Towards an Accessible Methodology in Precision Medicine: Methods for Censored Data and Non-regular Inferences

Gabrielle Simoneau

Doctor of Philosophy

Department of Epidemiology, Biostatistics and Occupational Health

McGill University
Montréal, Québec, Canada
September 2019

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

Dedication

This thesis is dedicated to maman, papa, Natalie, Charles, Martin, Julien, grand-maman Carmen, grand-maman Françoise, and Liao.

Acknowledgements

First and foremost, I would like to thank my co-supervisors Drs. Erica Moodie and Robert Platt for their constant support over the past four years. Erica, for her availability and her prompt response to my questions and requests, for always being up-to-date on the progress of my research and for her guidance and encouragements in more difficult times. Robert, for his enthusiasm and encouragements and for his mentoring beyond academic life. I am grateful to both of my supervisors for providing uncountable opportunities to disseminate my work and for financially supporting my numerous travels.

This work would not have been possible without the financial support provided by the Fonds de recherche du Québec – Nature et technologies, by my supervisors and by the Department of Epidemiology, Biostatistics and Occupational Health (EBOH) at McGill. I would also like to acknowledge the constant support offered by staff members of the EBOH department. Katherine Hayden, Deirdre Lavery and André Yves Gagnon, for considerably facilitating the progress of my research. Staffs at the Lady Davis Institute also helped in the accomplishment of this thesis, notably Marisa Mancini and Hui Yin.

The data analyzed in this thesis were kindly made available by Dr. Michael Kramer and the investigators of the PROmotion of Breastfeeding Intervention Trial, and by Dr. Jagtar Nijjar and the Scottish Early Rheumatoid Arthritis Inception Cohort Investigators. Data were also acquired through the Clinical Practice Research Datalink license of the McGill Pharmacoepidemiology Research Unit. I acknowledge the usage of computational facilities provided by Compute Canada and the Department of Statistics at McGill.

I am grateful to my colleagues at the EBOH department for nurturing a pleasant and motivating research environment, especially the resourceful and inspiring Drs. Sahir Bhatnagar and Maxime Turgeon. I also want to acknowledge my mentors, Drs. Andrea Benedetti, Antonio Ciampi, Paramita Saha Chaudhuri, and Alexandra Schmidt for providing insightful

advice during my time at McGill, and Drs. David Haziza and Alejandro Murua from Université de Montréal for setting me on this path.

I am forever grateful to my friends and family. To maman and papa, for continuously repeating how proud of me they are. To my little brothers, Charles, Martin and Julien, for making me smile and laugh. To Natalie and my friend Laura, who could understand my student life struggles and victories. And to my life partner, Liao, for his unconditional support and continuous encouragements, and for helping me in becoming a better person.

Preface: Contribution of Authors

This thesis includes original contributions and reviews of existing methods for the estimation of optimal dynamic treatment regimes. Details on the original contributions to knowledge in Chapters 3–6 are also given in their respective preambles.

Chapters 1 and 2 provide an original introduction and a critical review of existing methods and challenges in dynamic treatment regimes. Both chapters were written by Gabrielle Simoneau (GS) and edited by Erica EM Moodie (EEMM) and Robert W Platt (RWP).

Chapter 3 was conceptualized by GS and EEMM, following previous collaborative work between EEMM and Bibhas Chakraborty (BC). GS carried out the methodological work, programming, data analysis and writing, with the help of EEMM and RWP in designing the simulation study, interpreting interim results and editing the manuscript. BC and Michael S Kramer also reviewed the manuscript.

Chapter 4 was conceptualized by GS and EEMM. All methodological derivations were carried out by GS as well as the design of the simulation study, with the help of EEMM for guidance and troubleshooting. The programming and data analysis were done by GS. The manuscript was written by GS and corrected by EEMM, RWP and Jagtar S Nijjar.

Chapter 5 was conceptualized by GS and EEMM. All methodological work, programming and writing were carried out by GS. EEMM helped GS to design the simulation study. The manuscript was reviewed and edited by EEMM and RWP.

The case study in Chapter 6 was conceptualized by GS, Laurent Azoulay (LA) and RWP. The programming, data analysis, interpretation of the results and writing was carried out by GS. Hui Yin provided sample codes and definitions for building the study cohort. EEMM, LA and RWP helped to interpret interim and final results and edited the manuscript.

Chapter 7 includes an original summary of the work in this thesis and ideas for future work by GS. The chapter was written by GS and corrected by EEMM and RWP.

Preface: Ethics Approval

The analyses presented in this thesis use data previously collected from human subjects. Chapter 3 uses data from the PROmotion of Breastfeeding Intervention Trial, for which ethics approvals were obtained by the original study from the McGill University Health Center Research Ethics Board, the Human Subjects Committees at Harvard Pilgrim Health Care and the Avon Longitudinal Study of Parents and Children Law and Ethics Committee. Chapter 4 uses data from the Scottish Early Rheumatoid Arthritis Study, for which an ethics approval was obtained from the West of Scotland Research Ethics Service by the original study. Chapter 6 uses a cohort from the United Kingdom Clinical Research Practice Datalink (CPRD). Ethics approvals were obtained from the Independent Scientific Advisory Committee of the CPRD (protocol number 18_169) and from the Faculty of Medicine Institutional Review Board at McGill University.

Abstract

A dynamic treatment regime (DTR) formalizes the study of precision medicine in which treatment decisions across multiple stages of clinical intervention are tailored to evolving, patient-level information. Statistical methods for DTR are concerned with identifying an optimal DTR, that is, the sequence of treatment decisions that yields the best expected outcome for a population of ("similar") individuals.

Dynamic weighted ordinary least squares (dWOLS) offers an accessible and theoretically robust framework for estimation and inference of an optimal DTR. However, it suffers from several limitations. First, like other regression-based DTR estimation approaches, dWOLS can yield estimators with non-regular limiting distributions in the sense that the standard asymptotic theory does not hold, in turn leading to incorrect coverage of confidence intervals. A second limitation is that dWOLS only handles uncensored continuous outcomes. It is often the case that the clinical outcome of interest is a survival time, which is typically subject to right-censoring.

This thesis is composed of four manuscripts. The first manuscript compares the standard bootstrap to the m-out-of-n bootstrap with dWOLS when estimators suffer from non-regularity. An application to decision rules about an infant's diet on childhood outcomes six years later, in which estimators are likely to suffer from non-regularity, is presented.

In the second manuscript, we propose a novel method called dynamic weighted survival modeling (DWSurv) for estimation and inference of an optimal DTR with survival outcomes subject to right-censoring. An application to rheumatoid arthritis, in which a series of treatments is typically recommended to achieve remission, is presented.

The third manuscript describes an extensive simulation study to evaluate the finite sample properties of competing methods for constructing confidence intervals for the DWSurv parameters, including parametric and non-parametric bootstrap as well as methods based on asymptotic theory. The impact of non-regularity is also assessed.

The fourth and last manuscript showcases DWSurv in an illustrative example about the

treatment of type 2 diabetes, where the objective is to find an optimal sequence of treatments that maximizes the time until the occurrence of a cardiovascular event or death. The first stage compares the addition of sulfonylurea or dipeptidyl peptidase-4 inhibitors to metformin. Extensions to more than one stage are described. Data from a large observational database are used.

Abrégé

Les plans dynamiques de traitements (PDT) formalisent l'étude de la médecine de précision où les décisions de traitement à travers plusieurs phases d'intervention sont adaptées aux caractéristiques des patients. Les méthodes statistiques pour l'étude d'un PDT cherchent à identifier un PDT optimal, c'est-à-dire la séquence de décisions de traitement qui mène à la meilleure réponse espérée pour une population d'individus similaires.

