2^{(0,0)})$	0.006	0.254	0.017	0.034	0.045	0.511

Table 2.3: The *Mean* row shows the mean of the 5000 estimates of the expected potential outcomes. The *SE* row shows the standard error of the 5000 estimates of the expected potential outcomes. The *Bias* row shows the absolute value of the difference between the mean and the true value. See Table 2.2 for the description of the five columns: *SG*, *IPW*, *DR*, *Correct* and *Wrong*. See Table 3 in Manuscript I for the true values of the expected potential outcomes. A numerical problem occurred for the estimation when the π-models were correctly specified. This means that the *Mean*, the *SE* and the *Bias* are based on 4999 estimates instead of 5000 estimates.

Table 2.3 shows that all the estimates of $E(SDS_2^{(1,1)})$ are unbiased even when the μ-models and the π-models are misspecified. It may be a coincidence. The SG estimator did not show

any problems of estimating the different effects and all the estimates are also unbiased. We obtained biased estimates using the IPW estimator even though the π-models were correctly specified. This is a good example of the problem concerning the weights in the IPW estimator. The DR estimator shows that one of the estimates is biased when the μ-models are misspecified and the π-models are correctly specified. The estimator with this specific combination of the μ-models and the π-models should provide an unbiased estimate according to Table 2.1. The standard error of the estimate of $E(SDS_2^{(0,1)})$ is also the largest one. The π-models may cause the biased estimate. The DR estimator may inherit the downside of the IPW estimator that the weights may explode even if the π-models are correctly specified. The π-models may also cause the standard errors to be large. The numerical problem was a convergence problem for the π-models.

We will return to the $U(Z)$-function (2.13) and the $U_{AIPW}(Z)$-function (2.17) again in Chapter 4. We will not consider the $U_{IPW}(Z)$-function (2.15) in the rest of this thesis because we obtained biased estimates using the IPW estimator when the π-models were correctly specified. We will not discard the $U_{AIPW}(Z)$-function (2.17) despite that one of the estimates was biased. The number of the specific combination of the two exposures $(0, 1)$ is sparse in the simulated data. This may cause the estimate to be biased but the simulation study shows some concern about the estimator (2.16).

3 | Mediation

Mediation analysis lies in the interest of the direct effect of the exposure on the outcome as well as the transmitted effect via one or more intermediate measurements. The investigation of the transmitted effect is interesting even if the (total) causal effect is almost zero. It may happen that the direct effect may cancel out with the transmitted effect via the intermediated measurements. Section 3.1 revisits briefly the different assumptions that we need to assume to identify the different direct and indirect effects. Section 3.2 considers the causal estimands from Manuscript II. Section 3.3 and Section 3.4 (based on the PERFORM study) are two mediation analysis of the two simulation studies from Chapter 2.

3.1 Direct and indirect effects

Let A denote the exposure and let M denote the mediator. Let Y denote the continuous outcome variable. Let C denote some baseline measurements not affected by the exposure. See the DAG in Figure 3.1a. Let Y^a be the value that would have been observed if the exposure A is set to a. Let M^a be the value that would have been observed if the exposure A is set to a. Let Y^{am} be the value that would have been observed if the exposure A is set to a and the mediator M is set to m.

Pearl (2001) defined the controlled direct effect by $E(Y^{am}) - E(Y^{a^*m})$. It is the difference between two expected potential outcomes with two different values of the exposure, a and a^*, when the mediator is kept fixed at level m. Robins and Greenland (1992) and Pearl (2001) defined the natural direct effect by $E(Y^{aM^{a^*}}) - E(Y^{a^*M^{a^*}})$. It is the difference between two expected potential outcomes with two different values of the exposure, a and a^*, but the mediator is set to its natural level had A been set to a^*. Robins and Greenland (1992) and Pearl (2001) defined the natural indirect effect by $E(Y^{aM^a}) - E(Y^{aM^{a^*}})$. It is the effect of the exposure on the outcome via the mediator. The total causal effect is defined by $E(Y^{aM^a}) - E(Y^{a^*M^{a^*}})$ (Pearl (2001)). The sum of the natural direct effect and the natural indirect effect has the property that the sum is equal to the total causal effect. VanderWeele et al. (2014) defined the interventional direct effect and the interventional indirect effect. Let $G^{a|C}$ denote a random drawn from the distribution of the mediator among those with exposure status a conditional on C (VanderWeele et al. (2014)). The interventional direct effect $E(Y^{aG^{a^*|C}}) - E(Y^{a^*G^{a^*|C}})$ is

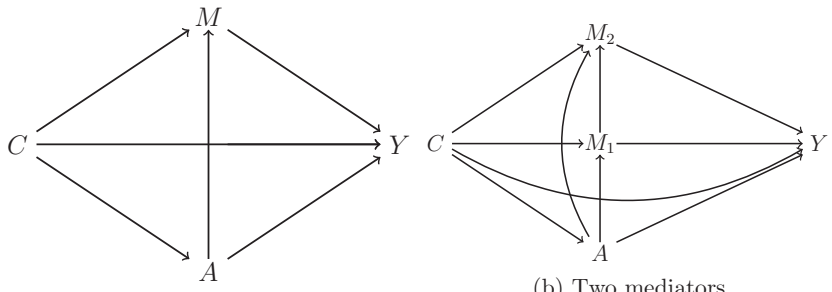
defined by

$$\int (E(Y^{am} \mid C = c) - E(Y^{a^*m} \mid C = c)) f_{M|A,C}(m \mid a^*, c) f_C(c) d(m, c).$$

and the interventional indirect effect $E(Y^{aG^{a|C}}) - E(Y^{aG^{a^*|C}})$ is defined by

$$\int E(Y^{am} \mid C = c)(f_{M|A,C}(m \mid a, c) - f_{M|A,C}(m \mid a^*, c)) f_C(c) d(m, c).$$

The sum of the interventional direct effect and the interventional indirect effect is defined by $E(Y^{aG^{a|C}}) - E(Y^{a^*G^{a^*|C}})$. However, the sum of the interventional effects may necessarily not be equal to the total causal effect. It will instead be called "the overall effect".



(a) One mediator (b) Two mediators

Figure 3.1: Let Y denote the continuous outcome variable. Let A denote the exposure and let C be some baseline measurements not affected by the exposure. This applies for the two DAGs. Figure 3.1a shows a DAG with one mediator M and Figure 3.1b shows a DAG with two mediators M_1 and M_2. The measurement M_1 is a mediator-outcome confounder.

The assumptions for the identification of the five effects are: (i) $Y^{am} \perp\!\!\!\perp A \mid C = c$ for all $(a, m, c) \in A, M, C$, (ii) $Y^{am} \perp\!\!\!\perp M \mid A = a, C = c$ for all $(a, m, c) \in A, M, C$, (iii) $M^a \perp\!\!\!\perp A \mid C = c$ for all $(a, m, c) \in A, M, C$ and (iv) $Y^{am} \perp\!\!\!\perp M^{a^*} \mid C = c$ for all $(a, a^*, m, c) \in A, M, C$. The controlled direct effect is identified if we assume the two assumptions: (i) and (ii). The interventional direct effect and the interventional indirect effect are identified if we assume the three assumptions: (i), (ii) and (iii). The natural direct effect and the natural indirect effect are identified if we assume all four assumptions: (i), (ii), (iii) and (iv). Assumption (iv) is the cross-world assumption (VanderWeele (2009a,b, 2011, 2016); Tchetgen Tchetgen and VanderWeele (2014)). See also Goetgeluk and Vansteelandt (2008) and VanderWeele and Tchetgen Tchetgen (2016). The interventional direct and indirect effects have the advantage of being meaningful even though the exposure is not manipulable (VanderWeele and Robinson (2014); Vansteelandt and Daniel (2017)). It is exemplified in Chapter 5 and the three manuscripts.

3.2 Multiple mediators

We consider two mediators for pedagogic purpose and the causal estimands can be extended to K mediators. Let M_1 and M_2 denote two ordered mediators, see the DAG in Figure 3.1b (Daniel et al. (2015); Vansteelandt and Daniel (2017)). We propose a definition of sequential mediation for multiple mediators in Manuscript II. The measurement M_1 is a mediator-outcome confounder and it violates the cross-world assumption (Avin et al. (2005)). Manuscript II defines the causal estimand for the interventional direct effect of the exposure on the outcome to be given by

$$\int \left(E(Y^{am_1m_2} \mid c) - E(Y^{a^*m_1m_2} \mid c)\right) f_{M_2^{a^*}|M_1^{a^*},C}(m_2 \mid m_1, c) f_{M_1^{a^*}|C}(m_1 \mid c) f_C(c) d(\overline{m}_2, c)$$

and it corresponds to the causal path: $A \to Y$. Let *dir* denote the interventional direct effect. The causal estimand for the interventional indirect effect of the exposure on the outcome via M_1 is given by

$$\int E(Y^{am_1m_2} \mid c) f_{M_2^{a^*}|M_1^{a^*},C}(m_2 \mid m_1, c) \left\{ f_{M_1^a|C}(m_1 \mid c) - f_{M_1^{a^*}|C}(m_1 \mid c) \right\} f_C(c) d(\overline{m}_2, c)$$

and the causal estimand corresponds to the sum of the two causal paths: $A \to M_1 \to Y$ and $A \to M_1 \to M_2 \to Y$. Let $indir_{M_1}$ denote the interventional indirect effect of the exposure on the outcome via M_1. The causal estimand for the interventional indirect effect of the exposure on the outcome via M_2 is given by

$$\int E(Y^{am_1m_2} \mid c) \left\{ f_{M_2^a|M_1^a,C}(m_2 \mid m_1, c) - f_{M_2^{a^*}|M_1^{a^*},C}(m_2 \mid m_1, c) \right\} f_{M_1^a|C}(m_1 \mid c) f_C(c) d(\overline{m}_2, c)$$

and the causal estimand corresponds to the causal path: $A \to M_2 \to Y$. Let $indir_{M_2}$ denote the interventional indirect effect of the exposure on the outcome via M_2. Manuscript II shows that the sum of the three causal estimands is equal to the total causal effect. The assumptions for identification of the causal estimands above (with two mediators) are (i') $Y^{a\overline{m}_2} \perp\!\!\!\perp A \mid C = c \quad \forall (a, \overline{m}_2, c) \in A, \overline{M}_2, C$, (ii') $Y^{a\overline{m}_2} \perp\!\!\!\perp (M_1, M_2) \mid A = a, C = c \quad \forall (a, \overline{m}_2, c) \in A, \overline{M}_2, C$ and (iii') $(M_2^a, M_1^a) \perp\!\!\!\perp A \mid C = c \quad \forall (a, c) \in A, C$. See Manuscript II for further information about the statistical estimands that are identified under the assumptions (i'), (ii') and (iii').

Estimator

Manuscript III rewrites the causal estimand for the interventional direct effect to be given by $\Gamma(a, a^*, a^*) - \Gamma(a^*, a^*, a^*)$ with

$$\Gamma(j, k, l) = \int E(Y^{jm_1m_2} \mid c) f_{M_2^k|M_1^k,C}(m_2 \mid m_1, c) f_{M_1^l|C}(m_1 \mid c) f_C(c) d(\overline{m}_2, c) \tag{3.1}$$

for different $j, k, l \in \{a, a^*\}$. The causal estimand for the interventional indirect effect via M_1 is given by $\Gamma(a, a^*, a) - \Gamma(a, a^*, a^*)$ using (3.1). The causal estimand for the interventional indirect effect via M_2 is given by $\Gamma(a, a, a) - \Gamma(a, a^*, a)$ using (3.1). Manuscript III shows that the estimator $\hat{\Gamma}(j, k, l)$ for $\Gamma(j, k, l)$ is given by

$$\hat{\Gamma}(j, k, l) = \frac{1}{n} \sum_{i=1}^{n} \mu_{j,k,l}\{V_{0,i}, \hat{\boldsymbol{\gamma}}\} \tag{3.2}$$

with $m_j\{v_2, \boldsymbol{\xi}\} = E(Y \mid M_2 = m_2, M_1 = m_1, A = j, C = c)$, $\mu_{j,k}\{v_1, \boldsymbol{\gamma}\} = E(m_j\{V_2, \boldsymbol{\xi}\} \mid M_1 = m_1, A = k, C = c)$ and $\mu_{j,k,l}\{v_0, \boldsymbol{\gamma}\} = E(\mu_{j,k}\{V_1, \boldsymbol{\gamma}\} \mid A = l, C = c)$. The estimator $\hat{\Gamma}(j, k, l)$ solves the estimating equation $0 = \sum_{i=1}^{n} U_{j,k,l}(Z_i)$ with

$$U_{j,k,l}(Z_i) = \mu_{j,k,l}\{V_{0,i}, \boldsymbol{\gamma}_0\} - \Gamma(j, k, l). \tag{3.3}$$

Let the $m_j\{V_2, \boldsymbol{\xi}_0\}$-model denote the true model with the vector of true parameter values $\boldsymbol{\xi}_0$. Let the $\mu_{j,k}\{V_1, \boldsymbol{\gamma}_0\}$-model and let the $\mu_{j,k,l}\{V_0, \boldsymbol{\gamma}_0\}$-model denote the true models with the vector of true parameter values $\boldsymbol{\gamma}_0$. Let $\hat{m}_j(V_2)$ denote the $m_j\{V_2, \hat{\boldsymbol{\xi}}\}$-model, let $\hat{\mu}_{j,k}(V_1)$ denote the $\mu_{j,k}\{V_1, \hat{\boldsymbol{\gamma}}\}$-model and let $\hat{\mu}_{j,k,l}(V_0)$ denote the $\mu_{j,k,l}\{V_0, \hat{\boldsymbol{\gamma}}\}$-model to simplify the notation. All the hats indicate predicted values from the specified models that have been used for the estimation and the predicted values are plugged into the estimator. We will refer to the $m_j\{v_2, \boldsymbol{\xi}\}$-model, the $\mu_{j,k}\{v_1, \boldsymbol{\gamma}\}$-model and the $\mu_{j,k,l}\{v_0, \boldsymbol{\gamma}\}$-model as the $\mu_{j,k,l}$-models. See Manuscript III for further information. The estimator $\widehat{dir}$ for the interventional direct effect, $\Gamma(a, a^*, a^*) - \Gamma(a^*, a^*, a^*)$, is given by

$$\widehat{dir} := \frac{1}{n} \sum_{i=1}^{n} (\hat{\mu}_{a,a^*,a^*}(V_{0,i}) - \hat{\mu}_{a^*,a^*,a^*}(V_{0,i})).$$

The estimator $\widehat{indir}_{M_1}$ for the interventional indirect effect via M_1, $\Gamma(a, a^*, a) - \Gamma(a, a^*, a^*)$, is given by

$$\widehat{indir}_{M_1} := \frac{1}{n} \sum_{i=1}^{n} (\hat{\mu}_{a,a^*,a}(V_{0,i}) - \hat{\mu}_{a,a^*,a^*}(V_{0,i}))$$

and the estimator $\widehat{indir}_{M_2}$ for the interventional indirect effect via M_2, $\Gamma(a, a, a) - \Gamma(a, a^*, a)$, is given by

$$\widehat{indir}_{M_2} := \frac{1}{n} \sum_{i=1}^{n} (\hat{\mu}_{a,a,a}(V_{0,i}) - \hat{\mu}_{a,a^*,a}(V_{0,i})).$$

3.3 Simulation study

We conduct a mediation analysis of the simulated data from Section 2.4. We want to estimate the interventional direct effect of PDQ_b on SDS_2. We also want to estimate the interventional indirect effect of PDQ_b on SDS_2 via SDS_b and the interventional indirect effect of PDQ_b on

SDS_2 via PHQ_2. Let pdq_b be equal to 1 and let pdq_b^* be equal to 0 in the estimation of the three different effects of PDQ_b on SDS_2. Let pdq_2 be equal to 0 and let pdq_2^* be equal to 0 in the estimation of the three different effects of PDQ_b on SDS_2. We use the marginal structural model (2.18) to obtain the equality

$$E(SDS_2^{(pdq_b,0)}) - E(SDS_2^{(pdq_b^*,0)}) = \beta_1(pdq_b - pdq_b^*)$$

and we have plugged the values of pdq_2 and pdq_2^* into the marginal structural model. We have

$$E\left(SDS_2^{(pdq_b,0)}\right) = E\left(SDS_2^{(pdq_b SDS_b^{pdq_b} PHQ_2^{pdq_b},0)}\right)$$

which means that the total causal effect for the exposure PDQ_b is given by

$$E\left(SDS_2^{(pdq_b SDS_b^{pdq_b} PHQ_2^{pdq_b},0)}\right) - E\left(SDS_2^{(pdq_b^* SDS_b^{pdq_b^*} PHQ_2^{pdq_b^*},0)}\right) = \beta_1(pdq_b - pdq_b^*).$$

We compare the overall effect from Table 3.1 with the coefficient β_1 from the SG column from Table 2.2 when we compare the overall effect to the total causal effect. We use the estimator (3.2) to estimate the interventional direct and indirect effects and we only use the true $\mu_{j,k,l}$-models with respect to the data for estimating the interventional direct and indirect effects. We denote the estimator (3.2) with the letters SSM (Simple Sequential Mediation formula). The estimation is evaluated by the mean and the standard error of the 3000 estimates of $\eta = (dir_b, indir_{SDS_b}, indir_{PHQ_2}, Overall)$. Table 3.1 shows the results of the estimation of η.

	dir_b	$indir_{SDS_b}$	$indir_{PHQ_2}$	$Overall$
Mean	0.504	0.300	0.200	1.003
SE	0.110	0.039	0.034	0.096

Table 3.1: Let $\eta = (dir_b, indir_{SDS_b}, indir_{PHQ_2}, Overall)$. The *Mean* row shows the mean of the 3000 estimates of η and the *SE* row shows the standard error of the 3000 estimates of η. The dir_b column is the interventional direct effect of PDQ_b on SDS_2. The $indir_{SDS_b}$ column is the interventional indirect effect of PDQ_b on SDS_2 via SDS_b. The $indir_{PHQ_2}$ column is the interventional indirect effect of PDQ_b on SDS_2 via PHQ_2. The $Overall$ column shows the sum of the three interventional effects (the overall effect). See the estimate of the coefficient β_1 in the *SG* column in Table 2.2 to see the total causal effect. The true direct effect is 0.5. The true mediated effect via SDS_b is 0.3 and the true mediated effect via PHQ_2 is 0.2.

Table 3.1 shows that we are able to obtain the true direct and indirect effects with our definition and we are also able to obtain the overall effect to be equal to the total causal effect with our definition. The results are in line with Manuscript II.

3.4 Simulation study based on the PERFORM study

We also conduct an analysis of the simulated data from Section 2.5 (based on the PERFORM study). Table 3.2 shows the estimates of the causal effects $\boldsymbol{\beta}$ of the MSM (2.18) and the estimates from the mediation analysis. We compare the overall effect to the coefficient β_1 when we compare the overall effect to the total causal effect. We use the same arguments from Section 3.3 for the comparison between the overall effect and the coefficient β_1. We use the SG estimator (2.12) and we only use the true μ-models with respect to the data for estimating the causal effects. We use the SSM estimator (3.2) and we only use the true $\mu_{j,k,l}$-models with respect to the data for estimating the interventional direct and indirect effects. Table 3.2 shows the mean of the 5000 estimates of $\eta = (dir_b, indir_{SDS_b}, indir_{PHQ_2}, Overall)$, the standard error of the 5000 estimates of η and the absolute value of bias (the difference between the mean and the true value).

	SG				SSM			
	β_I	β_1	β_2	β_3	dir_b	$indir_{SDS_b}$	$indir_{PHQ_2}$	$Overall$
Mean	9.823	3.754	3.578	-1.762	1.392	2.468	-0.106	3.754
SE	0.909	1.123	1.359	1.571	0.996	0.360	0.540	1.123
Bias	0.006	0.014	0.023	0.026	0.010	0.005	0.001	0.014

Table 3.2: Let $\eta = (dir_b, indir_{SDS_b}, indir_{PHQ_2}, Overall)$. The *SG* column shows the estimates of the causal effects $\boldsymbol{\beta}$ obtained using the SG estimator (2.12). The *SSM* column shows the estimates of the mediated effects obtained using the SSM estimator (3.2). The *Mean* row shows the mean of the 5000 estimates of η, the *SE* row shows the standard error of the 5000 estimates of η and the *Bias* row shows the absolute value of the difference between the mean and the true value. See Table 3.1 for the description of the four columns: dir_b, $indir_{SDS_b}$, $indir_{PHQ_2}$ and $Overall$. The true causal effects are $\boldsymbol{\beta} = (\beta_I, \beta_1, \beta_2, \beta_3) = (9.8, 3.8, 3.6, -1.8)$. The true direct effect is 1.4. The true mediated effect via SDS_b is 2.5 and the true mediated effect via PHQ_2 is -0.1. The true effects are rounded.

Table 3.2 shows that we are able to obtain the true direct and indirect effects with our definition and we are also able to obtain the overall effect to be equal to the total causal effect with our definition. The results are in line with Manuscript II.

4 | Missing observations

Data may contain vectors/individuals/patients with missing observations. Section 4.1 introduces data with missing observations. Section 4.2 introduces the adaptive estimator to analyse longitudinal data containing missing observations that follow a monotone pattern. An estimator for mediation analysis with data containing missing observations that follow a monotone pattern is also introduced. Section 4.3 and Section 4.4 finish the two simulation studies from Chapter 2. In Section 4.5, we argue for the choice of the $U(Z)$-function that has been used in the three manuscripts. In Section 4.6, we sketch an estimator with data containing missing observations that follow a nonmonotone pattern.

4.1 Missing observations in data

Let *full data* denote data that would have been collected on all the vectors in the sample. Let *observed data* denote the actually observed data containing vectors with missing observations. Let *complete data* denote a subset of the observed data with no missing observations. If the missingness mechanism is *missing completely at random* (MCAR) then the probability of being observed is independent of the data. If the missingness mechanism is *missing at random* (MAR) then the probability of missingness depends only on the observed data. If the missingness mechanism is *missing not at random* (MNAR) then the probability of missingness may also depend on the unobserved data (Tsiatis (2006)). Multiple imputation is one solution for data containing missing observations (Bartlett et al. (2015)). Reducing data to complete data is another solution but we may obtain biased estimates. However, if the missingness mechanism is missing completely at random then using only complete data will not cause biased estimates.

Coarsened data

Let Z be defined in Section 2.2. Let $\mathcal{C} \in \{1, \ldots, \mathbf{c}\} \cup \{\infty\}$ denote a random variable. If $\mathcal{C}$ is equal to 1 then L_0 is the only observed variable in the ordered sequence. If $\mathcal{C}$ is equal to 2 then L_0 and A_0 are the only two observed variables in the ordered sequence. If $\mathcal{C}$ is equal to $\mathbf{c}$ then it is only the outcome that is missing from Z. Note that $\mathbf{c}$ is an integer and it is *not* equal to c. If

the variable $\mathcal{C}$ is equal to ∞ then Z is fully observed. See Manuscript I for further information. Tsiatis (2006) defines a map G_r and it maps a vector from the full data to the observed data. The map is given by $G_r : Z \to g_r$ where g_r denote the observed vector in the data with r observed variables for $r = 1, \ldots, \mathbf{c}$. *Coarsening completely at random* (CCAR) corresponds to $\mathcal{C}$ is independent of the data like MCAR. *Coarsening at random* (CAR) corresponds to $\mathcal{C}$ depends on the observed data like MAR. *Coarsening not at random* (CNAR) is a function that may depend on the not observed data like MNAR. We also need to distinguish between *monotone* and *nonmonotone* missingness.

Monotone missingness

The vectors in the *observed data* are denoted by the iid random quantities given by

$$\{G_{\mathcal{C}_1}(Z_1), \mathcal{C}_1\}, \ldots, \{G_{\mathcal{C}_n}(Z_n), \mathcal{C}_n\}$$

where $\{G_{\mathcal{C}_i}(Z_i), \mathcal{C}_i\}$ denote the i-th vector in the observed data. A *complete case* is a vector denoted by $\{G_\infty(Z), \infty\}$. Complete data contains only complete cases ($\{G_\infty(Z), \infty\}$). The pattern of missingness in the data can be described by the following vectors: $G_1(Z) = (L_0)$, $G_2(Z) = (L_0, A_0)$, $G_3(Z) = (L_0, A_0, L_1)$, ..., $G_{\mathbf{c}}(Z) = (L_0, A_0, \ldots, L_T, A_T)$ and $G_\infty(Z) = Z$. In the pattern described above, if L_t is observed then $\overline{L}_{t-1}$ and $\overline{A}_{t-1}$ are necessarily also observed. This pattern is known as *monotone* missingness (Tsiatis (2006)). See Manuscript I for further information about monotone missingness. The pattern of nonmonotone missingness allows a variable between two observed variables in an ordered sequence to be missing. As an example, L_0 and L_1 are the only two observed variables in the ordered sequence $Z = (L_0, A_0, L_1, \ldots, L_T, A_T, Y)$ and the exposure A_0 is missing. The vector in the data is given by $G_2(Z) = (L_0, L_1)$ and the variable $\mathcal{C}$ is equal to 2. We will not look further into nonmonotone missingness but Section 4.6 sketches an estimator for nonmonotone missingness. We refer to monotone coarsening or monotone missingness. We assume the conditional probability of observing a complete vector given Z is strictly greater than zero, i.e. that:

$$\varpi\{\infty, Z, \boldsymbol{\psi}_0\} = P(\mathcal{C} = \infty \mid Z) > 0.$$

The probability $\varpi\{\infty, Z, \boldsymbol{\psi}_0\}$ denotes the true model with the vector of true parameter values. We sometimes refer to the probability $\varpi\{\infty, Z, \boldsymbol{\psi}_0\}$ as the ϖ-model. Let $\lambda_r\{G_r(Z), \boldsymbol{\psi}_0\} = P(\mathcal{C} = r \mid \mathcal{C} \geq r, Z)$ for $r \neq \infty$ denote the probability of stopping the observing of additional observations given r observed. Let $\boldsymbol{\psi}_0$ denote the true model with the vector of true parameter values. We refer to the $\lambda_r\{G_r(Z), \boldsymbol{\psi}\}$-models as the λ-models. Let $K_r\{G_r(Z), \boldsymbol{\psi}\}$ be equal to the product $\prod_{j=1}^{r}(1 - \lambda_j\{G_j(Z), \boldsymbol{\psi}\})$ and $K_{\mathbf{c}}\{G_{\mathbf{c}}(Z), \boldsymbol{\psi}\}$ is equal to the probability $\varpi\{\infty, Z, \boldsymbol{\psi}\}$ (Tsiatis (2006)). We assume the missingness mechanism is CAR which means that the coarsening probabilities only depend on the data as a function of the observed data. The coarsening probabilities are given by $\varpi\{r, G_r(Z), \boldsymbol{\psi}\} = \lambda_r\{G_r(Z), \boldsymbol{\psi}\}K_{r-1}\{G_{r-1}(Z), \boldsymbol{\psi}\}$ (see Tsiatis (2006) for further information). The ϖ-model is correctly specified if all the λ-models are correctly specified. The ϖ-model is misspecified if one of the λ-models is misspecified. We

assume the probability $\lambda_r\{G_r(Z), \boldsymbol{\psi}\}$ is given by

$$\lambda_r\{G_r(Z), \boldsymbol{\psi}\} = \frac{\exp\left(\psi_{I,r} + G_r(Z)\boldsymbol{\psi}_r\right)}{1 + \exp\left(\psi_{I,r} + G_r(Z)\boldsymbol{\psi}_r\right)},$$

where the column vector $\boldsymbol{\psi}_r$ has the same dimension as the row vector $G_r(Z)$ (Tsiatis (2006)). Let the vector of parameter values be given by $\boldsymbol{\psi} = (\psi_{I,r}, \boldsymbol{\psi}_r')$ where the coefficient $\psi_{I,r}$ denotes the intercept and $\boldsymbol{\psi}_r'$ is the transpose row vector of $\boldsymbol{\psi}_r$.

The IPW estimator for missing observations: The inverse probability weighted (IPW) estimator (e.g. with the $U(Z)$-function (2.13)) is a suitable estimator to solve the issue of missing observations (Seaman and White (2013); Li et al. (2013)). We assume that the pattern of the missing observations is monotone and the missingness mechanism is CAR. We also assume that the μ-models in the $U(Z)$-function (2.13) are correctly specified. The IPW estimator with the $U(Z)$-function (2.13) is given by

$$\frac{1}{n}\sum_{i=1}^{n}\frac{I(\mathcal{C}_i = \infty)}{\varpi\{\infty, Z_i, \hat{\boldsymbol{\psi}}\}}\hat{\mu}(V_{0,i}) \to E\left(\frac{I(\mathcal{C} = \infty)}{\varpi\{\infty, Z, \boldsymbol{\psi}\}}\mu(V_0)\right).$$

The estimator is unbiased if the probability $\varpi\{\infty, Z, \hat{\boldsymbol{\psi}}\}$ is correctly specified. Unfortunately, the $\varpi\{\infty, Z, \boldsymbol{\psi}\}$-model is most likely unknown and we may specify the $\varpi\{\infty, Z, \boldsymbol{\psi}\}$-model wrong. The next Section shows how we can handle the issue of the misspecification of the $\varpi\{\infty, Z, \boldsymbol{\psi}\}$-model.

4.2 Adaptive estimator

Let $U(Z_i)$ be written in the form $U(Z_i) = h(Z_i) - E(Y^{\overline{a}_T})$ for the estimating equation $0 = \sum_{i=1}^{n} U(Z_i)$ and fulfilling at the same time $E(U(Z)) = 0$. Let $h(Z_i)$ be a function of the data and let $E(Y^{\overline{a}_T})$ denote a constant. The estimator $\hat{E}(Y^{\overline{a}_T})$ for $E(Y^{\overline{a}_T})$ is given by

$$\begin{aligned}\hat{E}\left(Y^{\overline{a}_T}\right) = \frac{1}{n}\sum_{i=1}^{n}\Bigg(&\frac{I(\mathcal{C}_i = \infty)}{\varpi\{\infty, Z_i, \hat{\boldsymbol{\psi}}\}}\hat{h}(Z_i) \\ &+ \sum_{r=1}^{\mathbf{c}}\frac{I(\mathcal{C}_i = r) - \lambda_r\{G_r(Z_i), \hat{\boldsymbol{\psi}}\}I(\mathcal{C}_i \geq r)}{K_r\{G_r(Z_i), \hat{\boldsymbol{\psi}}\}}E(h(Z) \mid G_r(Z_i), \hat{\boldsymbol{\zeta}})\Bigg)\end{aligned} \tag{4.1}$$

for data containing missing observations that follow a monotone pattern and the missingness mechanism is CAR. We show in the Appendix of Manuscript I how we have derived the estimator at (4.1). All the hats indicate predicted values from the specified models that have been used for the estimation and the predicted values are plugged into the estimator. The estimator (4.1) is a doubly robust estimator. Let $h(Z)$ be correctly specified with respect to the distribution of Z. The estimator (4.1) is unbiased if either the conditional expectations

$E(h(Z) \mid G_r(Z), \boldsymbol{\zeta})$ are correctly specified with respect to the distribution of Z or the λ-models relating to the missingness mechanism are correctly specified. The estimator (4.1) solves the estimating equation

$$0 = \sum_{i=1}^{n} \left(\frac{I(\mathcal{C}_i = \infty)}{\varpi\{\infty, Z_i, \hat{\boldsymbol{\psi}}\}} U(Z_i) + \sum_{r=1}^{\mathbf{c}} \frac{I(\mathcal{C}_i = r) - \lambda_r\{G_r(Z_i), \hat{\boldsymbol{\psi}}\} I(\mathcal{C}_i \geq r)}{K_r\{G_r(Z_i), \hat{\boldsymbol{\psi}}\}} E(U(Z) \mid G_r(Z_i), \hat{\boldsymbol{\zeta}}) \right).$$

Let $\hat{\varpi}(\infty, Z)$ denote the probability $\varpi\{\infty, Z, \hat{\boldsymbol{\psi}}\}$ to simplify the notation, let $\hat{\lambda}_r(G_r(Z))$ denote the probability $\lambda_r\{G_r(Z), \hat{\boldsymbol{\psi}}\}$ to simplify the notation and let $\hat{K}_r(G_r(Z))$ denote the probability $K_r\{G_r(Z), \hat{\boldsymbol{\psi}}\}$ to simplify the notation.

The causal effect with data containing missing observations

Returning to the $U(Z)$-function (2.13) and the $U_{AIPW}(Z)$-function (2.17) from Chapter 2 and the $U_{j,k,l}(Z)$-function (3.3) from Chapter 3. Pick $r' \in \{1, \ldots, \mathbf{c}\}$. Manuscript I shows that the conditional expectation of the $U(Z)$-function (2.13) is given by $E\left(U(Z) \mid G_{r'}(Z), \boldsymbol{\zeta}_0\right) = E\left(\mu\{V_0, \boldsymbol{\gamma}\} \mid G_{r'}(Z), \boldsymbol{\zeta}_0\right) - E\left(Y^{\overline{a}_T}\right)$ with the set $G_{r'}(Z)$. The vector $\boldsymbol{\zeta}_0$ of true parameter values denotes the true model with respect to the distribution of Z. The estimator (4.1) with the $U(Z)$-function (2.13) for estimating $E(Y^{\overline{a}_T})$ is given by

$$\begin{aligned} \hat{E}\left(Y^{\overline{a}_T}\right) = \frac{1}{n} \sum_{i=1}^{n} \Bigg[& \frac{I(\mathcal{C}_i = \infty)}{\hat{\varpi}(\infty, Z_i)} \hat{\mu}(V_{0,i}) \\ & + \sum_{r=1}^{\mathbf{c}} \frac{I(\mathcal{C}_i = r) - \hat{\lambda}_r(G_r(Z_i)) I(\mathcal{C}_i \geq r)}{\hat{K}_r(G_r(Z_i))} E\left(\mu(V_0) \mid G_r(Z_i), \hat{\boldsymbol{\zeta}}\right) \Bigg]. \end{aligned} \tag{4.2}$$

Manuscript I denotes the estimator (4.2) with the letters DRMGf (*Doubly Robust estimator for Monotone missingness for the G-formula*). See Manuscript I for further information about the estimator. The difference between the SG estimator (2.12) and the DRMGf estimator (4.2) will become small if the included covariates are poor at predicting drop-out. We also show in the Appendix in Manuscript I that the estimator (4.2) is asymptotically normally distributed in the situation when T is equal to 1.

The conditional expectation of the $U_{AIPW}(Z)$-function (2.17) is given by

$$E(U_{AIPW}(Z) \mid G_{r'}(Z), \boldsymbol{\zeta}_0) = E(\Theta(Z) | G_{r'}(Z), \boldsymbol{\zeta}_0) - E\left(Y^{\overline{a}_T}\right)$$

with

$$\Theta(Z) = Y \prod_{t=0}^{T} \frac{I(a_t)}{\pi_t(V_t)} + \sum_{t=0}^{T} \prod_{k=0}^{t-1} \frac{I(a_k)}{\pi_k(V_k)} \left(1 - \frac{I(a_t)}{\pi_t(V_t)}\right) \mu(V_t)$$

with the set $G_{r'}(Z)$. The estimator (4.1) with the $U_{AIPW}(Z)$-function (2.17) for estimating

$E(Y^{\bar{a}_T})$ is given by

$$\hat{E}\left(Y^{\bar{a}_T}\right) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{I(\mathcal{C}_i=\infty)}{\hat{\varpi}(\infty,Z_i)}\left(Y_i\prod_{t=0}^{T}\frac{I(a_t)}{\hat{\pi}_t(V_{t,i})}+\sum_{t=0}^{T}\prod_{k=0}^{t-1}\frac{I(a_k)}{\hat{\pi}_k(V_{k,i})}\left(1-\frac{I(a_t)}{\hat{\pi}_t(V_{t,i})}\right)\hat{\mu}(V_{t,i})\right)\right.$$
$$\left.+\sum_{r=1}^{\mathbf{c}}\frac{I(\mathcal{C}_i=r)-\hat{\lambda}_r(G_r(Z_i))I(\mathcal{C}_i\geq r)}{\hat{K}_r(G_r(Z_i))}E\left(\Theta(Z)\mid G_r(Z_i),\hat{\boldsymbol{\zeta}}\right)\right]. \tag{4.3}$$

All the hats indicate predicted values from the specified models that have been used for the estimation and the predicted values are plugged into the estimator (4.3). The estimator is unbiased if the μ-models are correctly specified. The estimator is also unbiased if both the π-models and the λ-models are correctly specified simultaneously. We denote the estimator (4.3) by the name *Bang and Robins for data with Monotone Missingness* (BRMM). Table 4.1 shows the combinations of the μ-, the π- and the λ-models to obtain an unbiased estimator. We show in Appendix A that the BRMM estimator (4.3) is unbiased with a specific combination of the μ-, the π- and the λ-models (Table 4.1).

μ	Correct				Wrong			
π	Correct		Wrong		Correct		Wrong	
λ	Correct	Wrong	Correct	Wrong	Correct	Wrong	Correct	Wrong
Unbiased	✔	✔	✔	✔	✔	✗	✗	✗

Table 4.1: The combination of the μ-, the π- and the λ-models to obtain an unbiased BRMM estimator (4.3). See Table 2.1 for the description of the *Unbiased* row, the two symbols ✔ and ✗ and the *Correct* and *Wrong* columns.

Mediation analysis with data containing missing observations

Pick $r' \in \{1,\ldots,\mathbf{c}\}$ again. Manuscript III shows that the conditional expectation of the $U_{j,k,l}(Z)$-function (3.3) is given by $E\left(U_{j,k,l}(Z)\mid G_{r'}(Z)\right) = E\left(\mu_{j,k,l}\{V_0,\boldsymbol{\gamma}\}\mid G_{r'}(Z),\boldsymbol{\xi}\right) - \Gamma(j,k,l)$ with the set $G_{r'}(Z)$. The estimator $\hat{\Gamma}(j,k,l)$ for $\Gamma(j,k,l)$ is given by

$$\hat{\Gamma}(j,k,l) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{I(\mathcal{C}_i=\infty)}{\hat{\varpi}(\infty,Z_i)}\mu_{j,k,l}\{V_{0,i},\hat{\boldsymbol{\gamma}}\}\right.$$
$$\left.+\sum_{r=1}^{\mathbf{c}}\frac{I(\mathcal{C}_i=r)-\hat{\lambda}_r(G_r(Z_i))I(\mathcal{C}_i\geq r)}{\hat{K}_r(G_r(Z_i))}E\left(\mu_{j,k,l}\{V_{0,i},\boldsymbol{\gamma}\}\mid G_r(Z_i),\hat{\boldsymbol{\zeta}}\right)\right] \tag{4.4}$$

for data containing missing observations that follow a monotone pattern. All the hats indicate predicted values from the specified models that have been used for the estimation and the predicted values are plugged into the estimator (4.4). Manuscript III denotes the estimator (4.4) with the letters DRMSM (*Doubly Robust estimator for Monotone missingness for Sequential Mediation*). See Manuscript III for further information about the estimator.