DWOLS est une méthode statistique théoriquement robuste, accessible et facile à appliquer pour l'estimation et l'inférence d'un PDT optimal. Cependant, la méthode comporte plusieurs limitations. Premièrement, comme d'autres approches d'estimation de PDT basées sur la régression, les estimateurs dWOLS peuvent avoir des distributions limites non-regulières dans le sens où la théorie asymptotique standard ne s'applique pas, menant ainsi à des intervalles de confiance avec des couvertures incorrectes. Une deuxième limitation est que dWOLS prend seulement en compte les réponses continues non-censurées. Il arrive souvent que les réponses d'intérêt clinique soient des temps de survie typiquement sujets à la censure.

Cette thèse est composée de quatre manuscrits. Le premier manuscrit compare le bootstrap standard au bootstrap *m-out-of-n* (m parmi n) avec dWOLS lorsque les estimateurs souffrent de non-regularité. Une application concernant des règles de décisions pour la diète d'un nourrisson sur des réponses métaboliques mesurées durant l'enfance six ans plus tard, contexte dans lequel les estimateurs sont probablement non-réguliers, est présentée.

Dans le deuxième manuscrit, nous proposons une nouvelle méthode appelée DWSurv pour l'estimation et l'inférence d'un PDT optimal avec des temps de survie sujets à la censure comme réponse. Une application à l'arthrite rhumatoïde, une maladie chronique pour laquelle une séquence de traitements est typiquement recommandée pour atteindre la rémission, est présentée.

Le troisième manuscrit décrit une étude de simulation de grande ampleur pour évaluer les propriétés d'échantillon fini de différentes méthodes pour construire des intervalles de confiance

pour les paramètres de DWSurv, incluant le bootstrap paramétrique et non-paramétrique ainsi que des méthodes basées sur la théorie asymptotique. L'impact de la non-regularité est aussi étudié.

Le quatrième et dernier manuscrit démontre l'utilité de DWSurv dans une étude de cas sur le traitement du diabète de type 2 pour laquelle l'objectif est de trouver une séquence optimale de traitements qui maximise le temps jusqu'à la survenance d'un évènement cardiovasculaire ou la mort. La première phase compare l'addition du sulfonylurea ou des inhibiteurs de la dipeptidyl peptidase-4 à metformin. L'extension à plus d'une phase de traitements est décrite. Des données provenant d'une grande base de données observationnelles sont utilisées.

Table of contents

List of Tables

List of Figures

Abbreviations

ACR American College of Rheumatology

AFT accelerated failure time

BMI body mass index

CPRD Clinical Practice Research Datalink

CRAN the comprehensive R archive network

CRP C-reactive protein level

DAS28 Disease Activity Score 28

DMARD disease-modifying antirheumatic drug

DPP-4i dipeptidyl peptidase-4 inhibitors

DTR dynamic treatment regime

dWOLS dynamic weighted ordinary least squares

DWSurv dynamic weighted survival modeling

ESR erythrocyte sedimentation rate

GEE generalized estimating equations

GLP-1 glucagon-like peptide-1 receptor agonists

GVHD graft-versus-host disease

HbA1c glycated hemoglobin

HNW Huang et al. (2014)

i.i.d. independent and identically distributed

IPCW inverse probability of censoring weights

IPTW inverse probability of treatment weights

IPW inverse probability weighted

kg kilograms

MLE maximum likelihood estimator

mo months

OLS ordinary least squares

PROBIT PROmotion of Breastfeeding Intervention Trial

RA rheumatoid arthritis

SERA Scottish Early Rheumatoid Arthritis

SES socio-economic status

SNFTM structural nested failure time models

SUTVA stable unit treatment value assumption

SVM support vector machines

T2D type 2 diabetes

TNF tumor necrosis factor

UA undifferentiated arthritis

UK United Kingdom

UNICEF United Nations Children's Fund

WHO World Health Organization

ZIPI zeroing instead of plugging in

Chapter 1

Introduction

In this thesis, we consider the estimation of and inference for an optimal dynamic treatment regime (DTR) with continuous outcomes, censored or not, using observational data. A DTR is a set of treatment decision rules, each rule corresponding to a decision time point, that dictates the treatment action to be taken at that point as a function of (time-dependent) individual characteristics. For example, treatment decisions may be made at different stages of a disease or at routine clinical visits and the decisions may depend on the patient's specific condition at the time of decision-making, including his response to previous treatments. This framework is especially relevant for chronic diseases, such as rheumatoid arthritis and type 2 diabetes (T2D), where the patient's condition is changing over time and treatments must correspondingly be altered. The study of precision medicine, which aims to identify "the right treatment for the right patient", is concerned with the discovery of an optimal DTR, that is, the sequence of treatment rules that leads to the best expected outcome for a population of patients sharing similar characteristics.

Identifying an optimal DTR is a sequential decision-making problem in which short- and long-term treatment effects may be hard to disentangle. The sequence of treatments found by optimizing separately each decision may not correspond to the sequence of treatments that

actually leads to the best long-term outcome. For example, a less effective first-line treatment may increase adherence to second-line treatments and lead to a better outcome while a more aggressive first-line treatment would appear optimal in the short term but actually decrease adherence to second-line treatments and yield a suboptimal outcome. Determining an optimal sequence of treatments by first comparing first-line treatments would unknowingly set patients on treatment pathways that include only inferior long-term outcomes; thus the need for statistical methods that jointly estimate an optimal DTR across decision time points.

In this thesis, we build on an existing method for optimal DTR with continuous uncensored outcomes named dynamic weighted ordinary least squares (dWOLS) (Wallace & Moodie, 2015). DWOLS is theoretically robust yet its simple framework based on a series of linear regressions is accessible to non-statisticians. It is implemented in the `R` package `DTRreg` along with tools for model checking. DWOLS estimates an optimal DTR from experimental data (e.g. data arising from sequential multiple assignment randomized trials (Murphy, 2005)) or from non-experimental data (e.g. observational study, registry data, claims data). The estimation of an optimal DTR from non-experimental data must deal with the absence of treatment randomization which may lead to confounding; that is, the treatment assignment and the outcome share common causes which prevent estimating an unbiased treatment effect. For this, dWOLS relies on a broad class of weights, including the well-known inverse probability of treatment weights (IPTW), to balance the distribution of the confounders across treatment groups.

DWOLS currently suffers from two main limitations. First, dWOLS yields non-regular estimators; that is, estimators whose distribution does not converge uniformly over the parameter space (Robins, 2004). A negative consequence of non-regularity is that confidence intervals for the parameters used to construct the treatment decision rules may not have nominal coverage. It is only under certain data generating mechanisms that non-regularity

will be problematic; thus the importance of studying why and when non-regular estimation have an impact in practice. Second, dWOLS is currently restricted to uncensored continuous outcomes. Often, the criterion of optimality is survival time. For example, in T2D, an optimal sequence of treatments could maximize the time until the occurrence of severe complications. With survival outcomes being subject to right-censoring, another layer of statistical complexity is introduced because the survival times of all individuals are not necessarily observed nor do all individuals undergo the same number of stages of clinical intervention (some individuals will be censored or experienced an event early and will not receive second- or third-line treatments).

In this thesis, we use dWOLS as a starting point to study the impact of non-regular estimation in practice and to propose a novel multi-stage optimal DTR method that can handle survival outcomes. In Chapter 2, we provide a critical review of the literature on the problem of non-regular estimation and on optimal DTR methods for censored data. In Chapter 3, we propose using the m-out-of-n bootstrap to alleviate the negative impact of non-regular estimation with dWOLS, supporting our demonstration with a computationally intensive simulation study and a case study on the impact of infant food intake on childhood metabolic outcomes. In Chapter 4, we develop dynamic weighted survival modeling (DWSurv) to estimate an optimal DTR with survival outcomes. In Chapter 5, we explore the finite sample performance of different methods for constructing confidence intervals for the parameters in the treatment decision rules with DWSurv, including scenarios that focus on non-regular estimation. In Chapter 6, we present a case study on the treatment of T2D using data from a large registry database and DWSurv. Chapters 3 to 6 were written as stand-alone manuscripts and despite our efforts to maintain consistent notation and minimize overlap of information throughout the thesis, there is necessarily some repeated information between the literature review and each manuscript as well as some variations in the notation across the chapters of this thesis. Chapter 3 has been published in *Biostatistics*. Chapter 4 has been published in the *Journal of the American Statistical Association*. Chapters 5 and 6

have been submitted for publication. The `DTRreg` package has been updated and published on the comprehensive R archive network (CRAN) with the methods presented in Chapters 3–5 (https://CRAN.R-project.org/package=DTRreg).

Chapter 2

Literature Review

The literature review is comprised of five sections. The first introduces concepts and methods regularly used in the DTR literature, including dWOLS. The second section describes the problem of non-regular estimation in the context of DTRs and reviews solutions proposed in the literature. The third section focuses on optimal DTRs with censored data, describing additional challenges and providing an extensive review of existing methods. The last section introduces current guidelines on the management of T2D and reviews (the lack of) evidence for identifying optimal DTRs.

2.1 Dynamic Treatment Regimes: General Framework

A DTR is a set of treatment decision rules implemented over time, each rule inputting individual characteristics and outputting a recommended treatment for that individual at the time of decision-making. A DTR is partitioned into stages of clinical intervention where each treatment rule is applied at the beginning of the corresponding stage. For clarity, our descriptions often concern a DTR with two stages, with the understanding that generalization to more than two stages is straightforward once the two-stage setting is established.

Upper and lower case letters respectively correspond to random variables and their realizations. We suppose that the available data come from a non-experimental study with n participants identified with the subscript i, often omitted for clarity, although all methods presented in this thesis could be applied to experimental data. Treatment options at the beginning of stage j are denoted with A_j where we allow the treatment options in the second stage to depend on the treatment received in the first stage. We assume that treatments are binary options so that $A_j \in \{0,1\} \equiv \mathcal{A}_j$. Let $\boldsymbol{X_1}$ denote the individual characteristics (covariates) measured at the beginning of the first stage before initiating treatment A_1 and $\boldsymbol{X_2}$ denote covariates measured at the beginning of stage 2 before initiating treatment A_2. $\boldsymbol{X_2}$ includes time-varying covariates which may depend on the treatment received in the first stage and, with a slight abuse of notation, $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ may represent different subsets of measured covariates. The overall outcome Y is defined as the sum of stage-specific outcomes Y_1 and Y_2 and is a continuous uncensored variable, although this requirement will be relaxed in Section 2.3. Sometimes, $Y_1 \equiv 0$ and $Y_2 = Y$ is measured only at the end of the second and last stage of intervention. Without loss of generality, we assume that larger values of the outcome are preferred. Individual data are represented with a longitudinal trajectory $(\boldsymbol{X_1}, A_1, Y_1, \boldsymbol{X_2}, A_2, Y_2)$ where the ordering of the variables corresponds to the order in which they are recorded. The data are conveniently grouped into history $\boldsymbol{H_j} \in \mathcal{H}_j$ which corresponds to accrued treatment and covariate information available at the beginning of stage j but not including the stage j treatment, so $\boldsymbol{H_1} = \boldsymbol{X_1}$ and $\boldsymbol{H_2} = (\boldsymbol{X_1}, A_1, \boldsymbol{X_2})$. A two-stage DTR consists of the treatment decision rules $\boldsymbol{d} = \{d_1(\boldsymbol{h_1}), d_2(\boldsymbol{h_2})\}$ where $d_j(\boldsymbol{h_j}) : \mathcal{H}_j \to \mathcal{A}_j$ is the decision rule at the beginning of stage j which inputs the history $\boldsymbol{H_j}$ and outputs a recommended treatment A_j.

Statistical models to estimate an optimal DTR can be framed in terms of potential outcomes, also called counterfactual outcomes (Rubin, 1974). Let Y^{a_1,a_2} denote an individual's potential outcome at the end of the second stage if, possibly contrary to the fact, he receives treatments $\boldsymbol{a} = (a_1, a_2)$. The axiom of consistency, which states that the actual outcome Y and potential

outcome $Y^{\boldsymbol{a}}$ are equal when the regime $\boldsymbol{a}$ is actually received, must be satisfied. The two following assumptions are required for the estimation of an optimal DTR:

(A1) The stable unit treatment value assumption, which states that an individual's outcome is not influenced by the treatment allocation of other individuals (Rubin, 1980).

(A2) Sequential ignorability, which extends the no unmeasured confounders assumption to longitudinal settings and requires that, for any regime (a_1, a_2), $A_j \perp\!\!\!\perp Y^{a_1,a_2} \mid \boldsymbol{H_j}$ for $j = 1, 2$ (Robins, 1997).

An optimal DTR is defined as the sequence of treatment decision rules $\boldsymbol{d}^{\text{opt}} = \{d_1^{\text{opt}}(\boldsymbol{h_1}), d_2^{\text{opt}}(\boldsymbol{h_2})\}$, respectively corresponding to optimal treatments $\boldsymbol{a}^{\text{opt}} = (a_1^{\text{opt}}, a_2^{\text{opt}})$, that maximizes the expected potential outcome $\mathbb{E}[Y^{\boldsymbol{d}}]$. The expectation $\mathbb{E}[Y^{\boldsymbol{d}}]$ defines the population average outcome had all individuals received treatments according to $\boldsymbol{d}$. The potential outcome $Y^{\boldsymbol{d}}$ is also called the value of a regime and denoted as $V(\boldsymbol{d})$.

There are two general approaches to estimate an optimal DTR, regression-based estimation and value search estimation, which we describe in a single-stage setting for clarity. Regression-based estimation posits and estimates a (parametric) model for the outcome given the treatment a_1 and covariate $\boldsymbol{h_1} = (1, x_1)$ e.g.

$$\mathbb{E}(Y| = \boldsymbol{H_1} = \boldsymbol{h_1}, A_1 = a_1; \boldsymbol{\beta}, \boldsymbol{\psi}) = \beta_0 + \beta_1 x_1 + a_1(\psi_0 + \psi_1 x_1) \tag{2.1}$$

and finds the optimal treatment rule $d^{\text{opt}}(\boldsymbol{h_1})$ by maximizing this quantity with respect to the treatment i.e. $d^{\text{opt}}(\boldsymbol{h_1}) = \arg\max_{a_1 \in \mathcal{A}_1} \mathbb{E}(Y|\boldsymbol{H_1} = \boldsymbol{h_1}, A_1 = a_1; \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\psi}})$. The parametric specification of the outcome model defines a class of possible regimes $\mathcal{D}_\psi$ indexed by a parameter ψ. For example, the parametrization (2.1) defines linear decision rules of the form $d_{\boldsymbol{\psi}}(\boldsymbol{h_1}) = \mathbb{I}(\psi_0 + \psi_1 x_1 > 0)$. Regression-based estimation typically requires that the model for the outcome is correctly specified in order to consistently estimate the parameters $\boldsymbol{\psi}$ used to construct the decision rule and further requires that the true optimal decision

rule belongs to the class of regimes defined by the chosen parametrization for the outcome model. Doubly-robust alternatives have been proposed, offering additional protection against misspecification of the outcome model by incorporating a model for the probability of treatment in the estimation procedure (e.g. Robins, 2004; Wallace & Moodie, 2015). Q-learning, G-estimation and dWOLS are important regression-based estimation methods which are described in Section 2.1.2. Bayesian approaches have also been proposed, including the recent Bayesian machine learning method (Murray et al., 2018).