4.3 Simulation study

We simulate missing observations to the data from Section 2.4 and the pattern is monotone missingness. The monotone missingness is simulated with the probabilities as follows: $\text{logit}(\lambda_1(G_1(Z_b))) = -2.5 + 0.4PHQ_b$, $\text{logit}(\lambda_2(G_2(Z_b))) = -1.9 + 0.9PHQ_b - 0.9PDQ_b$, $\text{logit}(\lambda_3(G_3(Z_b))) = -1.9+0.9PHQ_b-0.9PDQ_b-0.8SDS_b+1.6SDS_bPDQ_b$, $\text{logit}(\lambda_4(G_4(Z_b))) = -1.9+0.9PHQ_b-0.9PDQ_b-0.9SDS_b+0.5PHQ_2+0.6SDS_bPDQ_b$ and $\text{logit}(\lambda_5(G_5(Z_b))) = -2.0 + 0.9PHQ_b - 0.9PDQ_b - 0.6SDS_b + 0.3PHQ_2 + 0.9PDQ_2$. To analyse the simulated data, we only use the estimator for the simpler g-formula (SG) and our DRMGf estimator for estimating the causal effects $\boldsymbol{\beta}$. Both estimators are using the true μ-models with respect to the data. The SG estimator uses only complete data. We use the mean and the standard error to evaluate the estimation with the two estimators. Table 4.2 shows the results of the two estimators. We also use the BRMM estimator (4.3) for the estimation with the different combinations of the μ-, the π- and the λ-models. We also use the mean and the standard error to evaluate the estimation using the BRMM estimator (4.3) and Table 4.3 shows the results.

	SG				DRMGf			
	β_I	β_1	β_2	β_3	β_I	β_1	β_2	β_3
Mean	9.443	0.786	1.708	1.253	8.998	1.006	2.003	0.991
SE	0.095	0.133	0.167	0.220	0.090	0.129	0.154	0.216

Table 4.2: See Table 2.2 for the description of the two rows: *Mean* and *SE* and the true causal effects $\boldsymbol{\beta} = (\beta_I, \beta_1, \beta_2, \beta_3)$. The *SG* column shows the estimation with the SG estimator (2.12). The *DRMGf* column shows the estimation with our DRMGf estimator (4.2).

Table 4.2 shows that the estimates obtained using the SG estimator (2.12) are biased. The estimator shows weakness in estimating the parameters of interest when data contains missing observations that follow a monotone pattern. However, the DRMGf estimator (4.2) is able to estimate all four causal effects and the estimates are unbiased.

μ		Correct				Wrong			
π		Correct		Wrong		Correct		Wrong	
λ		Correct	Wrong	Correct	Wrong	Correct	Wrong	Correct	Wrong
Mean	β_I	8.998	8.998	8.997	8.997	9.021	9.322	8.782	9.121
	β_1	1.003	1.003	1.003	1.002	0.980	0.664	1.210	0.858
	β_2	2.003	2.003	2.001	2.002	1.985	1.766	2.479	2.180
	β_3	0.994	0.994	0.997	0.997	1.015	1.413	0.609	1.067
SE	β_I	0.144	0.114	0.161	0.124	0.549	0.305	0.685	0.387
	β_1	0.203	0.183	0.245	0.214	0.566	0.339	0.710	0.428
	β_2	0.330	0.242	0.257	0.198	0.822	0.475	0.776	0.433
	β_3	0.399	0.338	0.353	0.315	0.916	0.625	0.839	0.551

Table 4.3: See Table 2.2 for the description of the two rows: *Mean* and *SE* and the true causal effects $\boldsymbol{\beta} = (\beta_I, \beta_1, \beta_2, \beta_3)$. See Table 4.1 for the description of μ, π, λ, *Correct* and *Wrong*.

Table 4.3 shows the expected performance of the BRMM estimator (4.3) since the estimates in the first five columns are unbiased and it is in line with Table 4.1. All the standard errors are higher compared to the standard errors in Table 4.2 (our DRMGf estimator). The π-models may cause the standard errors to be higher.

The mediation analysis of the simulation study.

We use the estimator for the simpler sequential mediation formula (SSM) and our DRMSM estimator (4.4). The SSM estimator (3.2) uses only complete data. Both estimators are using the true $\mu_{j,k,l}$-models with respect to the simulated data. Table 4.4 shows the results of the estimation of the interventional direct and indirect effects for both estimators. We use the mean and the standard error to evaluate the two estimators.

	SSM				DRMSM			
	dir_b	$indir_{SDS_b}$	$indir_{PHQ_2}$	$Overall$	dir_b	$indir_{SDS_b}$	$indir_{PHQ_2}$	$Overall$
Mean	0.211	0.332	0.332	0.786	0.507	0.300	0.200	1.006
SE	0.162	0.068	0.068	0.133	0.148	0.056	0.046	0.129

Table 4.4: The *SSM* column shows the estimates obtained using the estimator for the simpler sequential mediation formula (SSM) and the *DRMSM* column shows the estimates obtained using our DRMSM estimator. See Table 2.2 for the description of the two rows: *Mean* and *SE*. See Table 3.1 for the true direct effect of PDQ_b on SDS_2, the true indirect effect of PDQ_b on SDS_2 via SDS_b and the true indirect effect of PDQ_b on SDS_2 via PHQ_2. See also Table 3.1 for the description of the four columns: dir_b, $indir_{SDS_b}$, $indir_{PHQ_2}$ and $Overall$.

Table 4.4 shows that our DRMSM estimator provides unbiased estimates and the SSM estimator shows weakness in estimating the parameters of interest when data contains missing observations that follow a monotone pattern. The results are in line with Manuscript III.

4.4 Simulation study based on the PERFORM study

We simulate missing observations to the data from Section 2.5 and the pattern is monotone missingness. The probabilities generating the missing observations are shown in Manuscript I. The simulated data are identical to the simulated data used in Manuscript I and Manuscript III. We use the BRMM estimator (4.3) for estimating the four expected potential outcomes: $E(SDS_2^{(1,1)})$, $E(SDS_2^{(0,1)})$, $E(SDS_2^{(1,0)})$ and $E(SDS_2^{(0,0)})$. The ϖ-model is only correctly specified. It is not possible to specify the λ-models wrong because the predictors for the probabilities are only linear. Table 4.5 shows the results of the estimation using the BRMM estimator (4.3). The estimation with the BRMM estimator is evaluated by the mean, the standard error (SE) and the absolute value of bias.

	μ	Correct		Wrong	
	π	Correct	Wrong	Correct	Wrong
Mean	$E(SDS_2^{(1,1)})$	15.422	15.420	15.405	15.411
	$E(SDS_2^{(0,1)})$	13.623	13.628	12.895	12.874
	$E(SDS_2^{(1,0)})$	13.671	13.650	13.610	13.701
	$E(SDS_2^{(0,0)})$	9.721	9.243	10.206	9.411
SE	$E(SDS_2^{(1,1)})$	3.748	3.756	3.657	3.685
	$E(SDS_2^{(0,1)})$	32.705	24.537	25.138	19.366
	$E(SDS_2^{(1,0)})$	6.878	6.333	8.030	6.899
	$E(SDS_2^{(0,0)})$	19.600	45.947	26.838	67.539
Bias	$E(SDS_2^{(1,1)})$	0.024	0.021	0.006	0.012
	$E(SDS_2^{(0,1)})$	0.205	0.210	0.523	0.544
	$E(SDS_2^{(1,0)})$	0.085	0.064	0.024	0.115
	$E(SDS_2^{(0,0)})$	0.096	0.574	0.389	0.406

Table 4.5: The estimates are obtained using the BRMM estimator (4.3) with the true ϖ-model. See Table 2.3 for the description of the three rows: *Mean*, *SE* and *Bias*, the two columns: *Correct* and *Wrong* and the true effects. Numerical problems occurred for the estimation when the π-models were correctly specified. It caused a reduction of 80 estimates. It means that the *Mean*, the *SE* and the *Bias* are based on 4920 estimates instead of 5000 estimates.

Table 4.5 shows that the estimates obtained using the BRMM estimator (4.3) are biased even though the μ-models and the π-models are correctly specified. The standard errors obtained using the BRMM estimator are also bigger compared to the standard errors (Table 3 in Manuscript I) obtained using our DRMGf estimator. The numerical problems were convergence problems for the π-models.

4.5 The choice of the U-function

We use the $U(Z)$-function (2.13) in the three manuscripts because we want the possibility of letting the λ-models be misspecified and still obtain an unbiased estimator. We will not use the $U_{AIPW}(Z)$-function (2.17) because it is possible to obtain biased estimates even though the μ-, the π- and the λ-models are correctly specified. The simulation study from Section 4.4 is a good example of biased estimates even when the μ-, the π- and the λ-models are correctly specified. The π-models and the λ-models need to be correctly specified simultaneously in the BRMM estimator (4.3) to obtain an unbiased estimator which means that we lose the opportunity for letting the λ-models be misspecified. We also obtain smaller standard errors with the $U(Z)$-function (2.13) compared to the $U_{AIPW}(Z)$-function (2.17). The $U_{IPW}(Z)$-function was already discarded in Chapter 2 because the IPW estimator (2.14) showed weakness in estimating the parameters of interest in the simulation study from Section 2.5. The estimates obtained using the IPW estimator were biased even though the π-models were correctly specified.

4.6 Nonmonotone missingness

We sketch an estimator to analyse data containing missing observations that follow a nonmonotone pattern. Let the missing observations be CAR. We will use Hilbert $\mathcal{H}$ spaces but we will not go into details about Hilbert $\mathcal{H}$ spaces. See Tsiatis (2006) for further information about Hilbert $\mathcal{H}$ spaces. Let $\mathcal{H}^F$ denote the full-data Hilbert space. Let $U^F(Z)$ denote the $U(Z)$-function on the full-data. Tsiatis (2006) defines a linear operator $\mathcal{L}$ that maps from the full-data Hilbert space to the observed-data Hilbert space and the linear operator $\mathcal{L}$ is defined by

$$\mathcal{L}\{U^F(Z), \boldsymbol{\zeta}\} = \sum_{r=1}^{\infty} I(\mathcal{C} = r) E\left(U^F(Z) \mid G_r(Z), \boldsymbol{\zeta}\right).$$

Tsiatis (2006) defines furthermore a linear operator $\mathcal{M}$ that maps from the full-data Hilbert space to the full-data Hilbert space. The linear operator $\mathcal{M}\{U^F(Z), \boldsymbol{\psi}, \boldsymbol{\zeta}\}$ is defined by $E(\mathcal{L}\{U^F(Z), \boldsymbol{\zeta}\} \mid Z)$ and the operator is given by

$$\mathcal{M}\{U^F(Z), \boldsymbol{\psi}, \boldsymbol{\zeta}\} = \sum_{r=1}^{\infty} \varpi\{r, G_r(Z), \boldsymbol{\psi}\} E\left(U^F(Z) \mid G_r(Z), \boldsymbol{\zeta}\right).$$

Tsiatis (2006) shows that the inverse operator $\mathcal{M}^{-1}\{U(Z), \boldsymbol{\psi}, \boldsymbol{\zeta}\}$ exists and the inverse operator is defined by $d^F(Z, \boldsymbol{\psi}, \boldsymbol{\zeta}) = \mathcal{M}^{-1}\{U(Z), \boldsymbol{\psi}, \boldsymbol{\zeta}\}$.

We need to solve the estimating equation

$$0 = \sum_{i=1}^{n} \left(\frac{I(\mathcal{C}_i = \infty)}{\varpi\{\infty, Z_i, \hat{\boldsymbol{\psi}}\}} U(Z_i) + L_2^*\{\mathcal{C}_i, G_{\mathcal{C}_i}(Z_i), \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\xi}}\} \right) \tag{4.5}$$

with

$$\begin{aligned} L_2^*\{\mathcal{C}_i, G_{\mathcal{C}_i}(Z_i), \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\zeta}}\} = &- \frac{I(\mathcal{C}_i = \infty)}{\varpi\{\infty, Z_i, \hat{\boldsymbol{\psi}}\}} \sum_{r \neq \infty} \varpi\{r, G_r(Z_i), \hat{\boldsymbol{\psi}}\} E\left(d^F(Z, \boldsymbol{\psi}, \boldsymbol{\zeta}) \mid G_r(Z_i), \hat{\boldsymbol{\zeta}}\right) \\ &+ \sum_{r \neq \infty} I(\mathcal{C}_i = r) E\left(d^F(Z, \boldsymbol{\psi}, \boldsymbol{\zeta}) \mid G_r(Z_i), \hat{\boldsymbol{\zeta}}\right). \end{aligned}$$

Unfortunately, it may be difficult to compute the inverse operator $\mathcal{M}^{-1}$ for nonmonotone missingness. An iterative approach is to approximate the inverse operator $\mathcal{M}^{-1}$ by the approximation $d^F_{(j)}(Z, \boldsymbol{\psi}, \boldsymbol{\zeta})$ after (j) iterations. Tsiatis (2006) defines

$$\begin{aligned} L_{2(j)}^*\{\mathcal{C}_i, G_{\mathcal{C}_i}(Z_i), \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\zeta}}\} = &- \frac{I(\mathcal{C}_i = \infty)}{\varpi\{\infty, Z_i, \hat{\boldsymbol{\psi}}\}} \sum_{r \neq \infty} \varpi\{r, G_r(Z_i), \hat{\boldsymbol{\psi}}\} E\left(d^F_{(j)}(Z, \boldsymbol{\psi}, \boldsymbol{\zeta}) \mid G_r(Z_i), \hat{\boldsymbol{\zeta}}\right) \\ &+ \sum_{r \neq \infty} I(\mathcal{C}_i = r) E\left(d^F_{(j)}(Z, \boldsymbol{\psi}, \boldsymbol{\zeta}) \mid G_r(Z_i), \hat{\boldsymbol{\zeta}}\right) \end{aligned}$$

and we replace $L_2^*\{\mathcal{C}_i, G_{\mathcal{C}_i}(Z_i), \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\zeta}}\}$ with $L_{2(j)}^*\{\mathcal{C}_i, G_{\mathcal{C}_i}(Z_i), \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\zeta}}\}$ in the estimating equation (4.5). We need to solve the following estimating equation

$$0 = \sum_{i=1}^{n} \left(\frac{I(\mathcal{C}_i = \infty)}{\varpi\{\infty, Z_i, \hat{\boldsymbol{\psi}}\}} U(Z_i) + L_{2(j)}^*\{\mathcal{C}_i, G_{\mathcal{C}_i}(Z_i), \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\xi}}\} \right). \tag{4.6}$$

The element $L_{2(j)}^*\{\mathcal{C}_i, G_{\mathcal{C}_i}(Z_i), \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\zeta}}\}$ lies in the Augmentation space even though the approximation $d_{(j)}^F(Z, \boldsymbol{\psi}, \boldsymbol{\zeta})$ is not equal to $\mathcal{M}^{-1}\{U(Z), \boldsymbol{\psi}, \boldsymbol{\zeta}\}$ (Tsiatis (2006)). See Tsiatis (2006) for further information about the Augmentation space. The estimating equation (4.6) is doubly robust if the conditional expectations $E(d^F(Z, \boldsymbol{\psi}, \boldsymbol{\zeta}) \mid G_r(Z), \boldsymbol{\zeta})$ are misspecified with respect to the distribution of Z or the models relating to the missingness mechanism are misspecified.

The inverse operator $\mathcal{M}^{-1}$ for monotone missingness

Tsiatis (2006) shows that the inverse operator $\mathcal{M}^{-1}$ for monotone missingness is given by

$$\mathcal{M}^{-1}\{U(Z), \boldsymbol{\psi}, \boldsymbol{\zeta}\} = U(Z) + \sum_{r \neq \infty} \frac{\lambda\{G_r(Z), \boldsymbol{\psi}\}}{K_r\{G_r(Z), \boldsymbol{\psi}\}} \left(U(Z) - E\left(U(Z) \mid G_r(Z), \boldsymbol{\zeta} \right) \right).$$

The conditional expectation of the inverse operator $E(d^F(Z, \boldsymbol{\psi}, \boldsymbol{\zeta}) \mid G_r(Z), \boldsymbol{\zeta})$ is equal to $E(U(Z) \mid G_r(Z), \boldsymbol{\zeta})$ because the sum is equal to zero. The sum is equal to zero because $\lambda_r\{G_r(Z), \boldsymbol{\psi}\}$, $K_r\{G_r(Z), \boldsymbol{\psi}\}$ and $E(U(Z) \mid G_r(Z), \boldsymbol{\zeta})$ are measurable with respect to the set $G_r(Z)$. The element $L_2^*\{\mathcal{C}, G_{\mathcal{C}}(Z), \boldsymbol{\psi}, \boldsymbol{\zeta}\}$ is given by

$$\begin{aligned}
L_2^*\{\mathcal{C}, G_{\mathcal{C}}(Z), \boldsymbol{\psi}, \boldsymbol{\zeta}\} = & - \frac{I(\mathcal{C} = \infty)}{\varpi\{\infty, Z, \boldsymbol{\psi}\}} \sum_{r \neq \infty} \varpi\{r, G_r(Z), \boldsymbol{\psi}\} E\left(U(Z) \mid G_r(Z), \boldsymbol{\zeta} \right) \\
& + \sum_{r \neq \infty} I(\mathcal{C} = r) E\left(U(Z) \mid G_r(Z), \boldsymbol{\zeta} \right) \\
= & \sum_{r \neq \infty} \left(I(\mathcal{C} = r) - \frac{I(\mathcal{C} = \infty)\varpi\{r, G_r(Z), \boldsymbol{\psi}\}}{\varpi\{\infty, Z, \boldsymbol{\psi}\}} \right) E\left(U(Z) \mid G_r(Z), \boldsymbol{\zeta} \right) \\
= & \sum_{r \neq \infty} \frac{I(\mathcal{C} = r) - \lambda\{G_r(Z), \boldsymbol{\psi}\} I(\mathcal{C} \geq r)}{K_r\{G_r(Z), \boldsymbol{\psi}\}} E\left(U(Z) \mid G_r(Z), \boldsymbol{\zeta} \right).
\end{aligned}$$

The estimating equation (4.5) is equal to the estimating equation that was used to obtain the estimator at (4.1). Tsiatis (2006) shows the equality

$$\sum_{r \neq \infty} \left(I(\mathcal{C} = r) - \frac{I(\mathcal{C} = \infty)\varpi\{r, G_r(Z), \boldsymbol{\psi}\}}{\varpi\{\infty, Z, \boldsymbol{\psi}\}} \right) = \sum_{r \neq \infty} \frac{I(\mathcal{C} = r) - \lambda\{G_r(Z), \boldsymbol{\psi}\} I(\mathcal{C} \geq r)}{K_r\{G_r(Z), \boldsymbol{\psi}\}}.$$

5 | Analysing the PERFORM study

Section 5.1 sketches the statistical analysis of the PERFORM study from Manuscript I and Manuscript III. Section 5.2 summaries briefly the results from Manuscript I and Manuscript III. An analysis of the three middle time points ($t = 2$, $t = 6$ and $t = 12$) of the PERFORM study is also conducted with the same methods that were introduced and used in Manuscript I and Manuscript III to analyse the two time points $t = b$ and $t = 18$. The results of the new analysis of the PERFORM study are then compared to the results from Manuscript I and Manuscript III. The reason for the additional analysis is caused by the clinical interest of the course of the disease over the two years. Analysing the three middle time points will connect the dots between the two separate analysis for the early ($t = b$) and the later ($t = 18$) time points and the course of the acute phase and the maintenance phase of the disease for patients with MDD may be revealed.

5.1 Statistical analysis

Patients are included in the analysis if depression severity at the prior time point before time t (PHQ_{pt}) is observed for $t \in \{2, 6, 12\}$. If the pattern of the missing observations is nonmonotone then the missing observations are forced to follow a monotone pattern. The missing observations are forced to follow a monotone pattern by setting the subsequent measurements to be missing in the ordered sequence. We use the dichotomized version of PDQ_t. If the original global score of PDQ_t is less than or equal to 5 then it corresponds to having no or minimal cognitive symptoms at time $t \in \{2, 6, 12, 18\}$. If the original global score of PDQ_t is strictly greater than 5 then it corresponds to having cognitive symptoms at time $t \in \{2, 6, 12, 18\}$. Let PDQ_t be equal to 0 if the patient has no or minimal cognitive symptoms and let PDQ_t be equal to 1 if the patient has cognitive symptoms at time $t \in \{2, 6, 12, 18\}$. We assume that the sequential conditional exchangeability is given by

$$SDS_{st}^{(pdq_t, pdq_{st})} \perp\!\!\!\perp PDQ_t \mid PHQ_t, SDS_{pt}, PDQ_{pt}, PHQ_{pt}$$

and

$$SDS_{st}^{(pdq_t, pdq_{st})} \perp\!\!\!\perp PDQ_{st} \mid PHQ_{st}, SDS_t, PDQ_t, PHQ_t$$

for $t \in \{2, 6, 12\}$ (Manuscript I). The marginal structural model is given by

$$E\left(SDS_{st}^{(pdq_t, pdq_{st})}\right) = \beta_{I,t} + \beta_{1,t} pdq_t + \beta_{2,t} pdq_{st} + \beta_{3,t} pdq_t pdq_{st}$$

and let $\boldsymbol{\beta}_t = (\beta_{I,t}, \beta_{1,t}, \beta_{2,t}, \beta_{3,t})$ denote the vector of causal effects at time $t \in \{2, 6, 12\}$. See Manuscript I for further information. Let $Z_{2,i} = (W_{b,i}, W_{2,i}, W_{6,i})$, let $Z_{6,i} = (W_{2,i}, W_{6,i}, W_{12,i})$ and let $Z_{12,i} = (W_{6,i}, W_{12,i}, W_{18,i})$ with $W_{t,i} = (PHQ_{t,i}, PDQ_{t,i}, SDS_{t,i})$. Our DRMGf estimator is compared to the *naïve* estimator, the *LSmeans* (SAS Institute Inc (2012)) estimator and the estimator for the simpler *g-formula* (SG) (Robins (1986)). We use all four estimators that were used in Manuscript I. See Manuscript I for the description of the naïve estimator, the LSmeans estimator, the estimator for the simpler g-formula (SG) and our DRMGf estimator. We also compare our DRMSM estimator to the estimator for the simpler sequential mediation formula (SSM) for the mediation analysis of $Z_{t,i}$ for $t \in \{2, 6, 12\}$. We use all four estimators that were used in Manuscript III (the two estimators for the mediation analysis and the two estimators that were used to estimate the causal effects $\boldsymbol{\beta}_t = (\beta_{I,t}, \beta_{1,t}, \beta_{2,t}, \beta_{3,t})$). See Manuscript III for the description of the four estimators. All the patients with the pattern $G_4(Z_t)$ and $G_7(Z_t)$ for $t \in \{2, 6, 12\}$ are removed from the data. The vector $G_4(Z_t)$ is the four measurements given by $(PHQ_{pt}, PDQ_{pt}, SDS_{pt}, PHQ_t)$ and the vector $G_7(Z_t)$ is the seven measurements given by $(W_{pt}, PHQ_t, PDQ_t, SDS_t, PHQ_{st})$ with $W_{pt} = (PHQ_{pt}, PDQ_{pt}, SDS_{pt})$. The patients are removed from the data because we want to avoid numerical problems for the estimation using the μ-models and the λ-models. The two λ-models for $G_4(Z_t)$ and $G_7(Z_t)$ are given by $\lambda_4(G_4(Z_t)) = \lambda_7(G_7(Z_t)) = 0$ for $t \in \{2, 6, 12\}$. The confidence intervals are obtained using 1000 bootstraps. However, the analysis for $t = 2$ had some numerical problems. These numerical problems were caused by the bootstrap sample was not large enough to estimate the parameter of the interaction between PDQ_t and PDQ_{st}. The confidence intervals in Figure 5.1a (the analysis for $t = 2$) are obtained using 927 bootstraps. All four estimators failed in 73 bootstraps. The confidence intervals in Table 5.2 and Table B.4 for $t = 2$ are obtained using 954 bootstraps because all four estimators failed in 46 bootstraps.

5.2 Result

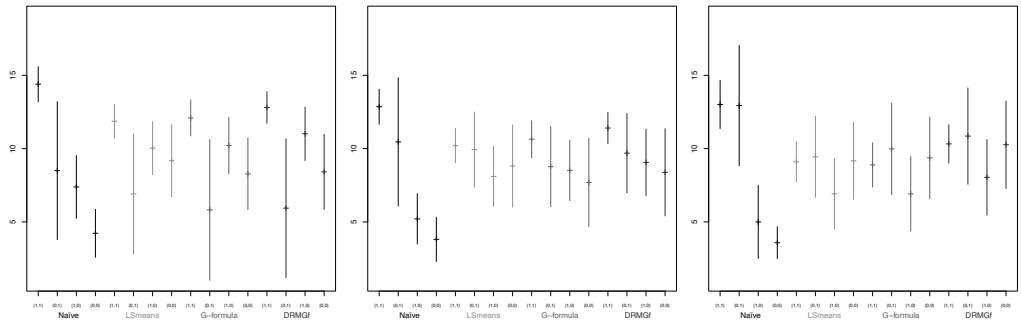
Manuscript I shows the results of the expected value of the four different counterfactual levels for both the early ($t = b$) and the later ($t = 18$) time points of the PERFORM study with four different estimators (the naïve estimator, the LSmeans estimator, the estimator for the simpler g-formula (SG) and our DRMGf estimator). The analysis for both time points ($t = b$ and $t = 18$) shows that patients with cognitive symptoms at both visits have worse functioning than patients with no or with minimal cognitive symptoms at both visits. It applies for all four estimators. The differences in the expected outcomes between the four groups of exposures shrink when the estimators adjust for confounding. The pattern is most pronounced for the naïve estimator compared to the other three estimators. The analysis also shows that the difference between the estimator for the simpler g-formula (SG) and our DRMGf estimator is

surprisingly small. The included covariates are poor at predicting drop-out. This is causing the estimates to be almost similar for the two estimators. See Manuscript I for further information.

Manuscript III shows the results of the mediation analysis of the PERFORM study using the estimator for the simpler sequential mediation formula (SSM) and our DRMSM estimator. The results show that the estimates are almost similar for the two estimators. The similarities of the estimates for the two estimators may be caused by the included covariates are poor at predicting drop-out. The direct effect of cognitive symptoms on functional impairment at a later time is positive for $t = b$. This applies for both estimators (SSM and DRMSM). The direct effect of cognitive symptoms on functional impairment at a later time is negative for $t = 18$. This applies for both estimators (SSM and DRMSM). The negative direct effect for the later time point ($t = 18$) may be caused by many patients are doing very well after 18 months and we see the effect that patients with cognitive symptoms are the only ones who can improve their functioning. The scales do not allow for further improvement among the patients if the patients already have the lowest score on the scale. See Manuscript III for further information.

The results of expected score of functional impairment

Figure 5.1 shows the results of the estimates of the expected value of the four different counterfactual levels for the three time points, $t \in \{2, 6, 12\}$. Each plot shows the estimates with the confidence intervals. See Manuscript I for further information. The actual numbers are shown in Table B.1, Table B.2 and Table B.3 in Appendix B.



(a) Analysis for $t = 2$. (b) Analysis for $t = 6$. (c) Analysis for $t = 12$.

Figure 5.1: The horizontal lines are the estimates. The verticals lines are the 95% confidence intervals. The confidence intervals are obtained using 1000 bootstraps. *The confidence intervals for the analysis for* $t = 2$ *are obtained using* 927 *bootstraps since all four estimators had numerical problems.* The actual numbers are shown in Appendix B. See Figure 2 in Manuscript I for further information.

The results of the analysis of the three time points $t \in \{2, 6, 12\}$ are almost similar to the results of the analysis for t equal to b and 18. The large variations in the expected outcomes

between the four combinations of cognitive deficits for the naïve estimator are similar to the results in Manuscript I. The differences in the expected outcomes between the four groups of exposures shrink when the estimators adjust for confounding similar to the analysis in Manuscript I. All four estimators suggest that patients with cognitive symptoms at both visits have worse functioning than patients with no or with minimal cognitive symptoms at both visits. However, this does not apply for the analysis for t equal to 12. The three estimators: LSmeans, g-formula (SG) and DRMGf show (the analysis $t = 12$) that patients having no or minimal cognitive symptoms at visit 12 and having cognitive symptoms at the subsequent visit have worse functioning than patients having cognitive symptoms at both visits. The results show that the impact of cognition on functioning is worse for the patient if the patient experiences a relapse of cognitive symptoms compared to have stable cognitive symptoms over the two years.

The analysis shows that the presence of cognitive symptoms will cause poor functioning and the pattern is most pronounced for the naïve estimator. The results are not surprising since the analysis from Manuscript I shows similar results.

The results from the estimator for the simpler g-formula (SG) and our DRMGf estimator are also not surprising. We know from the analysis in Manuscript I that the included covariates are poor at predicting drop-out and this is causing the small difference between the two estimators. Table 5.1 shows the minimum and maximum of the predicted values of each $\lambda_r\{G_r(Z), \hat{\boldsymbol{\psi}}\}$ for the three time points $t \in \{2, 6, 12\}$.

Analysis	Range	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	ϖ
$t = 2$	Minimum	0.200	0.219	0.171	0	0.060	0.042	0	0.041	0.112
	Maximum	0.225	0.253	0.314	0	0.290	0.533	0	0.385	0.391
$t = 6$	Minimum	0.158	0.207	0.199	0	0.065	0.068	0	0.039	0.112
	Maximum	0.173	0.330	0.347	0	0.378	0.369	0	0.533	0.352
$t = 12$	Minimum	0.159	0.171	0.155	0	0.064	0.086	0	0.036	0.135
	Maximum	0.227	0.354	0.361	0	0.481	0.413	0	0.334	0.386

Table 5.1: The minimum and maximum of the predicted values of each $\lambda_r\{G_r(Z), \hat{\boldsymbol{\psi}}\}$ for the three time points $t \in \{2, 6, 12\}$. See Table 2 in Manuscript I for further information.

Table 5.1 shows almost similar ranges of the minimum and maximum of the predicted values of $\lambda_r\{G_r(Z_t), \boldsymbol{\psi}\}$ as Table 2 in Manuscript I. The narrow ranges cause the estimator for the simpler g-formula (SG) and our DRMGf estimator to provide almost the same estimates. The similarity of the two estimators have been explored with simulation studies. The results of the analysis with the three time points are in line with Manuscript I.

The results of the mediation analysis

Table 5.2 shows the estimates of the interventional direct effect and the interventional indirect effects of cognitive symptoms on functional impairment at a later time with the confidence intervals for the three time points ($t \in \{2, 6, 12\}$). The estimate of the causal effect $\beta_{1,t}$ with

the confidence interval is also shown in the Table for the three time points ($t \in \{2, 6, 12\}$). The coefficient $\beta_{1,t}$ is the causal effect of cognitive symptoms on functional impairment at a later time for the time point $t \in \{2, 6, 12\}$. We use the same arguments from Section 3.3 when we compare the overall effect to $\beta_{1,t}$ for the comparison between the overall effect and the total causal effect. We have only shown the estimate of the coefficient $\beta_{1,t}$ in Table 5.2 because we only need $\beta_{1,t}$ for the comparison between the overall effect and the total causal effect. Table B.4 (in Appendix B) shows the estimates of the causal effects $\boldsymbol{\beta}_t = (\beta_{I,t}, \beta_{1,t}, \beta_{2,t}, \beta_{3,t})$ with the confidence intervals.

		SSM/SG			DRMSM/DRMGf		
Analysis		Effect	SE	95%-CI	Effect	SE	95%-CI
Z_2	dir_2	1.006	1.383	(-1.704 ; 3.715)	1.197	1.447	(-1.639 ; 4.033)
	$indir_{SDS_2}$	0.271	0.608	(-0.920 ; 1.462)	0.910	0.492	(-0.054 ; 1.873)
	$indir_{PHQ_6}$	0.658	0.595	(-0.507 ; 1.824)	0.491	0.590	(-0.666 ; 1.648)
	$Overall$	1.935	1.573	(-1.148 ; 5.018)	2.598	1.601	(-0.541 ; 5.736)
	$\beta_{1,2}$	1.935	1.573	(-1.148 ; 5.018)	2.598	1.601	(-0.541 ; 5.736)
Z_6	dir_6	-0.571	1.414	(-3.343 ; 2.202)	-0.769	1.487	(-3.683 ; 2.145)
	$indir_{SDS_6}$	0.637	0.438	(-0.221 ; 1.495)	0.706	0.365	(-0.009 ; 1.422)
	$indir_{PHQ_{12}}$	0.762	0.694	(-0.599 ; 2.123)	0.736	0.655	(-0.548 ; 2.020)
	$Overall$	0.828	1.692	(-2.489 ; 4.145)	0.673	1.698	(-2.654 ; 4.001)
	$\beta_{1,6}$	0.828	1.692	(-2.489 ; 4.145)	0.673	1.698	(-2.654 ; 4.001)
Z_{12}	dir_{12}	-2.437	1.651	(-5.674 ; 0.800)	-2.447	1.877	(-6.127 ; 1.232)
	$indir_{SDS_{12}}$	0.442	0.227	(-0.003 ; 0.887)	0.377	0.214	(-0.043 ; 0.796)
	$indir_{PHQ_{18}}$	-0.464	0.556	(-1.553 ; 0.625)	-0.154	0.516	(-1.165 ; 0.857)
	$Overall$	-2.459	1.651	(-5.694 ; 0.777)	-2.225	1.885	(-5.919 ; 1.470)
	$\beta_{1,12}$	-2.459	1.651	(-5.694 ; 0.777)	-2.225	1.885	(-5.919 ; 1.470)

Table 5.2: The *SSM/SG* column shows the estimates obtained using the estimator for the simpler sequential mediation formula (SSM) and the estimator for the simpler g-formula (SG). The *DRMSM/DRMGf* column shows the estimates obtained using our two estimators DRMSM and DRMGf. The *Effect* column shows the estimated effects. The *SE* column shows the standard error for each estimate. See Table 3.2 for further information about the five rows: dir_t, $indir_{SDS_t}$, $indir_{PHQ_{st}}$, $Overall$ and $\beta_{1,t}$ for $t \in \{2, 6, 12\}$. The standard errors are obtained using 1000 bootstraps. The 95%-*CI* column shows the 95% confidence intervals. *The confidence intervals for the analysis for* $t = 2$ *are obtained using* 954 *bootstraps because the estimators had numerical problems.*

Table 5.2 shows that the estimates for the two estimators (the estimator for the simpler sequential mediation formula (SSM) and our DRMSM estimator) are almost similar. The difference between the two estimators is not surprising since the included covariates are poor at predicting drop-out. This causes the estimates for the two estimators to be almost similar. We notice that the overall effect is equal to the total causal effect for all six analysis. The results are similar to the results in Manuscript III. At first sight, the negative effects can be surprising since a negative effect indicates that patients with cognitive symptoms at time t (equal to 2, 6 and 12) are more likely to improve directly their functioning in contrast to patients with no or with minimal cognitive symptoms. We had expected positive estimates. However, the negative

effects may indicate the ceiling and floor effect that more patients become better over time and it becomes more difficult for patients with no or with minimal cognitive symptoms to further improve if the patients already have the lowest score on the scale. The results are in line with Manuscript III.

6 | Discussion

This thesis contributes to the statistical methodology of causal inference to analyse longitudinal data with repeated measurements and missing observations.

The aim of this thesis was to develop an estimator to analyse longitudinal data with time-dependent confounding and missing observations that follow a monotone pattern. An estimator for mediation analysis for multiple mediators and missing observations that follow a monotone pattern was also needed. The aim was to utilize data better instead of using only complete cases and reducing bias of the estimates when data contains missing observations that follow a monotone pattern. The work in this thesis was motivated by the PERFORM study since the study has substantial many missing observations. We were interested in the effect of cognitive symptoms on functional impairment at a later time. An analysis using only complete cases was not a satisfactory analysis of the PERFORM study. We have therefore proposed a doubly robust estimator (DRMGf) in Manuscript I to estimate the causal effect of a time-varying exposure in the presence of time-dependent confounding when data contains missing observations that follow a monotone pattern. We have also proposed a doubly robust estimator for sequential mediation (DRMSM) in Manuscript III when data contains missing observations that follow a monotone pattern. The overall effect obtained using the DRMSM estimator is equal to the total causal effect. The estimator is based on the new definition of sequential mediation proposed in Manuscript II. The two estimators and the new definition were applied to the observational cohort study PERFORM consisting of patients with depression.

Our DRMGf estimator and three existing estimators from the literature were applied to the PERFORM study. The difference of the expected outcomes between the four exposure groups became smaller as the confounders were included in the models. Our DRMGf estimator was compared to the estimator for the simpler g-formula (SG). The results of our DRMGf estimator and the estimator for the simpler g-formula (SG) were surprisingly similar. The similarities of the two estimators were caused by the included covariates were poor at predicting drop-out. The robustness of our estimator was not revealed in the analysis of the PERFORM study. We also did a mediation analysis of the PERFORM study. The doubly robust estimator (DRMSM) for sequential mediation was used to analyse the PERFORM study. Our DRMSM estimator was compared to the estimator for the simpler sequential mediation formula (SSM) in the analysis of the PERFORM study. The robustness of our two estimators DRMGf and DRMSM were not revealed in the analysis of the PERFORM study. However, the simulation studies revealed

the robustness of our two estimators (DRMGf and DRMSM). We must remember that we have only considered a simplified version of the PERFORM study. We might have seen a bigger difference between our DRMGf estimator and the SG estimator or our DRMSM estimator and the SSM estimator if we had included age, gender, disease history and other relevant predictors. The results from the analysis for all $t \in \{b, 2, 6, 12, 18\}$ reflect that the treatment goal in the first months of treatment (the acute phase) is to relieve the depressive symptoms while the goal later in treatment (the maintenance phase) shifts towards stabilisation and relapse prevention. It appears from the analysis of the expected value of the four different counterfactual levels that the impact of cognition on functioning is worse for a patient if the patient experiences a relapse of cognitive symptoms compared to having stable cognitive symptoms. The mediation analysis may indicate that the ceiling and floor effect occur in the PERFORM study since the study is scale based and each scale has a finite range. It may happen that almost all patients have the highest score in the beginning of the study which cause the ceiling effect and then some months later almost all patients have the lowest score which cause the floor effect. It is easier to see an improvement among the patients if almost all/many patients have a high score and it becomes more difficult to detect any improvement if almost all/many patients have a low score. We are not able to see any improvement of the patient if the patient's score is already on the lower boundary of the scale compared to a patient with a higher score. The scales do not leave any room for further improvement if the patient is already doing well because the scales have a lower finite limit. It is a limitation for scales based studies.

It is now tempting to only use the estimator for the simpler g-formula (SG) instead of our DRMGf estimator when data contains missing observations. There was almost no difference between the two estimators in the analysis of the PERFORM study. The estimator for the simpler g-formula (SG) is also easier to use compared to our DRMGf estimator. However, the simulation studies (this thesis, Manuscript I and Manuscript III) have shown that we obtain biased estimates using only complete cases if the included covariates are strong at predicting drop-out. The simulation studies with the missing observations have revealed the robustness of our DRMGf estimator. The estimates obtained using the estimator for the simpler g-formula (SG) were biased when the included covariates were strong at predicting drop-out. The overall interpretation of the simulation studies is that our DRMGf estimator is better of estimating the parameters of interest than the estimator for the simpler g-formula (SG). The simulation studies with the missing observations have also revealed that our DRMGf estimator is better of estimating the parameters of interest than the BRMM estimator (4.3) (the extended version of the doubly robust estimator introduced by Bang and Robins (2005) when data contains missing observations). The estimates obtained using the BRMM estimator (4.3) were biased even though the μ-, the π- and the λ-models were correctly specified. This is not surprising because the doubly robust estimator (2.16) showed already a bit of weakness compared to the estimator for the simpler g-formula (SG) with full data in Section 2.5. The conclusion is that the BRMM estimator (4.3) should be avoided because it is possible to obtain biased estimates even when the μ-models and the π-models are correctly specified. The simulation studies with the missing observations have also revealed the robustness of our DRMSM estimator. The estimates obtained using the estimator for the simpler sequential mediation formula (SSM) were biased

when the included covariates were strong at predicting drop-out. The overall interpretation of the simulation studies is that our DRMSM estimator is better of estimating the parameters of interest than the estimator for the simpler sequential mediation formula (SSM).