Regression-based methods focus on estimation and inference for the parameters $\boldsymbol{\psi}$ used to construct the decision rules. The uncertainty about the decision rules can be characterized using well-established inferential principles, which is essential if one wants to use the estimated DTR to inform clinical decisions or future research. Standard model checking tools for regression models can also be used. Regression-based estimation suffers a few limitations. In DTRs with two stages or more, the estimators $\hat{\boldsymbol{\psi}}$ have non-regular limiting distributions which may negatively impact the construction of confidence intervals. This issue is discussed in detail in Section 2.2. Regression-based methods are also criticized for offering a weak framework for nonlinear decision rules (Laber et al., 2014a; Qian & Murphy, 2011). Methods have been proposed to accommodate flexible decision rules at the cost of losing interpretability of the rules (Laber et al., 2014a; Moodie et al., 2014).

Value search estimation alleviates the need to specify a model for the outcome by directly maximizing an estimator of the value of a regime expressed as a function of the decision rule, $\hat{V}(d_\psi)$. A restricted class of regimes $d_\psi(\boldsymbol{h_1}) \in \mathcal{D}_\psi$ is also considered. Most value search methods build on the inverse probability weighted (IPW) estimator for $V(d_\psi)$ (Robins, 2000a) given by

$$\hat{V}(d_\psi) = \frac{1}{n} \sum \frac{Y \mathbb{I}(A_1 = d_\psi(\boldsymbol{h_1}))}{P(A_1 = a_1 | \boldsymbol{h_1}; \hat{\boldsymbol{\alpha}})} \tag{2.2}$$

in which a propensity score model $P(A_1 = 1 | \boldsymbol{h_1}; \boldsymbol{\alpha})$ needs to be specified and estimated. The value is maximized with respect to ψ, $\psi^{\text{opt}} = \arg\max_\psi \hat{V}(d_\psi)$, and the optimal decision rule

is obtained as $d_{\psi^{\text{opt}}}(\boldsymbol{h_1})$. The consistency of estimators derived from value search methods relies on the correct specification of the propensity score model as opposed to the correct specification of the outcome model in regression-based estimation. Doubly-robust value search methods have also been proposed and more efficient versions of the IPW estimator have been developed by adding augmentation terms to $\hat{V}(d_\psi)$ (Wahed & Tsiatis, 2004; B. Zhang et al., 2012b, 2013). Dynamic marginal structural modeling (Orellana et al., 2010) is another approach to direct maximization. Value search estimation has been reformulated into a weighted classification problem (B. Zhang et al., 2012a; Y. Q. Zhao et al., 2012): identifying the optimal decision rule by maximizing (2.2) is equivalent to finding $d_{\psi^{\text{opt}}}(\boldsymbol{h_1})$ that minimizes $\mathbb{E}\left[\frac{Y\mathbb{I}(A_1 \neq d_\psi(\boldsymbol{h_1}))}{P(A_1=a_1|\boldsymbol{h_1};\hat{\boldsymbol{\alpha}})}\right]$. The latter formulation can be viewed as solving a weighted misclassification problem and (non-parametric) statistical learning methods for classification problems can be used to find the optimal decision rule (Y. Zhang et al., 2015; Zhou et al., 2017). For example, backward outcome weighted learning (Y. Q. Zhao et al., 2015) is an important method that applies to multi-stage DTRs, using support vector machines (SVM) to solve the minimization problem.

Value search methods focus on estimation and inference about the value of a regime instead of inferences about the parameters $\boldsymbol{\psi}$ and yield decision rules that are difficult to use in clinical practice. Their performance is often evaluated by assessing how close the estimated value is to the true optimal value but disregards inferences about the parameters in the decision rule. Value search methods that adopt a classification perspective offer more flexibility than regression-based estimation and non-classification value search estimation in terms of specifying a form for the decision rules (B. Zhang et al., 2012b) at the cost of providing uninterpretable decision rules and a lack of tools for inference (Laber & Zhao, 2015; Y. Zhang et al., 2015; Y. Q. Zhao & Laber, 2014a; Y. Q. Zhao et al., 2015). The main purpose of value search estimation is thus often to generate hypotheses and inform future research (Y. Q. Zhao et al., 2015), for example, in exploratory analyses of clinical trials.

Regression-based estimation is the focus of this thesis, with dWOLS being the starting point for our contribution. DWOLS unites the strengths of two well-established regression-based methods, Q-learning and G-estimation, into a theoretically robust yet accessible framework to estimate an optimal DTR for uncensored continuous outcomes. The method is thoroughly described in Section 2.1.2 after an introduction of Q-learning and G-estimation in the next section.

2.1.1 Q-learning and G-estimation

An overview of Q-learning and G-estimation along with important DTR concepts is given in this section. To avoid unnecessarily involved notation, we continue to focus on DTRs with two stages although both methods can accommodate a larger number of stages.

Q-learning (Murphy, 2003; Watkins & Dayan, 1992) offers a simple framework for estimating an optimal DTR via a series of regressions. It relies on the principle of backward induction in that the estimation is carried out backward in time, starting with the optimization of the treatment in the last stage and moving recursively through stages to optimize each previous treatment. At each stage, Q-learning views the outcome Y_j as a reward from receiving treatment A_j given the history $\boldsymbol{H_j}$ and the objective is to identify the treatment option that yields the largest expected reward. Because Q-learning works backward in time, the optimal stage j treatment is that which maximizes a pseudo-outcome defined as the stage j reward plus the sum of future rewards had future treatments been optimal.

For a two-stage DTR, Q-learning defines the Q-functions

$$Q_2(\boldsymbol{h_2}, a_2) = \mathbb{E}[Y_2|\boldsymbol{h_2}, a_2]$$
$$Q_1(\boldsymbol{h_1}, a_1) = \mathbb{E}[Y_1 + \max_{a_2 \in \mathcal{A}_2} Q_2(\boldsymbol{h_2}, a_2)|\boldsymbol{h_1}, a_1]$$

and the optimal treatment in stage j is $a_j^{\text{opt}} = \arg\max_{a_j \in \mathcal{A}_j} Q_j(\boldsymbol{h_j}, a_j)$. Q_1 models the stage 1

pseudo-outcome $\tilde{Y}_1 = Y_1 + \max_{a_2 \in \mathcal{A}_2} Q_2(\boldsymbol{h_2}, a_2)$ which represents the (counterfactual) overall outcome had an individual received his optimal treatment in stage 2. The second term, $\max_{a_2 \in \mathcal{A}_2} Q_2(\boldsymbol{h_2}, a_2)$, can be decomposed as

$$\max_{a_2 \in \mathcal{A}_2} Q_2(\boldsymbol{h_2}, a_2) = \mathbb{E}[Y_2 | \boldsymbol{h_2}, a_2] + \mu_2(\boldsymbol{h_2}, a_2) = \mathbb{E}[Y_2 | \boldsymbol{h_2}, a_2] + \mathbb{E}[Y_2^{a_1, a_2^{\text{opt}}} - Y_2^{a_1, a_2} | \boldsymbol{h_2}, a_2].$$

The quantity $\mu_j(\boldsymbol{h_j}, a_j)$, called the regret (Murphy, 2003), is always non-negative and represents the expected loss from receiving treatment a_j in stage j instead of the optimal treatment a_j^{opt}, assuming optimal treatments are received thereafter. This formulation provides an alternative implementation of Q-learning where one would minimize the regrets instead of maximizing the Q-functions (Murphy, 2003).

Q-learning relies on a series of regressions to estimate the optimal DTR. A common parametrization of the Q-functions is a linear function

$$Q_j(\boldsymbol{h_j}, a_j; \boldsymbol{\beta_j}, \boldsymbol{\psi_j}) = \boldsymbol{\beta_j^T h_j} + a_j \boldsymbol{\psi_j^T h_j}, \tag{2.3}$$

which can be fitted with standard forms of regression such as ordinary least squares (OLS). With OLS and the linear model (2.3), the algorithm follows the three steps below:

1. Estimate the stage 2 parameters $(\boldsymbol{\beta_2}, \boldsymbol{\psi_2})$ by solving

$$(\hat{\boldsymbol{\beta}}_{\boldsymbol{2}}, \hat{\boldsymbol{\psi}}_{\boldsymbol{2}}) = \underset{(\boldsymbol{\beta_2}, \boldsymbol{\psi_2})}{\arg\min} \frac{1}{n} \sum \left(Y_2 - Q_2(\boldsymbol{h_2}, a_2; \boldsymbol{\beta_2}, \boldsymbol{\psi_2})\right)^2$$

   and derive the optimal stage 2 treatment $a_2^{\text{opt}} = \mathbb{I}(\hat{\boldsymbol{\psi}}_{\boldsymbol{2}}^{\boldsymbol{T}} \boldsymbol{h_2} > 0)$.

2. Construct the stage 1 pseudo-outcome $\tilde{Y}_1 = Y_1 + \hat{\boldsymbol{\beta}}_{\boldsymbol{2}}^{\boldsymbol{T}} \boldsymbol{h_2} + |\hat{\boldsymbol{\psi}}_{\boldsymbol{2}}^{\boldsymbol{T}} \boldsymbol{h_2}|$.

3. Estimate the stage 1 parameters $(\boldsymbol{\beta_1}, \boldsymbol{\psi_1})$ by solving

$$(\hat{\boldsymbol{\beta}}_{\boldsymbol{1}}, \hat{\boldsymbol{\psi}}_{\boldsymbol{1}}) = \underset{(\boldsymbol{\beta_1}, \boldsymbol{\psi_1})}{\arg\min} \frac{1}{n} \sum \left(\tilde{Y}_1 - Q_1(\boldsymbol{h_1}, a_1; \boldsymbol{\beta_1}, \boldsymbol{\psi_1})\right)^2$$

and derive the optimal stage 1 treatment $a_1^{\text{opt}} = \mathbb{I}(\hat{\boldsymbol{\psi}}_{\mathbf{1}}^{\boldsymbol{T}} \boldsymbol{h_1} > 0)$.

The parametrization in (2.3) defines a restricted class of regimes of the form $\mathbb{I}(\boldsymbol{\psi}_{\boldsymbol{j}}^{\boldsymbol{T}} \boldsymbol{h_j} > 0)$. This class of regimes yields decision rules that are easy to understand although they may not always be realistic (Moodie et al., 2014; Y. Zhang et al., 2015). With Q-learning, the Q-functions must be correctly specified to consistently estimate the parameters $\boldsymbol{\psi}$.

G-estimation (Robins, 2004) is a more robust alternative to Q-learning. It also uses backward induction to find an optimal DTR across multiple stages. Unlike Q-learning, it offers robustness against misspecification of the outcome model at the cost of additional modeling steps and increased complexity. G-estimation requires solving a series of estimating equations expressed in terms of a model for the probability of treatment and a model for the outcome, decomposed into a treatment-free component and a treatment component called the blip function.

The blip function expresses the difference in expected counterfactual outcomes of subjects receiving different levels of treatment. In a two-stage DTR, the blip functions are

$$\gamma_1(a_1, \boldsymbol{h_1}) = \mathbb{E}[Y^{a_1, a_2^{\text{opt}}} | \boldsymbol{h_1}] - \mathbb{E}[Y^{0, a_2^{\text{opt}}} | \boldsymbol{h_1}] \text{ and } \gamma_2(a_2, \boldsymbol{h_2}) = \mathbb{E}[Y^{a_1, a_2} | \boldsymbol{h_2}] - \mathbb{E}[Y^{a_1, 0} | \boldsymbol{h_2}].$$

At each stage, $\gamma_j(a_j, \boldsymbol{h_j})$ represents the difference between the expected outcome of an individual who received treatment a_j and the expected outcome of the same individual had he received some reference treatment $a_j = 0$, assuming that the individual goes on to receive optimal treatment in subsequent stages. In the last stage (here, the second stage), there is no subsequent treatment so the blip is only a difference in expected outcomes for individuals receiving treatment a_2 or reference treatment $a_2 = 0$. The blip functions must satisfy $\gamma_j(0, \boldsymbol{h_j}) = 0$. With a dichotomous treatment coded as $\{0, 1\}$, it necessarily takes the form $\gamma_j(a_j, \boldsymbol{h_j}) = a_j g(\boldsymbol{h_j})$ with $g(\cdot)$ an arbitrary function of the history $\boldsymbol{h_j}$. The blip function is central in the estimation of an optimal DTR as it is only through the blip that the treatment affects the outcome. The optimal stage j treatment is then estimated as

$a_j^{\text{opt}} = \arg\max_{a_j \in \mathcal{A}_j} \gamma_j(a_j, \boldsymbol{h_j})$. A linear function $\gamma_j(a_j, \boldsymbol{h_j}; \boldsymbol{\psi_j}) = a_j \boldsymbol{\psi_j^T h_j}$ is a common choice of parametrization. This parametrization defines a restricted class of regimes also of the form $\mathbb{I}(\boldsymbol{\psi_j^T h_j} > 0)$. There is a one-to-one correspondence between blip functions and the regret functions: $\mu_j(a_j, \boldsymbol{h_j}) = \gamma_j(a_j^{\text{opt}}, \boldsymbol{h_j}) - \gamma_j(a_j, \boldsymbol{h_j})$.

G-estimation further relies on functions $G_j(\boldsymbol{\psi_j})$ defined as $G_1(\boldsymbol{\psi_1}) = Y_1 - \gamma_1(a_1, \boldsymbol{h_1}; \boldsymbol{\psi_1}) + Y_2 + \mu_2(a_2, \boldsymbol{h_2}; \boldsymbol{\psi_2})$ in the first stage and $G_2(\boldsymbol{\psi_2}) = Y_2 - \gamma_2(a_2, \boldsymbol{h_2}; \boldsymbol{\psi_2})$ in the second stage. At each stage j, the G-function is the sum of observed outcomes from stage j onwards with the effect of treatment a_j removed and, if applicable, the expected loss due to receiving suboptimal treatments in subsequent stages added. The G-function represents the treatment-free component of the outcome and depends on the unknown blip parameters $\boldsymbol{\psi}$. In a two-stage DTR, G-estimation takes the following steps:

1. Propose a model for the blip function $\gamma_2(a_2, \boldsymbol{h_2}; \boldsymbol{\psi_2})$ and set $S_2(A_2) = \frac{\partial}{\partial \boldsymbol{\psi_2}} \gamma_2$.
2. Propose a model for the treatment-free component $\mathbb{E}[G_2(\boldsymbol{\psi_2})|\boldsymbol{h_2}; \boldsymbol{\beta_2}]$ and use the data to estimate $\hat{\boldsymbol{\beta}}_{\boldsymbol{2}}(\boldsymbol{\psi_2})$.
3. Propose a treatment model $\mathbb{E}[A_2|\boldsymbol{h_2}; \boldsymbol{\alpha_2}]$ and use the data to estimate $\hat{\boldsymbol{\alpha}}_{\boldsymbol{2}}$.
4. Obtain estimates $\hat{\boldsymbol{\psi}}_{\boldsymbol{2}}$ by solving the estimating equations

$$U_2^{\text{gest}}(\boldsymbol{\psi_2}, \hat{\boldsymbol{\beta}}_{\boldsymbol{2}}(\boldsymbol{\psi_2}); \hat{\boldsymbol{\alpha}}_{\boldsymbol{2}}) = \sum_{i=1}^{n} (S_2(A_2) - \mathbb{E}[A_2|\boldsymbol{h_2}; \hat{\boldsymbol{\alpha}}_{\boldsymbol{2}}]) \Big(G_2(\boldsymbol{\psi_2}) - \mathbb{E}[G_2(\boldsymbol{\psi_2})|\boldsymbol{h_2}; \hat{\boldsymbol{\beta}}_{\boldsymbol{2}}(\boldsymbol{\psi_2})] \Big) = 0.$$