We are able to obtain the overall effect to be equal to the total causal effect using our new definition of sequential mediation. The simulation studies have revealed that our definition is able to obtain the true effects (Manuscript II). Our definition is able to obtain the interventional direct effect and the interventional indirect effects for multiple mediators so that the overall effect is equal to the total causal effect. Our definition allows the included models to have interactions between the different measurements/variables. Our definition was compared to an already existing definition by VanderWeele and Vansteelandt (2013) in three simulation studies and our definition was better of estimating the parameters of interest compared to the existing definition.

The conclusion is that our two estimators DRMGf and DRMSM protect against biased estimates compared to the two simpler estimators (the estimator for the simpler g-formula (SG) and the estimator for the simpler sequential mediation formula (SSM)). Our two estimators (DRMGf and DRMSM) also utilize data better compared to the two simpler estimators (SG and SSM). The assumption about the missingness is weaker for our two estimators (DRMGf and DRMSM) compared to the two simpler estimators. Our two estimators (DRMGf and DRMSM) hinge on the assumption missing at random which is less strict than the assumption missing completely at random (the two estimators SG and SSM hinge on the assumption missing completely at random). The two estimators SG and SSM show weakness in estimating the parameters of interest when data contains missing observations that follow a monotone pattern. The cost that we may have to pay with our two estimators is larger standard errors. Our two estimators DRMGf and DRMSM share the same advantages, disadvantages and limitations.

6.1 Perspectives and limitations

The consistency assumption and the conditional exchangeability assumption for no unmeasured confounding are a limitation for the two estimators and the definition. These two assumptions are untestable. Another limitation for the two estimators (DRMGf and DRMSM) and the definition of sequential mediation is the ordering of the variables because it may be a natural assumption for some studies. However, the order of the different domains in the PERFORM study is partially clear because the three domains were measured simultaneously for each patient. The causal ordering between the different time points is introduced by time itself but the causal ordering between the different domains within the same time point is less clear. The causal ordering between the different domains within the same time point is based on clinical insight (a change in depression severity causes a change in cognitive performance, which in turn causing a change in functioning). The assumption of monotone missingness is a realistic assumption since patients tend to drop-out of studies but it does not allow for intermittent missing data. It may be unrealistic to assume that the missing observations in the PERFORM

study follow a monotone pattern and the missing observations are missing at random. However, to analyse a subset of the PERFORM study with only fully observed patients is not less critical because an analysis using only complete cases depends on the assumption missing completely at random.

The interpretation of the results hinges on our assumptions but they are not guaranteed and the data will not provide us with any information whether the assumptions are correct or not. The assumption of monotone missingness leaves room for further research for extending the two estimators for data containing missing observations that follow a nonmonotone pattern. Further research is to obtain the influence function of the g-formula instead of the $U(Z)$-function (2.13) that has been used in this thesis. The influence function of the g-formula will make it possible to obtain the asymptotic variance instead of estimating the variance by bootstrapping.

7 | Summary of manuscripts

7.1 Summary of Manuscript I

A doubly robust estimator for monotone missing data in the presence of time-dependent confounding

This manuscript presents a doubly robust estimator for estimating the causal effect in the presence of time-dependent confounding while data contains missing observations. The estimator allows the models relating to the missingness mechanism in the data to be misspecified and the estimator will still be unbiased. The estimator is an extension of the g-formula. Our estimator utilizes data better compared to an estimator using only complete data since our estimator allows partially observed vectors to be included in the analysis without reducing data to a subset of complete cases. The observational study Prospective Epidemiological Research on Functioning Outcomes Related to Major depressive disorder (PERFORM) was the motivation to develop the estimator. Our doubly robust estimator was used to estimate the causal effect of cognitive symptoms on functional impairment at a later time while data contains missing observations. However, the analysis of the PERFORM study did not reveal the robustness of our estimator.

Manuscript I is submittable.

7.2 Summary of Manuscript II

Sequential mediation analysis with multiple mediators

This manuscript proposes a new definition of sequential mediation to obtain the interventional direct effect and the interventional indirect effects for multiple mediators. The overall effect with our definition is equal to the total causal effect. The definition is inspired by Vansteelandt and Daniel (2017) for decomposing the total causal effect into the interventional direct and

indirect effects for multiple mediators. The observational study Prospective Epidemiological Research on Functioning Outcomes Related to Major depressive disorder (PERFORM) was the motivation to develop the definition since we are interested in the direct effect of cognitive symptoms on functional impairment at a later time, and we prefer that the overall effect is equal to the total causal effect for the sake of the interpretation.

Manuscript II is submittable.

7.3 Summary of Manuscript III

Sequential mediation analysis with multiple mediators for data with missing observations

This manuscript presents a doubly robust estimator for estimating the interventional direct effect and the interventional indirect effects for multiple mediators while data contains missing observations. We have developed an estimator that allows the models relating to the missingness mechanism in the data to be misspecified and the estimator will still be unbiased. The estimator is an extension of the definition of sequential mediation from Manuscript II. Our estimator utilizes data better compared to an estimator using only complete data since our estimator allows partially observed vectors to be included in the analysis. The observational study Prospective Epidemiological Research on Functioning Outcomes Related to Major depressive disorder (PERFORM) was the motivation to develop the estimator. Our doubly robust estimator was used for estimating the direct effect of cognitive symptoms on functional impairment at a later time and the mediated effects. The analysis of the PERFORM study did not reveal the robustness of our estimator.

Manuscript III is submittable.

A | The doubly robust estimator

The estimator (2.16) by Bang and Robins (Part of Chapter 2): We show that the estimator (2.16) is a doubly robust estimator but we begin the Appendix by showing that the mean $E(\Upsilon_T(Z))$ is equal to zero. We assume that the μ-models and the π-models are correctly specified. We also assume conditional exchangeability and consistency. We have the equality

$$E\left(\left(1-\frac{I(a_T)}{\pi_T(V_T)}\right)m\{V_T,\boldsymbol{\xi}\}\middle|\overline{L}_T,\overline{A}_{T-1}\right)=E\left(\left(1-\frac{I(a_T)}{\pi_T(V_T)}\right)E(Y^{\bar{a}_T}|\overline{L}_T,\overline{A}_{T-1})\middle|\overline{L}_T,\overline{A}_{T-1}\right)$$

because we have assumed conditional exchangeability and consistency. We notice the following equalities

$$\begin{aligned}E\left(\frac{I(a_T)}{\pi_T(V_T)}E(Y^{\bar{a}_T}|\overline{L}_T,\overline{A}_{T-1})\middle|\overline{L}_T,\overline{A}_{T-1}\right)&=\frac{E(I(a_T)\mid\overline{L}_T,\overline{A}_{T-1})}{\pi_T(V_T)}E(Y^{\bar{a}_T}|\overline{L}_T,\overline{A}_{T-1})\\&=\frac{\pi_T(V_T)}{\pi_T(V_T)}E(Y^{\bar{a}_T}|\overline{L}_T,\overline{A}_{T-1})\end{aligned}$$

where the conditional expectation $E(I(a_T)\mid\overline{L}_T,\overline{A}_{T-1})$ is equal to $\pi_T(V_T)$ such that

$$E\left(\left(1-\frac{I(a_T)}{\pi_T(V_T)}\right)m\{V_T,\boldsymbol{\xi}\}\middle|\overline{L}_T,\overline{A}_{T-1}\right)=0.$$

Now, we can show that the mean $E(\Upsilon_T(Z))$ is equal to zero by the following calculations

$$\begin{aligned}E\left(\Upsilon_T(Z)\right)&=E\left(E\left(\prod_{t=0}^{T-1}\frac{I(a_t)}{\pi_t(V_t)}\left(1-\frac{I(a_T)}{\pi_T(V_T)}\right)m(V_T)+\Upsilon_{T-1}(Z)\middle|\overline{L}_T,\overline{A}_{T-1}\right)\right)\\&=E\left(\prod_{t=0}^{T-1}\frac{I(a_t)}{\pi_t(V_t)}E\left(\left(1-\frac{I(a_T)}{\pi_T(V_T)}\right)m(V_T)\middle|\overline{L}_T,\overline{A}_{T-1}\right)+\Upsilon_{T-1}(Z)\right)\\&=E\left(\Upsilon_{T-1}(Z)\right).\end{aligned}$$

The mean $E(\Upsilon_T(Z))$ is equal to zero because we have shown that the mean $E(\Upsilon_T(Z))$ is equal to $E(\Upsilon_{T-1}(Z))$ and it is possible to show the equality $T-1$ times more and the last mean $E(\Upsilon_0(Z))$ is equal to zero. To show that the estimator (2.16) is doubly robust then

we need to show that the estimator is unbiased when either the μ-models or the π-models are correctly specified. We assume that the π-models are correctly specified and we let the μ-models be misspecified denoted by μ^*. We let $\Upsilon_T^{\mu^*}(Z)$ denote the Υ-term when the μ-models are misspecified. The mean $E(U_{AIPW}(Z))$ is given by

$$\begin{aligned} E(U_{AIPW}(Z)) &= E\left(Y\prod_{t=0}^{T}\frac{I(a_t)}{\pi_t(V_t)} + \sum_{t=0}^{T}\prod_{k=0}^{t-1}\frac{I(a_k)}{\pi_k(V_k)}\left(1-\frac{I(a_t)}{\pi_t(V_t)}\right)\mu^*(V_t)\right) - E\left(Y^{\bar{a}_T}\right) \\ &= E(\Upsilon_T^{\mu^*}(Z)) \end{aligned}$$

which is zero because the expectation $E(\Upsilon_T^{\mu^*}(Z))$ is equal to zero since the π-models are correctly specified. Showing that the mean $E(\Upsilon_T^{\mu^*}(Z))$ is equal to zero are almost the same as showing that the mean $E(\Upsilon_T(Z))$ is equal to zero (replace μ with μ^* in the calculations above). We assume that the μ-models are correctly specified and we let the π-models be misspecified denoted by π^*. The mean $E(U_{AIPW}(Z))$ is given by

$$\begin{aligned} E(U_{AIPW}(Z)) &= E\left(Y\prod_{t=0}^{T}\frac{I(a_t)}{\pi_t^*(V_t)} + \sum_{t=0}^{T}\prod_{k=0}^{t-1}\frac{I(a_k)}{\pi_k^*(V_k)}\left(1-\frac{I(a_t)}{\pi_t^*(V_t)}\right)\mu(V_t)\right) - E\left(Y^{\bar{a}_T}\right) \\ &= E\left(Y\prod_{t=0}^{T}\frac{I(a_t)}{\pi_t^*(V_t)} + \sum_{t=1}^{T}\prod_{k=0}^{t-1}\frac{I(a_k)}{\pi_k^*(V_k)}\mu(V_t) - \sum_{t=0}^{T}\prod_{k=0}^{t}\frac{I(a_k)}{\pi_k^*(V_k)}\mu(V_t)\right) \\ &= E\left(\prod_{t=0}^{T}\frac{I(a_t)}{\pi_t^*(V_t)}(Y-\mu(V_T)) + \sum_{t=1}^{T}\prod_{k=0}^{t-1}\frac{I(a_k)}{\pi_k^*(V_k)}\mu(V_t) - \sum_{t=0}^{T-1}\prod_{k=0}^{t}\frac{I(a_k)}{\pi_k^*(V_k)}\mu(V_t)\right) \\ &= E\left(\prod_{t=0}^{T}\frac{I(a_t)}{\pi_t^*(V_t)}(Y-m(V_T)) + \sum_{t=1}^{T}\prod_{k=0}^{t-1}\frac{I(a_k)}{\pi_k^*(V_k)}(\mu(V_t)-\mu(V_{t-1}))\right) \end{aligned}$$

which is zero using the law of iterated conditional expectations.

The BRMM estimator (Part of Chapter 4): We show that the BRMM estimator (4.3) is robust in case of misspecification of the included models (the μ-, the π- and the λ-models) according to Table 4.1. We assume that the μ-models or the π-models are correctly specified for now. We need to show that the expectation

$$E\left(Y\prod_{t=0}^{T}\frac{I(a_t)}{\pi_t\{V_t,\boldsymbol{\alpha}\}} + \Upsilon_T(Z) - E\left(Y^{\bar{a}_T}\right) + \sum_{r=1}^{\mathbf{c}}\frac{I(\mathcal{C}=r)-\lambda_r\{G_r(Z),\boldsymbol{\psi}\}I(\mathcal{C}\geq r)}{K_r\{G_r(Z),\boldsymbol{\psi}\}}\Xi_T(Z)\right) \quad \text{(A.1)}$$

is equal to zero for a specific combination of the included models (the μ-, the π- and the λ-models) with

$$\Xi_T(Z) = E\left(Y\prod_{t=0}^{T}\frac{I(a_t)}{\pi_t\{V_t,\boldsymbol{\alpha}\}} + \Upsilon_T(Z)\middle| G_r(Z),\boldsymbol{\zeta}\right) - \left(Y\prod_{t=0}^{T}\frac{I(a_t)}{\pi_t\{V_t,\boldsymbol{\alpha}\}} + \Upsilon_T(Z)\right).$$

It is possible to rewrite the $\Xi_T(Z)$-term to the expression $E(U_{AIPW}(Z) \mid G_r(Z), \boldsymbol{\zeta}) - U_{AIPW}(Z)$. The expectation of the part before the sum $\sum_{r=1}^{\mathbf{c}}(\ldots)$ in (A.1) is equal to zero by construction because the μ-models or the π-models are correctly specified. We need to show that the expectation of the sum $\sum_{r=1}^{\mathbf{c}}(\ldots)$ in (A.1) is equal to zero. We define the set $\mathcal{F}_r = \sigma(I(\mathcal{C} = 1), \ldots, I(\mathcal{C} = r-1), Z)$ and we obtain the equality $E(I(\mathcal{C} = r) \mid \mathcal{F}_r) = \lambda_r\{G_r(Z), \boldsymbol{\psi}\} I(\mathcal{C} \geq r)$ (Tsiatis (2006)). We obtain by conditioning on the set $\mathcal{F}_r$ the following equality

$$E\left(\frac{I(\mathcal{C} = r) - \lambda_r\{G_r(Z), \boldsymbol{\psi}^*\} I(\mathcal{C} \geq r)}{K_r\{G_r(Z), \boldsymbol{\psi}^*\}} \Xi_T(Z)\right) = E\left(\mathcal{S}\left(G_r(Z)\right) I(\mathcal{C} \geq r) \Xi_T(Z)\right)$$

with

$$\mathcal{S}\left(G_r(Z)\right) = \frac{\lambda_r\{G_r(Z), \boldsymbol{\psi}_0\} - \lambda_r\{G_r(Z), \boldsymbol{\psi}^*\}}{K_r\{G_r(Z), \boldsymbol{\psi}^*\}}.$$

If the models relating to the missingness mechanism are correctly specified $\boldsymbol{\psi}^* = \boldsymbol{\psi}_0$ then the $\mathcal{S}(G_r(Z))$ is equal to zero for $r = 1, \ldots, \mathbf{c}$ which means that the sum $\sum_{r=1}^{\mathbf{c}}(\ldots)$ in (A.1) is equal to zero. If the models relating to the missingness mechanism are misspecified $\boldsymbol{\psi}^* \neq \boldsymbol{\psi}_0$ then the $\mathcal{S}(G_r(Z))$ is not equal to zero for $r = 1, \ldots, \mathbf{c}$ which means that the sum $\sum_{r=1}^{\mathbf{c}}(\ldots)$ in (A.1) is not equal to zero. Now, we assume that the μ-models are correctly specified and we are allowing the π-models to be either correctly specified or misspecified. Pick $r \in \{1, \ldots, \mathbf{c}\}$. We have the conditional expectation

$$E\left(\Xi_T(Z) \mid I(\mathcal{C} \geq r), G_r(Z)\right) = E\left(U_{AIPW}(Z) \mid G_r(Z), \boldsymbol{\zeta}\right) - E\left(U_{AIPW}(Z) \mid I(\mathcal{C} \geq r), G_r(Z)\right)$$

and the conditional expectation $E\left(U_{AIPW}(Z) \mid \cdot\right)$ is rewritten to

$$E\left(\prod_{t=0}^{T} \frac{I(a_t)}{\pi_t\{V_t, \boldsymbol{\alpha}\}} \left(Y - \mu\{V_T, \boldsymbol{\gamma}\}\right) + \sum_{j=1}^{T} \prod_{t=0}^{j-1} \frac{I(a_t)}{\pi_t\{V_t, \boldsymbol{\alpha}\}} \left(\mu\{V_j, \boldsymbol{\gamma}\} - \mu\{V_{j-1}, \boldsymbol{\gamma}\}\right) \middle| \cdot\right) - E\left(Y^{\overline{a}_T}\right)$$

and all the parts with the μ-models are equal to zero because the μ-models are correctly specified. This means that the conditional expectation $E(\Xi_T(Z) \mid I(\mathcal{C} \geq r), G_r(Z))$ is equal to zero because the conditional expectation $E\left(U_{AIPW}(Z) \mid G_r(Z), \boldsymbol{\zeta}\right)$ is equal to $E\left(Y^{\overline{a}_T}\right)$ and the conditional expectation $E\left(U_{AIPW}(Z) \mid I(\mathcal{C} \geq r), G_r(Z)\right)$ is equal to $E\left(Y^{\overline{a}_T}\right)$. We have the equality

$$E\left\{\mathcal{S}\left(G_r(Z)\right) I(\mathcal{C} \geq r) \Xi_T(Z)\right\} = 0$$

because

$$E\left\{\mathcal{S}\left(G_r(Z)\right) I(\mathcal{C} \geq r) \Xi_T(Z)\right\} = E\left\{\mathcal{S}\left(G_r(Z)\right) I(\mathcal{C} \geq r) E\left(\Xi_T(Z) \mid I(\mathcal{C} \geq r), G_r(Z)\right)\right\}$$

and the conditional expectation $E(\Xi_T(Z) \mid I(\mathcal{C} \geq r), G_r(Z))$ is equal to zero. This implies that the expectation of the sum $\sum_{r=1}^{\mathbf{c}}(\ldots)$ in (A.1) is equal to zero because the μ-models are correctly specified. We must recall that the $\mu\{V_T, \boldsymbol{\gamma}\}$-model is equal to the $m\{V_T, \boldsymbol{\xi}\}$-model.

B | Tables from the analysis of the PERFORM study

Tables with the actual numbers for Figure 5.1: Table B.1 shows the estimates with the confidence intervals for Figure 5.1a (the analysis for $t = 2$). Table B.2 shows the estimates with the confidence intervals for Figure 5.1b (the analysis for $t = 6$). Table B.3 shows the estimates with the confidence intervals for Figure 5.1c (the analysis for $t = 12$).

Estimator		Estimate	SE	95%-CI	Width of CI
Naïve	$(1,1)$	14.397	0.610	(13.202 ; 15.593)	2.392
	$(0,1)$	8.500	2.401	(3.795 ; 13.205)	9.411
	$(1,0)$	7.387	1.096	(5.240 ; 9.535)	4.295
	$(0,0)$	4.225	0.836	(2.586 ; 5.864)	3.278
LSmeans	$(1,1)$	11.869	0.590	(10.712 ; 13.026)	2.314
	$(0,1)$	6.901	2.091	(2.802 ; 10.999)	8.198
	$(1,0)$	10.034	0.923	(8.225 ; 11.843)	3.618
	$(0,0)$	9.177	1.260	(6.709 ; 11.646)	4.937
G-formula	$E(SDS_6^{(1,1)})$	12.094	0.627	(10.864 ; 13.323)	2.459
	$E(SDS_6^{(0,1)})$	5.822	2.450	(1.020 ; 10.624)	9.605
	$E(SDS_6^{(1,0)})$	10.215	0.978	(8.299 ; 12.131)	3.832
	$E(SDS_6^{(0,0)})$	8.280	1.250	(5.830 ; 10.729)	4.899
DRMGf	$E(SDS_6^{(1,1)})$	12.8060	0.552	(11.724 ; 13.889)	2.165
	$E(SDS_6^{(0,1)})$	5.9400	2.417	(1.203 ; 10.676)	9.473
	$E(SDS_6^{(1,0)})$	11.0140	0.930	(9.191 ; 12.837)	3.646
	$E(SDS_6^{(0,0)})$	8.4160	1.307	(5.853 ; 10.979)	5.125

Table B.1: Figure 5.1a is based on the actual numbers. The analysis for $t = 2$. The *Estimator* column shows the estimator that was used to obtain the estimates. The *Estimate* column shows the estimates obtained using the different estimators. The *SE* column shows the standard errors obtained using 927 bootstraps. The 95%-*CI* column shows the 95% confidence intervals. The *Width of CI* column shows the width of the confidence intervals.

Estimator		Estimate	SE	95%-CI	Width of CI
Naïve	$(1,1)$	12.858	0.608	(11.666 ; 14.051)	2.385
	$(0,1)$	10.462	2.231	(6.089 ; 14.834)	8.745
	$(1,0)$	5.208	0.876	(3.492 ; 6.925)	3.433
	$(0,0)$	3.813	0.771	(2.302 ; 5.323)	3.020
LSmeans	$(1,1)$	10.207	0.598	(9.036 ; 11.379)	2.343
	$(0,1)$	9.936	1.307	(7.375 ; 12.497)	5.122
	$(1,0)$	8.104	1.046	(6.053 ; 10.155)	4.102
	$(0,0)$	8.811	1.429	(6.010 ; 11.612)	5.603
G-formula	$E(SDS_{12}^{(1,1)})$	10.643	0.646	(9.378 ; 11.909)	2.531
	$E(SDS_{12}^{(0,1)})$	8.771	1.394	(6.039 ; 11.504)	5.465
	$E(SDS_{12}^{(1,0)})$	8.520	1.053	(6.456 ; 10.584)	4.127
	$E(SDS_{12}^{(0,0)})$	7.692	1.536	(4.682 ; 10.701)	6.019
DRMGf	$E(SDS_{12}^{(1,1)})$	11.404	0.544	(10.338 ; 12.471)	2.133
	$E(SDS_{12}^{(0,1)})$	9.691	1.385	(6.975 ; 12.406)	5.430
	$E(SDS_{12}^{(1,0)})$	9.059	1.157	(6.792 ; 11.326)	4.535
	$E(SDS_{12}^{(0,0)})$	8.386	1.521	(5.405 ; 11.366)	5.961

Table B.2: Figure 5.1b is based on the actual numbers. The analysis for $t = 6$. The *Estimator* column shows the estimator that was used to obtain the estimates. The *Estimate* column shows the estimates obtained using the different estimators. The *SE* column shows the standard errors obtained using 1000 bootstraps. The 95%-*CI* column shows the 95% confidence intervals. The *Width of CI* column shows the width of the confidence intervals.

Estimator		Estimate	SE	95%-CI	Width of CI
Naïve	$(1,1)$	13.010	0.839	(11.367 ; 14.654)	3.287
	$(0,1)$	12.937	2.097	(8.827 ; 17.048)	8.222
	$(1,0)$	5.000	1.276	(2.500 ; 7.500)	5.000
	$(0,0)$	3.579	0.554	(2.493 ; 4.665)	2.172
LSmeans	$(1,1)$	9.097	0.698	(7.729 ; 10.466)	2.737
	$(0,1)$	9.439	1.412	(6.672 ; 12.206)	5.534
	$(1,0)$	6.921	1.223	(4.524 ; 9.317)	4.793
	$(0,0)$	9.168	1.339	(6.545 ; 11.792)	5.247
G-formula	$E(SDS_{18}^{(1,1)})$	8.889	0.771	(7.378 ; 10.400)	3.022
	$E(SDS_{18}^{(0,1)})$	9.979	1.597	(6.849 ; 13.109)	6.260
	$E(SDS_{18}^{(1,0)})$	6.913	1.301	(4.363 ; 9.462)	5.100
	$E(SDS_{18}^{(0,0)})$	9.371	1.417	(6.593 ; 12.149)	5.556
DRMGf	$E(SDS_{18}^{(1,1)})$	10.324	0.670	(9.010 ; 11.637)	2.626
	$E(SDS_{18}^{(0,1)})$	10.856	1.675	(7.573 ; 14.138)	6.565
	$E(SDS_{18}^{(1,0)})$	8.042	1.313	(5.468 ; 10.616)	5.148
	$E(SDS_{18}^{(0,0)})$	10.267	1.528	(7.272 ; 13.262)	5.989

Table B.3: Figure 5.1c is based on the actual numbers. The analysis for $t = 12$. See Table B.2 for the description of the five columns: *Estimator*, *Estimate*, *SE*, 95%-*CI* and *Width of CI*.

Table with the results of the causal effects: Table B.4 shows all the estimates and the confidence intervals from the analysis of the causal effects $\boldsymbol{\beta}_t = (\beta_{I,t}, \beta_{1,t}, \beta_{2,t}, \beta_{3,t})$ for t equal to 2, 6 and 12.

		SG			DRMGf		
Analysis		Effect	SE	95%-CI	Effect	SE	95%-CI
Z_2	$\beta_{I,2}$	8.280	1.291	(5.750 ; 10.810)	8.416	1.313	(5.842 ; 10.990)
	$\beta_{1,2}$	1.935	1.573	(-1.148 ; 5.018)	2.598	1.601	(-0.541 ; 5.736)
	$\beta_{2,2}$	-2.457	2.591	(-7.536 ; 2.621)	-2.476	2.445	(-7.269 ; 2.316)
	$\beta_{3,2}$	4.336	2.881	(-1.310 ; 9.982)	4.269	2.733	(-1.088 ; 9.626)
Z_6	$\beta_{I,6}$	7.692	1.501	(4.749 ; 10.634)	8.386	1.513	(5.421 ; 11.351)
	$\beta_{1,6}$	0.828	1.692	(-2.489 ; 4.145)	0.673	1.698	(-2.654 ; 4.001)
	$\beta_{2,6}$	1.079	1.530	(-1.920 ; 4.079)	1.305	1.661	(-1.951 ; 4.560)
	$\beta_{3,6}$	1.044	1.543	(-1.981 ; 4.068)	1.040	1.588	(-2.072 ; 4.153)
Z_{12}	$\beta_{I,12}$	9.371	1.478	(6.475 ; 12.267)	10.267	1.617	(7.098 ; 13.435)
	$\beta_{1,12}$	-2.459	1.651	(-5.694 ; 0.777)	-2.225	1.885	(-5.919 ; 1.470)
	$\beta_{2,12}$	0.608	1.715	(-2.754 ; 3.970)	0.589	1.836	(-3.010 ; 4.188)
	$\beta_{3,12}$	1.369	1.946	(-2.445 ; 5.182)	1.693	2.047	(-2.319 ; 5.704)

Table B.4: The estimation of the causal effects $\boldsymbol{\beta}_t = (\beta_{I,t}, \beta_{1,t}, \beta_{2,t}, \beta_{3,t})$ for t equal to 2, 6 and 12. The *SG* column shows the estimates obtained using the estimator for the simpler g-formula (SG). The *DRMGf* column shows the estimates obtained using our DRMGf estimator. The *Effect* column shows the estimated effects. The *SE* column shows the standard errors. The standard errors are obtained using 1000 bootstraps. The 95%-*CI* column shows the 95% confidence intervals. *The standard errors are obtained using* 954 *bootstraps for* $t = 2$ *because the estimators had numerical problems.* See Table 1 in Manuscript III for further information.

List of Abbreviations

Symbols

G_r 22
λ-models 22
c 21
$\mathcal{C}$ 21
μ-models 8
π-models 9
ϖ-model 22
dir 17
$indir_{M_1}$ 17
$indir_{M_2}$ 17
pt 3
st 3
$\boldsymbol{\beta}$ 11
$\boldsymbol{\beta}_t$ 32
$\boldsymbol{\psi}$ 23

B

BRMM 25

C

CAR 22
CCAR 22
CNAR 22

D

DAG 3
DR 11
DRMGf 24
DRMSM 25

G

g-formula 6, 7

H

Hilbert $\mathcal{H}$ space 29

I

iid 6
IPW 6, 9

M

MAR 21
MCAR 21
MDD 1
MNAR 21
MSM 6, 10

P

PDQ 2
PERFORM 1
PHQ 2

S

SDS 2
SG 11
SSM 19

Bibliography

Chen Avin, Ilya Shpitser, and Judea Pearl. Identifiability of path-specific effects. pages 357–363, June 2005. URL `http://www.escholarship.org/uc/item/45x689gq`.

Heejung Bang and James M. Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–973, December 2005. doi: 10.1111/j.1541-0420.2005.00377.x.

Jonathan W Bartlett, Shaun R. Seaman, Ian R. White, and James R. Carpenter. Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model. *Statistical Methods in Medical Research*, 24(4):462–487, 2015. doi: 10.1177/0962280214521348.

Pratap Chokka, Joanna Bougie, Jean Proulx, Anders Holmegaard Tvistholm, and Anders Ettrup. Long-term functioning outcomes are predicted by cognitive symptoms in working patients with major depressive disorder treated with vortioxetine: results from the atworc study (book review). *CNS Spectrums*, pages 1–12, 2019. ISSN 1092-8529. doi: 10.1017/S1092852919000786.

Stephen Cole, R. and Constantine Frangakis, E. The consistency statement in causal inference: A definition or an assumption? *Epidemiology*, 20(1):3–5, 2009. ISSN 1044-3983.

Stephen R. Cole and Miguel Ángel Hernán. Fallibility in estimating direct effects. *International Journal of Epidemiology*, 31(1):163–165, February 2002.

Stephen R. Cole and Miguel A. Hernán. Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology*, 168(6):656–664, 2008.

Rhian M. Daniel, M.G. Kenward, S.N. Cousens, and B.I. De Stavola. Using causal diagrams to guide analysis in missing data problems. 21(3):243–56, 2012. doi: 10.1177/0962280210394469.

Rhian M. Daniel, S.N. Cousens, B.L. De Stavola, M.G. Kenward, and J. A.C. Sterne. Methods for dealing with time-dependent confounding. *Statistics in Medicine*, 32(9):1584–1618, April 2013. doi: 10.1002/sim.5686.

Rhian M. Daniel, B. L. De Stavola, S. N. Cousens, and S. Vansteelandt. Causal mediation analysis with multiple mediators. *Biometrics*, 71(1):1–14, March 2015. doi: 10.1111/biom.12248.

Vanessa Didelez, A. Philip Dawid, and Sara Geneletti. Direct and indirect effects of sequential treatments. pages 138–146, June 2012.

Peng Ding and Tyler VanderWeele. Sensitivity analysis without assumptions. July 2015.

Sylvie Goetgeluk and Stijn Vansteelandt. Estimation of controlled direct effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5):1049–1066, November 2008. doi: 10.1111/j.1467-9868.2008.00673.x.

Lene Hammer-Helmich, Josep Maria Haro, Bengt Jönsson, Audrey Tanguy Melac, Sylvie Di Nicola, Julien Chollet, Dominique Milea, Benoît Rive, and Delphine Saragoussi. Functional impairment in patients with major depressive disorder: the 2-year perform study. *Neuropsychiatric Disease and Treatment*, 14:239–249, January 2018. doi: 10.2147/NDT.S146098.

Josep Maria Haro, Lene Hammer-Helmich, Delphine Saragoussi, Anders Ettrup, and Klaus Groes Larsen. Patient-reported depression severity and cognitive symptoms as determinants of functioning in patients with major depressive disorder: a secondary analysis of the 2-year prospective perform study. *Neuropsychiatric Disease and Treatment*, 15:2313–2323, 2019.

M.A Hernán and J.M. Robins. *Causal Inference*. "Boca Raton: Chapman and Hall/CRC", 2017. For free.

Kosuke Imai, Luke Keele, and Teppei Yamamoto. Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, 25(1):51–71, February 2010. doi: 10.1214/10-STS321.

Marshall M. Joffe and Tom Greene. Related causal frameworks for surrogate outcomes. *Biometrics*, 65(2):530–538, June 2009.

Jay S. Kaufman, Richard F. Maclehose, and Sol Kaufman. A further critique of the analytic strategy of adjusting for covariates to identify biologic mediation. *Epidemiologic perspectives & innovations : EP+I*, 1(1):4, 2004. ISSN 1742-5573.

Ruth H. Keogh, Rhian M. Daniel, Tyler J. VanderWeele, and Stijn Vansteelandt. Analysis of longitudinal studies with repeated outcome measures: Adjusting for time-dependent confounding using conventional methods. *American journal of epidemiology*, 187(5), May 2018. ISSN 1476-6256. doi: 10.1093/aje/kwx311.

Noémi Kreif, Linh Tran, Richard Grieve, Bianca De Stavola, Robert C. Tasker, and Maya Petersen. Estimating the comparative effectiveness of feeding interventions in the pediatric intensive care unit: A demonstration of longitudinal targeted maximum likelihood estimation. *American Journal of Epidemiology*, 186(12):1370–1379, 2017.

Kurt Kroenke and Robert Spitzer. The phq-9: A new depression diagnostic and severity measure. *Psychiatric Annals*, 32(9):509–515, 2002. ISSN 00485713. URL `http://search.proquest.com/docview/217052580/`.

Lingling Li, Changyu Shen, Xiaochun Li, and James M. Robins. On weighting approaches for missing data. *Statistical Methods in Medical Research*, 22(1):14–30, 2013. doi: 10.1177/0962280211403597.

Wei Li and Xiao,ÄêHua Zhou. Identifiability and estimation of causal mediation effects with missing data. *Statistics in Medicine*, 36(25):3948–3965, 2017. ISSN 0277-6715.

P. M. Llora, R. Jain, Maelys Touya, Kenneth Simonsen, and Lene Hammer-Helmich. Improvement in functioning after treatment of major depression with vortioxetine. interim results from the "real-life effectiveness of vortioxetine" (relieve) study. 2019.

Judith J. Lok and Victor DeGruttola. Impact of time to start treatment following infection with application to initiating haart in hiv-positive patients. *Biometrics*, 68, September 2012. doi: 10.1111/j.1541-0420.2011.01738.x.

Judea Pearl. Direct and indirect effects. *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, (10):411–420, 2001.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press., 2nd edition, 2009. ISBN 9780521895606.

Judea Pearl. An introduction to causal inference. *The international journal of biostatistics*, 6 (2), February 2010. doi: 10.2202/1557-4679.1203.

James Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect (book review). *Mathematical Modelling*, 7(9-12):1393–1512, 1986. ISSN 02700255.

James Robins, Miguel Ángel Hernán, and Andrea Rotnitzky. Invited commentary: Effect modification by time-varying covariates. *American Journal of Epidemiology*, 166(9):994–1002, November 2007. doi: 10.1093/aje/kwm231. URL `http://search.proquest.com/docview/235814401/`.

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

James M. Robins and T. S. Richardson. Alternative graphical causal models and the identification of direct effects. *To appear in Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures. P. Shrout, Editor. Oxford University Press*, 2010.

James M. Robins and Andrea Rotnitzky. Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association*, 90(429):122–129, 1995. doi: 10.1080/01621459.1995.10476494.

James M. Robins and Larry A. Wasserman. Estimation of effects of sequential treatments by reparameterizing directed acyclic graphs. February 2013.

James M. Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866, 1994. doi: 10.1080/01621459.1994.10476818.

James M. Robins, Miguel Ángel Hernánl, and Babette Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560, September 2000.

Donald B. Rubin. Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, 6(1):34–58, 1978.

Donald B. Rubin. Direct and indirect causal effects via potential outcomes*. *Scandinavian Journal of Statistics*, 31(2):161–170, June 2004.

SAS Institute Inc. "`LSMEANS` statement." in sas/stat 9.3 user's guide. `http://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#statug_introcom_a0000003362.htm`, 2012.

Shaun R Seaman and Ian R White. Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research*, 22(3):278–295, 2013. doi: 10.1177/0962280210395740.

Eric J. Tchetgen Tchetgen and Tyler J. VanderWeele. Identification of natural direct effects when a confounder of the mediator is directly affected by exposure. *Epidemiology*, 25(2):282–291, 2014.

Sengwee Toh and Miguel Ángel Hernán. Causal inference from longitudinal studies with baseline randomization. *The International Journal of Biostatistics*, 4(1):1117–1117, October 2008. ISSN 1557-4679. doi: 10.2202/1557-4679.1117. URL `http://search.proquest.com/docview/35308841/`.

Anastasios Athanasios Tsiatis, Anastasios A. *Semiparametric theory and missing data*. Springer series in statistics. Springer, 2006. ISBN 9780387373454.

Tyler VanderWeele and Stijn Vansteelandt. Mediation analysis with multiple mediators. *Epidemiologic Methods*, 2(1):95–115, 2013. doi: 10.1515/em-2012-0010. URL `http://search.proquest.com/docview/1704262638/`.

Tyler J. VanderWeele. Marginal structural models for the estimation of direct and indirect effects. *Epidemiology*, 20(1):18–26, January 2009a. doi: 0.1097/EDE.0b013e31818f69ce.

Tyler J. VanderWeele. Mediation and mechanism. *European Journal of Epidemiology*, 24(5):217–224, May 2009b. doi: 10.1007/s10654-009-9331-1.

Tyler J. VanderWeele. Concerning the consistency assumption in causal inference. *Epidemiology*, 20(6):880–883, 2009c. doi: 10.1097/EDE.0b013e3181bd5638.

Tyler J. VanderWeele. Controlled direct and mediated effects: Definition, identification and bounds. *Scandinavian Journal of Statistics*, 38(3):551–563, September 2011.

Tyler J. VanderWeele. Policy-relevant proportions for direct effects. *Epidemiology*, 24(1):175–176, 2013. ISSN 1044-3983. doi: 10.1097/EDE.0b013e3182781410.

Tyler J. VanderWeele. Mediation analysis: A practitioner's guide. *Annual Review of Public Health*, 37, 2016. ISSN 0163-7525.

Tyler J. VanderWeele and Whitney R. Robinson. On the causal interpretation of race in regressions adjusting for confounding and mediating variables. *Epidemiology*, 25(4):473–484, July 2014. doi: 10.1097/EDE.0000000000000105.

Tyler J. VanderWeele and Eric J. Tchetgen Tchetgen. Mediation analysis with time varying exposures and mediators. *Royal Statistical Society*, 79(3):917–938, June 2016. ISSN 1369-7412. doi: 10.1111/rssb.12194.

Tyler J. VanderWeele, Stijn Vansteelandt, and James M. Robins. Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology*, 25(2):300–306, March 2014. doi: 10.1097/EDE.0000000000000034.

Stijn Vansteelandt and Rhian M. Daniel. Interventional effects for mediation analysis with multiple mediators. *Epidemiology (Cambridge, Mass.)*, 28(2):258–265, March 2017. doi: 10.1097/EDE.0000000000000596.

Stijn Vansteelandt and Arvid Sjolander. Revisiting g-estimation of the effect of a time-varying exposure subject to time-varying confounding. *Epidemiologic Methods*, 5(1):37–56, January 2016. doi: 10.1515/em-2015-0005. URL `http://search.proquest.com/docview/1857880751/`.

E. J. Williamson, A. Forbes, and R. Wolfe. Doubly robust estimators of causal exposure effects with missing data in the outcome, exposure or a confounder. *Statistics in Medicine*, 31(30):4382–4400, 2012.

Manuscript I

A doubly robust estimator for monotone missing data in the presence of time-dependent confounding

Research Article

Statistics
in Medicine

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

A doubly robust estimator for monotone missing data in the presence of time-dependent confounding

Thomas Maltesen[ab*], Torben Martinussen[b], Klaus Groes Larsen[c] and Lene Hammer-Helmich[d]

Patients in observational and interventional studies tend to drop-out, which leads to data with missing observations. Missing observations may complicate the analysis of a longitudinal study with repeated measures over time with time-dependent confounding. Standard methods fail in the presence of time-dependent confounding and reducing the data to fully observed vectors can cause biased estimates. We propose an augmented inverse probability weighted (AIPW) estimator to estimate the causal effect of a binary time-varying exposure in the presence of time-dependent confounding with a continuous outcome subject to missingness. The estimator is robust regarding misspecification of the parametric model for the monotone missingness in the data under the assumption that the missingness is missing at random (MAR). Our estimator requires only partially observed vectors to be included in the analysis. We use the proposed estimator on the observational study Prospective Epidemiological Research on Functioning Outcomes Related to Major depressive disorder (PERFORM), which is a longitudinal study with time-dependent confounding and missing observations. Copyright © 0000 John Wiley & Sons, Ltd.

Keywords: causal inference, g-formula, time-dependent confounding, doubly robust estimator, monotone missingness

1. Introduction

Major depressive disorder (MDD) is a multidimensional disease characterized by emotional, physical and cognitive symptoms. Treatment of cognitive symptoms may hold the key to achieving functional recovery in MDD, but the relationship between cognitive symptoms and functional impairment is not well understood [1]. The observational study Prospective Epidemiological Research on Functioning Outcomes Related to Major depressive disorder (PERFORM) (NCT01427439) is an observational cohort study undertaken to better understand the course of a depressive episode and its impact on patient functioning over two years in outpatients with MDD [2]. Functional impairment, cognitive symptoms and depression severity have been measured for each patient in the PERFORM study. The measurements were based on self-reported scales. Data were collected at different time points: at baseline and after 2, 6, 12, 18 and 24 months [2]. A naïve regression analysis of functional impairment on cognitive symptoms may lead to a biased estimate since depression severity may impose time-dependent confounding [3]. The data contains a substantial number of patients with missing observations. Reducing data with missing observations to a subset of fully observed vectors may also result in biased estimates.

Methods have been proposed to handle data with missing observations e.g. inverse probability weighting (IPW) where fully observed vectors are up-weighted to represent full data. Multiple imputation (MI) is another popular method to

[a] *Health Economics and Epidemiology Statistics, H.Lundbeck A/S, 2500 Valby, Denmark*
[b] *Section of Biostatistics, University of Copenhagen, 1014 Copenhagen K, Denmark*
[c] *Statistics, H.Lundbeck A/S, 2500 Valby, Denmark*
[d] *Medical Affairs, Value Evidence, H.Lundbeck A/S, 2500 Valby, Denmark*
[*] *Correspondence to: Thomas Maltesen, H.Lundbeck A/S, 2500 Valby, Denmark. E-mail: thomas.maltesen@protonmail.com*

handle data with missing observations. MI has two main advantages over IPW. First, (unless it is monotone missing or a more complicated MI-model is used) the IPW uses only the fully observed data. Second, the MI is more efficient than IPW, but IPW is less technical, easier to understand and easier to explain to collaborators [4, 5]. Robins, J. M. and Rotnitzky, A., among others, have considered methods of semiparametric models with inverse probability weighted estimators when data contains missing observations [6, 7, 8]. Bang, H. and Robins, J. M., for binary exposure, introduced a doubly robust estimator for data with missing outcomes and an augmented inverse probability weighted (AIPW) estimator for causal inference models with time-dependent confounding [9]. Williamson, E. J., Forbes, A., and Wolfe, R. introduced a doubly robust estimator in case of the (causal) exposure of interest, the confounder or the outcome being missing. Confounding and missingness in the data may occur simultaneously and the model for the missingness mechanism in the data is typically uncertain. A doubly robust estimator is desirable to overcome the issue of the missingness in the data [10]. The properties of these methods are appealing; however, these methods are mainly attractive if only one variable is missing at a time or data do not have any time-dependent confounding while data contains missing observations. Missingness in the data can occur for outcome, exposure and time-dependent confounding at the same time.

The PERFORM study is the motivation to develop an estimator for longitudinal data with missing observations while adjusting for time-dependent confounding at the same time. The previously mentioned methods cannot be used for our data since more than one measurement is missing at a time and we also have time-dependent confounding. Therefore, we need an estimator for a time-varying exposure that is robust of misspecification of the parametric model for the missingness mechanism and at the same time can adjust for time-dependent confounding. We will first consider the estimator derived from the g-formula for estimating the effect of the time-varying exposure in the presence of time-dependent confounding. Then, we will extend the estimator to handle partially observed vectors using the techniques developed in Tsiatis [11]. The estimator includes as many vectors as possible and the estimator is not restricted to only one missing variable at a time. The article is organized as follows. Section 2 revisits the estimator of the g-formula for continuous outcome and binary exposure. Section 3 considers data containing missing observations and then extends the g-formula to handle missing observations using the techniques developed in Tsiatis [11]. We denote our estimator by DRMGf (Doubly Robust estimator for Monotone missingness for the G-formula). In Section 4, we apply our DRMGf estimator from Section 3 to analyse the PERFORM study. Section 5 shows a simulation study with data simulated with parameters obtained from the PERFORM study to show the performance of the DRMGf estimator. Section 6 is a discussion of our findings.

2. The estimator for estimating the causal effects in full data

Suppose that our data comprises n identical and independent distributed realizations of random variables $Z_1, \ldots, Z_n$ with Z_i denoting the i-th vector in the data (we suppress the index i for simplicity). The vector Z defines an ordered sequence of variables $(L_0, A_0, L_1, A_1, \ldots, L_T, A_T, Y)$. The outcome variable Y is assumed to be continuous and we let A_t denote the binary exposure at time $t \in \{0, \ldots, T\}$. Define $\overline{A}_t$ to be a vector of exposures up to time t $(A_0, \ldots, A_t)$. Let $\overline{A}_T$ denote $(A_0, \ldots, A_T)$. Let $Y^{\overline{a}_T}$ be the potential outcome that would have been observed if $\overline{A}_T$ had been set to $\overline{a}_T$. Let L_t be a set of measured potential confounders at time $t \in \{0, \ldots, T\}$ [3]. Let $\overline{L}_T$ denote the vector $(L_0, \ldots, L_T)$ and let $\overline{l}_T$ denote the vector $(l_0, \ldots, l_T)$. The outcome Y may be causally influenced by the whole history of $\overline{A}_T$ and $\overline{L}_T$. We define $V_t = (\overline{L}_t, \overline{A}_t)$ and $v_t = (\overline{l}_t, \overline{a}_t)$. We assume sequential conditional exchangeability

$$Y^{\overline{a}_T} \perp\!\!\!\perp A_t \mid \overline{L}_t, \overline{A}_{t-1} \quad \forall \overline{a}_T \in \overline{A}_T, \forall t \in \{0, \ldots, T\}.$$

The g-formula [12] is given by

$$E\left(Y^{\overline{a}_T}\right) = \int E(Y \mid \overline{L}_T = \overline{l}_T, \overline{A}_T = \overline{a}_T) \prod_{t=0}^{T} f_{L_t|\overline{A}_{t-1},\overline{L}_{t-1}}(l_t \mid \overline{a}_{t-1}, \overline{l}_{t-1}) d\overline{l}_t \tag{1}$$

and can be rewritten as a series of iterated conditional expectations given by

$$E\left(Y^{\overline{a}_T}\right) = E(\cdots E(E(Y \mid \overline{L}_T, \overline{A}_T = \overline{a}_T) \mid \overline{L}_{T-1}, \overline{A}_{T-1} = \overline{a}_{T-1}) \cdots \mid \overline{L}_0, \overline{A}_0 = \overline{a}_0)$$

[13]. From the series of iterated conditional expectations we may then estimate $E(Y^{\overline{a}_T})$ with full data by the estimator

$$\hat{E}\left(Y^{\overline{a}_T}\right) = \frac{1}{n} \sum_{i=1}^{n} \mu\{V_{0,i}; \hat{\boldsymbol{\gamma}}\} \tag{2}$$

with the first model $m\{v_T, \boldsymbol{\xi}\} = E(Y \mid \overline{L}_T = \overline{l}_T, \overline{A}_T = \overline{a}_T)$. The next models are given by $\mu\{v_{T-1}, \boldsymbol{\gamma}\} = E(m\{V_T, \boldsymbol{\xi}\} \mid \overline{L}_{T-1} = \overline{l}_{T-1}, \overline{A}_{T-1} = \overline{a}_{T-1})$ and $\mu\{v_t, \boldsymbol{\gamma}\} = E(\mu\{V_{t+1}, \boldsymbol{\gamma}\} \mid \overline{L}_t = \overline{l}_t, \overline{A}_t = \overline{a}_t)$. The last model is given by $\mu\{v_0, \boldsymbol{\gamma}\} =$

$E(\mu\{V_1, \boldsymbol{\gamma}\} \mid L_0 = l_0, A_0 = a_0)$. We refer to the $m\{v_T, \boldsymbol{\xi}\}$-model and all the $\mu\{v_t, \boldsymbol{\gamma}\}$-models as the μ-models. Let $m\{v_T, \boldsymbol{\xi}_0\}$ denote the true model with the vector of true parameter values $\boldsymbol{\xi}_0$ and let $\mu\{v_t, \boldsymbol{\gamma}_0\}$ denote the true models with the vector of true parameter values $\boldsymbol{\gamma}_0$. All the μ-models in (2) contain hats to indicate that we have plugged in the predicted values from the specified μ-models that have been used for the estimation. The estimator (2) is unbiased if the μ-models are correctly specified with respect to the underlying process that has generated the data. See Kreif et al. [13] for the steps of the estimation. The estimator (2) solves the estimating equation

$$\sum_{i=1}^{n} U(Z_i) = 0$$

with

$$U(Z_i) = \mu\{V_{0,i}, \boldsymbol{\gamma}_0\} - E\left(Y^{\overline{a}_T}\right). \tag{3}$$

The $\mu\{V_t, \boldsymbol{\gamma}\}$-model can be extended if the confounder $L_t = (L_{t_1}, \ldots, L_{t_q})$ is q-dimensional at time $t \in \{0, \ldots T\}$. See Daniel et al. [3] for further information. Section 4 shows an example where the time-dependent confounder is multivariate ($q = 2$). The estimator (2) is asymptotically normally distributed with the mean $E(Y^{\overline{a}_T})$ and a variance. We show in Appendix B that the estimator (2) is asymptotically normally distributed in the situation when T is equal to 1.

3. Vectors with missing observations

Define $\mathcal{C}$ to be a random variable and it takes positive integers or infinity $\mathcal{C} \in \{1, \ldots, \mathbf{c}\} \cup \{\infty\}$. Let $G_{\mathcal{C}}(Z) \subseteq Z$ denote a vector and let $\{G_{\mathcal{C}_i}(Z_i), \mathcal{C}_i\}$ denote the i-th vector in the observed data. If $\mathcal{C}$ is equal to 1 then it corresponds to have only observed L_0 in Z and the vector is denoted by $G_1(Z) = (L_0)$. If $\mathcal{C} = 2$ then L_0 and A_0 are the only two observed variables in Z and $G_2(Z) = (L_0, A_0)$. We mean by $\mathcal{C}$ equal to $\mathbf{c}$ that only the outcome is missing from Z. Note that $\mathbf{c}$ is an integer and it is equal to $2(T + 1)$. If $\mathcal{C}$ is equal to infinity then a vector is complete and $G_\infty(Z) = (Z)$. In the pattern described above, if L_t is observed then $\overline{L}_{t-1}$ and $\overline{A}_{t-1}$ are also necessarily observed, and if A_t is observed then $\overline{L}_t$ and $\overline{A}_{t-1}$ are also necessarily observed. Such a pattern is known as *monotone* missingness [11]. Complete cases (CC) are a subset of the observed data. Complete cases contain only vectors of the form $G_\infty(Z)$. We assume the conditional probability of observing a complete vector given Z is strictly greater than zero, i.e. that:

$$P(\mathcal{C} = \infty \mid Z) > 0$$

and $\varpi\{\infty, Z, \boldsymbol{\psi}\}$ denotes the probability $P(\mathcal{C} = \infty \mid Z)$ [11]. Let

$$\lambda_r\{G_r(Z), \boldsymbol{\psi}\} = P(\mathcal{C} = r \mid \mathcal{C} \geq r, Z)$$

for $r \neq \infty$ denote the probability of stopping the observing additional observations given r observed. Tsiatis defines

$$K_r\{G_r(Z), \boldsymbol{\psi}\} = \prod_{j=1}^{r} (1 - \lambda_j\{G_j(Z), \boldsymbol{\psi}\})$$

where $K_{\mathbf{c}}\{G_{\mathbf{c}}(Z), \boldsymbol{\psi}\}$ is the probability $\varpi\{\infty, Z, \boldsymbol{\psi}\}$ [11]. We assume that data are coarsened at random (CAR) which means the coarsening probabilities only depend on the data as a function of the observed data and the coarsening probabilities are given by $\varpi\{r, G_r(Z), \boldsymbol{\psi}\} = \lambda_r\{G_r(Z), \boldsymbol{\psi}\} K_{r-1}\{G_{r-1}(Z), \boldsymbol{\psi}\}$. We assume $\lambda_r\{G_r(Z), \boldsymbol{\psi}\}$ is given by

$$\lambda_r\{G_r(Z), \boldsymbol{\psi}\} = \frac{\exp\left(\psi_{I,r} + G_r(Z)\boldsymbol{\psi}_r\right)}{1 + \exp\left(\psi_{I,r} + G_r(Z)\boldsymbol{\psi}_r\right)}, \tag{4}$$

where the column vector $\boldsymbol{\psi}_r$ has the same dimension as the row vector $G_r(Z)$. Let $\boldsymbol{\psi} = (\psi_{I,r}, \boldsymbol{\psi}_r')$ where the coefficient $\psi_{I,r}$ denotes the intercept and $\boldsymbol{\psi}_r'$ is the transposed row vector of $\boldsymbol{\psi}_r$. We refer to the $\lambda_r\{G_r(Z), \boldsymbol{\psi}\}$-models as the λ-models. To distinguish between the models we let $\varpi\{\infty, Z, \boldsymbol{\psi}_0\}$ denote the true function of the missing mechanism with the true parameter values $\boldsymbol{\psi}_0$. Tsiatis [11] shows that the augmented inverse probability weighted (AIPW) estimator for $E(Y^{\overline{a}_T})$ is obtained by solving the estimating equation given by

$$0 = \sum_{i=1}^{n} \left(\frac{I(\mathcal{C}_i = \infty) U(Z_i)}{\varpi\{\infty, Z_i, \hat{\boldsymbol{\psi}}\}} + \sum_{r=1}^{\mathbf{c}} \frac{I(\mathcal{C}_i = r) - \lambda_r\{G_r(Z_i), \hat{\boldsymbol{\psi}}\} I(\mathcal{C}_i \geq r)}{K_r\{G_r(Z_i), \hat{\boldsymbol{\psi}}\}} E\left(U(Z) \mid G_r(Z_i), \hat{\boldsymbol{\zeta}}\right) \right)$$

where the probabilities of the models relating to the missingness mechanism are given by $K_r\{G_r(Z), \hat{\psi}\} = \prod_{j=1}^{r}(1 - \lambda_j\{G_j(Z), \hat{\psi}\})$ and the probability of observing a complete vector is given by $\varpi\{\infty, Z, \hat{\psi}\} = K_{\mathbf{c}}(G_{\mathbf{c}}(Z), \hat{\psi})$. The estimates $\hat{\psi}$ are obtained using maximum likelihood estimation according to the specific model of $\lambda_r\{G_r(Z), \psi\}$. The hat indicates that we have chosen a model to estimate and we have afterwards used the model to predict the values with respect to the set $G_r(Z)$ for $r = 1, \ldots, \mathbf{c}$. The conditional expectation $E(U(Z) \mid G_r(Z), \hat{\boldsymbol{\zeta}})$ of (3) is given by

$$E\left(U(Z) \mid G_r(Z), \hat{\boldsymbol{\zeta}}\right) = E\left(\mu\{V_0, \boldsymbol{\gamma}_0\} \mid G_r(Z), \hat{\boldsymbol{\zeta}}\right) - E(Y^{\overline{a}_T}).$$

We need to evaluate the conditional expectation for every set of $G_r(Z)$ for $r = 1, \ldots, \mathbf{c}$; this is exemplified in Section 4. We use $\hat{\boldsymbol{\zeta}}$ to denote that we have modelled the conditional expectation of $U(Z)$ with respect to the set $G_r(Z)$. The estimator for $E(Y^{\overline{a}_T})$ is given by

$$\hat{E}\left(Y^{\overline{a}_T}\right) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{I(\mathcal{C}_i = \infty)\mu\{V_{0,i}, \hat{\boldsymbol{\gamma}}\}}{\varpi\{\infty, Z_i, \hat{\psi}\}} + \sum_{r=1}^{\mathbf{c}}\frac{I(\mathcal{C}_i = r) - \lambda_r\{G_r(Z_i), \hat{\psi}\}I(\mathcal{C}_i \geq r)}{K_r\{G_r(Z_i), \hat{\psi}\}}E\left(\mu\{V_0, \boldsymbol{\gamma}\} \mid G_r(Z_i), \hat{\boldsymbol{\zeta}}\right)\right] \quad (5)$$

and it is doubly robust of misspecification. Let the μ-models be correctly specified. The estimator (5) is unbiased if the $E(\mu\{V_0, \boldsymbol{\gamma}\} \mid G_r(Z), \hat{\boldsymbol{\zeta}})$-models are correctly specified with respect to the distribution of Z and the λ-models relating to the missingness mechanism may be misspecified. The estimator (5) is also unbiased if the λ-models relating to the missingness mechanism are correctly specified and the $E(\mu\{V_0, \boldsymbol{\gamma}\} \mid G_r(Z), \hat{\boldsymbol{\zeta}})$-models may be misspecified with respect to the distribution of Z. Let the estimator (5) be denoted by DRMGf (Doubly Robust estimator for Monotone missingness for the G-formula). In the estimator (5) we have plugged in the predicted values using all the μ-models and the λ-models. All the conditional expectations are evaluated and afterwards used to predict the values with respect to the set $G_r(Z)$. We show in Appendix A how the estimator (5) is derived. In the next Section we refer to a vector in the data as a patient. The estimator (5) is asymptotically normally distributed with the mean $E(Y^{\overline{a}_T})$ and a variance. We show in Appendix B that the estimator (5) is asymptotically normally distributed in the situation when T is equal to 1.

4. Application to the PERFORM study

4.1. Study design and variables

We apply our DRMGf estimator to the PERFORM study consisting of 1090 patients. We are interested in the causal effect of cognitive symptoms on functional impairment at a later time. The functional impairment was measured by the Sheehan Disability Scale (SDS) consisting of 3 items with each item ranging from 0 to 10 with a global score ranging from 0 to 30. The Scale describes patients' work/school, social life/leisure activities and family life/home duties. The cognitive symptoms (memory, concentration and executive function) were measured by the Perceived Deficit Questionnaire (PDQ-5). The scale consists of 5 items with each item ranging from 0 to 4 with a global score ranging from 0 to 20 (we suppress the '-5' in the name PDQ-5 to simplify the notation). The depression severity of the patient was measured by the Patient Health Questionnaire (PHQ-9) and the scale consists of 9 items with each item ranging from 0 to 3 with a global score ranging from 0 to 27 (we suppress the '-9' in the name PHQ-9 to simplify the notation). A greater score for all three scales correspond to being more constrained, suffering greater severity of their cognitive symptoms and more severe depression. All three scales were measured over two years repeatedly. We assume that depression severity affects both cognitive symptoms and functional impairment and that cognitive symptoms affect functional impairment. We further assume that the present measurements affect all the future measurements at the next time point. We also assume that the present measurements do not affect the past measurements [14]. The process is indicated by the directed acyclic graph (DAG) in Figure 1 for all six time points over the two years.

If the global score of PDQ is less than or equal to 5, then it corresponds to patients having no or minimal cognitive symptoms and if the score is strictly greater than 5 then it corresponds to patients having cognitive symptoms. For simplicity, and since this analysis is mainly for illustration, we therefore dichotomized PDQ to be 0 if the original global score of PDQ is less than or equal to 5 and 1 otherwise. Let SDS_t denote SDS at time t to simplify the notation. Let PDQ_t denote PDQ at time t to simplify the notation. Let PHQ_t denote PHQ at time t to simplify the notation. Let $t = b$ denote the baseline, and let t be equal to 2, 6, 12, 18 and 24 (months) which denotes the measurement time points since baseline. We use b to denote the time point for baseline instead of 0 because we want to avoid any confusion when we refer to the time point and the true parameter values. Let W_t denote the vector of all three measurements at time $t \in \{b, 2, 6, 12, 18, 24\}$, $W_t = (PHQ_t, PDQ_t, SDS_t)$. We define pt to denote the *prior* time point before t, t denotes the present time point and st denotes the *subsequent* time point after t in the subscript of PHQ, PDQ and SDS. Hence, if t is equal to 6 (month 6) then pt is equal to 2 (month 2) and st is equal to 12 (month 12). We assume the subsequent effect

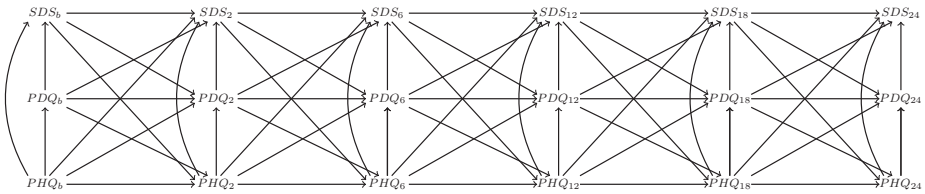


Figure 1. The DAG displays the measurements of the three scales over the two years. The node SDS_t denotes the Sheehan Disability Scale (SDS) at time t, the node PDQ_t denotes the Perceived Deficit Questionnaire (PDQ) at time t and the node PHQ_t denotes the Patient Health Questionnaire (PHQ) at time t. We let $t = b$ denote the baseline and let t be equal to 2, 6, 12, 18 and 24 (months) which denotes the measurement time points since baseline.

of depression severity, cognitive symptoms and functional impairment are conditionally independent of the prior effect of depression severity, cognitive symptoms and functional impairment given the present effect of depression severity, cognitive symptoms and functional impairment. It means we have assumed that the effect of $(PHQ_{pt}, PDQ_{pt}, SDS_{pt})$ does not affect $(PHQ_{st}, PDQ_{st}, SDS_{st})$ directly but only via (PHQ_t, PDQ_t, SDS_t).

Let PDQ_t denote the exposure of cognitive symptoms at time $t \in \{b, 2, 6, 12, 18, 24\}$. Let SDS_{st} denote the outcome of functional impairment at time $st \in \{2, 6, 12, 18, 24\}$. We assume according to the DAG the sequential conditional exchangeability and it is generalized to

$$SDS_{st}^{(pdq_t, pdq_{st})} \perp\!\!\!\perp PDQ_t \mid PHQ_t, SDS_{pt}, PDQ_{pt}, PHQ_{pt}$$

and

$$SDS_{st}^{(pdq_t, pdq_{st})} \perp\!\!\!\perp PDQ_{st} \mid PHQ_{st}, SDS_t, PDQ_t, PHQ_t$$

for $t \in \{b, 2, 6, 12, 18\}$. We assume the confounder L_t at time t is defined by $(PHQ_{pt}, PDQ_{pt}, SDS_{pt}, PHQ_t)$ for $t \in \{b, 2, 6, 12, 18\}$. We also assume that the confounder L_{st} at time st is defined by $L_{st} = (SDS_t, PHQ_{st})$. The set V_t is given by the confounder and the exposure at time t, $V_t = (L_t, PDQ_t)$ and the set V_{st} is defined by the confounders and the exposures up to time st, $V_{st} = (\overline{L}_{st}, \overline{PDQ}_{st})$ where $\overline{L}_{st} = (L_t, L_{st})$ and $\overline{PDQ}_{st} = (PDQ_t, PDQ_{st})$. A Table in the Supplementary material displays the two confounders L_t and L_{st} and the outcome SDS_{st} for different t. We define $Z_{t,i}$ to be the set $(W_{pt,i}, W_{t,i}, W_{st,i})$ in the analysis for $i = 1, \ldots, 1090$ and $t \in \{b, 2, 6, 12, 18\}$ with W_{pb} being the empty set since it corresponds to the vector of the measurements before baseline. It means that $Z_{b,i}$ denotes $(PHQ_{b,i}, PDQ_{b,i}, SDS_{b,i}, PHQ_{2,i}, PDQ_{2,i}, SDS_{2,i})$, $Z_{2,i}$ denotes $(PHQ_{b,i}, PDQ_{b,i}, SDS_{b,i}, PHQ_{2,i}, PDQ_{2,i}, SDS_{2,i}, PHQ_{6,i}, PDQ_{6,i}, SDS_{6,i})$ and $Z_{6,i}$ denotes $(PHQ_{2,i}, PDQ_{2,i}, SDS_{2,i}, PHQ_{6,i}, PDQ_{6,i}, SDS_{6,i}, PHQ_{12,i}, PDQ_{12,i}, SDS_{12,i})$, etc. for $i = 1, \ldots, 1090$. See the Supplementary material for further information about the different vectors $Z_{t,i}$ for $i = 1, \ldots, 1090$ and $t \in \{b, 2, 6, 12, 18\}$.

4.2. Statistical methods

We use the following marginal structural model (MSM)

$$E\left(SDS_{st}^{(pdq_t, pdq_{st})}\right) = \beta_{I,t} + \beta_{1,t} pdq_t + \beta_{2,t} pdq_{st} + \beta_{3,t} pdq_t pdq_{st} \tag{6}$$

for $t \in \{b, 2, 6, 12, 18\}$. The t index of the coefficients $\boldsymbol{\beta}_t = (\beta_{I,t}, \beta_{1,t}, \beta_{2,t}, \beta_{3,t})$ indicates which period of $t \in \{b, 2, 6, 12, 18\}$ we analyse. Let $\boldsymbol{\beta}_{0,t}$ denote the true vector of the causal parameters. We are interested in what could be achieved if an effective therapy was developed that could relieve cognitive symptoms. Model (6) presents four combinations: $(pdq_t, pdq_{st}) = (1, 1)$ corresponds to having cognitive symptoms at visit t and st, $(pdq_t, pdq_{st}) = (0, 1)$ corresponds to having no or minimal cognitive symptoms at visit t and having cognitive symptoms at visit st, $(pdq_t, pdq_{st}) = (1, 0)$ corresponds to having cognitive symptoms at visit t and having no or minimal cognitive symptoms at visit st and $(pdq_t, pdq_{st}) = (0, 0)$ corresponds to having no or minimal cognitive symptoms at visit t and st. The score of functional impairment at 0 corresponds to being unimpaired and 30 corresponds to being impaired. The $E(SDS_{st}^{(1,1)})$ is the expected score of functional impairment that would be seen if the patient had cognitive symptoms at the two time points t and st. The $E(SDS_{st}^{(0,1)})$ is the expected score of functional impairment that would be seen if the patient had no or minimal cognitive symptoms at the time point t and had cognitive symptoms at the time point st. The $E(SDS_{st}^{(1,0)})$ is the expected score of functional impairment that would be seen if the patient had cognitive symptoms at the time point t and had no or minimal cognitive symptoms at the time point st. The $E(SDS_{st}^{(0,0)})$ is the expected score of functional

impairment that would be seen if the patient had no or minimal cognitive symptoms at the two time points t and st. We define the $m\{V_{st}, \boldsymbol{\xi}_t\}$-model and the $\mu\{V_t, \boldsymbol{\gamma}_t\}$-model for $t \in \{b, 2, 6, 12, 18\}$ with the main effects and let the interactions between depression severity and cognitive symptoms be included in the $m\{V_{st}, \boldsymbol{\xi}_t\}$-model. The $m\{V_{st}, \boldsymbol{\xi}_t\}$-model is given by

$$\begin{aligned} m\{V_{st}, \boldsymbol{\xi}_t\} =& E(SDS_{st} \mid PDQ_{st}, PHQ_{st}, SDS_t, PDQ_t, PHQ_t) \\ =& \xi_{I,t} + \xi_{1,t} PHQ_t + \xi_{2,t} PDQ_t + \xi_{3,t} SDS_t + \xi_{4,t} PHQ_{st} + \xi_{5,t} PDQ_{st} + \xi_{6,t} PDQ_t PHQ_t + \\ & \xi_{7,t} PDQ_{st} PHQ_t + \xi_{8,t} PDQ_t PHQ_{st} + \xi_{9,t} PDQ_{st} PHQ_{st} + \xi_{10,t} PDQ_t PDQ_{st} \end{aligned} \tag{7}$$

where the t index of the coefficients $\xi_{\cdot,t}$ indicate which time point $t \in \{b, 2, 6, 12, 18\}$ is considered. Let $\xi_{I,t}$ denote the intercept. We must recall that the main effects of depression severity (PHQ), cognitive symptoms (PDQ) and functional impairment (SDS) at time pt are not included in the $m\{V_{st}, \boldsymbol{\xi}_t\}$-model because we assume that functional impairment (SDS) at time st is conditionally independent of W_{pt} given by depression severity (PHQ), cognitive symptoms (PDQ) and functional impairment (SDS) at time t, $(SDS_{st} \perp\!\!\!\perp W_{pt} \mid W_t)$. The confounder L_{st} at time st has two measurements on the causal path which means that the following $\mu\{V_t, \boldsymbol{\gamma}_t\}$-model is given by:

$$\begin{aligned} \mu_{t_2}\{V_t, \boldsymbol{\gamma}_{t_2}\} =& E(m\{V_{st}, \boldsymbol{\xi}_t\} \mid SDS_t, PDQ_t, PHQ_t) \\ =& \gamma_{I,t_2} + \gamma_{1,t_2} PHQ_t + \gamma_{2,t_2} PDQ_t + \gamma_{3,t_2} SDS_t \end{aligned} \tag{8}$$

and

$$\begin{aligned} \mu_{t_1}\{V_t, \boldsymbol{\gamma}_{t_1}\} =& E(\mu_{t_2}(V_t, \boldsymbol{\gamma}_{t_2}) \mid PDQ_t, PHQ_t, SDS_{pt}, PDQ_{pt}, PHQ_{pt}) \\ =& \gamma_{I,t_1} + \gamma_{1,t_1} PHQ_{pt} + \gamma_{2,t_1} PDQ_{pt} + \gamma_{3,t_1} SDS_{pt} + \gamma_{4,t_1} PHQ_t + \gamma_{5,t_1} PDQ_t \end{aligned} \tag{9}$$

for $t \in \{b, 2, 6, 12, 18\}$. Let γ_{I,t_2} and γ_{I,t_1} denote the intercepts. The two subscripts of t (t_1 and t_2) indicate the order of the two mediators of the causal path between the exposure PDQ_t and the outcome SDS_{st}. The model at (8) is obtained from the equality between the two following equations

$$E(m\{V_{st}, \boldsymbol{\xi}_t\} \mid SDS_t, PDQ_t, PHQ_t, SDS_{pt}, PDQ_{pt}, PHQ_{pt}) = E(m\{V_{st}, \boldsymbol{\xi}_t\} \mid SDS_t, PDQ_t, PHQ_t)$$

since we are assuming that the future is conditional independent of the past given the present $m\{V_{st}, \boldsymbol{\xi}_t\} \perp\!\!\!\perp W_{pt} \mid W_t$. The $U(Z_i)$ at (3) for the analysis of the PERFORM study is given by

$$U(Z_{t,i}) = \mu_{t_1}\{V_{t,i}, \boldsymbol{\gamma}_{t_1}\} - E\left(SDS_{st}^{(pdq_t, pdq_{st})}\right)$$

with respect to either $Z_{b,i}$, $Z_{2,i}$, $Z_{6,i}$, $Z_{12,i}$ or $Z_{18,i}$ for $i = 1, \ldots, 1090$ patients. In the analysis for $t = b$ or $t = 2$ we use all the patients in the data who have observed depression severity at baseline (PHQ_b). In the analysis for $t = 6$ we use all the patients who have observed depression severity at month 2 (PHQ_2), and so on. If the patient has a nonmonotone pattern (see Tsiatis [11] for further information) then the patient is modified to follow a monotone pattern by setting the subsequent variables to be missing as well as in the ordered sequence.

t	$G_1(Z_t)$	$G_2(Z_t)$	$G_3(Z_t)$	$G_4(Z_t)$	$G_5(Z_t)$	$G_6(Z_t)$	$G_7(Z_t)$	$G_8(Z_t)$	$G_\infty(Z_t)$	n_t
b	200	176	126	11	86				341	940
2	200	176	126	11	86	76	1	38	226	940
6	132	199	115	4	58	56	5	31	205	805
12	135	147	105	5	64	60	3	25	196	740
18	89	162	105	4	52	39	1	34	215	701

Table 1. The number of patients in the data fulfilling the monotone pattern. The number of complete cases for a specific analysis are displayed by the $G_\infty(Z_t)$ column and the Total (n_t) column displays the number of patients who follow the monotone pattern for a specific t.

We define the conditional probabilities $\lambda_r\{G_r(Z_t), \psi\}$ by the hazard function at (4) to model the mechanism relating to the missingness in the data. The λ-models include only the main effects without any interactions or quadratic terms. We need to model the hazard function $\lambda_r\{G_r(Z_b), \psi\}$ five times for $t = b$ because a patient at time $t = b$ has five possible sets of ordered measurements without including the outcome, see Table 1. The different patients in the data for the analysis with $t = b$ are given by $G_1(Z_b) = (PHQ_b)$, $G_2(Z_b) = (PHQ_b, PDQ_b)$, $G_3(Z_b) = (PHQ_b, PDQ_b, SDS_b)$, $G_4(Z_b) = (PHQ_b, PDQ_b, SDS_b, PHQ_2)$, $G_5(Z_b) = (PHQ_b, PDQ_b, SDS_b, PHQ_2, PDQ_2)$ and $G_\infty(Z_b) = (Z_b)$. We have, according to Table 1 for $t = b$, 200 patients in the data who only have the first measurement

observed ($G_1(Z_b) = (PHQ_b)$). An additional 176 patients in the data contain only the first two observed measurements ($G_2(Z_b) = (PHQ_b, PDQ_b)$). Table 1 shows that the analysis for $t = b$ only includes 341 patients in the data who are complete cases. We need to model the hazard function $\lambda_r\{G_r(Z_t), \boldsymbol{\psi}\}$ eight times for each $t \in \{2, 6, 12, 18\}$ because it is possible to have eight different sets of ordered measurements without including the outcome, see Table 1. The different patients in the data for $t \in \{2, 6, 12, 18\}$ are given by $G_1(Z_t) = (PHQ_{pt})$, $G_2(Z_t) = (PHQ_{pt}, PDQ_{pt})$, $G_3(Z_t) = (W_{pt})$, $G_4(Z_t) = (W_{pt}, PHQ_t)$, $G_5(Z_t) = (W_{pt}, PHQ_t, PDQ_t)$, $G_6(Z_t) = (W_{pt}, W_t)$, $G_7(Z_t) = (W_{pt}, W_t, PHQ_{st})$, $G_8(Z_t) = (W_{pt}, W_t, PHQ_{st}, PDQ_{st})$ and $G_\infty(Z_t) = (Z_t)$. We must recall that W_{pt} denotes the vector before time t and W_t denotes the vector at time t. We have, according to Table 1 e.g. for $t = 18$, 89 patients in the data who only have PHQ_{12} observed ($G_1(Z_{18}) = (PHQ_{12})$). An additional 162 patients in the data contain only the two observed measurements PHQ_{12} and PDQ_{12} ($G_2(Z_{18}) = (PHQ_{12}, PDQ_{12})$). Table 1 shows that the analysis for $t = 18$ only includes 215 patients in the data who are complete cases. Let n_t denote the sample size for a specific t, see Table 1.