5. Repeat steps 1–3 with $\gamma_1(a_1, \boldsymbol{h_1}; \boldsymbol{\psi_1})$, $G_1(\boldsymbol{\psi_1})$ and $\mathbb{E}[A_1|\boldsymbol{h_1}; \boldsymbol{\alpha_1}]$, respectively.
6. Obtain estimates $\hat{\boldsymbol{\psi}}_{\boldsymbol{1}}$ by solving the estimating equations

$$U_1^{\text{gest}}(\boldsymbol{\psi_1}, \hat{\boldsymbol{\beta}}_{\boldsymbol{1}}(\boldsymbol{\psi_1}); \hat{\boldsymbol{\alpha}}_{\boldsymbol{1}}) = \sum_{i=1}^{n} (S_1(A_1) - \mathbb{E}[A_1|\boldsymbol{h_1}; \hat{\boldsymbol{\alpha}}_{\boldsymbol{1}}]) \Big(G_1(\boldsymbol{\psi_1}) - \mathbb{E}[G_1(\boldsymbol{\psi_1})|\boldsymbol{h_1}; \hat{\boldsymbol{\beta}}_{\boldsymbol{1}}(\boldsymbol{\psi_1})] \Big) = 0.$$

G-estimation can accommodate nonlinear specification of the G-functions by modeling $g(\mathbb{E}[G_j(\boldsymbol{\psi_j})|\boldsymbol{h_j}; \boldsymbol{\beta_j}])$ linearly with $g(\cdot)$ a link function. For example, the identity link yields the linear model $\mathbb{E}[G_j(\boldsymbol{\psi_j})|\boldsymbol{h_j}; \boldsymbol{\beta_j}] = \boldsymbol{\beta_j^T h_j}$. G-estimation is doubly-robust as it yields consistent estimators of $\boldsymbol{\psi}$ provided that either $\mathbb{E}[G_j(\boldsymbol{\psi_j})|\boldsymbol{h_j}]$ or $\mathbb{E}[A_j|\boldsymbol{h_j}]$ is correctly modeled.

Despite G-estimation having several theoretical advantages over competing regression-based or value search methods, it is not routinely used in practice (Vansteelandt & Joffe, 2014). This may be explained by the fact that discussions around G-estimation often focus on theory rather than applications or demonstrative case studies. Another explanation may be that G-estimation can be computationally intensive and off-the-shelf software for its application is still lacking (Vansteelandt & Joffe, 2014). Only recently have efforts been made to render G-estimation more accessible thanks to the derivation of a simplified theory and to the development of the `R` package `DTRreg` (Wallace et al., 2014, 2017c).

2.1.2 Dynamic Weighted Ordinary Least Squares

Q-learning and G-estimation carry their respective strengths and limitations. The implementation of Q-learning is simple via standard regressions but the method lacks robustness to misspecification of the outcome model. While G-estimation offers the double-robustness property, it is harder to translate into practice. DWOLS unites the strengths of the two methods into one simple yet robust framework, introduced below.

DWOLS borrows from the framework set by Q-learning except that it relies on *weighted* OLS and thus additionally requires constructing and estimating weights. The steps for estimating a two-stage optimal DTR are:

1. Propose a model for treatment in the second stage $\mathbb{E}[A_2|\boldsymbol{h_2};\boldsymbol{\alpha_2}]$ and use the data to estimate $\hat{\boldsymbol{\alpha}}_{\boldsymbol{2}}$.

2. Choose weights $w_2(a_2, \boldsymbol{h_2})$ that satisfy the balancing property

$$\pi(\boldsymbol{h_2})w_2(1, \boldsymbol{h_2}) = (1 - \pi(\boldsymbol{h_2}))w_2(0, \boldsymbol{h_2}) \tag{2.4}$$

   where $\pi(\boldsymbol{h_2}) = \mathbb{E}[A_2|\boldsymbol{h_2}]$. Calculate the weights for each individual using $\hat{\boldsymbol{\alpha}}_{\boldsymbol{2}}$.

3. Propose a model for the expected outcome $\mathbb{E}[Y|\boldsymbol{h_2}, a_2; \boldsymbol{\beta_2}, \boldsymbol{\psi_2}] = f_2(\boldsymbol{h_2}; \boldsymbol{\beta_2}) + a_2\boldsymbol{\psi_2^T h_2}$ and estimate $(\hat{\boldsymbol{\beta}}_{\boldsymbol{2}}, \hat{\boldsymbol{\psi}}_{\boldsymbol{2}})$ by solving

$$U_2(\boldsymbol{\psi_2}, \boldsymbol{\beta_2}) = \frac{1}{n}\sum \hat{w}_2 \begin{bmatrix} \boldsymbol{h_2} \\ a_2\boldsymbol{h_2} \end{bmatrix} \left(Y_2 - f_2(\boldsymbol{h_2}; \boldsymbol{\beta_2}) - a_2\boldsymbol{\psi_2^T h_2}\right) = 0.$$

Derive the stage 2 optimal decision rule as $a_2^{\text{opt}} = \mathbb{I}(\hat{\boldsymbol{\psi}}_{\boldsymbol{2}}^{\boldsymbol{T}}\boldsymbol{h_2} > 0)$.

4. Construct the pseudo-outcome using $\hat{\boldsymbol{\psi}}_{\boldsymbol{2}}$ as

$$\tilde{Y} = Y + (a_2^{\text{opt}} - a_2)\hat{\boldsymbol{\psi}}_{\boldsymbol{2}}^{\boldsymbol{T}}\boldsymbol{h_2}. \tag{2.5}$$

5. Repeat steps 1 and 2 for the stage 1 treatment model $\mathbb{E}[A_1|\boldsymbol{h_1}; \boldsymbol{\alpha_1}]$ and weights $w_1(a_1, \boldsymbol{h_1}; \hat{\boldsymbol{\alpha}}_{\boldsymbol{1}})$.

6. Propose a model for the expected pseudo-outcome $\mathbb{E}[\tilde{Y}|\boldsymbol{h_1}, a_1; \boldsymbol{\beta_1}, \boldsymbol{\psi_1}] = f_1(\boldsymbol{h_1}; \boldsymbol{\beta_1}) + a_1\boldsymbol{\psi_1^T h_1}$ and estimate $(\hat{\boldsymbol{\beta}}_{\boldsymbol{1}}, \hat{\boldsymbol{\psi}}_{\boldsymbol{1}})$ by solving

$$U_1(\boldsymbol{\psi_1}, \boldsymbol{\beta_1}; \hat{\boldsymbol{\psi}}_{\boldsymbol{2}}) = \frac{1}{n}\sum \hat{w}_1 \begin{bmatrix} \boldsymbol{h_1} \\ a_1\boldsymbol{h_1} \end{bmatrix} \left(\tilde{Y} - f_1(\boldsymbol{h_1}; \boldsymbol{\beta_1}) - a_1\boldsymbol{\psi_1^T h_1}\right) = 0.$$

Derive the stage 1 optimal decision rule as $a_1^{\text{opt}} = \mathbb{I}(\hat{\boldsymbol{\psi}}_{\boldsymbol{1}}^{\boldsymbol{T}}\boldsymbol{h_1} > 0)$.