The estimator for the analysis of the PERFORM study is given by

$$\hat{E}\left(SDS_{st}^{(pdq_t, pdq_{st})}\right) = \frac{1}{n_t}\sum_{i=1}^{n_t}\left[\frac{I(\mathcal{C}_i = \infty)\mu_{t_1}\{V_{t,i}, \hat{\boldsymbol{\gamma}}_{t_1}\}}{\varpi\{\infty, Z_{t,i}, \hat{\boldsymbol{\psi}}\}} + \sum_{r=1}^{\mathbf{c}}\frac{I(\mathcal{C}_i = r) - \lambda_r\{G_r(Z_{t,i}), \hat{\boldsymbol{\psi}}\}I(\mathcal{C}_i \geq r)}{K_r\{G_r(Z_{t,i}), \hat{\boldsymbol{\psi}}\}}E\left(\mu_{t_1}\{V_t, \boldsymbol{\gamma}_{t_1}\} \mid G_r(Z_{t,i}), \hat{\boldsymbol{\zeta}}_t\right)\right] \tag{10}$$

when $\mathbf{c}$ is equal to 5 for $t = b$ and $\mathbf{c}$ is equal to 8 for $t \in \{2, 6, 12, 18\}$. Furthermore, we need to model the conditional expectations in (10) and afterwards use the model for predicting values according to the different sets of $G_r(Z_{t,i})$ for $r = 1, \ldots, \mathbf{c}$, $i = 1, \ldots, n_t$ and $t \in \{b, 2, 6, 12, 18\}$. The conditional expectations $E(\mu_{b_1}\{V_b, \boldsymbol{\gamma}_{b_1}\} \mid G_r(Z_b), \hat{\boldsymbol{\zeta}}_b)$ for $t = b$ are modelled according to

$$\begin{cases} E(\mu_{b_1}\{V_b, \boldsymbol{\gamma}_{b_1}\} \mid G_r(Z_b), \hat{\boldsymbol{\zeta}}_b) & \text{for } r = 1 \\ \mu_{b_1}\{V_b, \boldsymbol{\gamma}_{b_1}\} & \text{for } r \in \{2, 3, 4, 5\} \end{cases}$$

and we model the conditional expectation $E(\mu_{b_1}(V_b) \mid G_1(Z_b), \hat{\boldsymbol{\zeta}}_b)$ with the main effect only without any quadratic terms and it is afterwards used to predict values. The conditional expectations $E(\mu_{t_1}\{V_t, \boldsymbol{\gamma}_{t_1}\} \mid G_r(Z_t), \hat{\boldsymbol{\zeta}}_t)$ for $t \in \{2, 6, 12, 18\}$ are modelled according to

$$\begin{cases} E(\mu_{t_1}\{V_t, \boldsymbol{\gamma}_{t_1}\} \mid G_r(Z_t), \hat{\boldsymbol{\zeta}}_t) & \text{for } r \in \{1, 2, 3, 4\} \\ \mu_{t_1}\{V_t, \boldsymbol{\gamma}_{t_1}\} & \text{for } r \in \{5, 6, 7, 8\} \end{cases}$$

and we model the conditional expectations $E(\mu_{t_1}(V_t) \mid G_r(Z_t), \hat{\boldsymbol{\zeta}}_t)$ with the main effects only without any interactions or quadratic terms and they are afterwards used to predict values. We compare our DRMGf estimator to the *naïve* estimator (described below), *LSmeans* [15] estimator and the estimator for the the *g-formula* [12] (described below). We only use complete cases for the analysis with the naïve estimator, LSmeans estimator and the estimator for the simpler g-formula. We will sometimes refer to the g-formula as the simpler g-formula since our DRMGf estimator is an extended g-formula. The $m\{v_{st}, \boldsymbol{\xi}_t\}$-model at (7) that is used in our DRMGf estimator is also used for the LSmeans estimator and the estimator for the simpler g-formula. The *naïve* estimator is a regression of the outcome, SDS_{st} on both exposures PDQ_t and PDQ_{st} and the interaction between the two exposures PDQ_tPDQ_{st} to estimate the coefficients $\tilde{\boldsymbol{\xi}}_t = (\tilde{\xi}_{I,t}, \tilde{\xi}_{1,t}, \tilde{\xi}_{2,t}, \tilde{\xi}_{3,t})$. Let $\tilde{\xi}_{I,t}$ denote the intercept. The estimated coefficients $\hat{\tilde{\boldsymbol{\xi}}}_t$ are used to predict four pseudo outcomes with the model

$$\hat{\tilde{\xi}}_{I,t} + \hat{\tilde{\xi}}_{1,t}pdq_t + \hat{\tilde{\xi}}_{2,t}pdq_{st} + \hat{\tilde{\xi}}_{3,t}pdq_tpdq_{st}$$

with respect to the four pairs $(pdq_t, pdq_{st}) = (1, 1)$, $(pdq_t, pdq_{st}) = (0, 1)$, $(pdq_t, pdq_{st}) = (1, 0)$ and $(pdq_t, pdq_{st}) = (0, 0)$. The naïve estimator has causal interpretation if there are no confounding at all. The *LSmeans* works in the following way: Estimate all the coefficients from the $m\{v_{st}, \boldsymbol{\xi}_t\}$-model and then the estimated coefficients (that are indicated by hats) are multiplied to the average of the different measurements in the model and it is given by

$$\begin{aligned} &\hat{\xi}_{I,t} + \hat{\xi}_{1,t}avg(PHQ_t) + \hat{\xi}_{2,t}pdq_t + \hat{\xi}_{3,t}avg(SDS_t) + \hat{\xi}_{4,t}avg(PHQ_{st}) + \hat{\xi}_{5,t}pdq_{st} + \hat{\xi}_{6,t}avg(pdq_tPHQ_t) \\ &+ \hat{\xi}_{7,t}avg(pdq_{st}PHQ_t) + \hat{\xi}_{8,t}avg(pdq_tPHQ_{st}) + \hat{\xi}_{9,t}avg(pdq_{st}PHQ_{st}) + \hat{\xi}_{10,t}pdq_tpdq_{st}. \end{aligned} \tag{11}$$

It is used to predict four pseudo outcomes with respect to either $(pdq_t, pdq_{st}) = (1, 1)$, $(pdq_t, pdq_{st}) = (0, 1)$, $(pdq_t, pdq_{st}) = (1, 0)$ or $(pdq_t, pdq_{st}) = (0, 0)$ and $avg(\cdot)$ denotes the average of a specific measurement. The estimator for the simpler *g-formula* with the estimator (2) is given by

$$\hat{E}\left(SDS_{st}^{(pdq_t, pdq_{st})}\right) = \frac{1}{\tilde{n}_t}\sum_{i=1}^{\tilde{n}_t}\mu_{t_1}\{V_{t,i}, \boldsymbol{\gamma}_{t_1}\}$$

where the $m\{v_{st}, \boldsymbol{\xi}_t\}$-model at (7) and the two μ-models that are given by (8) and (9) are used for the estimation and prediction to obtain the estimates of the four expected potential outcomes given by $E(SDS_{st}^{(1,1)})$, $E(SDS_{st}^{(0,1)})$, $E(SDS_{st}^{(1,0)})$ and $E(SDS_{st}^{(0,0)})$. The $\tilde{n}_t$ denotes the number of complete cases. The confidence intervals for all four estimators are obtained by using 1000 bootstraps. Table 1 shows the number of patients who fulfil the monotone missingness pattern. The $G_\infty(Z_t)$ column in Table 1 shows the number of complete cases. If the analysis is based on complete cases then it only utilizes about 18% to 31% of the sample size. We have removed some of the patients from the data to avoid numerical problems for the estimation with the μ-models and the λ-models used for the analysis with $t = b$ and $t = 18$. It applies for all the patients with the partially observed vector $G_4(Z_t)$ for $t \in \{b, 18\}$ and all the patients with the partially observed vector $G_7(Z_{18})$. The λ_4-model is given by $\lambda_4(G_4(Z_t)) = 0$ for $t \in \{b, 18\}$ and the λ_7-model is given by $\lambda_7(G_7(Z_{18})) = 0$. The sample size n_b is reduced from 940 to 929. The sample size n_{18} is reduced from 701 to 696. We utilize about 64% to 85% of the sample size when we use patients with monotone missingness.

4.3. Results

Here, we present the results for all four estimators for $t = b$ and $t = 18$. The MSM at (6) for $t = b$ is given by

$$E\left(SDS_2^{(pdq_b, pdq_2)}\right) = \beta_{I,b} + \beta_{1,b} pdq_b + \beta_{2,b} pdq_2 + \beta_{3,b} pdq_b pdq_2$$

and for $t = 18$ it is given by

$$E\left(SDS_{24}^{(pdq_{18}, pdq_{24})}\right) = \beta_{I,18} + \beta_{1,18} pdq_{18} + \beta_{2,18} pdq_{24} + \beta_{3,18} pdq_{18} pdq_{24}.$$

This means that the expected value at the four different counterfactual levels are modelled freely without restrictions. The estimates of the four means are shown in Figure 2 for an early ($t = b$) and a later ($t = 18$) time point with confidence intervals. The actual numbers are in two Tables in the Supplementary material.

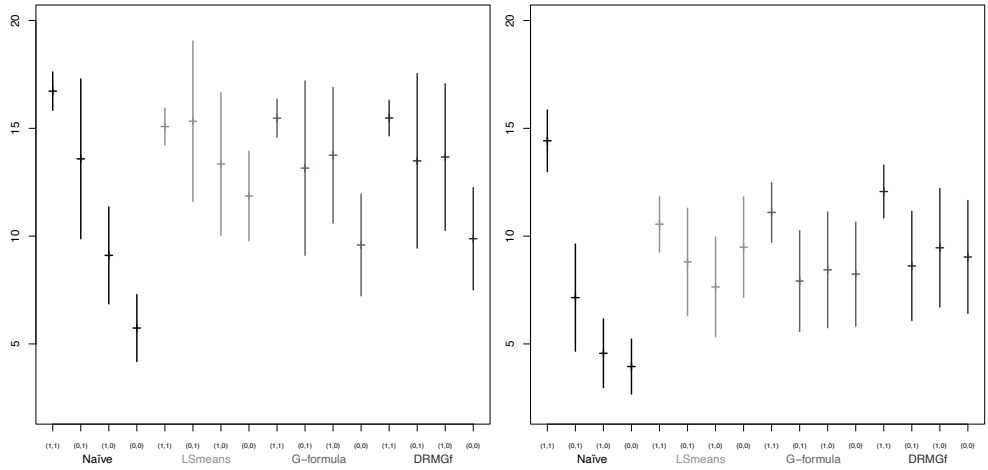


Figure 2. The plot on the left hand side is the estimation for $t = b$ with the two exposures PDQ_b and PDQ_2. The plot on the right hand side is the estimation for $t = 18$ with the two exposures PDQ_{18} and PDQ_{24}. The y-axis/range in the two plots show the score of the functional impairment. The x-axis/domain in the two plots show the four different estimators and the different pairs (pdq_t, pdq_{st}) for each estimator. The combinations of visits are given by $(pdq_t, pdq_{st}) = (1, 1)$, $(pdq_t, pdq_{st}) = (0, 1)$, $(pdq_t, pdq_{st}) = (1, 0)$ and $(pdq_t, pdq_{st}) = (0, 0)$. The horizontal lines are the estimates. The vertical lines are the corresponding 95%-confidence intervals. The 95%-confidence intervals are obtained using 1000 bootstraps.

For the analysis at both the earlier and the later time points, the naïve estimator shows a larger variation in the expected outcome between the four combinations of cognitive deficits at the two visits than the other three estimators. All four estimators suggest that patients with cognitive symptoms at both visits have worse functioning than patients with no or with minimal cognitive symptoms at both visits. In the analysis of the early time point, the expected outcome, if patients present cognitive symptoms at one of the visits but not the other, lies between the other two, while for the late time point

patients with cognitive symptoms at one but not the other of the two visits have similar expected outcome to those with no or with minimal cognitive symptoms at both visits. Note that the length of the confidence interval reflects the actual proportion of patients with the combination of cognitive symptoms observed in the data, as expected.

The general pattern of differences in the expected outcome between the "exposure" groups as defined by the presentation of cognitive symptoms at the two latest visits are really to be expected: presence of more cognitive symptoms is a precursor for poor functioning. The pattern is most pronounced for the naïve estimator. This is also not surprising, because the naive estimator fails to account for any confounders, such as depression severity at both visits and functioning at the earlier visit. The estimators based on counterfactuals, the g-formula and the DRMGf estimators, both allow for taking observed confounders into account through standardization, and thereby key confounding variables may be accounted for. Doing so, the trend across the groups stays intact, though differences between "exposure" groups markedly shrink, when compared to the naïve estimator. Table 2 shows the minimum and maximum of the predicted values of each $\lambda_r\{G_r(Z_t), \hat{\psi}\}$ for both analysis of $t \in \{b, 18\}$.

Analysis	Range	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	ϖ
$t = b$	Minimum	0.200	0.219	0.171	0	0.059				0.319
	Maximum	0.224	0.252	0.313	0	0.291				0.447
$t = 18$	Minimum	0.086	0.188	0.192	0	0.065	0.065	0	0.023	0.172
	Maximum	0.224	0.415	0.261	0	0.375	0.239	0	0.464	0.413

Table 2. Displays the minimum and maximum of the predicted values of $\lambda_r\{G_r(Z_t), \psi\}$. We let λ_r denote the probabilities $\lambda_r\{G_r(Z_t), \psi\}$ for $r = 1, \ldots, 8$ for both analysis of $t \in \{b, 18\}$. We let ϖ denote the probabilities for observing a complete vector for both analysis of $t \in \{b, 18\}$.

The estimator for the simpler g-formula and the DRMGf estimator yield surprisingly similar estimates. This is likely because the included covariates are poor at predicting drop-out, which seems to be the case. The weights in the estimator (10) have a low contribution to the estimation because the values of probabilities $\lambda_r\{G_r(Z_t), \psi\}$ displayed in Table 2 are low. While the naïve estimator ignores confounding altogether and thereby supposedly overestimates differences between exposure groups, the LSmeans estimator adjusts for confounders. Although the estimates are formally predicted values at average values of the covariates, these are closer to the standardization-based estimates of the g-formula and the DRMGf than to the naïve estimates. Some differences between the LSmeans and the two latter estimators should be noted: at the early time point, LSmeans estimates indicate smaller differences between groups, and - as the only case - patients presenting cognitive problems at both visits prior are not predicted to have the worst functioning subsequently, which would be difficult to explain from a clinical point of view; at the later time point, the ordering of the four cognition combinations is different for the LSmeans estimates compared with the other two, which are in mutual agreement.

The similarity of the g-formula and DRMGf estimates prompts for a further exploration of scenarios, where a correct handling of missing data due to drop-out would be crucial for the interpretation of the data and whether the DRMGf estimator better recovers the parameters of interest in that setting. This is explored further in the simulation study in Section 5. The simulation study shows how the probabilities regarding the missingness mechanism have an impact on the estimates. It will cause the difference between the simpler g-formula and our DRMGf estimator.

5. Simulation study

The purpose of the simulation study is to investigate the proposed DRMGf estimator within a situation that is similar to the PERFORM study with a substantial amount of drop-out, but - contrary to the PERFORM data - where the ranges of the predicted values of $\lambda_r\{G_r(Z_b), \psi\}$ are broader than the ones we obtained in the PERFORM study. The simulation study is based on the estimated coefficients from the first two vectors (W_b, W_2) that we observed in the PERFORM study. The sample size for each data is 1000 and the data are replicated 5000 times. Data are simulated as follows: $PHQ_b \sim \text{Normal}(\eta_{phq_b}, 5.314^2)$, $PDQ_b \sim \text{Bernoulli}(\varkappa_{pdq_b})$, $SDS_b \sim \text{Normal}(\eta_{sds_b}, 5.058^2)$, $PHQ_2 \sim \text{Normal}(\eta_{phq_2}, 5.215^2)$, $PDQ_2 \sim \text{Bernoulli}(\varkappa_{pdq_2})$ and $SDS_2 \sim \text{Normal}(\eta_{sds_2}, 4.663^2)$ where the means are

given by

$$\begin{aligned}
\eta_{phq_b} :=& 17.615,\\
\eta_{sds_b} :=& 3.349 + 0.693PHQ_b + 4.373PDQ_b,\\
\eta_{phd_2} :=& 2.475 + 0.471PHQ_b - 0.113PDQ_b + 0.086SDS_b \text{ and}\\
\eta_{sds_2} :=& 0.144 - 0.387PHQ_b + 1.179PDQ_b + 0.484SDS_b + 0.74PHQ_2 + 0.67PDQ_2 - 0.131PDQ_bPHQ_b\\
&+ 0.308PDQ_2PHQ_b + 0.209PDQ_bPHQ_2 - 0.206PDQ_2PHQ_2 - 1.734PDQ_bPDQ_2
\end{aligned}$$

and the probabilities are given by

$$\begin{aligned}
\text{logit}(\varkappa_{pdq_b}) :=& -2.512 + 0.298PHQ_b \text{ and}\\
\text{logit}(\varkappa_{pdq_2}) :=& -3.092 - 0.441PHQ_b + 2.144PDQ_b + 0.198SDS_b + 0.445PHQ_2 + 0.42PDQ_bPHQ_b\\
&- 0.176PDQ_bSDS_b - 0.161PDQ_bPHQ_2
\end{aligned}$$

where $\text{logit}(x) = \log(x) - \log(1-x)$. We have included all the main effects in the $\text{logit}(\varkappa_{pdq_2})$-model and the interactions between PDQ_b and the two measurements for depression severity at the two time points are also included. We want to include the effect of the interactions between depression severity and cognitive symptoms. The interaction between PDQ_b and SDS_b is also included. We have chosen more extreme values of ψ for the models relating to the missing mechanism to make the robustness of our DRMGf estimator more clear. The probabilities to simulate the monotone missingness in the data are given by $\text{logit}(\lambda_1(G_1(Z_b))) := -12.3 + 0.5PHQ_b$, $\text{logit}(\lambda_2(G_2(Z_b))) := -9.2 + 0.5PHQ_b - 0.8PDQ_b$, $\text{logit}(\lambda_3(G_3(Z_b))) := -3.6 + 0.7PHQ_b - 0.6PDQ_b - 0.6SDS_b$, $\text{logit}(\lambda_4(G_4(Z_b))) := -3.2 + 0.5PHQ_b - 0.6PDQ_b - 0.6SDS_b + 0.4PHQ_2$ and $\text{logit}(\lambda_5(G_5(Z_b))) := -2.5 + 0.4PHQ_b + 0.6PDQ_b - 0.6SDS_b + 0.4PHQ_2 + 0.5PDQ_2$ where $\lambda_r(G_r(Z_b))$ is $P(\mathcal{C} = r \mid \mathcal{C} \geq r, Z_b)$. We have used all four estimators from the previous Section on the simulated data. The four estimators are: the naïve estimator, the LSmeans estimator, the estimator for the simpler the g-formula and our DRMGf estimator for the estimation. We use the same models given in Subsection 4.2 for $t = b$ to analyse the simulated data. We have the results of the estimation in Table 3. The simulation study is evaluated by the mean of the 5000 estimates obtained across the replicated data, the empirical standard error (SE) of the 5000 estimates obtained across the replicated data, the absolute value of bias (the difference between the empirical mean and the true value), the ratio between the absolute value of bias and the empirical standard error scaled 100 times and the mean squared error (MSE) is also displayed [10].

Estimator		True	Mean	SE	Bias	$\frac{\text{Bias}}{\text{SE}}100$	MSE
Naïve	$(1,1)$	15.398	15.317	0.424	0.081	19.196	0.186
	$(0,1)$	13.418	13.540	1.731	0.122	7.062	3.013
	$(1,0)$	13.586	7.732	0.872	5.854	671.577	35.030
	$(0,0)$	9.817	6.046	1.023	3.771	368.655	15.266
LSmeans	$(1,1)$	15.398	13.582	0.432	1.817	420.380	3.488
	$(0,1)$	13.418	14.079	1.991	0.661	33.214	4.401
	$(1,0)$	13.586	12.112	0.897	1.473	164.171	2.976
	$(0,0)$	9.817	10.874	1.244	1.057	84.999	2.664
G-formula	$E(SDS_2^{(1,1)})$	15.398	13.858	0.437	1.540	352.193	2.564
	$E(SDS_2^{(0,1)})$	13.418	12.515	2.142	0.903	42.137	5.405
	$E(SDS_2^{(1,0)})$	13.586	12.415	0.917	1.171	127.734	2.211
	$E(SDS_2^{(0,0)})$	9.817	9.149	1.364	0.668	49.007	2.306
DRMGf	$E(SDS_2^{(1,1)})$	15.398	15.399	0.511	0.000	0.040	0.261
	$E(SDS_2^{(0,1)})$	13.418	13.390	2.452	0.028	1.135	6.014
	$E(SDS_2^{(1,0)})$	13.586	13.608	1.239	0.022	1.793	1.536
	$E(SDS_2^{(0,0)})$	9.817	9.810	1.907	0.007	0.359	3.635

Table 3. The True column displays the true values. The Mean column displays the mean of the 5000 estimates obtained across the replicated data. The SE column displays the standard error of the 5000 estimates obtained across the replicated data. The Bias column displays the absolute value of the difference between the empirical mean and the true value. The $\frac{\text{Bias}}{\text{SE}}100$ column displays the ratio between the absolute value of bias and the standard error scaled 100 times. The MSE column is the mean square error obtained by $\text{Bias}^2 + \text{SE}^2$.

Table 3 shows that the naïve estimator as well as the LSmeans estimator are biased and should be avoided. The estimates from the simulation study have similar sizes as the estimates from the analysis of the PERFORM study for $t = b$ for both the naïve estimator and the LSmeans estimator. Table 3 shows that the simpler g-formula does not perform very well compared to our DRMGf estimator when the parameter values of ψ for probabilities relating to the missing mechanism in the data are more extreme than the ones in the PERFORM study. The g-formula sometimes shows an even poorer performance of estimating the effects compared to the naïve and LSmeans estimator. Our DRMGf estimator shows that the bias of the estimates are low compared to the three other estimators. The simulation study has also shown that our DRMGf estimator will protect against biased estimates compared to the estimator for the simpler g-formula and our DRMGf estimator should be used for analysing data containing missing observations. The price we pay for unbiased estimates is potentially larger standard errors.

6. Discussion

Motivated by the PERFORM study, this manuscript proposed a robust estimator (the DRMGf estimator) for the estimation of the effect of a time-varying exposure in the presence of time-dependent confounding and missing data. The proposed estimator was applied to data from an observational study (PERFORM) on patients with depression. The example showed that accounting for the time-varying confounders shrunk the difference between the exposure groups, which was to be expected from a clinical point of view, as disease severity accounted for some of the differences in exposure levels while at the same time influenced the level of functioning among the patients. This property was shared with the estimator for the simpler g-formula, which does not take the missing data into account. The similarities in the estimates were because there were no strong predictors of patients dropping out in the example. Thereby, the presumed robustness of the DRMGf did not show. Contrary to what was seen in the data example, the simulation study revealed that if variables predicting drop-out were present, then the estimator for the simpler g-formula was biased, while the DRMGf was not. Another advantage of the proposed estimator is that it allows for a better use of the data as it utilizes all data points and not only so-called complete cases. In the PERFORM example, this means 2.5 to 4 times as many patients in the sample. This comes at the price of specific model assumptions that need to be addressed, but a complete case analysis hinges on assumptions that are not less critical and even less plausible (such as missing completely at random). One limitation of the DRMGf estimator is that the model framework assumes monotone missing data, and does not in itself allow for intermittent missing data. Consequently, outcome variables need to be ordered. While this may often be a natural assumption, in the PERFORM study such ordering was only partly clear. This is because several patient relevant domains were measured simultaneously. In this example, disease severity, cognitive performance and functioning were measured at the same time points up to six occasions throughout two years. While time itself induced a causal ordering of measurements at different time points, the ordering between domains within time points were less clear, and assumptions had to be made based on clinical insights (change in disease severity causing change in cognitive performance, which in turn causes change in functioning). Obviously, any interpretation of results hinge on these assumptions, which cannot be verified in the data. Based on a missing at random argumentation, intermittent missing data is often considered less of a problem than monotone missing data in indications such as depression (PERFORM), as it may be overcome, e.g. by multiple imputation methods. However, in other disease areas this may not be the case, and new methodology would be needed for such situations.

Acknowledgments

The PERFORM study was funded by Lundbeck SAS. The authors thank Rhian M. Daniel (Division of Population Medicine, Cardiff University, UK) for her advice and ideas.

Conflict of interest

Thomas Maltesen (temporary employee), Klaus Groes Larsen and Lene Hammer-Helmich are full-time employees of H. Lundbeck A/S. Torben Martinussen is a full-time employee at the University of Copenhagen and he has not received any honoraria. The authors report no conflict of interest.

A. Monotone missingness

The estimating equation for monotone missingness is given by

$$
\begin{aligned}
0 &= \sum_{i=1}^{n}\left(\frac{I(\mathcal{C}_i=\infty)U(Z_i)}{\varpi\{\infty,Z_i,\hat{\psi}\}}+\sum_{r=1}^{\mathbf{c}}\frac{I(\mathcal{C}_i=r)-\lambda_r\{G_r(Z_i),\hat{\psi}\}I(\mathcal{C}_i\geq r)}{K_r\{G_r(Z_i),\hat{\psi}\}}E\left(U(Z)\mid G_r(Z_i),\hat{\zeta}\right)\right)\\
&=\sum_{i=1}^{n}\left(\frac{I(\mathcal{C}_i=\infty)\left(\mu\{V_{0,i},\gamma_0\}-E\left(Y^{\overline{a}_T}\right)\right)}{\varpi\{\infty,Z_i,\hat{\psi}\}}\right.\\
&\qquad\left.+\sum_{r=1}^{\mathbf{c}}\frac{I(\mathcal{C}_i=r)-\lambda_r\{G_r(Z_i),\hat{\psi}\}I(\mathcal{C}_i\geq r)}{K_r\{G_r(Z_i),\hat{\psi}\}}\left(E\left(\mu\{V_0,\gamma_0\}\mid G_r(Z_i),\hat{\zeta}\right)-E\left(Y^{\overline{a}_T}\right)\right)\right)\\
&=\sum_{i=1}^{n}\left(\frac{I(\mathcal{C}_i=\infty)\mu\{V_{0,i},\gamma_0\}}{\varpi\{\infty,Z_i,\hat{\psi}\}}+\sum_{r=1}^{\mathbf{c}}\frac{I(\mathcal{C}_i=r)-\lambda_r\{G_r(Z_i),\hat{\psi}\}I(\mathcal{C}_i\geq r)}{K_r\{G_r(Z_i),\hat{\psi}\}}E\left(\mu\{V_0,\gamma_0\}\mid G_r(Z_i),\hat{\zeta}\right)-E\left(Y^{\overline{a}_T}\right)\right)
\end{aligned}
$$

and the estimator $\hat{E}(Y^{\overline{a}_T})$ at (5) solves the equation above.

B. The asymptotic properties of the estimators

B.1. The estimator with full data

We show that the estimator (2) is asymptotically normally distributed in the situation when T is equal to 1. The two μ-models in (2) are given by $m\{v_1,\boldsymbol{\xi}\}=E(Y\mid\overline{A}_1=\overline{a}_1,\overline{L}_1=\overline{l}_1)$ and $\mu\{v_0,\boldsymbol{\gamma}\}=E(m\{V_1,\boldsymbol{\xi}\}\mid A_0=a_0,L_0=l_0)$. The Z vector defines an ordered sequence of the variables (L_0,A_0,L_1,A_1,Y). The outcome variable Y is assumed to be continuous and we let A_t denote the binary exposure at time $t\in\{0,1\}$. Define $\overline{A}_1$ to be the vector of the two exposures (A_0,A_1) and let $\overline{a}_1$ be the vector (a_0,a_1). Let $Y^{(a_0,a_1)}$ be the potential outcome that would have been observed if (A_0,A_1) had been set to (a_0,a_1). Let L_t be a potential confounder at time $t\in\{0,1\}$ and let $\overline{L}_1$ denote the vector of (L_0,L_1). The outcome Y may be causally influenced by the whole history of $\overline{A}_1$ and $\overline{L}_1$. We define the function

$$\Psi(Z,\boldsymbol{\xi},\boldsymbol{\gamma})=E_{\boldsymbol{\gamma}}(E_{\boldsymbol{\xi}}(Y\mid\overline{A}_1=\overline{a}_1,\overline{L}_1)\mid A_0=a_0,L_0)-E(Y^{(a_0,a_1)})$$

according to Stefanski and Boos [16]. The function $\Psi(Z,\boldsymbol{\xi},\boldsymbol{\gamma})$ fulfills

$$E_F(\Psi(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0))=\int\Psi(z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)dF(z)=0.$$

Define a G_n function to be given by

$$G_n(Z,\hat{\boldsymbol{\xi}},\hat{\boldsymbol{\gamma}})=\frac{1}{n}\sum_{i=1}^{n}E_{\hat{\boldsymbol{\gamma}}}(E_{\hat{\boldsymbol{\xi}}}(Y\mid\overline{A}_{1,i}=\overline{a}_1,\overline{L}_{1,i})\mid A_{0,i}=a_0,L_{0,i})-E(Y^{(a_0,a_1)})$$

so that

$$G_n(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)=\frac{1}{n}\sum_{i=1}^{n}E_{\boldsymbol{\gamma}_0}(E_{\boldsymbol{\xi}_0}(Y\mid\overline{A}_{1,i}=\overline{a}_1,\overline{L}_{1,i})\mid A_{0,i}=a_0,L_{0,i})-E(Y^{(a_0,a_1)}).$$

Let $\hat{E}(Y^{(a_0,a_1)})$ denote the estimator given by

$$\hat{E}(Y^{(a_0,a_1)})=\frac{1}{n}\sum_{i=1}^{n}E_{\hat{\boldsymbol{\gamma}}}(E_{\hat{\boldsymbol{\xi}}}(Y\mid\overline{A}_{1,i}=\overline{a}_1,\overline{L}_{1,i})\mid A_{0,i}=a_0,L_{0,i}).$$

The Taylor expansion of $G_n(Z,\hat{\boldsymbol{\xi}},\hat{\boldsymbol{\gamma}})$ is given by

$$\sqrt{n}G_n(Z,\hat{\boldsymbol{\xi}},\hat{\boldsymbol{\gamma}})=\sqrt{n}G_n(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)+\dot{G}_{\boldsymbol{\xi},n}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)\sqrt{n}(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi}_0)+\dot{G}_{\boldsymbol{\gamma},n}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)\sqrt{n}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)+\sqrt{n}R_n \tag{12}$$

where $\dot{G}_{\boldsymbol{\xi},n}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)$ and $\dot{G}_{\boldsymbol{\gamma},n}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)$ denote

$$\dot{G}_{\boldsymbol{\xi},n}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)=\left.\frac{\partial G_n(Z,\boldsymbol{\xi},\boldsymbol{\gamma})}{\partial\boldsymbol{\xi}^T}\right|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0,\boldsymbol{\gamma}=\boldsymbol{\gamma}_0}\text{ and }\dot{G}_{\boldsymbol{\gamma},n}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)=\left.\frac{\partial G_n(Z,\boldsymbol{\xi},\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^T}\right|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0,\boldsymbol{\gamma}=\boldsymbol{\gamma}_0}$$

respectively. The two partial derivatives $\dot{G}_{\boldsymbol{\xi},n}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)$ and $\dot{G}_{\boldsymbol{\gamma},n}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)$ are given by

$$\dot{G}_{\boldsymbol{\xi},n}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)=\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial}{\partial\boldsymbol{\xi}^T}E_{\boldsymbol{\gamma}}(E_{\boldsymbol{\xi}}(Y\mid\overline{A}_{1,i}=\overline{a}_1,\overline{L}_{1,i})\mid A_{0,i}=a_0,L_{0,i})\right]\Bigg|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0,\boldsymbol{\gamma}=\boldsymbol{\gamma}_0}$$

and

$$\dot{G}_{\boldsymbol{\gamma},n}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)=\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial}{\partial\boldsymbol{\gamma}^T}E_{\boldsymbol{\gamma}}(E_{\boldsymbol{\xi}}(Y\mid\overline{A}_{1,i}=\overline{a}_1,\overline{L}_{i})\mid A_{0,i}=a_0,L_{0,i})\right]\Bigg|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0,\boldsymbol{\gamma}=\boldsymbol{\gamma}_0}$$

respectively. The two partial derivatives $\dot{G}_{\boldsymbol{\xi},n}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)$ and $\dot{G}_{\boldsymbol{\gamma},n}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)$ are two row vectors. If the two row vectors $A_{\boldsymbol{\xi}}(\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)$ and $A_{\boldsymbol{\gamma}}(\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)$ exist then we have by the weak law of large numbers (WLLN)

$$\dot{G}_{\boldsymbol{\xi},n}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)\xrightarrow{p}A_{\boldsymbol{\xi}}(\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)\text{ and }\dot{G}_{\boldsymbol{\gamma},n}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)\xrightarrow{p}A_{\boldsymbol{\gamma}}(\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)$$

where $A_{\boldsymbol{\xi}}(\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)$ is equal to $E(\dot{\Psi}_{\boldsymbol{\xi}}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0))$ with

$$\dot{\Psi}_{\boldsymbol{\xi}}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)=\left[\frac{\partial}{\partial\boldsymbol{\xi}^T}E_{\boldsymbol{\gamma}}(E_{\boldsymbol{\xi}}(Y\mid\overline{A}_1=\overline{a}_1,\overline{L}_1)\mid A_0=a_0,L_0)\right]\Bigg|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0,\boldsymbol{\gamma}=\boldsymbol{\gamma}_0}$$

and $A_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_0)$ being equal to $E(\dot{\Psi}_{\boldsymbol{\gamma}}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0))$ with

$$\dot{\Psi}_{\boldsymbol{\gamma}}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)=\left[\frac{\partial}{\partial\boldsymbol{\gamma}^T}E_{\boldsymbol{\gamma}}(E_{\boldsymbol{\xi}}(Y\mid\overline{A}_1=\overline{a}_1,\overline{L}_1)\mid A_0=a_0,L_0)\right]\Bigg|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0,\boldsymbol{\gamma}=\boldsymbol{\gamma}_0}.$$

We have that $\sqrt{n}G_n(Z,\hat{\boldsymbol{\xi}},\hat{\boldsymbol{\gamma}})$ in (12) converges in distribution to a normal distribution with the mean zero and a variance [16]. This implies that the estimator $\hat{E}(Y^{(a_0,a_1)})$ is asymptotically normally distributed with the mean $E(Y^{(a_0,a_1)})$ and a variance. The empirical variance is obtained using the estimator given by

$$s(a_0,a_1)=\sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\Psi}(Z_i,\hat{\boldsymbol{\xi}},\hat{\boldsymbol{\gamma}})+\dot{G}_{\boldsymbol{\xi},n}(Z,\hat{\boldsymbol{\xi}},\hat{\boldsymbol{\gamma}})m(Z_i,\hat{\boldsymbol{\xi}})+\dot{G}_{\boldsymbol{\gamma},n}(Z,\hat{\boldsymbol{\xi}},\hat{\boldsymbol{\gamma}})k(Z_i,\hat{\boldsymbol{\gamma}})\right\}^2}{n}}$$

with

$$\hat{\Psi}(Z_i,\hat{\boldsymbol{\xi}},\hat{\boldsymbol{\gamma}})=E_{\hat{\boldsymbol{\gamma}}}(E_{\hat{\boldsymbol{\xi}}}(Y\mid\overline{A}_{1,i}=\overline{a}_1,\overline{L}_{1,i})\mid A_{0,i}=a_0,L_{0,i})-\hat{E}(Y^{(a_0,a_1)}).$$

We have that $m(Z_i,\hat{\boldsymbol{\xi}})$ denotes the influence function with the estimated parameter $\hat{\boldsymbol{\xi}}$ obtained using the influence function $m(Z_i,\boldsymbol{\xi})$

$$\sqrt{n}(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi}_0)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}m(Z_i,\boldsymbol{\xi})$$

and $k(Z_i,\hat{\boldsymbol{\gamma}})$ denotes the influence function with the estimated parameter $\hat{\boldsymbol{\gamma}}$ obtained using the influence function $k(Z_i,\boldsymbol{\gamma})$

$$\sqrt{n}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}k(Z_i,\boldsymbol{\gamma}).$$

B.2. The estimator with data with monotone missingness

We continue with the example being T is equal to 1 and the two μ-models are given by $m\{v_1,\boldsymbol{\xi}\}=E(Y\mid\overline{A}_1=\overline{a}_1,\overline{L}_1=\overline{l}_1)$ and $\mu\{v_0,\boldsymbol{\gamma}\}=E(m\{V_1,\boldsymbol{\xi}\}\mid A_0=a_0,L_0=l_0)$. The value of $\mathbf{c}$ is equal to 4. The $G_1(Z)$ vector contains only the first variable L_0 and the $G_2(Z)$ vector contains the first two variables L_0 and A_0. The $G_3(Z)$ vector contains the three variables L_0, A_0 and L_1 and the $G_4(Z)$ vector contains the four variables L_0, A_0, L_1 and A_1. Define a M_n function to be given by

$$\begin{aligned}M_n(Z,\hat{\boldsymbol{\xi}},\hat{\boldsymbol{\gamma}})=&\frac{1}{n}\sum_{i=1}^{n}\frac{I(\mathcal{C}_i=\infty)}{\varpi\{\infty,Z_i,\hat{\boldsymbol{\psi}}\}}E_{\hat{\boldsymbol{\gamma}}}(E_{\hat{\boldsymbol{\xi}}}(Y\mid\overline{A}_{1,i}=\overline{a}_1,\overline{L}_{1,i})\mid A_{0,i}=a_0,L_{0,i})\\&+\sum_{r=1}^{4}\frac{I(\mathcal{C}_i=r)-\lambda_r\{G_r(Z_i),\hat{\boldsymbol{\psi}}\}I(\mathcal{C}_i\geq r)}{K_r\{G_r(Z_i),\hat{\boldsymbol{\psi}}\}}E\left(E_{\boldsymbol{\gamma}}(E_{\boldsymbol{\xi}}(Y\mid\overline{A}_1=\overline{a}_1,\overline{L}_1)\mid A_0=a_0,L_0)\mid G_r(Z_i),\hat{\boldsymbol{\zeta}}\right)\\&-E(Y^{(a_0,a_1)}).\end{aligned}$$

Let $\hat{E}(Y^{(a_0,a_1)})$ denote the estimator to be given by

$$\hat{E}(Y^{(a_0,a_1)}) = \frac{1}{n}\sum_{i=1}^{n}\frac{I(\mathcal{C}_i=\infty)}{\varpi\{\infty,Z_i,\hat{\psi}\}}E_{\hat{\gamma}}(E_{\hat{\xi}}(Y \mid \overline{A}_{1,i}=\overline{a}_1,\overline{L}_{1,i}) \mid A_{0,i}=a_0,L_{0,i}) + \sum_{r=1}^{4}\frac{I(\mathcal{C}_i=r)-\lambda_r\{G_r(Z_i),\hat{\psi}\}I(\mathcal{C}_i\geq r)}{K_r\{G_r(Z_i),\hat{\psi}\}}E\left(E_{\gamma}(E_{\xi}(Y \mid \overline{A}_1=\overline{a}_1,\overline{L}_1) \mid A_0=a_0,L_0) \mid G_r(Z_i),\hat{\zeta}\right).$$

The Taylor expansion of $M_n(Z,\hat{\xi},\hat{\gamma})$ is given by

$$\sqrt{n}M_n(Z,\hat{\xi},\hat{\gamma}) = \sqrt{n}M_n(Z,\xi_0,\gamma_0) + \dot{M}_{\xi,n}(Z,\xi_0,\gamma_0)\sqrt{n}(\hat{\xi}-\xi_0) + \dot{M}_{\gamma,n}(Z,\xi_0,\gamma_0)\sqrt{n}(\hat{\gamma}-\gamma_0) + \sqrt{n}R_n \quad (13)$$

where $\dot{M}_{\xi,n}(Z,\xi_0,\gamma_0)$ and $\dot{M}_{\gamma,n}(Z,\xi_0,\gamma_0)$ denote

$$\dot{M}_{\xi,n}(Z,\xi_0,\gamma_0) = \left.\frac{\partial M_n(Z,\xi,\gamma)}{\partial \xi^T}\right|_{\xi=\xi_0,\gamma=\gamma_0} \text{ and } \dot{M}_{\gamma,n}(Z,\xi_0,\gamma_0) = \left.\frac{\partial M_n(Z,\xi,\gamma)}{\partial \gamma^T}\right|_{\xi=\xi_0,\gamma=\gamma_0}$$

respectively. The two partial derivatives $\dot{M}_{\xi,n}(Z,\xi_0,\gamma_0)$ and $\dot{M}_{\gamma,n}(Z,\xi_0,\gamma_0)$ are given by

$$\dot{M}_{\xi,n}(Z,\xi_0,\gamma_0) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{I(\mathcal{C}_i=\infty)}{\varpi\{\infty,Z_i,\hat{\psi}\}}\frac{\partial}{\partial \xi^T}E_{\gamma}(E_{\xi}(Y \mid \overline{A}_{1,i}=\overline{a}_1,\overline{L}_{1,i}) \mid A_{0,i}=a_0,L_{0,i}) + \sum_{r=1}^{4}\frac{I(\mathcal{C}_i=r)-\lambda_r\{G_r(Z_i),\hat{\psi}\}I(\mathcal{C}_i\geq r)}{K_r\{G_r(Z_i),\hat{\psi}\}}\times E\left(\frac{\partial}{\partial \xi^T}E_{\gamma}(E_{\xi}(Y \mid \overline{A}_1=\overline{a}_1,\overline{L}_1) \mid A_0=a_0,L_0) \mid G_r(Z_i),\hat{\zeta}\right)\right]\Bigg|_{\xi=\xi_0,\gamma=\gamma_0}$$

and

$$\dot{M}_{\gamma,n}(Z,\xi_0,\gamma_0) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{I(\mathcal{C}_i=\infty)}{\varpi\{\infty,Z_i,\hat{\psi}\}}\frac{\partial}{\partial \gamma^T}E_{\gamma}(E_{\xi}(Y \mid \overline{A}_{1,i}=\overline{a}_1,\overline{L}_{1,i}) \mid A_{0,i}=a_0,L_{0,i}) + \sum_{r=1}^{4}\frac{I(\mathcal{C}_i=r)-\lambda_r\{G_r(Z_i),\hat{\psi}\}I(\mathcal{C}_i\geq r)}{K_r\{G_r(Z_i),\hat{\psi}\}}\times E\left(\frac{\partial}{\partial \gamma^T}E_{\gamma}(E_{\xi}(Y \mid \overline{A}_1=\overline{a}_1,\overline{L}_1) \mid A_0=a_0,L_0) \mid G_r(Z_i),\hat{\zeta}\right)\right]\Bigg|_{\xi=\xi_0,\gamma=\gamma_0}$$

respectively. The two partial derivatives $\dot{M}_{\xi,n}(Z,\xi_0,\gamma_0)$ and $\dot{M}_{\gamma,n}(Z,\xi_0,\gamma_0)$ are two row vectors.