At each stage j, the dWOLS algorithm models the (pseudo-)outcome as a function of a term that does not depend on treatment A_j, the treatment-free component $f_j(\boldsymbol{h_j}; \boldsymbol{\beta_j})$, and a term that depends on the treatment, the blip component $a_j\boldsymbol{\psi_j^T h_j}$ (the two components do not necessarily depend on the same subset of the history $\boldsymbol{h_j}$). This decomposition is akin to G-estimation which also separates the outcome model into the same two components. The parameters $\boldsymbol{\beta_j}$ and $\boldsymbol{\psi_j}$ are respectively called the treatment-free and blip parameters and the covariates $\boldsymbol{h_j}$ in the blip are called tailoring variables. The blip parameters are the

focus of the estimation because it is only through the blip that the optimal decision rule is derived as $a_j^{\text{opt}} = \arg\max_{a_j} \gamma_j(a_j, \boldsymbol{h_j}; \hat{\boldsymbol{\psi}}_{\boldsymbol{j}})$. DWOLS constructs a pseudo-outcome in the first stage which represents the (counterfactual) outcome had the optimal stage 2 treatment been received. DWOLS takes advantage of the fact that this counterfactual is observed for individuals who have indeed received their estimated optimal stage 2 treatment and thus estimates the pseudo-outcome (2.5) only for individuals who have not received it. This places less burden on predictions as opposed to Q-learning which estimates a pseudo-outcome for all individuals, regardless if they received or not their optimal treatment (see step 2 of the Q-learning algorithm), and requires correct specification of the outcome model to obtain such predictions.

Like G-estimation, dWOLS is doubly-robust as it yields consistent estimators of the blip parameters if one or both the treatment-free and treatment models are correctly specified under assumptions (A1), (A2) and positivity of the treatment $P(A_j = a_j|\boldsymbol{h_j}) > 0$ for $a_j \in \mathcal{A}_j$. The double-robustness property is obtained using a weighting argument. Theorem 1 in Wallace & Moodie (2015) states that, under assumptions (A1) and (A2) and assuming that $\mathbb{E}[Y|\boldsymbol{h}; \boldsymbol{\beta}, \boldsymbol{\psi}] = f(\boldsymbol{h}; \boldsymbol{\beta}) + a\boldsymbol{\psi}^{\boldsymbol{T}}\boldsymbol{h}$ for some function f, a weighted OLS of y on $(\boldsymbol{h}, a\boldsymbol{h})$ yields consistent estimators of $\boldsymbol{\psi}$ if the weights satisfy the balancing property (2.4). The balancing property defines an entire family of weights that remove any confounding effect of treatment in a weighted OLS. For example, IPTW $w(a, \boldsymbol{h}) = [P(A = a|\boldsymbol{h})]^{-1}$ satisfy (2.4) since

$$\pi(\boldsymbol{h})w(1, \boldsymbol{h}) = \pi(\boldsymbol{h})/\pi(\boldsymbol{h}) = (1 - \pi(\boldsymbol{h}))/(1 - \pi(\boldsymbol{h})) = (1 - \pi(\boldsymbol{h}))w(0, \boldsymbol{h}).$$

On the one hand, if the treatment model is correctly specified and the weights satisfy (2.4), then any confounding between the treatment and outcome is removed and the dWOLS algorithm yields consistent blip estimators. On the other hand, if the treatment-free model is correctly specified, the estimating functions U_1 and U_2 have expectation zero and the blip estimators are consistent.

Wallace & Moodie (2015) propose dWOLS weights of the form $w(a, \boldsymbol{h}) = |a - \mathbb{E}[A|\boldsymbol{h}]|$ which weigh each individual proportionally to the probability of receiving the opposite treatment. They showed that, when treatment is binary, the estimators $(\hat{\boldsymbol{\beta}}_{\boldsymbol{j}}, \hat{\boldsymbol{\psi}}_{\boldsymbol{j}})$ obtained with dWOLS weights are also a solution to the estimating equations U_1^{gest} and U_2^{gest} in G-estimation. A simulation study suggested that the dWOLS weights yield the most efficient estimators of the blip parameters when compared to three alternatives, including IPTW (Wallace & Moodie, 2015). Li et al. (2018) extensively discuss weights that satisfy (2.4), which they refer to as balancing weights because they balance the distributions of covariates $\boldsymbol{h}$ between the two treatment groups. They refer to weights of the form $w(a, \boldsymbol{h}) = |a - \mathbb{E}[A|\boldsymbol{h}]|$ as overlap weights [1] and show that the overlap weights indeed minimize the asymptotic variance of the weighted average treatment effect. Balancing weights allow defining different target populations through different choices of weights. For example, the overlap weights place more emphasis on individuals with propensity score close to 1/2 relative to individuals with propensity score closer to 0 or 1. This choice of weights yields a target population of individuals in equipoise between treatments i.e. with a combination of characteristics such that they could be assigned to either treatment with approximately equal probability. More research around this target population may be needed as it represents patients for whom the optimal treatment choice is unclear (Li et al., 2018).

The theory developed by Robins (2004) allows deriving an estimator for the asymptotic variance of $\hat{\boldsymbol{\psi}}$ which must adjust for the estimation of the treatment model and, if applicable, for the substitution (i.e. "plug-in") estimators in the pseudo-outcome. The following derivations hold under standard regularity conditions that allow interchanging sums and integrals and calculating first and second derivatives of the estimating functions U_1 and U_2. In the second stage, adjusted estimating functions $U_{\text{adj},2}$ are defined by performing a first-order Taylor

[1] We interchangeably use overlap weights and dWOLS weights to refer to weights of the form $w(a, \boldsymbol{h}) = |a - \mathbb{E}[A|\boldsymbol{h}]|$

expansion of U_2 about the limiting values of the nuisance parameters $\boldsymbol{\alpha_2}$, leading to

$$U_{\text{adj},2}(\boldsymbol{\beta_2}, \boldsymbol{\psi_2}) = U_2(\boldsymbol{\beta_2}, \boldsymbol{\psi_2}; \hat{\boldsymbol{\alpha}}_{\mathbf{2}}) - \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\alpha_2}} U_2(\boldsymbol{\beta_2}, \boldsymbol{\psi_2}; \hat{\boldsymbol{\alpha}}_{\mathbf{2}})\right] \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\alpha_2}} s_\alpha(\hat{\boldsymbol{\alpha}}_{\mathbf{2}})\right]^{-1} s_\alpha(\hat{\boldsymbol{\alpha}}_{\mathbf{2}}) \quad (2.6)$$

where $s_\alpha(\hat{\boldsymbol{\alpha}}_{\mathbf{2}})$ is the score function of the treatment model $\mathbb{E}[A_2|\boldsymbol{h_2}; \boldsymbol{\alpha_2}]$ evaluated at $\hat{\boldsymbol{\alpha}}_{\mathbf{2}}$. Using the delta method, the asymptotic variance of $\hat{\boldsymbol{\psi}}_{\mathbf{2}}$ is the lower-right square submatrix of dimension $\dim(\hat{\boldsymbol{\psi}}_{\mathbf{2}})$ of

$$\mathbb{E}\left[\left\{\left(\mathbb{E}\left[\frac{\partial}{\partial(\boldsymbol{\beta_2}, \boldsymbol{\psi_2})} U_{\text{adj},2}(\boldsymbol{\beta_2}, \boldsymbol{\psi_2})\right]\right)^{-1} U_{\text{adj},2}(\boldsymbol{\beta_2}, \boldsymbol{\psi_2})\right\}^{\otimes 2}\right] \quad (2.7)$$

where $\mathbb{E}[X^{\otimes 2}] = \mathbb{E}[XX^T]$. The variance of the first-stage estimators $\hat{\boldsymbol{\psi}}_{\mathbf{1}}$ is derived in a similar fashion except that the adjusted estimating functions further account for the plug-in estimators $\hat{\boldsymbol{\psi}}_{\mathbf{2}}$ used to construct the pseudo-outcome (2.5) as

$$\begin{aligned} U^{\epsilon}_{\text{adj},1}(\boldsymbol{\beta_1}, \boldsymbol{\psi_1}) =& U_{\text{adj},1}(\boldsymbol{\beta_1}, \boldsymbol{\psi_1}) - \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\psi_2}} U_1(\boldsymbol{\beta_1}, \boldsymbol{\psi_1}; \hat{\boldsymbol{\alpha}}_{\mathbf{1}}, \hat{\boldsymbol{\psi}}_{\mathbf{2}})\right] \left(\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\psi_2}} U_{\text{adj},2}(\boldsymbol{\beta_2}, \boldsymbol{\psi_2}; \hat{\boldsymbol{\alpha}}_{\mathbf{2}})\right]\right)^{-1} \\ & \times U_{\text{adj},2}(\boldsymbol{\beta_2}, \boldsymbol{\psi_2}; \hat{\boldsymbol{\alpha}}_{\mathbf{2}}) \end{aligned}$$

where $U_{\text{adj},1}$ is derived as $U_{\text{adj},2}$ but with the stage 1 models and parameters. We note that, at each stage, the variance of $(\hat{\boldsymbol{\beta}}_{\boldsymbol{j}}, \hat{\boldsymbol{\psi}}_{\boldsymbol{j}})$ depends on the choice of weights. Wallace & Moodie (2015) provide sample derivations with the dWOLS weights.