Let the $\lambda_r\{G_r(Z),\psi^*\}$-models relating to the missingness mechanism be correctly specified (such as $\psi^*=\psi_0$) or let the $E(E_{\gamma}(E_{\xi}(Y \mid \overline{A}=\overline{a},\overline{L}) \mid A_0=a_0,L_0) \mid G_r(Z),\zeta^*)$-models with respect to the distribution of Z be correctly specified (such as $\zeta^*=\zeta_0$) so that we then have

$$E\left[\sum_{r=1}^{4}\frac{I(\mathcal{C}=r)-\lambda_r\{G_r(Z),\psi^*\}I(\mathcal{C}\geq r)}{K_r\{G_r(Z),\psi^*\}}\left(\dot{\Psi}_{\xi}(Z,\xi_0,\gamma_0)-E\left(\dot{\Psi}_{\xi}(Z,\xi_0,\gamma_0) \mid G_r(Z),\zeta^*\right)\right)\right]=0 \quad (14)$$

and

$$E\left[\sum_{r=1}^{4}\frac{I(\mathcal{C}=r)-\lambda_r\{G_r(Z),\psi^*\}I(\mathcal{C}\geq r)}{K_r\{G_r(Z),\psi^*\}}\left(\dot{\Psi}_{\gamma}(Z,\xi_0,\gamma_0)-E\left(\dot{\Psi}_{\gamma}(Z,\xi_0,\gamma_0) \mid G_r(Z),\zeta^*\right)\right)\right]=0. \quad (15)$$

The expectation (14) implies that

$$E\left[\dot{\Psi}_{\xi}(Z,\xi_0,\gamma_0)-\sum_{r=1}^{4}\frac{I(\mathcal{C}=r)-\lambda_r\{G_r(Z),\psi^*\}I(\mathcal{C}\geq r)}{K_r\{G_r(Z),\psi^*\}}\left(\dot{\Psi}_{\xi}(Z,\xi_0,\gamma_0)-E\left(\dot{\Psi}_{\xi}(Z,\xi_0,\gamma_0) \mid G_r(Z),\zeta^*\right)\right)\right] = E\left[\dot{\Psi}_{\xi}(Z,\xi_0,\gamma_0)\right]$$

and the expectation (15) implies that

$$E\left[\dot{\Psi}_{\boldsymbol{\gamma}}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)-\sum_{r=1}^{4}\frac{I(\mathcal{C}=r)-\lambda_r\{G_r(Z),\boldsymbol{\psi}^*\}I(\mathcal{C}\geq r)}{K_r\{G_r(Z),\boldsymbol{\psi}^*\}}\left(\dot{\Psi}_{\boldsymbol{\gamma}}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)-E\left(\dot{\Psi}_{\boldsymbol{\gamma}}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)\mid G_r(Z),\boldsymbol{\zeta}^*\right)\right)\right]$$
$$=E\left[\dot{\Psi}_{\boldsymbol{\gamma}}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)\right].$$

If the two row vectors $A_{\boldsymbol{\xi}}(\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)$ and $A_{\boldsymbol{\gamma}}(\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)$ exist then we have by the WLLN

$$\dot{M}_{\boldsymbol{\xi},n}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)\xrightarrow{p}A_{\boldsymbol{\xi}}(\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)\text{ and }\dot{M}_{\boldsymbol{\gamma},n}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)\xrightarrow{p}A_{\boldsymbol{\gamma}}(\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)$$

where $A_{\boldsymbol{\xi}}(\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0)$ is given by $E(\dot{\Psi}_{\boldsymbol{\xi}}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0))$ and $A_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}_0,\boldsymbol{\gamma}_0)$ is given by $E(\dot{\Psi}_{\boldsymbol{\gamma}}(Z,\boldsymbol{\xi}_0,\boldsymbol{\gamma}_0))$. We have that $\sqrt{n}M_n(Z,\hat{\boldsymbol{\xi}},\hat{\boldsymbol{\gamma}})$ in (13) converges in distribution to a normal distribution with the mean zero and a variance [16]. This implies that the estimator $\hat{E}(Y^{(a_0,a_1)})$ is asymptotically normally distributed with the mean $E(Y^{(a_0,a_1)})$ and a variance. The empirical variance is obtained using the estimator given by

$$s(a_0,a_1)=\sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\Psi}(Z_i,\hat{\boldsymbol{\xi}},\hat{\boldsymbol{\gamma}})+\dot{M}_{\boldsymbol{\xi},n}(Z,\hat{\boldsymbol{\xi}},\hat{\boldsymbol{\gamma}})m(Z_i,\hat{\boldsymbol{\xi}})+\dot{M}_{\boldsymbol{\gamma},n}(Z,\hat{\boldsymbol{\xi}},\hat{\boldsymbol{\gamma}})k(Z_i,\hat{\boldsymbol{\gamma}})\right\}^2}{n}}$$

with

$$\hat{\Psi}(Z_i,\hat{\boldsymbol{\xi}},\hat{\boldsymbol{\gamma}})=E_{\hat{\boldsymbol{\gamma}}}(E_{\hat{\boldsymbol{\xi}}}(Y\mid\overline{A}_{1,i}=\overline{a}_1,\overline{L}_{1,i})\mid A_{0,i}=a_0,L_{0,i})-\hat{E}(Y^{(a_0,a_1)}).$$

We have that $m(Z_i,\hat{\boldsymbol{\xi}})$ denotes the influence function with the estimated parameter $\hat{\boldsymbol{\xi}}$ obtained using the influence function $m(Z_i,\boldsymbol{\xi})$ such that

$$\sqrt{n}(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi}_0)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}m(Z_i,\boldsymbol{\xi})$$

and $k(Z_i,\hat{\boldsymbol{\gamma}})$ denotes the influence function with the estimated parameter $\hat{\boldsymbol{\gamma}}$ obtained using the influence function $k(Z_i,\boldsymbol{\gamma})$ such that

$$\sqrt{n}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}k(Z_i,\boldsymbol{\gamma}).$$

We have now shown that the estimator (2) and the estimator (5) are asymptotically normally distributed when T is equal to 1 but this can also be shown for a larger T. The Taylor expansion at (12) and (13) will contain more parameters for a larger T. We conduct a simulation study in the next Subsection.

B.3. Simulation study

We consider only full data as an illustrative example. All the models are correctly specified in the simulation study. We consider data with a sample size of 1000 and the data are replicated 2000 times. The data are simulated as follows: $L_0\sim\text{Normal}(\eta_{l_0},1^2)$, $A_0\sim\text{Bernoulli}(\varkappa_{a_0})$, $L_1\sim\text{Normal}(\eta_{l_1},1^2)$, $A_1\sim\text{Bernoulli}(\varkappa_{a_1})$ and $Y\sim\text{Normal}(\eta_y,1^2)$ where the means and the probabilities are given by $\eta_{l_0}:=0$, $\text{logit}(\varkappa_{a_0}):=-0.2L_0$, $\eta_{l_1}:=1.5+0.4L_0+A_0$, $\text{logit}(\varkappa_{a_1}):=-0.2L_1-L_0+A_0$ and $\eta_y:=0.4L_0+2.7A_0+L_1+A_1$.

We consider the coverage of the estimator with full data given by

$$E(Y^{(a_0,a_1)})\in\left(\hat{E}(Y^{(a_0,a_1)})-z_{0.975}s(a_0,a_1),\hat{E}(Y^{(a_0,a_1)})+z_{0.975}s(a_0,a_1)\right)$$

where $z_{0.975}$ is equal to the value of the 97.5 percentile point of the standard normal distribution. We count the number of times when the expected potential outcome is included in their confidence interval.

Table 4 shows the percentage of the 2000 times of the coverage using the estimator with full data, the average of the 2000 estimates of the empirical standard deviation and the empirical standard deviation of the 2000 estimates of $\hat{E}(Y^{(a_0,a_1)})$.

Table 4 shows that the coverages of the estimator are close to 95% and the average of the empirical standard deviations are almost equal to the empirical standard deviations of $\hat{E}(Y^{(a_0,a_1)})$. The simulation study shows and supports the theory that the estimator with full data is asymptotically normally distributed.

	$a_0 = 0, a_1 = 0$	$a_0 = 1, a_1 = 0$	$a_0 = 0, a_1 = 1$	$a_0 = 1, a_1 = 1$
Coverage	95.40%	94.60%	94.90%	94.95%
$s(a_0, a_1)$	0.075	0.081	0.079	0.074
$sd(\hat{E}(Y^{(a_0,a_1)}))$	0.074	0.079	0.079	0.073

Table 4. The estimator with full data. The *Coverage* row shows the percentage of the coverage. The $s(a_0, a_1)$ row shows the average of the 2000 estimates of the empirical standard deviation when (a_0, a_1) is set to $(0, 0)$, $(1, 0)$, $(0, 1)$ or $(1, 1)$. The $sd(\hat{E}(Y^{(a_0,a_1)}))$ row shows the empirical standard deviation of the 2000 estimates of $\hat{E}(Y^{(a_0,a_1)})$.

References

[1] Chokka P, Bougie J, Proulx J, Tvistholm AH, Ettrup A. Long-term functioning outcomes are predicted by cognitive symptoms in working patients with major depressive disorder treated with vortioxetine: results from the atworc study (book review). *CNS Spectrums* 2019;:1–12.

[2] Hammer-Helmich L, Haro JM, Jönsson B, *et al.* Functional impairment in patients with major depressive disorder: the 2-year perform study. *Neuropsychiatric Disease and Treatment* 2018;**14**:239–249.

[3] Daniel RM, Cousens S, Stavola BD, Kenward M, Sterne JA. Methods for dealing with time-dependent confounding. *Statistics in Medicine* 2013;**32**(9):1584–1618.

[4] Bartlett JW, Seaman SR, White IR, Carpenter JR. Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model. *Statistical Methods in Medical Research* 2015;**24**(4):462–487.

[5] Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research* 2013;**22**(3):278–295.

[6] Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* 1994;**89**(427):846–866.

[7] Robins JM, Rotnitzky A. Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association* 1995;**90**(429):122–129.

[8] Rotnitzky A, Robins J. Analysis of semiparametric regression models with nonignorable nonresponse. *Statistics in Medicine* 1997;**16**(1):81–102.

[9] Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. *Biometrics* 2005; **61**(4):962–973.

[10] Williamson EJ, Forbes A, Wolfe R. Doubly robust estimators of causal exposure effects with missing data in the outcome, exposure or a confounder. *Statistics in Medicine* 2012;**31**(30):4382–4400.

[11] Tsiatis AA Anastasios A. Semiparametric theory and missing data. Springer series in statistics. Springer, 2006.

[12] Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect (book review). *Mathematical Modelling* 1986;**7**(9-12):1393–1512.

[13] Kreif N, Tran L, Grieve R, Stavola BD, Tasker RC, Petersen M. Estimating the comparative effectiveness of feeding interventions in the pediatric intensive care unit: A demonstration of longitudinal targeted maximum likelihood estimation. *American Journal of Epidemiology* 2017;**186**(12):1370–1379.

[14] Haro JM, Hammer-Helmich L, Saragoussi D, Ettrup A, Larsen KG. Patient-reported depression severity and cognitive symptoms as determinants of functioning in patients with major depressive disorder: a secondary analysis of the 2-year prospective perform study. *Neuropsychiatric Disease and Treatment* 2019;**15**:2313–2323.

[15] SAS Institute Inc. "LSMEANS statement." in sas/stat 9.3 user's guide. http://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#statug_introcom_a0000003362.htm, 2012.

[16] Stefanski LA, Boos DD. The calculus of m-estimation. *The American statistician* 2002;**56**(1):29–38.

[17] Keogh RH, Daniel RM, VanderWeele TJ, Vansteelandt S. Analysis of longitudinal studies with repeated outcome measures: Adjusting for time-dependent confounding using conventional methods. *American journal of epidemiology* 2018;**187**(5).

[18] Hernán M, Robins J. Causal Inference. "Boca Raton: Chapman and Hall/CRC", 2017. For free.

[19] Pearl J. An introduction to causal inference. *The international journal of biostatistics* 2010;**6**(2).

[20] Pearl J. Causality: Models, Reasoning, and Inference. Cambridge: Cambridge University Press., 2nd edition, 2009.

[21] Robins JM, Ángel Hernánl M, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000;**11**(5):550–560.

Research Article

Statistics
in Medicine

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

Supplementary material for "A doubly robust estimator for monotone missing data in the presence of time-dependent confounding"

Thomas Maltesen[ab*], Torben Martinussen[b], Klaus Groes Larsen[c] and Lene Hammer-Helmich[d]

1. Analysis of the PERFORM study

Table 1 displays the different vectors that contribute to a specific Z_t for $t \in \{b, 2, 6, 12, 18\}$.

t	Z_t	W_{pt}	W_t	W_{st}
b	Z_b		(PHQ_b, PDQ_b, SDS_b)	(PHQ_2, PDQ_2, SDS_2)
2	Z_2	(PHQ_b, PDQ_b, SDS_b)	(PHQ_2, PDQ_2, SDS_2)	(PHQ_6, PDQ_6, SDS_6)
6	Z_6	(PHQ_2, PDQ_2, SDS_2)	(PHQ_6, PDQ_6, SDS_6)	$(PHQ_{12}, PDQ_{12}, SDS_{12})$
12	Z_{12}	(PHQ_6, PDQ_6, SDS_6)	$(PHQ_{12}, PDQ_{12}, SDS_{12})$	$(PHQ_{18}, PDQ_{18}, SDS_{18})$
18	Z_{18}	$(PHQ_{12}, PDQ_{12}, SDS_{12})$	$(PHQ_{18}, PDQ_{18}, SDS_{18})$	$(PHQ_{24}, PDQ_{24}, SDS_{24})$

Table 1. The t column displays the five periods and Z_t displays the five different possible combinations of vectors. The W_{pt}, W_t and W_{st} columns indicate which vectors that contribute to a specific Z_t.

Table 2 displays the time-varying exposure, the time-dependent confounding and the outcome for all five periods.

t	L_t	PDQ_t	L_{st}	PDQ_{st}	SDS_{st}
b	(PHQ_b)	PDQ_b	(SDS_b, PHQ_2)	PDQ_2	SDS_2
2	$(PHQ_b, PDQ_b, SDS_b, PHQ_2)$	PDQ_2	(SDS_2, PHQ_6)	PDQ_6	SDS_6
6	$(PHQ_2, PDQ_2, SDS_2, PHQ_6)$	PDQ_6	(SDS_6, PHQ_{12})	PDQ_{12}	SDS_{12}
12	$(PHQ_6, PDQ_6, SDS_6, PHQ_{12})$	PDQ_{12}	(SDS_{12}, PHQ_{18})	PDQ_{18}	SDS_{18}
18	$(PHQ_{12}, PDQ_{12}, SDS_{12}, PHQ_{18})$	PDQ_{18}	(SDS_{18}, PHQ_{24})	PDQ_{24}	SDS_{24}

Table 2. The t column displays the five periods and L_t displays the confounder at time t. The PDQ_t column displays the exposure at time t. The L_{st} column displays the confounder at time st. The PDQ_{st} column displays the exposure at time st. The SDS_{st} column displays the outcome at time st.

[a] *Health Economics and Epidemiology Statistics, H.Lundbeck A/S, 2500 Valby, Denmark*
[b] *Section of Biostatistics, University of Copenhagen, 1014 Copenhagen K, Denmark*
[c] *Statistics, H.Lundbeck A/S, 2500 Valby, Denmark*
[d] *Medical Affairs, Value Evidence, H.Lundbeck A/S, 2500 Valby, Denmark*
[*] *Correspondence to: Thomas Maltesen, H.Lundbeck A/S, 2500 Valby, Denmark. E-mail: thomas.maltesen@protonmail.com*

Prepared using simauth.cls *[Version: 2010/03/10 v3.00]*

Table 3 displays the estimates and the confidence intervals of the four estimators for $t = b$ and Table 4 displays the estimates and the confidence intervals of the four estimators for $t = 18$.

Estimator		Estimate	SE	95%-CI	Width of CI
Naïve	$(1,1)$	16.724	0.455	(15.833 ; 17.616)	1.784
	$(0,1)$	13.583	1.892	(9.876 ; 17.291)	7.415
	$(1,0)$	9.108	1.147	(6.860 ; 11.356)	4.495
	$(0,0)$	5.737	0.794	(4.181 ; 7.293)	3.112
LSmeans	$(1,1)$	15.080	0.437	(14.223 ; 15.937)	1.713
	$(0,1)$	15.330	1.900	(11.607 ; 19.054)	7.447
	$(1,0)$	13.345	1.688	(10.037 ; 16.654)	6.616
	$(0,0)$	11.862	1.061	(9.782 ; 13.941)	4.159
G-formula	$E(SDS_2^{(1,1)})$	15.471	0.453	(14.583 ; 16.358)	1.775
	$E(SDS_2^{(0,1)})$	13.154	2.063	(9.110 ; 17.198)	8.088
	$E(SDS_2^{(1,0)})$	13.753	1.606	(10.606 ; 16.900)	6.293
	$E(SDS_2^{(0,0)})$	9.588	1.204	(7.229 ; 11.948)	4.720
DRMGf	$E(SDS_2^{(1,1)})$	15.477	0.419	(14.656 ; 16.299)	1.643
	$E(SDS_2^{(0,1)})$	13.491	2.063	(9.447 ; 17.535)	8.088
	$E(SDS_2^{(1,0)})$	13.672	1.734	(10.273 ; 17.071)	6.798
	$E(SDS_2^{(0,0)})$	9.880	1.211	(7.507 ; 12.253)	4.746

Table 3. The plot on the left hand side in Figure 2 is based on the actual numbers. The Estimator column displays the estimator that has been used to obtain the estimates. The Estimate column displays the estimates from the different estimators. The SE column displays the standard errors obtained using 1000 bootstraps. The 95%-CI column displays the confidence intervals for the estimates. The Width of CI column displays the width of the confidence intervals.

Estimator		Estimate	SE	95%-CI	Width of CI
Naïve	$(1,1)$	14.423	0.728	(12.997 ; 15.850)	2.853
	$(0,1)$	7.148	1.267	(4.665 ; 9.631)	4.966
	$(1,0)$	4.565	0.812	(2.973 ; 6.158)	3.185
	$(0,0)$	3.953	0.648	(2.680 ; 5.222)	2.542
LSmeans	$(1,1)$	10.554	0.660	(9.262 ; 11.847)	2.585
	$(0,1)$	8.806	1.268	(6.321 ; 11.291)	4.970
	$(1,0)$	7.642	1.177	(5.336 ; 9.949)	4.613
	$(0,0)$	9.485	1.188	(7.156 ; 11.814)	4.658
G-formula	$E(SDS_{24}^{(1,1)})$	11.100	0.707	(9.715 ; 12.485)	2.770
	$E(SDS_{24}^{(0,1)})$	7.916	1.192	(5.580 ; 10.252)	4.672
	$E(SDS_{24}^{(1,0)})$	8.436	1.369	(5.752 ; 11.120)	5.367
	$E(SDS_{24}^{(0,0)})$	8.238	1.230	(5.827 ; 10.649)	4.822
DRMGf	$E(SDS_{24}^{(1,1)})$	12.069	0.623	(10.847 ; 13.291)	2.444
	$E(SDS_{24}^{(0,1)})$	8.618	1.293	(6.083 ; 11.153)	5.070
	$E(SDS_{24}^{(1,0)})$	9.458	1.402	(6.709 ; 12.207)	5.498
	$E(SDS_{24}^{(0,0)})$	9.031	1.338	(6.410 ; 11.653)	5.243

Table 4. The plot on the right hand side in Figure 2 is based on the actual numbers. See Table 3 for the description of the columns.

Manuscript II

Sequential mediation analysis with multiple mediators

Research Article

Statistics
in Medicine

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

Sequential mediation analysis with multiple mediators

Thomas Maltesen[ab*], Klaus Groes Larsen[c] and Lene Hammer-Helmich[d]

The natural direct effect and the natural indirect effect from causal inference are attractive since the sum of the two effects are equal to the total causal effect. Unfortunately, identification of these effects relies on the cross-world assumption which is violated if there exists a mediator-outcome relationship. The interventional direct effect and the interventional indirect effect avoid the cross-world assumption but the overall effect is not necessarily equal to the total causal effect. We suggest a new definition for mediation analysis where the overall effect is equal to the total causal effect. We use different simulation scenarios to show that our definition can include models with interactions with the overall effect equal to the total causal effect. We use our definition to analyse the observational cohort study Prospective Epidemiological Research on Functioning Outcomes Related to Major depressive disorder (PERFORM). The (PERFORM) study includes two mediators on the causal path between exposure and outcome.

Keywords: causal inference, direct and indirect effects, longitudinal data, marginal structural model, sequential mediation

1. Introduction

Causal mediation analysis is applied in different scientific disciplines e.g. epidemiology, political science, psychology and sociology, whenever interest lies in investigating the extent to which the effect of exposure on the outcome is transmitted via one or more intermediate variables [1]. Methods in social sciences have used structural equation modelling (SEM) to perform such analysis. The structural equation modelling (SEM) consists of a series of multivariate linear models that are combined in a single analysis. The model allows the scales at the same time point to be correlated, however, these methods do not allow for interactions in the models. The indirect effect of the exposure on the outcome may also be obtained by subtracting the direct effect from the total effect but the strategy will fail if an interaction between the mediator and the exposure on the outcome exists [2]. The direct and indirect effects can be defined and the total effect can also be dismantled into direct and indirect effects involving interactions in the models [3].

Robins, Greenland [4] and Pearl [5] proposed the natural direct and indirect effects. These effects are used to obtain the direct effect of the exposure on the outcome and the transmitted effect via one or more intermediate measurements between the exposure on the outcome. The natural direct and indirect effects have the attractive property to sum the total causal effect but it relies on the cross-world assumption. The assumption does not necessarily hold in experimental data [6, 7]. VanderWeele, Vansteelandt and Robins [8] define the interventional direct and indirect effects between the exposure and the outcome. The identifications of the interventional effects avoid the cross-world assumption, but one drawback of the interventional direct and indirect effects is that the sum may not be equal to the total causal effect. Therefore it is called the overall effect instead.

An observational cohort study, Prospective Epidemiological Research on Functioning Outcomes Related to Major depressive disorder (PERFORM), was conducted to better understand the course of a depressive episode and its impact

[a] *Health Economics and Epidemiology Statistics, H.Lundbeck A/S, 2500 Valby, Denmark*
[b] *Section of Biostatistics, University of Copenhagen, 1014 Copenhagen K, Denmark*
[c] *Statistics, H.Lundbeck A/S, 2500 Valby, Denmark*
[d] *Medical Affairs, Value Evidence, H.Lundbeck A/S, 2500 Valby, Denmark*
[*] *Correspondence to: Thomas Maltesen, H.Lundbeck A/S, 2500 Valby, Denmark. E-mail: thomas.maltesen@protonmail.com*

on patient functioning over 2 years in outpatients with major depressive disorder (MDD). Depression severity, cognitive symptoms and functional impairment were measured for each patient at six time points. The measurements at the first two time points (baseline and after 2 months) are used as an example for our definition. The measurement of functional impairment at baseline and depression severity at month 2 are the two mediators on the causal path between the exposure (cognitive symptoms at baseline) and the outcome (functional impairment at month 2). See Hammer-Helmich et al. [9] for further information about the PERFORM study. The treatment of cognitive symptoms may hold the key to achieving functional recovery in MDD [10]. Unfortunately, using the g-formula in the presence of time-dependent confounding will only give the (total) causal effect of the cognitive symptoms and it does not facilitate understanding of how cognition may act on functionality via mediators. We are interested in the direct effect of the exposure (cognitive symptoms at baseline) on the outcome (functional impairment at month 2) and for the sake of the interpretation we would prefer if the overall effect is equal to the total causal effect.

In this paper, we are considering the definition proposed by Vansteelandt and Daniel [11] for defining the interventional direct and indirect effects for multiple mediators. The overall effect is equal to the total causal effect by introducing a mediated dependence term. Their definition does not require that the variables are in an ordered sequence. Section 2 revisits briefly the setup with only one mediator and the interpretation of the effects under different assumptions. Section 3 reviews the definition proposed by Vansteelandt and Daniel [11] to obtain the interventional direct and indirect effects for multiple mediators. Section 4 compares our definition of sequential mediation to a definition by VanderWeele and Vansteelandt [12] using three different simulation studies. Section 5 contains the analysis of the PERFORM study with our definition. Section 6 concludes with a discussion of our findings.

2. Mediation and the corresponding effects

Let Y denote the outcome, which we assume to be continuous, and let A and M denote the exposure and the mediator respectively. Let C be some baseline measurements not affected by the exposure, see the directed acyclic graph (DAG) denoted by (a) in Figure 1. Let Y^a and M^a be the values that Y and M would be if the exposure A is set to a respectively. Let Y^{am} be the value that Y would be if the exposure A is set to a and the mediator M is set to m. Pearl [5] defined the controlled direct effect as the difference between the two expected potential outcomes $E(Y^{am})$ and $E(Y^{a^*m})$ for two different values of the exposure, a and a^*, when the mediator is kept fixed at level m. Robins, Greenland [4] and Pearl [5] define the natural direct effect by $E(Y^{aM^{a^*}}) - E(Y^{a^*M^{a^*}})$ for two different values of the exposure a and a^*, but the mediator set to its natural level of A had been set to a^*. Robins, Greenland [4] and Pearl [5] define the natural indirect effect by $E(Y^{aM^a}) - E(Y^{aM^{a^*}})$ which defines the effect of the exposure A on Y mediated via the mediator M. The total causal effect is defined by $E(Y^{aM^a}) - E(Y^{a^*M^{a^*}})$.

VanderWeele el al. [8] define the random variable $G^{a|C}$. The random $G^{a|C}$ denotes a random drawn mediator from the distribution among those with exposure status a conditional on C. VanderWeele el al. [8] define the interventional direct effect by

$$E\left(Y^{aG^{a^*|C}} - Y^{a^*G^{a^*|C}}\right) = \int \left(E(Y^{am} \mid C=c) - E(Y^{a^*m} \mid C=c)\right) f_{M|A,C}(m \mid a^*, c) f_C(c) d(m,c) \qquad (1)$$

and VanderWeele el al. [8] define the interventional indirect effect by

$$E\left(Y^{aG^{a|C}} - Y^{aG^{a^*|C}}\right) = \int E(Y^{am} \mid C=c) \left(f_{M|A,C}(m \mid a, c) - f_{M|A,C}(m \mid a^*, c)\right) f_C(c) d(m,c). \qquad (2)$$

The disadvantage of the interventional effects is that the sum of (1) and (2) will not necessarily be equal to the total causal effect. The sum is (sometimes) called the overall effect instead. The overall effect is defined by $E(Y^{aG^{a|C}}) - E(Y^{a^*G^{a^*|C}})$.

The assumptions for the identification of these five effects (the controlled effect, the interventional direct and indirect effects and the natural direct and indirect effects) with only one mediator are listed below:
The controlled effect is identified if we assume these two assumptions:

(i) $Y^{am} \perp\!\!\!\perp A \mid C = c \quad \forall (a,m,c) \in A, M, C$
(ii) $Y^{am} \perp\!\!\!\perp M \mid A = a, C = c \quad \forall (a,m,c) \in A, M, C.$

The interventional direct and indirect effects are identified if we assume the two assumptions (i) and (ii) and we also assume the following assumption:

(iii) $M^a \perp\!\!\!\perp A \mid C = c \quad \forall (a, m, c) \in A, M, C.$

The natural direct and indirect effects are identified if we assume the previous three assumptions (i), (ii) and (iii) and we also assume the following assumption:

(iv) $Y^{am} \perp\!\!\!\perp M^{a^*} \mid C = c \quad \forall (a, a^*, m, c) \in A, M, C.$

The last assumption is the cross-world assumption.

3. Interventional direct and indirect effects

We assume 2 mediators for simplicity. We also assume that the two mediators are ordered meaning that M_1 may affect M_2 but not the other way around. The two mediators may thus result in 4 causal paths: the direct effect of A on Y ($A \to Y$), the indirect effect of A on Y via M_1 only ($A \to M_1 \to Y$), via M_2 only ($A \to M_2 \to Y$), and the indirect effect of A on Y via both M_1 and M_2 ($A \to M_1 \to M_2 \to Y$) [13]. See the DAG denoted by (b) in Figure 1. Let $\overline{M}_2$ define the vector (M_1, M_2) and let $\overline{M}_2^a$ denote the vector (M_1^a, M_2^a).

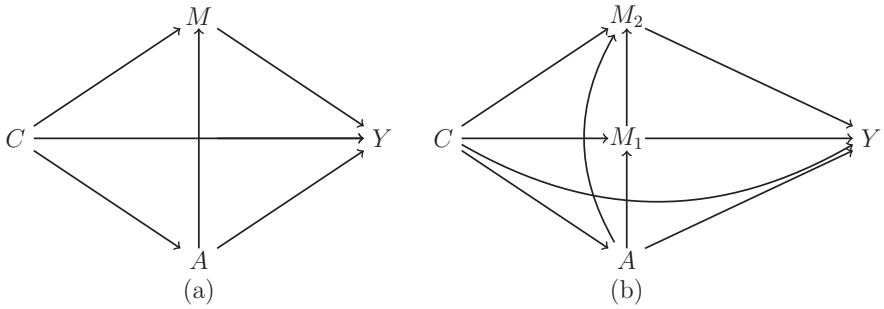


Figure 1. Let Y denote the outcome, which we assume to be continuous. Let A denote the exposure and let C be some baseline measurements that are not affected by the exposure. DAG (a) shows the causal paths with one mediator M. DAG (b) shows the causal paths with two mediators M_1 and M_2 ($\overline{M}_2 = (M_1, M_2)$).

3.1. The interventional effects for multiple mediators by Vansteelandt and Daniel

Vansteelandt and Daniel [11] do not require that the mediator variables are ordered in a sequence. Vansteelandt and Daniel define the interventional direct effect by the causal estimand

$$\int \Big(E(Y^{am_1m_2} \mid c) - E(Y^{a^*m_1m_2} \mid c) \Big) f_{\overline{M}_2^{a^*}|C}(\overline{m}_2 \mid c) f_C(c) d(\overline{m}_2, c). \tag{3}$$

Vansteelandt and Daniel define the causal estimand for the interventional indirect effect via M_1 to be given by

$$\int E(Y^{am_1m_2} \mid c) \Big(f_{M_1^a|C}(m_1 \mid c) - f_{M_1^{a^*}|C}(m_1 \mid c) \Big) f_{M_2^{a^*}|C}(m_2 \mid c) f_C(c) d(\overline{m}_2, c) \tag{4}$$

and define the causal estimand for the interventional indirect effect via M_2 to be given by

$$\int E(Y^{am_1m_2} \mid c) \Big(f_{M_2^a|C}(m_2 \mid c) - f_{M_2^{a^*}|C}(m_2 \mid c) \Big) f_{M_1^a|C}(m_1 \mid c) f_C(c) d(\overline{m}_2, c). \tag{5}$$

Vansteelandt and Daniel define the additional term by

$$\begin{aligned} \int E(Y^{am_1m_2} \mid c) \Big(& f_{\overline{M}_2^a|C}(\overline{m}_2 \mid c) - f_{M_2^a|C}(m_2 \mid c) f_{M_1^a|C}(m_1 \mid c) + \\ & f_{M_2^{a^*}|C}(m_2 \mid c) f_{M_1^{a^*}|C}(m_1 \mid c) - f_{\overline{M}_2^{a^*}|C}(\overline{m}_2 \mid c) \Big) f(c) d(\overline{m}_2, c) \end{aligned} \tag{6}$$

that captures the indirect effect of A on Y mediated via the dependence of M_2 on M_1. The effects (3) to (6) sum to the total causal effect [11]. The assumptions used to identify the estimands above (for two mediators) are

(i') $Y^{a\overline{m}_2} \perp\!\!\!\perp A \mid C = c \quad \forall (a, \overline{m}_2, c) \in A, \overline{M}_2, C$
(ii') $Y^{a\overline{m}_2} \perp\!\!\!\perp (M_1, M_2) \mid A = a, C = c \quad \forall (a, \overline{m}_2, c) \in A, \overline{M}_2, C$
(iii') $(M_2^a, M_1^a) \perp\!\!\!\perp A \mid C = c \quad \forall (a, c) \in A, C.$

The interventional direct and indirect effects have the advantage of being meaningful even when the variable is not manipulable [14, 11]. See the example in Section 5. The interpretation of these three assumptions (i'), (ii') and (iii') are given by: We have with the assumption (i') that the effect of the exposure A on the outcome Y is unconfounded conditional on C. With the assumption (ii') we have that the effect of the two mediators M_1 and M_2 on the outcome Y is unconfounded conditional on the exposure A and C. Finally, we have with the assumption (iii') that the effect of the exposure A on the two mediators is unconfounded conditional on C [11]. Recall that C represents some baseline measurements.

3.2. A new definition of the interventional indirect effects of sequential mediation

We suggest a new definition of sequential mediation for the path-specific effects with multiple mediators such that the overall effect is equal to the total causal effect. We assume the three assumptions (i'), (ii') and (iii') to identify the effects. Our definition is inspired by the definition proposed by Vansteelandt and Daniel [11]. The causal estimand for the interventional direct effect of the exposure on the outcome is defined by

$$\int \left(E(Y^{am_1m_2} \mid c) - E(Y^{a^*m_1m_2} \mid c) \right) f_{M_2^{a^*} \mid M_1^{a^*}, C}(m_2 \mid m_2, c) f_{M_1^{a^*} \mid C}(m_1 \mid c) f_C(c) d(\overline{m}_2, c) \tag{7}$$

and the causal estimand at (7) is identified by

$$\int \left(E(Y \mid \overline{M}_2 = \overline{m}_2, A = a, C = c) - E(Y \mid \overline{M}_2 = \overline{m}_2, A = a^*, C = c) \right) f_{\overline{M}_2 \mid A, C}(\overline{m}_2 \mid a^*, c) f_C(c) d(\overline{m}_2, c) \tag{8}$$

where $f(\overline{M}_2 \mid A = a^*, C)$ denotes the product of the two densities $f(M_2 \mid M_1, A = a^*, C) f(M_1 \mid A = a^*, C)$. The causal estimand at (7) is the same one as the causal estimand at (3).

The next two causal estimands are the interventional indirect effects of the exposure mediated via the two ordered mediators and the two causal estimands are defined by

$$\int E(Y^{am_1m_2} \mid c) f_{M_2^{a^*} \mid M_1^{a^*}, C}(m_2 \mid m_1, c) \left\{ f_{M_1^a \mid C}(m_1 \mid c) - f_{M_1^{a^*} \mid C}(m_1 \mid c) \right\} f_C(c) d(\overline{m}_2, c) \tag{9}$$

and

$$\int E(Y^{am_1m_2} \mid c) \left\{ f_{M_2^a \mid M_1^a, C}(m_2 \mid m_1, c) - f_{M_2^{a^*} \mid M_1^{a^*}, C}(m_2 \mid m_1, c) \right\} f_{M_1^a \mid C}(m_1 \mid c) f_C(c) d(\overline{m}_2, c). \tag{10}$$

They are identified by

$$\int E(Y \mid \overline{M}_2 = \overline{m}_2, A = a, C = c) f_{M_2 \mid M_1, A, C}(m_2 \mid m_1, a^*, c) \left\{ f_{M_1 \mid A, C}(m_1 \mid a, c) - f_{M_1 \mid A, C}(m_1 \mid a^*, c) \right\} f_C(c) d(\overline{m}_2, c) \tag{11}$$

and

$$\int E(Y \mid \overline{M}_2 = \overline{m}_2, A = a, C = c) \left\{ f_{M_2 \mid M_1, A, C}(m_2 \mid m_1, a, c) - f_{M_2 \mid M_1, A, C}(m_2 \mid m_1, a^*, c) \right\} f_{M_1 \mid A, C}(m_1 \mid a, c) f_C(c) d(\overline{m}_2, c), \tag{12}$$

respectively. The causal estimand at (9) is the interventional indirect effect via M_1 and it corresponds to the sum of the two causal paths $A \to M_1 \to Y$ and $A \to M_1 \to M_2 \to Y$. The causal estimand at (10) is the interventional indirect effect via M_2 and it corresponds to the path $A \to M_2 \to Y$. We assume the three assumptions (i'), (ii') and (iii') to identify the three statistical estimands (8), (11) and (12) from the three causal estimands (7), (9) and (10), respectively. The sum of the three causal estimands (7), (9) and (10) is given by

$$\begin{aligned}
(7) + (9) + (10) &= \int E(Y^{am_1m_2} \mid c) f_{M_2^a \mid M_1^a, C}(m_2 \mid m_1, c) f_{M_1^a \mid C}(m_1 \mid c) f_C(c) d(\overline{m}_2, c) \\
&\quad - \int E(Y^{a^*m_1m_2} \mid c) f_{M_2^{a^*} \mid M_1^{a^*}, C}(m_2 \mid m_1, c) f_{M_1^{a^*} \mid C}(m_1 \mid c) f_C(c) d(\overline{m}_2, c) \\
&= \int E(Y^{am_1m_2} \mid c) f_{\overline{M}_2^a \mid C}(\overline{m}_2 \mid c) f_C(c) d(\overline{m}_2, c) - \int E(Y^{a^*m_1m_2} \mid c) f_{\overline{M}_2^{a^*} \mid C}(\overline{m}_2 \mid c) f_C(c) d(\overline{m}_2, c) \\
&= \int E(Y^{aM_1^aM_2^a} \mid c) f(c) d(c) - \int E(Y^{a^*M_1^{a^*}M_2^{a^*}} \mid c) f(c) d(c)
\end{aligned}$$

and it is equal to the total causal effect of the exposure A on the outcome Y.

3.3. A definition of sequential mediation by VanderWeele and Vansteelandt

We are using a definition for sequential mediation analysis from VanderWeele and Vansteelandt [12]. The natural direct effect is $E(Y^{aM_1^{a^*}} - Y^{a^*M_1^{a^*}} \mid C)$ and it is possible to rewrite it to

$$E\left(Y^{aM_1^{a^*}} - Y^{aM_1^{a^*}M_2^{a^*}} \mid C\right) + E\left(Y^{aM_1^{a^*}M_2^{a^*}} - Y^{a^*M_1^{a^*}M_2^{a^*}} \mid C\right).$$

The first term in the sum is the mediated effect via M_2 ($A \to M_2 \to Y$) and it is identified by the statistical estimand at (14). The statistical estimand at (14) is identified from the causal estimand under the assumptions (i'), (ii') and (iii'). The last term in the sum is the direct effect and it is obtained by the causal estimand at (7). The statistical estimand at (8) is identified from the causal estimand at (7) under the assumptions (i'), (ii') and (iii'). The mediated effect of A on Y via M_1 and the additional effect via M_2 ($A \to M_1 \to Y$ and $A \to M_1 \to M_2 \to Y$) is defined by $E(Y^{aM_1^a} - Y^{aM_1^{a^*}} \mid C)$ which is identified by the statistical estimand at (13). The statistical estimand at (13) is identified from the causal estimand under the assumptions (i'), (ii') and (iii'). The statistical estimand for the interventional indirect effect via M_1 is given by

$$\int E(Y \mid M_1 = m_1, A = a, C = c)\left\{f_{M_1|A,C}(m_1 \mid a, c) - f_{M_1|A,C}(m_1 \mid a^*, c)\right\} f_C(c) d(m_1, c) \tag{13}$$

and the statistical estimand for the interventional indirect effect via M_2 is given by

$$\int \left\{E(Y \mid M_1 = m_1, A = a, C = c) - E(Y \mid \overline{M}_2 = \overline{m}_2, A = a, C = c) f_{M_2|M_1,A,C}(m_2 \mid m_1, a^*, c)\right\} f_{M_1|A,C}(m_1 \mid a^*, c) f_C(c) d(\overline{m}_2, c). \tag{14}$$

We are comparing our definition to the definition by VanderWeele and Vansteelandt [12] bacause our causal estimands and the causal estimands by VanderWeele and Vansteelandt [12] use the same three assumptions (i'), (ii') and (iii') to identify the effects.

4. Simulations

We consider three simulation studies with different models (simulation $\mathfrak{A}$, $\mathfrak{B}$ and $\mathfrak{C}$). The sample size is 500 and the data are replicated 2000 times. Let C follow a standard normal distribution. Let A be drawn with the probability $\text{logit}(P(A = 1 \mid C = c)) = \kappa_I + \kappa_1 c$ where $\text{logit}(x) = \log(x) - \log(1 - x)$.

Simulation $\mathfrak{A}$: Let the following three models be given by

$$E\left(Y \mid M_2 = m_2, M_1 = m_1, A = a, C = c\right) = \xi_I + \xi_1 c + \xi_2 a + \xi_3 m_1 + \xi_4 m_2, \tag{15}$$

$$E\left(M_2 \mid M_1 = m_1, A = a, C = c\right) = \alpha_I + \alpha_1 c + \alpha_2 a + \alpha_3 m_1 \tag{16}$$

and

$$E\left(M_1 \mid A = a, C = c\right) = \zeta_I + \zeta_1 c + \zeta_2 a \tag{17}$$

to simulate the data. The data are simulated as follows: $C \sim \text{Normal}(0, 1^2)$, $A \sim \text{Bernoulli}(\varkappa_a)$, $M_1 \sim \text{Normal}(\eta_{m_1}, 1^2)$, $M_2 \sim \text{Normal}(\eta_{m_2}, 1^2)$ and $Y \sim \text{Normal}(\eta_y, 1^2)$ where $\text{logit}(\varkappa_a) := -0.5C$ and the means are given by $\eta_{m_1} := -4C - 2A$, $\eta_{m_2} := -C + 2A - 2M_1$ and $\eta_y := 3C + 3A - M_1 - 2M_2$.

Simulation $\mathfrak{B}$: We assume that one interaction between the exposure and the first mediator will simulate the second mediator and it is given by

$$E\left(M_2 \mid M_1 = m_1, A = a, C = c\right) = \alpha_I + \alpha_1 c + \alpha_2 a + \alpha_3 m_1 + \alpha_4 a m_1. \tag{18}$$

We also assume the two models defined at (15) and (17) to simulate the data. The data are simulated as follows: $C \sim \text{Normal}(0, 1^2)$, $A \sim \text{Bernoulli}(\varkappa_a)$, $M_1 \sim \text{Normal}(\eta_{m_1}, 1^2)$, $M_2 \sim \text{Normal}(\eta_{m_2}, 1^2)$ and $Y \sim \text{Normal}(\eta_y, 1^2)$ where $\text{logit}(\varkappa_a) := -0.5C$ and the means are given by $\eta_{m_1} := -4C - 2A$, $\eta_{m_2} := -C + 2A - 2M_1 + 4AM_1$ and $\eta_y := 3C + 3A - M_1 - 2M_2$.

Simulation $\mathfrak{C}$: We assume three interactions between the exposure and the two mediators to simulate the outcome and it is given by

$$E\left(Y \mid M_2 = m_2, M_1 = m_1, A = a, C = c\right) = \xi_I + \xi_1 c + \xi_2 a + \xi_3 m_1 + \xi_4 m_2 + \xi_5 a m_1 + \xi_6 a m_2 + \xi_7 m_1 m_2, \quad (19)$$

and we also assume the two models defined at (16) and (17) to simulate the data. The data are simulated as follows: $C \sim \text{Normal}(0, 1^2)$, $A \sim \text{Bernoulli}(\varkappa_a)$, $M_1 \sim \text{Normal}(\eta_{m_1}, 1^2)$, $M_2 \sim \text{Normal}(\eta_{m_2}, 1^2)$ and $Y \sim \text{Normal}(\eta_y, 1^2)$ where $\text{logit}(\varkappa_a) := -0.5C$ and the means are given by $\eta_{m_1} := -4C - 2A$, $\eta_{m_2} := -C + 2A - 2M_1$ and $\eta_y := 3C + 3A - M_1 - 2M_2 + 2AM_1 - AM_2 - 3M_1M_2$.

The marginal structural model (MSM) for simulation $\mathfrak{A}$, $\mathfrak{B}$ and $\mathfrak{C}$ is given by $E(Y^a) = \beta_I + \beta_1 a$. Let a be equal to 1 and let a^* be equal to 0. The simulation studies are evaluated by the mean of the 2000 estimates of β_1 from the MSM, the empirical standard error (SE) of the 2000 estimates of β_1 and the mean squared error (MSE). Table 1 shows the evaluation of the estimates of the causal effects from the three simulation studies. Table 2 shows the evaluation of the estimates of the interventional direct and indirect effects with both definitions.

Simulation	True	Mean	SE	MSE
$\mathfrak{A}$	-7	-6.999	0.354	0.125
$\mathfrak{B}$	9	9.003	1.512	2.285
$\mathfrak{C}$	19	19.009	4.745	22.515

Table 1. The Simulation column shows the three different simulation studies. The True column shows the true causal effect. The Mean column shows the mean of the 2000 estimates of β_1. The SE column shows the standard error of the 2000 estimates of β_1. The MSE column shows the mean squared error. The estimation of β_I is not shown for the simulation studies.

Table 1 shows the estimates of the causal effect of A on Y for simulation $\mathfrak{A}$, $\mathfrak{B}$ and $\mathfrak{C}$. We will use the estimation of the causal effects for comparison with the overall effects that we obtain using the two definitions (our definition and the definition by VanderWeele and Vansteelandt [12]).

	Simulation $\mathfrak{A}$			Simulation $\mathfrak{B}$			Simulation $\mathfrak{C}$		
Effect	Mean	SE	MSE	Mean	SE	MSE	Mean	SE	MSE
Dir. eff.	3.000	0.161	0.026	2.999	0.134	0.018	2.997	0.751	0.564
Via M_1, (11)	-6.001	0.351	0.123	-6.001	0.356	0.127	10.018	4.280	18.322
Via M_2, (12)	-3.998	0.274	0.075	12.006	1.570	2.464	5.994	1.224	1.499
The overall eff. w. Our Def.	-6.999	0.354	0.125	9.003	1.512	2.285	19.009	4.745	22.515
Via M_1, (13)	-6.001	0.351	0.123	1.960	1.509	65.634	24.846	13.893	413.426
Via M_2, (14)	-3.998	0.274	0.075	0.258	2.206	142.733	71.249	14.307	4462.107
The overall eff. w. V and V	-6.999	0.354	0.125	5.217	1.653	17.048	99.092	9.643	6507.761

Table 2. Simulation $\mathfrak{A}$: The true direct effect is 3, the true indirect effect via M_1 is -6 and the true indirect effect via M_2 is -4. Simulation $\mathfrak{B}$: The true direct effect is 3, the true indirect effect via M_1 is -6 and the true indirect effect via M_2 is 12. Simulation $\mathfrak{C}$: The true direct effect is 3, the true indirect effect via M_1 is 10 and the true indirect effect via M_2 is 6. See Table 1 for the true total causal effect of β_1 and the description of the different columns: Mean, SE and MSE. *Dir. eff.* is an abbreviation for the interventional direct effect using the statistical estimand at (8). *Via M_1 $(\cdot)$* is an abbreviation for the interventional indirect effect via M_1 using the statistical estimand at either (11) or (13). *Via M_2 $(\cdot)$* is an abbreviation for the interventional indirect effect via M_2 using the statistical estimand at either (12) or (14). *The overall eff. w. Our Def.* is an abbreviation for the overall effect with our definition. *The overall eff. w. V and V* is an abbreviation for the overall effect with the definition by VanderWeele and Vansteelandt [12].

Table 2 shows that our definition is able to estimate the interventional direct effect and the interventional indirect effects in all three simulation studies (simulation $\mathfrak{A}$, $\mathfrak{B}$ and $\mathfrak{C}$). Table 2 shows that the overall effect is equal to the total causal effect (Table 1) in all three simulation studies (simulation $\mathfrak{A}$, $\mathfrak{B}$ and $\mathfrak{C}$) with our definition. Table 2 shows that the definition by VanderWeele and Vansteelandt [12] shows weakness when the models have interactions between the measurements. Their definition is not able to estimate the interventional indirect effects for the last two simulation studies (simulation $\mathfrak{B}$ and $\mathfrak{C}$) and the overall effects are not equal to the total causal effects for the last two simulation studies.

5. PERFORM

The PERFORM study was collected over 2 years. For each patient a baseline measurement was taken and they were measured again after 2, 6, 12, 18 and 24 months. The patients were measured on 3 self-reported scales. Functional impairment was measured by the Sheehan Disability Scale (SDS) consisting of 3 items. Each item ranges from 0 to 10 with a global score ranging from 0 to 30. Cognitive symptoms were measured by the Perceived Deficit Questionnaire (PDQ-5) consisting of 5 items. Each item ranges from 0 to 4 with a global score ranging from 0 to 20 (we suppress '-5' in the name PDQ-5 for simplicity). Depression severity was measured by the Patient Health Questionnaire (PHQ-9) consisting of 9 items. Each item ranges from 0 to 3 with a global score ranging from 0 to 27 (we suppress '-9' in the name PHQ-9 for simplicity). A higher score corresponds to the patient being more constrained, suffering greater severity of their cognitive symptoms and more severe depression. We assume that depression severity affects both cognitive symptoms and functional impairment and that cognitive symptoms affect functional impairment. We further assume that all the measurements at baseline affect all the measurements at month 2 [15]. See the DAG in Figure 2. We dichotomize the variable PDQ. Let PDQ be equal to 0 if the original global score is less than or equal to 5 and 1 otherwise. If PDQ is equal to 0 then it corresponds to the patient having no or minimal cognitive symptoms, and if PDQ is equal to 1 then it corresponds to the patient having cognitive symptoms. See Hammer-Helmich et al. [9] for further information about the PERFORM study. We focused only on the measurements at the first two time points (baseline and month 2) as an example for our definition. We considered only the subset of data with fully observed vectors (all six variables, see Figure 2). The number of fully observed vectors in the data is 341.

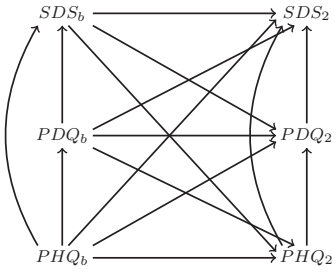


Figure 2. The nodes with the index b and 2 are the measurements at baseline and after 2 months respectively. The nodes PHQ_t, PDQ_t and SDS_t for $t \in \{b, 2\}$ represent the scales PHQ, PDQ and SDS respectively.

The MSM is given by

$$E\left(SDS_2^{(pdq_b, pdq_2)}\right) = \beta_I + \beta_1 pdq_b + \beta_2 pdq_2 + \beta_3 pdq_b pdq_2 \tag{20}$$

where $\boldsymbol{\beta} = (\beta_I, \beta_1, \beta_2, \beta_3)$ denotes the vector of causal effects. First, the aim is to estimate the (total) causal effect of the time-varying exposure on the outcome. The two measurements $\overline{M}_2 = (SDS_b, PHQ_2)$ are the two mediators for the exposure, PDQ_b (cognitive symptoms at baseline). We consider the g-formula [16] in the presence of time-dependent confounding. It is given by

$$E\left(SDS_2^{(pdq_b, pdq_2)}\right) = \int m\{v_2, \boldsymbol{\xi}\} f_{PHQ_2|SDS_b, PDQ_b, PHQ_b}(phq_2 \mid sds_b, pdq_b, phq_b) \\ f_{SDS_b|PDQ_b, PHQ_b}(sds_b \mid pdq_b, phq_b) f_{PDQ_b}(pdq_b) d(phq_2, sds_b, phq_b),$$

where

$$\begin{aligned} m\{V_2, \boldsymbol{\xi}\} &= E(SDS_2 \mid PDQ_2, PHQ_2, SDS_b, PDQ_b, PHQ_b) \\ &= \xi_I + \xi_1 PHQ_b + \xi_2 PDQ_b + \xi_3 SDS_b + \xi_4 PHQ_2 + \xi_5 PDQ_2 + \xi_6 PDQ_b PHQ_b + \\ &\quad \xi_7 PDQ_2 PHQ_b + \xi_8 PDQ_b PHQ_2 + \xi_9 PDQ_2 PHQ_2 + \xi_{10} PDQ_b PDQ_2 \end{aligned}$$

and V_2 denotes the set $(PDQ_2, PHQ_2, SDS_b, PDQ_b, PHQ_b)$. We include the interactions between depression severity and cognitive symptoms in the $m\{V_{st}, \boldsymbol{\xi}_t\}$-model with all the main effects. We assume that the conditional measures corresponding to the two densities $f_{PHQ_2|PHQ_b, PDQ_b, SDS_b}(phq_2 \mid phq_b, pdq_b, sds_b)$ and $f_{SDS_b|PHQ_b, PDQ_b}(sds_b \mid phq_b, pdq_b)$ are linear models without any interactions or quadratic terms. We use the three statistical estimands at (8), (11) and (12) with the model $m\{v_2, \boldsymbol{\xi}\}$. Let pdq_b be equal to 1 and let pdq_b^* be equal to 0 in the three statistical estimands of the interventional direct and indirect effects. Let pdq_2 be equal to 0 and let pdq_2^* be equal to 0 for the estimation of the

interventional direct and indirect effects of PDQ_b on SDS_2. See Kreif et al. [17] for further information about the series of iterated conditional expectations. Let $\beta_{1,dir}$ denote the interventional direct effect of PDQ_b on SDS_2. Let $\beta_{1,indir_{SDS_b}}$ and $\beta_{1,indir_{PHQ_2}}$ denote the two interventional indirect effects via SDS_b and PHQ_2 respectively. The confidence intervals are obtained using 1000 bootstraps.

5.1. Results

The results of the (total) causal effects $\boldsymbol{\beta} = (\beta_I, \beta_1, \beta_2, \beta_3)$, the interventional direct effect and the two interventional indirect effects are shown in Table 3.

Effect	Estimate	SE	95%-CI
β_I	9.588	1.243	(7.152 ; 12.025)
β_1	4.165	2.021	(0.204 ; 8.126)
β_2	3.566	2.026	(-0.404 ; 7.536)
β_3	-1.848	2.613	(-6.969 ; 3.272)
$\beta_{1,dir}$, Interventional direct of PDQ_b	1.385	1.711	(-1.968 ; 4.739)
$\beta_{1,indir_{SDS_b}}$, Interventional indirect of PDQ_b via SDS_b	2.490	0.705	(1.109 ; 3.870)
$\beta_{1,indir_{PHQ_2}}$, Interventional indirect of PDQ_b via PHQ_2	0.290	0.870	(-1.416 ; 1.995)
$\beta_{1,dir} + \beta_{1,indir_{SDS_b}} + \beta_{1,indir_{PHQ_2}}$, The overall effect	4.165	2.021	(0.204 ; 8.126)

Table 3. The Effect column shows the causal effects $\boldsymbol{\beta}$, the interventional direct effect of PDQ_b on SDS_2, the two interventional indirect effects of PDQ_b on SDS_2 and the overall effect. The SE column shows the standard errors. The standard error (SE) is obtained using 1000 bootstraps. The 95%-CI column shows the 95% confidence intervals.

Table 3 shows that β_1 is the only estimated coefficient of the (total) causal effects that is significant. The coefficient β_I corresponds to the expected score of functional impairment for patients having no or minimal cognitive symptoms at both visits. The coefficient β_I is $E(SDS_2^{(0,0)})$. The coefficient β_1 is the causal effect of cognitive symptoms at baseline on functional impairment at month 2. The coefficient β_1 corresponds to the additional effect we have to add to β_I for the expected score of functional impairment for patients having cognitive symptoms at baseline and having no or minimal cognitive symptoms at month 2. The sum of the two coefficients β_I and β_1 is $E(SDS_2^{(1,0)})$. The coefficient β_2 is the causal effect of cognitive symptoms at month 2 on functional impairment at month 2. The coefficient β_2 corresponds to the additional effect we have to add to β_I for the expected score of functional impairment for patients having no or minimal cognitive symptoms at baseline and having cognitive symptoms at month 2. The sum of the two coefficients β_I and β_2 is $E(SDS_2^{(0,1)})$. The coefficient β_3 is the causal effect of the interaction between cognitive symptoms at baseline and cognitive symptoms at month 2 on functional impairment at month 2. The sum of all four coefficients corresponds to the expected score of functional impairment for patients with cognitive symptoms at both visits. The sum of all four coefficients is $E(SDS_2^{(1,1)})$.

The estimates of the mediated effects are plausible from a clinical perspective since patients with cognitive symptoms at baseline will be more functionally impaired compared to patients with no or with minimal cognitive symptoms. The positive sign of the mediated effects do not conflict with the clinical expectation. The mediated effects indicate that patients with cognitive symptoms at baseline will not improve their functional impairment via one of the mediated effects, in contrast to patients with no or with minimal cognitive symptoms. More than half of the effect of the cognitive symptoms at baseline is transmitted via functional impairment at the same time point on functional impairment at a later time (the interventional indirect effect via SDS_b). It appears plausible that the interventional indirect effect via functional impairment at baseline has a certain proportion of the total causal effect since patients with cognitive symptoms will also be more functionally impaired at the same time point. It appears from the analysis that if the cognitive symptoms are relieved and the functioning improved at the same time point, then the patient functioning is more likely to improve at a later time.

6. Discussion

The motivation for this manuscript was to develop a definition so that the overall effect is equal to the total causal effect while at the same time avoiding the additional mediated dependence term. Our proposed definition has been worked through theoretically, and it has been applied on simulated data and real-world data on patients with MDD with the purpose to facilitate a better understanding of the role of cognition in reaching better functionality for the patients. We

have compared our definition to an already existing one with different simulation studies. We concluded that the simulation studies have revealed that our definition is better at obtaining the true effects of interest and the overall effect is equal to the total causal effect. Our proposed approach also encompasses models that have interactions between the different measurements. This is in contrast to the definition by VanderWeele and Vansteelandt [12], which shows weakness in the simulation studies for estimating the true effects. Their definition was not capable of including the interactions in the models between the different measurements that caused the overall effect not to be equal to the total causal effect.

Finally, the definition was applied to the observational cohort study PERFORM with patients having depression. The results from the analysis of the PERFORM study were in line with the expectations from a clinical perspective since the analysis indicates that patients with cognitive symptoms at baseline have worse functioning compared to patients with no or with minimal cognitive symptoms. A limitation for our definition is that we need the measurements to be ordered in a sequence. The causal ordering between the two time points in the PERFORM study is introduced by time itself. However, the causal ordering of the three measurements within the same time point is a limitation since all three measurements are measured at the same time point and the order is based on clinical insight. The interpretation of the results hinges on these assumptions. The assumptions are a limitation since we cannot verify them from the data. We have used the PERFORM study in this manuscript as an example and we have only used fully observed vectors from the data, therefore further research could focus on extending our definition to include vectors that are not fully observed.

Acknowledgements

The PERFORM study was funded by Lundbeck SAS. The authors thank Rhian M. Daniel (Division of Population Medicine, Cardiff University, UK) for her advice and discussions of the causal estimands.

Conflict of interest

Thomas Maltesen (temporary employee), Klaus Groes Larsen and Lene Hammer-Helmich are full-time employees of H. Lundbeck A/S. The authors report no conflict of interest.

References

[1] Imai K, Keele L, Yamamoto T. Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science* 2010;**25**(1):51–71.

[2] Kaufman JS, Maclehose RF, Kaufman S. A further critique of the analytic strategy of adjusting for covariates to identify biologic mediation. *Epidemiologic perspectives & innovations : EP+I* 2004;**1**(1):4.

[3] VanderWeele TJ. Marginal structural models for the estimation of direct and indirect effects. *Epidemiology* 2009; **20**(1):18–26.

[4] Robins JM, Greenland SM. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 1992; **3**(2):143–155.

[5] Pearl J. Direct and indirect effects. *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence* 2001;(10):411–420.

[6] Tchetgen Tchetgen EJ, JVanderWeele T. Identification of natural direct effects when a confounder of the mediator is directly affected by exposure. *Epidemiology* 2014;**25**(2):282–291.

[7] Robins JM, Richardson TS. Alternative graphical causal models and the identification of direct effects. *To appear in Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures P Shrout, Editor Oxford University Press* 2010;.

[8] VanderWeele TJ, Vansteelandt S, Robins JM. Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology* 2014;**25**(2):300–306.

[9] Hammer-Helmich L, Haro JM, Jönsson B, *et al.* Functional impairment in patients with major depressive disorder: the 2-year perform study. *Neuropsychiatric Disease and Treatment* 2018;**14**:239–249.

[10] Chokka P, Bougie J, Proulx J, Tvistholm AH, Ettrup A. Long-term functioning outcomes are predicted by cognitive symptoms in working patients with major depressive disorder treated with vortioxetine: results from the atworc study (book review). *CNS Spectrums* 2019;:1–12.

[11] Vansteelandt S, Daniel RM. Interventional effects for mediation analysis with multiple mediators. *Epidemiology (Cambridge, Mass)* 2017;**28**(2):258–265.

[12] VanderWeele T, Vansteelandt S. Mediation analysis with multiple mediators. *Epidemiologic Methods* 2013;**2**(1):95–115.

[13] Daniel RM, Stavola BLD, Cousens SN, Vansteelandt S. Causal mediation analysis with multiple mediators. *Biometrics* 2015;**71**(1):1–14.

[14] VanderWeele TJ, Robinson WR. On the causal interpretation of race in regressions adjusting for confounding and mediating variables. *Epidemiology* 2014;**25**(4):473–484.

[15] Haro JM, Hammer-Helmich L, Saragoussi D, Ettrup A, Larsen KG. Patient-reported depression severity and cognitive symptoms as determinants of functioning in patients with major depressive disorder: a secondary analysis of the 2-year prospective perform study. *Neuropsychiatric Disease and Treatment* 2019;**15**:2313–2323.

[16] Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect (book review). *Mathematical Modelling* 1986;**7**(9-12):1393–1512.

[17] Kreif N, Tran L, Grieve R, Stavola BD, Tasker RC, Petersen M. Estimating the comparative effectiveness of feeding interventions in the pediatric intensive care unit: A demonstration of longitudinal targeted maximum likelihood estimation. *American Journal of Epidemiology* 2017;**186**(12):1370–1379.

[18] VanderWeele TJ. Mediation and mechanism. *European Journal of Epidemiology* 2009;**24**(5):217–224.

[19] VanderWeele TJ. Policy-relevant proportions for direct effects. *Epidemiology* 2013;**24**(1):175–176.

[20] Coffman DL. Estimating causal effects in mediation analysis using propensity scores. *Structural Equation Modeling: A Multidisciplinary Journal* 2011;**18**(3):357–369.

[21] Pearl J. An introduction to causal inference. *The international journal of biostatistics* 2010;**6**(2).

[22] VanderWeele TJ. Controlled direct and mediated effects: Definition, identification and bounds. *Scandinavian Journal of Statistics* 2011;**38**(3):551–563.

[23] Avin C, Shpitser I, Pearl J. Identifiability of path-specific effects 2005;:357–363.

Manuscript III

Sequential mediation analysis with multiple mediators for data with missing observations

Research Article

Statistics
in Medicine

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

Sequential mediation analysis with multiple mediators for data with missing observations

Thomas Maltesen[ab*]

Causal mediation in both observational studies and interventional studies may be complicated by missing observations. Mediation analysis for multiple mediators with a mediator-outcome relationship will violate the cross-world assumption. This means that the identification of the natural direct effect and the natural indirect effect is not possible. I propose an augmented inverse probability weighted (AIPW) estimator to estimate both the interventional direct effect and the interventional indirect effects for multiple mediators with a continuous outcome, including partially observed vectors in the estimation. The estimator is robust regarding misspecification of the parametric model for the monotone missingness in the data, under the assumption that the missing observations are missing at random (MAR). The estimator is used on the observational study Prospective Epidemiological Research on Functioning Outcomes Related to Major depressive disorder (PERFORM), which is a longitudinal study with time-dependent confounding and missing observations. The causal paths between the exposure and the outcome contain multiple mediators and the causal paths also contain a mediator-outcome relationship. My estimator utilizes data better and it reduces bias when data contains missing observations compared to an estimator using only complete cases. Copyright © 0000 John Wiley & Sons, Ltd.

Keywords: causal inference, sequential mediation, multiple mediators, doubly robust estimator, monotone missingness, mediation with monotone missingness

1. Introduction

Causal mediation in longitudinal studies may be complicated by missing observations. The study Prospective Epidemiological Research on Functioning Outcomes Related to Major depressive disorder (PERFORM) (NCT01427439) is a good example as the study is longitudinal with a need to adjust for time-dependent confounding and a rich opportunity to study mediation. Missing observations across variables and drop-outs lead to a substantial reduction in observations when statistical analysis are based on complete cases. This may result in biased estimates. The study was conducted to better understand the course of a depressive episode and its impact on patient functioning over two years in outpatients with major depressive disorder (MDD). The treatment of cognitive symptoms may hold the key to achieving functional recovery in MDD [1]. See Hammer-Helmich et al. [2] for further information about the PERFORM study.

The natural direct and indirect effects [3, 4] are attractive to estimate since the sum of the effects is equal to the total causal effect. However, the natural direct and indirect effects are not possible to identify if there exists a mediator-outcome relationship [5]. Identification of the interventional direct and indirect effects avoids the cross-world assumption [6]. The interventional direct and indirect effects have the advantage of being meaningful even though the exposure variable is not manipulable [7, 8]. Vansteelandt and Daniel [8] have proposed a definition of causal estimands for the interventional direct effect and the interventional indirect effects for multiple mediators. The overall effect of the definition is equal to the total causal effect by introducing an additional mediated dependence term. The definition does not require the mediators to be ordered. VanderWeele and Vansteelandt [9] have a definition for sequential mediation, however the definition has issues by including models with interactions between the measurements. It may cause the overall effect not to be equal

[a] *Health Economics and Epidemiology Statistics, H.Lundbeck A/S, 2500 Valby, Denmark*
[b] *Section of Biostatistics, University of Copenhagen, 1014 Copenhagen K, Denmark*
[*] *Correspondence to: Thomas Maltesen, H.Lundbeck A/S, 2500 Valby, Denmark. E-mail: thomas.maltesen@protonmail.com*

to the total causal effect (Maltesen et al. [10]). Maltesen et al. [10] introduced a definition of sequential mediation for the interventional direct effect and the interventional indirect effects for multiple mediators. This definition avoids the additional mediated dependence term and the overall effect is equal to the total causal effect. This definition does however require the mediators to be ordered. The three definitions apply only to fully observed vectors. Li and Zhou [11] consider mediation analysis for data containing missing observations with one mediator and with the outcome possibly being missing. However, the PERFORM study has multiple mediators on the paths between the exposure and the outcome and the outcome variable is not the only variable missing. The effect of cognitive symptoms is also transmitted via a mediator-outcome relationship and violates the cross-world assumption. Maltesen et al. [12] consider an estimator to obtain the causal effect in longitudinal studies with time-dependent confounding while data contains missing observations. However, the estimator will not provide us with the direct effect if the causal path between the exposure and the outcome contains mediators.

The PERFORM study is the motivation to develop an estimator for sequential mediation when data contains missing observations since the before mentioned methods cannot be used in this analysis of the PERFORM study. The causal paths between the exposure and the outcome contain two mediators resulting in one direct causal path and three indirect causal paths. The estimation using the g-formula results only in the causal effect. The main interest is the direct effect of cognitive symptoms on functional impairment at a later time. However, the three indirect causal paths are also of interest because it may happen that the causal effect is almost zero. The indirect effects will provide us with the information if some of the effects cancel each other out. The manuscript is organized as follows: Section 2 revisits the causal estimands for multiple mediators from Maltesen et al. [10]. Section 3 considers data containing missing observations [13] and establishes an estimator for multiple mediators for data containing missing observations. Section 4 considers the PERFORM study. Section 5 considers a simulation study based on the PERFORM study. Section 6 finalizes with a discussion of the findings.

2. The estimator for the mediated effects

Suppose that our data comprises of n independent and identically distributed realizations of random variables $Z_1, \ldots, Z_n$ where Z_i denotes the i-th vector (the i index is suppressed to simplify the notation) [13]. I assume two mediators on the causal paths between the exposure and the outcome for simplicity. Let C denote some baseline measurements not affected by the exposure. Let A denote the binary exposure and let Y denote the outcome, which assumes to be continuous. Let M_1 and M_2 denote the two mediators. The mediator M_2 may be affected by the mediator M_1 but not the other way around. Let $\overline{M}_2$ denote the vector (M_1, M_2) and let $\overline{M}_2^a = (M_1^a, M_2^a)$. Let Y^a and $\overline{M}_2^a$ be the values that Y and $\overline{M}_2$ would be if the exposure A is set to a respectively. Let $Y^{a\overline{m}_2}$ be the value that Y would be if the exposure A is set to a and the vector of the two mediators $\overline{M}_2$ is set to $\overline{m}_2$. Let Z be defined by the ordered sequence (C, A, M_1, M_2, Y). Let V_0 denote the set (C, A) and let V_1 denote the set (C, A, M_1). Let V_2 denote the set (C, A, M_1, M_2). The outcome Y may be causally influenced by the whole history of (C, A, M_1, M_2). Maltesen et al. [10] define the causal estimand for the interventional direct effect with two mediators to be given by

$$\int \left\{ E(Y^{am_1m_2} \mid c) - E(Y^{a^*m_1m_2} \mid c) \right\} f_{M_2^{a^*}|M_1^{a^*},C}(m_2 \mid m_1, c) f_{M_1^{a^*}|C}(m_1 \mid c) f_C(c) d(\overline{m}_2, c).$$

The causal estimand for the interventional indirect effect via M_1 is given by

$$\int E(Y^{am_1m_2} \mid c) f_{M_2^{a^*}|M_1^{a^*},C}(m_2 \mid m_1, c) \left\{ f_{M_1^a|C}(m_1 \mid c) - f_{M_1^{a^*}|C}(m_1 \mid c) \right\} f_C(c) d(\overline{m}_2, c)$$

and the causal estimand for the interventional indirect effect via M_2 is given by

$$\int E(Y^{am_1m_2} \mid c) \left\{ f_{M_2^a|M_1^a,C}(m_2 \mid m_1, c) - f_{M_2^{a^*}|M_1^{a^*},C}(m_2 \mid m_1, c) \right\} f_{M_1^a|C}(m_1 \mid c) f_C(c) d(\overline{m}_2, c).$$

The assumptions needed to identify the three causal estimands above (for two mediators) are given by:

(i) $Y^{a\overline{m}_2} \perp\!\!\!\perp A \mid C = c \quad \forall (a, \overline{m}_2, c) \in A, \overline{M}_2, C$,
(ii) $Y^{a\overline{m}_2} \perp\!\!\!\perp (M_1, M_2) \mid A = a, C = c \quad \forall (a, \overline{m}_2, c) \in A, \overline{M}_2, C$ and
(iii) $(M_2^a, M_1^a) \perp\!\!\!\perp A \mid C = c \quad \forall (a, c) \in A, C$.

It is possible to rewrite the interventional direct effect to be given by $\Gamma(a, a^*, a^*) - \Gamma(a^*, a^*, a^*)$ where Γ is given by

$$\Gamma(j, k, l) = \int E(Y^{jm_1m_2} \mid c) f_{M_2^k|M_1^k,C}(m_2 \mid m_1, c) f_{M_1^l|C}(m_1 \mid c) f_C(c) d(\overline{m}_2, c) \tag{1}$$

for different $j, k, l \in \{a, a^*\}$. The two causal estimands for the interventional indirect effects via M_1 and M_2 can be written in a similar way. The interventional indirect effect via M_1 is given by $\Gamma(a, a^*, a) - \Gamma(a, a^*, a^*)$ and the interventional indirect effect via M_2 is given by $\Gamma(a, a, a) - \Gamma(a, a^*, a)$. The causal estimand $\Gamma(j, k, l)$ (1) is identified by the statistical estimand given by

$$\int E(Y \mid \overline{M}_2 = \overline{m}_2, A = j, C = c) f_{M_2|M_1,A,C}(m_2 \mid m_1, k, c) f_{M_1|A,C}(m_1 \mid l, c) f_C(c) d(\overline{M}_2, C)$$

and the statistical estimand is identified by the estimator given by

$$\hat{\Gamma}(j, k, l) = \frac{1}{n} \sum_{i=1}^{n} \mu_{j,k,l}\{V_{0,i}, \hat{\boldsymbol{\gamma}}\} \tag{2}$$

with the models $m_j\{v_2, \boldsymbol{\xi}\} = E(Y \mid M_2 = m_2, M_1 = m_1, A = j, C = c)$, $\mu_{j,k}\{v_1, \boldsymbol{\gamma}\} = E(m_j\{V_2, \boldsymbol{\xi}\} \mid M_1 = m_1, A = k, C = c)$ and $\mu_{j,k,l}\{v_0, \boldsymbol{\gamma}\} = E(\mu_{j,k}\{V_1, \boldsymbol{\gamma}\} \mid A = l, C = c)$. I will refer to the $m_j\{v_2, \boldsymbol{\xi}\}$-model, the $\mu_{j,k}\{v_1, \boldsymbol{\gamma}\}$-model and the $\mu_{j,k,l}\{v_0, \boldsymbol{\gamma}\}$-model as the $\mu_{j,k,l}$-models. The $\mu_{j,k,l}$-models have hats to indicate predicted values from the specified models that have been used for the estimation and the predicted values are plugged into the estimator. The estimator is unbiased if the $\mu_{j,k,l}$-models are correctly specified with respect to the data. The estimator (2) is obtained by solving the estimating equation $0 = \sum_{i=1}^{n} U_{j,k,l}(Z_i)$ with

$$U_{j,k,l}(Z_i) = \mu_{j,k,l}\{V_{0,i}, \boldsymbol{\gamma}_0\} - \int E(Y^{jm_1m_2} \mid c) f_{M_2^k|M_1^k,C}(m_2 \mid m_1, c) f_{M_1^l|C}(m_1 \mid c) f_C(c) d(\overline{m}_2, c). \tag{3}$$

The estimator for the interventional direct effect (dir) is given by $\widehat{dir} := \hat{\Gamma}(a, a^*, a^*) - \hat{\Gamma}(a^*, a^*, a^*)$ using the estimator (2). The estimator for the interventional indirect effect via M_1 ($indir_{M_1}$) is given by $\widehat{indir}_{M_1} := \hat{\Gamma}(a, a^*, a) - \hat{\Gamma}(a, a^*, a^*)$ using the estimator (2) and the estimator for the interventional indirect effect via M_2 ($indir_{M_2}$) is given by $\widehat{indir}_{M_2} := \hat{\Gamma}(a, a, a) - \hat{\Gamma}(a, a^*, a)$ using the estimator (2).

3. Vectors with missing observations in the data

Let $\mathcal{C}$ be a random variable that takes positive integers or infinity $\mathcal{C} \in \{1, \ldots, \mathbf{c}\} \cup \{\infty\}$. Let $\{G_{\mathcal{C}_i}(Z_i), \mathcal{C}_i\}$ denote the i-th vector in the observed data. If $\mathcal{C}$ is equal to 1 then it corresponds to only observe C in Z ($G_1(Z) = (C)$). If $\mathcal{C}$ is equal to 2 then C and A are the only two observed variables in Z ($G_2(Z) = (C, A)$). If $\mathcal{C}$ is equal to $\mathbf{c}$ ($= 4$) then it is only the outcome that is missing from Z. If $\mathcal{C}$ is equal to infinity then the vector is complete ($G_\infty(Z) = (Z)$). This pattern of missing observations is called monotone missingness. Complete cases (CC) are a subset of vectors containing only $G_\infty(Z)$ [12]. Note the distinction between the two letters $\mathbf{c}$ and c to avoid any confusion. I assume the probability for observing a complete vector is strictly greater than zero ($P(\mathcal{C} = \infty \mid Z) > 0$). Let $\varpi\{\infty, Z, \boldsymbol{\psi}_0\}$ denote the probability for observing a complete vector with the vector of true parameter values $\boldsymbol{\psi}_0$ [13]. Let

$$\lambda_r\{G_r(Z), \boldsymbol{\psi}\} = P(\mathcal{C} = r \mid \mathcal{C} \geq r, Z)$$

denote the probability of stopping the observing of additional observations given r observed. I assume that $\lambda_r\{G_r(Z), \boldsymbol{\psi}\}$ is given by

$$\lambda_r\{G_r(Z), \boldsymbol{\psi}\} = \frac{\exp\left(\psi_{I,r} + G_r(Z)\boldsymbol{\psi}_r\right)}{1 + \exp\left(\psi_{I,r} + G_r(Z)\boldsymbol{\psi}_r\right)}, \tag{4}$$

where the column vector $\boldsymbol{\psi}_r$ has the same dimension as the row vector $G_r(Z)$. Let $\boldsymbol{\psi}$ denote the vector $(\psi_{I,r}, \boldsymbol{\psi}_r')$ where the coefficient $\psi_{I,r}$ denotes the intercept [12, 13]. I assume that the missingness in the data are coarsened at random (CAR) which means that the coarsening probabilities only depend on the data as a function of the observed data. The coarsening probabilities are given by

$$\varpi\{r, G_r(Z), \boldsymbol{\psi}\} = \lambda_r\{G_r(Z), \boldsymbol{\psi}\} K_{r-1}\{G_{r-1}(Z), \boldsymbol{\psi}\}$$

where Tsiatis [13] defines

$$K_r\{G_r(Z), \boldsymbol{\psi}\} = \prod_{j=1}^{r} (1 - \lambda_j\{G_j(Z), \boldsymbol{\psi}\}).$$

I refer to the $\lambda_r\{G_r(Z), \boldsymbol{\psi}\}$-models as the λ-models and $K_{\mathbf{c}}\{G_{\mathbf{c}}(Z), \boldsymbol{\psi}\}$ is equal to the probability $\varpi\{\infty, Z, \boldsymbol{\psi}\}$. See Tsiatis [13] for further information about the CAR assumption and monotone missingness.

Tsiatis [13] shows that the adaptive doubly robust estimator is obtained by solving the estimating equation given by

$$0 = \sum_{i=1}^{n} \left(\frac{I(\mathcal{C}_i = \infty) U_{j,k,l}(Z_i)}{\varpi\{\infty, Z_i, \hat{\boldsymbol{\psi}}\}} + \sum_{r=1}^{\mathbf{c}} \frac{I(\mathcal{C}_i = r) - \lambda_r\{G_r(Z_i), \hat{\boldsymbol{\psi}}\} I(\mathcal{C}_i \geq r)}{K_r\{G_r(Z_i), \hat{\boldsymbol{\psi}}\}} E\left(U_{j,k,l}(Z) \mid G_r(Z_i), \hat{\boldsymbol{\zeta}} \right) \right)$$

with the probabilities for monotone missingness under the CAR assumption. The conditional expectation of $U_{j,k,l}(Z)$ (3) is given by

$$E\left(U_{j,k,l}(Z) \mid G_r(Z)\right) = E\left(\mu_{j,k,l}\{V_0, \boldsymbol{\gamma}_0\} \mid G_r(Z), \boldsymbol{\zeta}_0\right) - \Gamma(j,k,l).$$

The vector $\boldsymbol{\zeta}_0$ indicates the true model with the vector of true parameter values. I need to model the conditional expectation for every set of $G_r(Z)$ for $r = 1, \ldots, \mathbf{c}$. It is exemplified in Section 4. The estimator $\hat{\Gamma}(j,k,l)$ for $\Gamma(j,k,l)$ is given by

$$\begin{aligned} \hat{\Gamma}(j,k,l) = \frac{1}{n} \sum_{i=1}^{n} \Bigg[& \frac{I(\mathcal{C}_i = \infty) \mu_{j,k,l}\{V_0, \hat{\boldsymbol{\gamma}}\}}{\varpi\{\infty, Z_i, \hat{\boldsymbol{\psi}}\}} \\ & + \sum_{r=1}^{\mathbf{c}} \frac{I(\mathcal{C}_i = r) - \lambda_r\{G_r(Z_i), \hat{\boldsymbol{\psi}}\} I(\mathcal{C}_i \geq r)}{K_r\{G_r(Z), \hat{\boldsymbol{\psi}}\}} E\left(\mu_{j,k,l}\{V_0, \boldsymbol{\gamma}\} \mid G_r(Z_i), \hat{\boldsymbol{\zeta}} \right) \Bigg] \end{aligned} \tag{5}$$

for data with monotone missingness following the CAR assumption. Now, assume that the $\mu_{j,k,l}$-models are correctly specified. The estimator (5) is unbiased if the models of the conditional expectations $E(\mu_{j,k,l}\{V_0, \boldsymbol{\gamma}_0\} \mid G_r(Z), \hat{\boldsymbol{\zeta}})$ are correctly specified with respect to the distribution of Z and the λ-models relating to the missingness mechanism may be misspecified. The estimator (5) is also unbiased if the λ-models relating to the missingness mechanism are correctly specified and the models for the conditional expectations $E(\mu_{j,k,l}\{V_0, \boldsymbol{\gamma}_0\} \mid G_r(Z), \hat{\boldsymbol{\zeta}})$ may be misspecified with respect to the distribution of Z. I denote the estimator (5) by the name Doubly Robust estimator for Monotone missingness for Sequential Mediation (DRMSM). All the hats in the estimator (5) indicate predicted values that are plugged into the estimator. All the conditional expectations are evaluated and afterwards used to predict values with respect to the set $G_r(Z)$. We must recall that the probabilities regarding the models for the missingness are given by $K_r\{G_r(Z), \hat{\boldsymbol{\psi}}\} = \prod_{j=1}^{r}(1 - \lambda_j\{G_j(Z), \hat{\boldsymbol{\psi}}\})$ and $\varpi\{\infty, Z, \hat{\boldsymbol{\psi}}\} = K_{\mathbf{c}}\{G_{\mathbf{c}}(Z), \hat{\boldsymbol{\psi}}\}$. The estimates $\hat{\boldsymbol{\psi}}$ are obtained using maximum likelihood estimation according to the specific model for $\lambda_r\{G_r(Z), \boldsymbol{\psi}\}$. The interventional direct effect is estimated by $\widehat{dir} := \hat{\Gamma}(a, a^*, a^*) - \hat{\Gamma}(a^*, a^*, a^*)$ using the estimator (5). The interventional indirect effect via M_1 is estimated by $\widehat{indir}_{M_1} := \hat{\Gamma}(a, a^*, a) - \hat{\Gamma}(a, a^*, a^*)$ using the estimator (5) and the interventional indirect effect via M_2 is estimated by $\widehat{indir}_{M_2} := \hat{\Gamma}(a, a, a) - \hat{\Gamma}(a, a^*, a)$ using the estimator (5).

4. Analysing the PERFORM study

4.1. Study design and variables

The DRMSM estimator for mediation analysis is applied on the PERFORM study. Patients' functional impairment were measured by the Sheehan Disability Scale (SDS) consisting of 3 items with a global score ranging from 0 to 30. A score at 0 corresponds to being unimpaired and 30 corresponds to being impaired. The Scale describes the patients' work/school, social life/leisure activities and family life/home duties. Cognitive symptoms were measured by the Perceived Deficit Questionnaire (PDQ-5) consisting of 5 items with a global score ranging from 0 to 20 focusing on memory, concentration and executive function (the '-5' in the name PDQ-5 is suppressed to simplify the notation). The PDQ scale is dichotomized meaning that PDQ is 0 if the original global score of PDQ is less than or equal to 5 and 1 otherwise. If PDQ is equal to 0 then it corresponds to having no or minimal cognitive symptoms and 1 corresponds to having cognitive symptoms. The depression severity was measured by the Patient Health Questionnaire (PHQ-9) consisting of 9 items with a global score ranging from 0 to 27. The greater the score on the scale the more severe the depression (the '-9' in the name PHQ-9 is suppressed to simplify the notation). The sample size of the data is 1090. All three scales were measured over two years repeatedly. I assume that depression severity affects both cognitive symptoms and functional impairment and that cognitive symptoms affect functional impairment. I assume that the present measurements affect all the future measurements at the next time point. I also assume that the present measurements do not affect the past measurements. The process is indicated by a directed acyclic graph in Maltesen et al. [12]. Let SDS_t denote SDS at time $t \in \{b, 2, 6, 12, 18, 24\}$. Let PDQ_t denote PDQ at time $t \in \{b, 2, 6, 12, 18, 24\}$. Let PHQ_t denote PHQ at time $t \in \{b, 2, 6, 12, 18, 24\}$. Let W_t denote the vector $W_t = (PHQ_t, PDQ_t, SDS_t)$ for $t \in \{b, 2, 6, 12, 18, 24\}$. Let pt denote the prior time point before t, let t denote the present time point and let st denote the subsequent time point after t in the subscript of PHQ, PDQ and SDS. See

Maltesen et al. [12] for further information. I assume that $W_{pt} \perp\!\!\!\perp W_{st} \mid W_t$ and the sequential conditional exchangeability for $t \in \{b, 2, 6, 12, 18\}$ is given by

$$SDS_{st}^{(pdq_t, pdq_{st})} \perp\!\!\!\perp PDQ_t \mid PHQ_t, SDS_{pt}, PDQ_{pt}, PHQ_{pt}$$

and

$$SDS_{st}^{(pdq_t, pdq_{st})} \perp\!\!\!\perp PDQ_{st} \mid PHQ_{st}, SDS_t, PDQ_t, PHQ_t.$$

Let $Z_{b,i} = (W_{b,i}, W_{2,i})$ and let $Z_{18,i} = (W_{12,i}, W_{18,i}, W_{24,i})$ for $i = 1, \ldots, 1090$. Maltesen et al. [12] define the set $(PHQ_{pt}, PDQ_{pt}, SDS_{pt}, PHQ_t)$ to be the confounder L_t at time $t \in \{b, 18\}$ and the set (SDS_t, PHQ_{st}) to be the confounder L_{st} at time st. We must recall that the set $(PHQ_{pt}, PDQ_{pt}, SDS_{pt},)$ is empty for $t = b$. Let PDQ_t denote the exposure, cognitive symptoms at time $t \in \{b, 2, 18, 24\}$. The set V_t is given by the confounder and the exposure at time t, $V_t = (L_t, PDQ_t)$ and the set V_{st} is defined by the confounders and the exposures up to time st, $V_{st} = (\overline{L}_{st}, \overline{PDQ}_{st})$ where $\overline{L}_{st} = (L_t, L_{st})$ and $\overline{PDQ}_{st} = (PDQ_t, PDQ_{st})$.

4.2. *Statistical methods*

The marginal structural model (MSM) is given by

$$E\left(SDS_{st}^{(pdq_t, pdq_{st})}\right) = \beta_{I,t} + \beta_{1,t} pdq_t + \beta_{2,t} pdq_{st} + \beta_{3,t} pdq_t pdq_{st}$$

with the vector $\boldsymbol{\beta}_t = (\beta_{I,t}, \beta_{1,t}, \beta_{2,t}, \beta_{3,t})$ for $t \in \{b, 18\}$. Let $\boldsymbol{\beta}_{0,t}$ denote the true vector of causal parameter values. Maltesen et al. [12] define the $U(Z_{t,i})$-function to be given by $U(Z_{t,i}) = \mu_{t_1}\{V_t, \boldsymbol{\gamma}_{t_1}\} - E(SDS_{st}^{(pdq_t, pdq_{st})})$ for the analysis with respect to either $Z_{b,i}$ or $Z_{18,i}$ for $i = 1, \ldots, n_t$ and $t \in \{b, 18\}$. See the estimator (10) below. Maltesen et al. [12] define the $m\{v_{st}, \boldsymbol{\xi}_t\}$-model to be given by

$$\begin{aligned} m\{V_{st}, \boldsymbol{\xi}_t\} =& E(SDS_{st} \mid PDQ_{st}, PHQ_{st}, SDS_t, PDQ_t, PHQ_t) \\ =& \xi_{I,t} + \xi_{1,t} PHQ_t + \xi_{2,t} PDQ_t + \xi_{3,t} SDS_t + \xi_{4,t} PHQ_{st} + \xi_{5,t} PDQ_{st} + \xi_{6,t} PDQ_t PHQ_t + \\ & \xi_{7,t} PDQ_{st} PHQ_t + \xi_{8,t} PDQ_t PHQ_{st} + \xi_{9,t} PDQ_{st} PHQ_{st} + \xi_{10,t} PDQ_t PDQ_{st} \end{aligned} \tag{6}$$

and define the two μ-models to be given by

$$\begin{aligned} \mu_{t_2}\{V_t, \boldsymbol{\gamma}_{t_2}\} =& E(m\{V_{st}, \boldsymbol{\xi}_t\} \mid SDS_t, PDQ_t, PHQ_t) \\ =& \gamma_{I,t_2} + \gamma_{1,t_2} PHQ_t + \gamma_{2,t_2} PDQ_t + \gamma_{3,t_2} SDS_t \end{aligned} \tag{7}$$

and

$$\begin{aligned} \mu_{t_1}\{V_t, \boldsymbol{\gamma}_{t_1}\} =& E(\mu_{t_2}(V_t, \boldsymbol{\gamma}_{t_2}) \mid PDQ_t, PHQ_t, SDS_{pt}, PDQ_{pt}, PHQ_{pt}) \\ =& \gamma_{I,t_1} + \gamma_{1,t_1} PHQ_{pt} + \gamma_{2,t_1} PDQ_{pt} + \gamma_{3,t_1} SDS_{pt} + \gamma_{4,t_1} PHQ_t + \gamma_{5,t_1} PDQ_t \end{aligned} \tag{8}$$

since the confounder L_{st} consists of two measurements. All the observed patients having depression severity at baseline (PHQ_b) are used for the analysis with $t = b$ and all the observed patients having depression severity at month 12 (PHQ_{12}) are used for the analysis with $t = 18$. For example, if the missingness of the patient follows a nonmonotone pattern (see Tsiatis [13] for further information about nonmonotone pattern) then the missingness is forced to follow a monotone pattern. Maltesen et al. [12] define all the λ-models to include only the main effects without any interactions or quadratic terms. The hazard function $\lambda_r(G_r(Z_b))$ needs to be modelled five times for $r = 1, \ldots, 5$ and the hazard function $\lambda_r(G_r(Z_{18}))$ needs to be modelled eight times for $r = 1, \ldots, 8$. The probability $\lambda_4(G_4(Z_b))$ is set to 0 and all the vectors of $G_4(Z_b)$ are removed from the data because there are too few patients for the estimation. The two probabilities $\lambda_4(G_4(Z_{18}))$ and $\lambda_7(G_7(Z_{18}))$ are also set to 0 and all the vectors of $G_4(Z_{18})$ and $G_7(Z_{18})$ are removed from the data because there are too few patients for the estimation. See Maltesen et al. [12] for further information about how the pattern of missingness is forced to be monotone and the number of patients with the different vectors of $\{G_{\mathcal{C}}(Z_b), \mathcal{C}\}$ or $\{G_{\mathcal{C}}(Z_{18}), \mathcal{C}\}$ in the observed data. Let n_b denote the sample size for $t = b$ and that n_b is equal to 929. Let n_{18} denote the sample size for $t = 18$ and that n_{18} is equal to 696.

Let $\overline{M}_{2,t} = (SDS_t, PHQ_{st})$ denote the two mediators for $t \in \{b, 18\}$. Let pdq_t be equal to 1 and let pdq_t^* be equal to 0 for obtaining the interventional direct and indirect effects of PDQ_t on SDS_{st}. Let pdq_{st} be equal to 0 and let pdq_{st}^* also be equal to 0 for obtaining the interventional direct and indirect effects of PDQ_t on SDS_{st}. Let $m_j\{v_{st}, \boldsymbol{\xi}_t\} = E(SDS_{st} \mid PDQ_{st} = pdq_{st}, PHQ_{st} = phq_{st}, SDS_t = sds_t, PDQ_t = j, PHQ_t = phq_t)$ be defined by the $m\{v_{st}, \boldsymbol{\xi}_t\}$-model at (6). Let $\mu_{j,k_{t_2}}\{v_t, \boldsymbol{\gamma}_{t_2}\} = E(m_j\{V_{st}, \boldsymbol{\xi}_t\} \mid SDS_t = sds_t, PDQ_t = k, PHQ_t = phq_t)$ be defined by the model at (7) and let $\mu_{j,k,l_{t_1}}\{v_t, \boldsymbol{\gamma}_{t_1}\} = E(\mu_{j,k_{t_2}}\{V_t, \boldsymbol{\gamma}_{t_2}\} \mid PDQ_t = l, PHQ_t = phq_t, SDS_{pt} =$

$sds_{pt}, PDQ_{pt} = pdq_{pt}, PHQ_{pt} = phq_{pt}$) be defined by the model at (8). The estimator for analysing the data of the PERFORM study is given by

$$\hat{\Gamma}(j,k,l) = \frac{1}{n_t}\sum_{i=1}^{n_t}\left[\frac{I(\mathcal{C}_i=\infty)\mu_{j,k,l_{t_1}}\{V_t,\hat{\boldsymbol{\gamma}}_{t_1}\}}{\varpi\{\infty,Z_{t,i},\hat{\boldsymbol{\psi}}\}} + \sum_{r=1}^{\mathbf{c}}\frac{I(\mathcal{C}_i=r)-\lambda_r\{G_r(Z_{t,i},\hat{\boldsymbol{\psi}})\}I(\mathcal{C}_i\geq r)}{K_r\{G_r(Z_{t,i}),\hat{\boldsymbol{\psi}}\}}E\left(\mu_{j,k,l_{t_1}}\{V_t,\boldsymbol{\gamma}_{t_1}\}\mid G_r(Z_{t,i}),\hat{\boldsymbol{\zeta}}_t\right)\right] \tag{9}$$

and it is used to obtain the interventional direct effect of PDQ_t on SDS_{st} for $t\in\{b,18\}$ with the estimator $\widehat{dir}_t := \hat{\Gamma}(pdq_t,pdq_t^*,pdq_t^*) - \hat{\Gamma}(pdq_t^*,pdq_t^*,pdq_t^*)$. The interventional indirect effect of PDQ_t on SDS_{st} via SDS_t for $t\in\{b,18\}$ is obtained with the estimator $\widehat{indir}_{SDS_t} := \hat{\Gamma}(pdq_t,pdq_t^*,pdq_t) - \hat{\Gamma}(pdq_t,pdq_t^*,pdq_t^*)$ and the interventional indirect effect of PDQ_t on SDS_{st} via PHQ_{st} for $t\in\{b,18\}$ is obtained with the estimator $\widehat{indir}_{PHQ_{st}} := \hat{\Gamma}(pdq_t,pdq_t,pdq_t) - \hat{\Gamma}(pdq_t,pdq_t^*,pdq_t)$. The $\mathbf{c}$ is either 5 or 8 depending on the value of $t\in\{b,18\}$. The conditional expectations at (9) are modelled and afterwards used to predict values according to the different sets of $G_r(Z_{t,i})$ for $r=1,\ldots,\mathbf{c}$ and $i=1,\ldots,n_t$ for $t\in\{b,18\}$. The conditional expectations $E(\mu_{j,k,l_{b_1}}\{V_b,\boldsymbol{\gamma}_{b_1}\}\mid G_r(Z_b),\hat{\boldsymbol{\zeta}}_b)$ for $t=b$ are modelled according to

$$\begin{cases} E(\mu_{j,k,l_{b_1}}\{V_b,\boldsymbol{\gamma}_{b_1}\}\mid G_r(Z_b),\hat{\boldsymbol{\zeta}}_b) & \text{for } r=1 \\ \mu_{j,k,l_{b_1}}\{V_b,\boldsymbol{\gamma}_{b_1}\} & \text{for } r\in\{2,3,4,5\}\end{cases}$$

and the conditional expectation $E(\mu_{j,k,l_{b_1}}\{V_b,\boldsymbol{\gamma}_{b_1}\}\mid G_1(Z_b),\hat{\boldsymbol{\zeta}}_b)$ is only modelled with the main effect without any quadratic terms. The model is afterwards used to predict values. The conditional expectations $E(\mu_{j,k,l_{18_1}}\{V_{18},\boldsymbol{\gamma}_{18_1}\}\mid G_r(Z_{18}),\hat{\boldsymbol{\zeta}}_{18})$ for $t=18$ are modelled according to

$$\begin{cases} E(\mu_{j,k,l_{18_1}}\{V_{18},\boldsymbol{\gamma}_{18_1}\}\mid G_r(Z_{18}),\hat{\boldsymbol{\zeta}}_{18}) & \text{for } r\in\{1,2,3,4\} \\ \mu_{j,k,l_{18_1}}\{V_{18},\boldsymbol{\gamma}_{18_1}\} & \text{for } r\in\{5,6,7,8\}\end{cases}$$

and the conditional expectations $E(\mu_{j,k,l_{18_1}}\{V_{18},\boldsymbol{\gamma}_{18_1}\}\mid G_r(Z_{18}),\hat{\boldsymbol{\zeta}}_{18})$ for $r\in\{1,2,3,4\}$ are only modelled with the main effects without any interactions or quadratic terms. All the models are afterwards used to predict values. The DRMSM estimator is compared to the estimator for the simpler sequential mediation formula using complete cases. The estimator for the simpler sequential mediation formula for the interventional direct effect with two mediators is given by

$$\widehat{dir}_t := \frac{1}{\tilde{n}_t}\sum_{i=1}^{\tilde{n}_t}\left(\mu_{pdq_t,pdq_t^*,pdq_t^*}\{V_{t,i},\hat{\boldsymbol{\gamma}}\} - \mu_{pdq_t^*,pdq_t^*,pdq_t^*}\{V_{t,i},\hat{\boldsymbol{\gamma}}\}\right)$$

and the two estimators for the simpler sequential mediation formula for the interventional indirect effects via SDS_t and PHQ_{st} are given by

$$\widehat{indir}_{SDS_t} := \frac{1}{\tilde{n}_t}\sum_{i=1}^{\tilde{n}_t}\left(\mu_{pdq_t,pdq_t^*,pdq_t}\{V_{t,i},\hat{\boldsymbol{\gamma}}\} - \mu_{pdq_t,pdq_t^*,pdq_t^*}\{V_{t,i},\hat{\boldsymbol{\gamma}}\}\right)$$

and

$$\widehat{indir}_{PHQ_{st}} := \frac{1}{\tilde{n}_t}\sum_{i=1}^{\tilde{n}_t}\left(\mu_{pdq_t,pdq_t,pdq_t}\{V_{t,i},\hat{\boldsymbol{\gamma}}\} - \mu_{pdq_t,pdq_t^*,pdq_t}\{V_{t,i},\hat{\boldsymbol{\gamma}}\}\right)$$

respectively. Let $m_j\{v_{st},\boldsymbol{\xi}_t\} = E(SDS_{st}\mid PDQ_{st}=pdq_{st},PHQ_{st}=phq_{st},SDS_t=sds_t,PDQ_t=j,PHQ_t=phq_t)$ be defined by the $m\{v_{st},\boldsymbol{\xi}_t\}$-model at (6). Let $\mu_{j,k_{t_2}}\{v_t,\boldsymbol{\gamma}_{t_2}\} = E(m_j\{V_{st},\boldsymbol{\xi}_t\}\mid SDS_t=sds_t,PDQ_t=k,PHQ_t=phq_t)$ be defined by the model at (7) and let $\mu_{j,k,l_{t_1}}\{v_t,\boldsymbol{\gamma}_{t_1}\} = E(\mu_{j,k_{t_2}}\{V_t,\boldsymbol{\gamma}_{t_2}\}\mid PDQ_t=l,PHQ_t=phq_t,SDS_{pt}=sds_{pt},PDQ_{pt}=pdq_{pt},PHQ_{pt}=phq_{pt})$ be defined by the model at (8). The number $\tilde{n}_t$ denotes the number of complete cases. The number of complete cases for $t=b$ is equal to $(\tilde{n}_b=)341$ and the number of complete cases for $t=18$ is equal to $(\tilde{n}_{18}=)215$. The confidence intervals for both estimators are obtained using 1000 bootstraps.

4.2.1. Causal effect Maltesen et al. [12] show that the doubly robust DRMGf estimator (Doubly Robust estimator for Monotone missingness for the G-formula) of the g-formula for analysing the PERFORM study with missing observations is given by

$$\hat{E}\left(SDS_{st}^{(pdq_t,pdq_{st})}\right) = \frac{1}{n_t}\sum_{i=1}^{n_t}\left[\frac{I(\mathcal{C}_i=\infty)\mu_{t_1}\{V_{t,i},\hat{\boldsymbol{\gamma}}\}}{\varpi\{\infty,Z_{t,i},\hat{\boldsymbol{\psi}}\}} + \sum_{r=1}^{c}\frac{I(\mathcal{C}_i=r)-\lambda_r\{G_r(Z_{t,i},\hat{\boldsymbol{\psi}}\}I(\mathcal{C}_i\geq r)}{K_r\{G_r(Z_{t,i}),\hat{\boldsymbol{\psi}}\}}E\left(\mu_{t_1}\{V_t,\boldsymbol{\gamma}\}\mid G_r(Z_{t,i}),\hat{\boldsymbol{\zeta}}_t\right)\right] \tag{10}$$

when **c** equal to 5 for $t=b$ and **c** equal to 8 for $t=18$. It is also shown that both the estimator (10) and the estimator for the simpler g-formula below are asymptotically normally distributed in the situation when T is equal to 1 but this can also be shown for a larger T. The estimator uses the three models defined at (6), (7) and (8). The conditional expectations at (10) have to be modelled and afterwards used to predict values according to the different sets of $G_r(Z_{t,i})$ for $r=1,\ldots,\mathbf{c}$ and $i=1,\ldots,n_t$ for $t\in\{b,18\}$. The DRMGf estimator is compared to the estimator for the simper g-formula and it is given by

$$\hat{E}\left(SDS_{st}^{(pdq_t,pdq_{st})}\right) = \frac{1}{\tilde{n}_t}\sum_{i=1}^{\tilde{n}_t}\mu_{t_1}\{V_{t,i},\hat{\boldsymbol{\gamma}}_{t_1}\}$$

with the three models given at (6), (7) and (8). These three models are used for the estimation and the prediction. The number $\tilde{n}_t$ denotes the number of complete cases ($\tilde{n}_b=341$ and $\tilde{n}_{18}=215$). The confidence intervals for both estimators are obtained using 1000 bootstraps.

4.3. Results

The results of the causal effects are presented first followed by the results of the mediated effects. Table 1 shows the estimates of the causal effects $\boldsymbol{\beta}_t=(\beta_{I,t},\beta_{1,t},\beta_{2,t},\beta_{3,t})$ for $t\in\{b,18\}$ and Table 2 shows the estimates of the interventional direct effect and the interventional indirect effects with the two mediators. The coefficient $\beta_{I,t}$ corresponds to the expected score of functional impairment for patients having no or minimal cognitive symptoms at both visits for $t\in\{b,18\}$. The coefficient $\beta_{1,t}$ corresponds to the causal effect of cognitive symptoms on functional impairment at a later time for time $t\in\{b,18\}$. The coefficient $\beta_{2,t}$ corresponds to the causal effect of cognitive symptoms on functional impairment at the same time. The coefficient $\beta_{3,t}$ corresponds to the causal effect of the interaction between the two cognitive symptoms at the two time points for $t\in\{b,18\}$. The estimated effect of $\beta_{1,t}$ for $t\in\{b,18\}$ is the main interest because the causal paths between PDQ_t and SDS_{st} for $t\in\{b,18\}$ are the only causal paths containing mediators. The two mediators are SDS_t and PHQ_{st}. Both Tables show the results for the early (t=b) and the later (t=18) time points with the standard errors and the confidence intervals.

		G-formula			DRMGf		
Analysis		Effect	SE	95%-CI	Effect	SE	95%-CI
Z_b	$\beta_{I,b}$	9.588	1.243	(7.152 ; 12.025)	9.880	1.233	(7.463 ; 12.297)
	$\beta_{1,b}$	4.165	2.021	(0.204 ; 8.126)	3.791	1.892	(0.083 ; 7.500)
	$\beta_{2,b}$	3.566	2.026	(-0.404 ; 7.536)	3.611	2.076	(-0.458 ; 7.679)
	$\beta_{3,b}$	-1.848	2.613	(-6.969 ; 3.272)	-1.805	2.539	(-6.781 ; 3.171)
Z_{18}	$\beta_{I,18}$	8.238	1.225	(5.837 ; 10.638)	9.031	1.290	(6.504 ; 11.559)
	$\beta_{1,18}$	0.198	1.537	(-2.815 ; 3.212)	0.427	1.603	(-2.714 ; 3.568)
	$\beta_{2,18}$	-0.322	1.422	(-3.108 ; 2.465)	-0.413	1.595	(-3.540 ; 2.714)
	$\beta_{3,18}$	2.986	1.490	(0.066 ; 5.906)	3.024	1.502	(0.080 ; 5.969)

Table 1. The G-formula column shows the estimates obtained using the estimator for the simpler g-formula and the DRMGf column shows the estimates obtained using the DRMGf estimator. The Analysis column shows the analysis of the data with respect to either Z_b or Z_{18}. The Effect column shows the estimated effects. The SE column shows the standard errors for the estimates. The standard error is obtained using 1000 bootstraps. The 95%-CI column shows the confidence intervals for the estimates.

Both estimators provide almost the same estimates for the early and the later time points. The two estimators suggest that patients with cognitive symptoms at both visits have worse functioning than patients with no or with minimal cognitive symptoms at both visits. Maltesen et al. [12] have shown that the estimator for the simpler g-formula and the DRMGf

estimator will provide similar results. It is most likely caused by the included covariates that are poor at predicting drop-out that generate the pattern of monotone missingness in the data.

		Seq. mediation formula			DRMSM		
Analysis	Interventional	Effect	SE	95%-CI	Effect	SE	95%-CI
Z_b	dir_b	1.385	1.711	(-1.968 ; 4.739)	1.392	1.762	(-2.062 ; 4.846)
	$indir_{SDS_b}$	2.490	0.705	(1.109 ; 3.870)	2.461	0.610	(1.266 ; 3.657)
	$indir_{PHQ_2}$	0.290	0.870	(-1.416 ; 1.995)	-0.062	0.777	(-1.584 ; 1.461)
	The overall effect	4.165	2.021	(0.204 ; 8.126)	3.791	1.892	(0.083 ; 7.500)
Z_{18}	dir_{18}	-1.741	1.180	(-4.054 ; 0.573)	-1.771	1.309	(-4.338 ; 0.795)
	$indir_{SDS_{18}}$	0.318	0.370	(-0.408 ; 1.043)	0.680	0.404	(-0.112 ; 1.472)
	$indir_{PHQ_{24}}$	1.621	0.942	(-0.224 ; 3.467)	1.518	0.897	(-0.240 ; 3.277)
	The overall effect	0.198	1.537	(-2.815 ; 3.212)	0.427	1.603	(-2.714 ; 3.568)

Table 2. The Seq. mediation formula column shows the estimates obtained using the estimator for the simpler sequential mediation formula and the DRMSM column shows the estimates obtained using the DRMSM estimator. The dir_t row is the direct effect for $t \in \{b, 18\}$. The $indir_{SDS_t}$ row is the indirect effect via SDS_t for $t \in \{b, 18\}$. The $indir_{PHQ_{st}}$ row is the indirect effect via PHQ_{st} for $t \in \{b, 18\}$. The Interventional column shows the direct effect, indirect effects and the overall effect. See Table 1 for the description of the Analysis, Effect, SE and 95%-CI columns.

The results from the mediation analysis do not show a large difference between the two estimators. This is not surprising that the two estimators provide almost similar results since it is most likely caused by the included covariates that are poor at predicting drop-out. Despite the similar results with the two estimators, the results from the mediation analysis are slightly more surprising than the results from the estimation of the causal effects in Table 1. The results show a small difference between the two estimators in estimating the interventional indirect effects. The difference between the estimator for the simpler sequential mediation formula and the DRMSM estimator is more pronounced compared to the difference between the estimator for the simpler g-formula and the DRMGf estimator.

The negative sign of the coefficient of the direct effect appears counter-intuitive. It indicates that patients with cognitive symptoms are more likely to directly improve their functional impairment at a later time compared to patients with no or with minimal cognitive symptoms. I would have expected the opposite. However, the negative sign of the analysis with $t = 18$ may be caused by many patients after 18 months who are doing well. The room for improvement among the patients is smaller, and we then see the effect that patients with cognitive symptoms are more likely to improve their functioning compared to patients with no or with minimal cognitive symptoms since the scales have a lower finite limit.

I know from the simulation study in Maltesen et al. [12] that stronger predictors for the missing mechanism will create a larger difference between the two estimators. A simulation study will also be conducted here for further exploration of handling missing data due to drop-out and the interpretation of the data. This is explored in the simulation study in the next Section.

5. Simulation study

The purpose of the simulation study is to investigate the DRMSM estimator with similar data as the PERFORM study but the probabilities of the missingness mechanism will be more extreme compared to the missingness mechanism in the PERFORM study. The simulation study is the same one used in Maltesen et al. [12]. It is based on the first two vectors of the PERFORM study (W_b, W_2). The sample size of the data is 1000 and the data are replicated 5000 times. See Maltesen et al. [12] for further information about the simulation study. Table 3 shows the results of the estimation with respect to the estimator for the simpler sequential mediation formula and the DRMSM estimator. The models in Section 4.2 for $t = b$ are used for the estimation of the interventional direct and indirect effects of the simulated data. The simulation study is evaluated by the mean of the 5000 estimates of $\eta = (dir_b, indir_{SDS_b}, indir_{PHQ_2}, Overall)$, the empirical standard error (SE) of the 5000 estimates of η, the absolute value of bias (the difference between the empirical mean and the true value), the ratio between the absolute value of bias and SE scaled 100 times and the mean squared error (MSE) [12].

Table 3 shows the expected discrepancy between the estimator for the simpler sequential mediation formula and my DRMSM estimator. It is clear that my DRMSM estimator protects against biased estimates compared to the estimator for the simpler sequential mediation formula. The estimator for the simpler sequential mediation formula shows weakness in estimating the mediated effects. My DRMSM estimator should be used for estimating the mediated effects when data contains missing observations that follow a monotone pattern. The price for using my estimator may result in larger

Estimator	Interventional	True	Mean	SE	Bias	$\frac{\text{Bias}}{\text{SE}}100$	MSE
Seq. mediation formula	dir_b	1.402	1.076	1.386	0.326	23.521	2.029
	$indir_{SDS_b}$	2.473	2.391	0.655	0.083	12.642	0.436
	$indir_{PHQ_2}$	-0.107	-0.201	0.740	0.094	12.652	0.557
	$Overall$	3.769	3.266	1.622	0.503	30.977	2.885
DRMSM	dir_b	1.402	1.436	1.991	0.034	1.695	3.999
	$indir_{SDS_b}$	2.473	2.478	0.497	0.004	0.862	0.251
	$indir_{PHQ_2}$	-0.107	-0.116	0.683	0.009	1.315	0.476
	$Overall$	3.769	3.798	2.096	0.029	1.386	4.422

Table 3. Let η denote the vector $(dir_b, indir_{SDS_b}, indir_{PHQ_2}, Overall)$. The Seq. mediation formula row shows the estimates obtained using the estimator for the simpler sequential mediation formula. The DRMSM row shows the estimates obtained using the DRMSM estimator. The Interventional column shows the direct effect (dir_b), the indirect effects ($indir_{SDS_b}$ and $indir_{PHQ_2}$) and the overall effect ($Overall$). The True column shows the true effects. The Mean column is the mean of the 5000 estimates of η. The SE column is the standard error of the 5000 estimates of η. The Bias column shows the absolute value of the difference between the empirical mean and the true value. The $\frac{\text{Bias}}{\text{SE}}100$ column is the ratio between the absolute value of bias and the standard error scaled 100 times. The MSE column is the mean square error obtained by $\text{Bias}^2 + \text{SE}^2$.

standard errors in order to obtain unbiased estimates. The interpretation of Table 3 is that the missing observations in the data need to be addressed in mediation analysis otherwise mistakes may happen.

6. Discussion

This manuscript was motivated by the PERFORM study to develop a doubly robust estimator (DRMSM) for estimating the mediated effects of the exposure on the outcome while data contains missing observations that follow a monotone pattern. The proposed estimator was applied to the PERFORM study with patients suffering depression. My DRMSM estimator was compared to the estimator for the simpler sequential mediation formula, which did not take the missing data into account. The similarities in the estimates in the example are most likely caused by the included covariates that are poor at predicting drop-out. Thereby, the robustness of my DRMSM estimator was not shown in the results of the PERFORM study. However, the simulation study revealed that if the included covariates are strong at predicting drop-out, then the estimator for the simpler sequential mediation formula is biased, while my DRMSM estimator is not. My DRMSM estimator shares the same advantages, disadvantages and limitations as the DRMGf estimator. This means that my DRMSM estimator utilizes data better than an estimator using only complete cases. The missing at random assumption needs to be addressed, but an analysis only with complete cases relies on the assumption that the missingness is missing completely at random. This is less plausible than missing at random. The assumption regarding monotone missingness also needs to be addressed because the assumption does not allow data to have intermittent missing values. The assumption regarding an ordered sequence of variables is sometimes a natural assumption. The causal ordering of variables in the PERFORM study is only partially clear, as the causal ordering of the variables between the different time points is introduced by time itself. However, the assumption of the order between the three different domains within the same time point is not clear because the three domains were measured at the same time points at six occasions over two years. The assumption of the causal ordering between the three different domains within the same time points had to be made based on clinical insight such as a change in depression severity causing a change in cognitive performance, which in turn causes a change in functioning. The interpretation of the results hinges on these assumptions and yet the assumptions cannot be verified in the data [12]. Further research could be to extend the DRMSM estimator to also include vectors in the data that follow a nonmonotone pattern.

Acknowledgments

The PERFORM study was funded by Lundbeck SAS. The author thanks Rhian M. Daniel (Division of Population Medicine, Cardiff University, UK) for her advice and ideas. The author also thanks Lene Hammer-Helmich (Medical Affairs, Value Evidence, H.Lundbeck A/S, Denmark) for her advice and ideas.

Conflict of interest

Thomas Maltesen (temporary employee) is a full-time employee of H. Lundbeck A/S. The author reports no conflict of interest.

References

[1] Chokka P, Bougie J, Proulx J, Tvistholm AH, Ettrup A. Long-term functioning outcomes are predicted by cognitive symptoms in working patients with major depressive disorder treated with vortioxetine: results from the atworc study (book review). *CNS Spectrums* 2019;:1–12.

[2] Hammer-Helmich L, Haro JM, Jönsson B, *et al.* Functional impairment in patients with major depressive disorder: the 2-year perform study. *Neuropsychiatric Disease and Treatment* 2018;**14**:239–249.

[3] Pearl J. Direct and indirect effects. *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence* 2001;(10):411–420.

[4] Robins JM, Greenland SM. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 1992; **3**(2):143–155.

[5] Avin C, Shpitser I, Pearl J. Identifiability of path-specific effects 2005;:357–363.

[6] VanderWeele TJ, Vansteelandt S, Robins JM. Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology* 2014;**25**(2):300–306.

[7] VanderWeele TJ, Robinson WR. On the causal interpretation of race in regressions adjusting for confounding and mediating variables. *Epidemiology* 2014;**25**(4):473–484.

[8] Vansteelandt S, Daniel RM. Interventional effects for mediation analysis with multiple mediators. *Epidemiology (Cambridge, Mass)* 2017;**28**(2):258–265.

[9] VanderWeele T, Vansteelandt S. Mediation analysis with multiple mediators. *Epidemiologic Methods* 2013;**2**(1):95–115.

[10] Maltesen T, Larsen KG, Hammer-Helmich L. Sequential mediation analysis with multiple mediators. *Submittable* ;.

[11] Li W, Zhou X. Identifiability and estimation of causal mediation effects with missing data. *Statistics in Medicine* 2017;**36**(25):3948–3965.

[12] Maltesen T, Martinussen T, Larsen KG, Hammer-Helmich L. A doubly robust estimator for monotone missing data in the presence of time-dependent confounding. *Submittable* ;.

[13] Tsiatis AA Anastasios A. Semiparametric theory and missing data. Springer series in statistics. Springer, 2006.

[14] Pearl J. An introduction to causal inference. *The international journal of biostatistics* 2010;**6**(2).

[15] VanderWeele TJ. Marginal structural models for the estimation of direct and indirect effects. *Epidemiology* 2009; **20**(1):18–26.

[16] Kaufman JS, Maclehose RF, Kaufman S. A further critique of the analytic strategy of adjusting for covariates to identify biologic mediation. *Epidemiologic perspectives & innovations : EP+I* 2004;**1**(1):4.

[17] VanderWeele TJ. Mediation and mechanism. *European Journal of Epidemiology* 2009;**24**(5):217–224.

[18] VanderWeele TJ. Policy-relevant proportions for direct effects. *Epidemiology* 2013;**24**(1):175–176.