2.2 Non-regular Inference

In the previous section, we introduced important concepts and foundational methods for estimating an optimal DTR with continuous uncensored outcomes. DWOLS faces an important inferential challenge when estimating an optimal DTR with multiple stages: in all but the last stage, the blip estimators have non-regular limiting distributions (Hirano &

Porter, 2012; Laber et al., 2010; Robins, 2004). Non-regularity may negatively affect the performance of confidence intervals for the blip parameters. The problem of non-regularity has already been studied in the DTR literature as it also impacts related regression-based methods such as Q-learning and G-estimation. In this section, we explain the theoretical problem of non-regularity and its practical implication and review solutions that have been proposed in the literature to deal with non-regular inference.

2.2.1 Non-regular Estimators

In a two-stage DTR, the first-stage blip estimators obtained with dWOLS are non-regular because they depend on the pseudo-outcome which in turn depends on a non-differentiable (non-smooth) function of a plug-in estimator. Recall the form of the pseudo-outcome in dWOLS, $\tilde{Y}_1 = Y + (a_2^{\text{opt}} - a_2)\hat{\boldsymbol{\psi}}_{\mathbf{2}}^{\boldsymbol{T}}\boldsymbol{h_2}$, where a_2^{opt} is the indicator $\mathbb{I}(\hat{\boldsymbol{\psi}}_{\mathbf{2}}^{\boldsymbol{T}}\boldsymbol{h_2} > 0)$ that depends on the second stage blip estimators. The indicator function is non-smooth and non-differentiable at 0. Because the first-stage estimating functions depend on $\hat{\boldsymbol{\psi}}_{\mathbf{2}}$ plugged into the indicator function, the first-stage blip estimators $\hat{\boldsymbol{\psi}}_{\mathbf{1}}$ also depend on that non-smooth function. Consequently, standard asymptotic approximations to the sampling distribution of the estimators $\hat{\boldsymbol{\psi}}_{\mathbf{1}}$ cannot be used directly. More generally, in a DTR with J stages, the blip estimators in all but the last stage are non-regular because they depend on a pseudo-outcome defined with plug-in estimators of future blip parameters.

To assist with the understanding of non-regularity, we present a simple theory example used by Robins (2004) and revisited by others (Chakraborty et al., 2010; Moodie & Richardson, 2010). Consider the function $(x)_+ = x\mathbb{I}(x > 0) = \max(0, x)$ non-differentiable at 0. Suppose we wish to estimate $(\mu)_+$ from a sample of n independent and identically distributed (i.i.d.) observations X_i drawn from a $N(\mu, 1)$. The maximum likelihood estimator (MLE) of $(\mu)_+$ is $(\bar{X}_n)_+$ obtained by plugging the sample mean $\bar{X}_n$ into the function $(x)_+$. When $\mu \neq 0$, the limiting distribution of $\sqrt{n}((\bar{X}_n)_+ - (\mu)_+)$ is a standard Normal distribution left-truncated at

0 and $(\bar{X}_n)_+$ is asymptotically unbiased. An estimator is said to be asymptotically unbiased if $\sqrt{n}(\hat{\psi}-\psi)$ converges in distribution to a distribution F with $\mathbb{E}(F) = 0$, and asymptotically biased if $\mathbb{E}(F) \neq 0$. However, at the point of non-differentiability $\mu = 0$, it can be shown that $\sqrt{n}((\bar{X}_n)_+ - (\mu)_+)$ converges to a $(\frac{1}{2}, \frac{1}{2})$ mixture of a left-truncated standard Normal and a degenerate random variable with point mass at 0 and that the resulting mixture distribution has expectation $1/\sqrt{2\pi}$. Therefore, the estimator $(\bar{X}_n)_+$ is asymptotically normal and unbiased for $\mu \neq 0$ but asymptotically non-normal and biased for $\mu = 0$. Thus, we say that the estimator $(\bar{X}_n)_+$ is *non-regular* or that it has a *non-regular limiting distribution*. It is also useful to understand non-regularity as describing an estimator whose asymptotic distribution does not converge uniformly over the parameter space. In the previous example, the estimator $(\bar{X}_n)_+$ converges to different asymptotic distributions depending on the true value of the parameter μ, therefore it does not converge uniformly over the parameter space $\mu \in \mathbb{R}$.

The dWOLS estimators in all but the last stage are non-regular but it is only for *exceptional laws* that inference for the blip parameters is affected by non-regularity. Exceptional laws are defined as data generating mechanisms for which, at stage j, there is a positive probability that the true optimal decision for some individuals is not unique (Robins, 2004). In the case of a linear blip function $\gamma_j(a_j, \boldsymbol{h_j}; \boldsymbol{\psi_j}) = a_j \boldsymbol{\psi_j^T h_j}$, the optimal decision is non-unique when the linear combination $\boldsymbol{\psi_j^T h_j}$ is exactly zero, indicating that treatments $a_j = 0$ and $a_j = 1$ are equally good. This corresponds to the point of non-differentiability of the function $\mathbb{I}(\boldsymbol{\psi_j^T h_j} > 0)$ which defines the j^{th} optimal decision rule. The j^{th} optimal decision rule appears in the pseudo-outcomes defined in stages 1 to $j-1$ such that characterizing a law as exceptional in stage j affects the inferences for the blip parameters in the previous stages. Given the form one assumes for the blip model, two factors can lead to an exceptional law: (i) the true value of the blip parameters $\boldsymbol{\psi_j}$ or (ii) the mechanism that generated the tailoring variables $\boldsymbol{h_j}$ in the treatment rule. For (i), if the true effect of the treatment a_j and its interactions with tailoring covariates is null, the probability of a non-unique optimal

decision $P(\boldsymbol{\psi}_j^{\boldsymbol{T}} \boldsymbol{H}_j = 0)$ is necessarily one, regardless of the distribution of $\boldsymbol{H}_j$. For (ii), if $\mathcal{H}_j$ contains only discrete variables, exceptional laws occur when, given the true parameters $\boldsymbol{\psi}_j$, the probability of observing $\boldsymbol{h}_j$ leading to $\boldsymbol{\psi}_j^{\boldsymbol{T}} \boldsymbol{h}_j = 0$ is greater than zero. For example, for a blip of the form $a_1(\psi_1 + \psi_2 x_1)$ with X_1 taking discrete values, then $P(\psi_1 + \psi_2 X_1 = 0)$ is positive if the probability of generating $X_1 = -\psi_1/\psi_2$ is non-zero and the law is exceptional. Non-exceptional laws characterize data generating mechanisms in which the true optimal decision is unique for all individuals with probability one.

The definition of exceptional laws does not allow identifying a law as exceptional in practice because the true parameter values and the mechanism that generated the history are both unknown. Robins (2004) suggests a practical solution to detect exceptional laws with G-estimation: first, estimate the blip parameters across all stages and derive Wald confidence intervals about each parameter using the variance calculation described at the end of Section 2.1.2. Second, at each stage j, calculate the proportion $\hat{p}_j$ of individuals for whom the optimal decision rule recommends both treatments a_j when considering all values in the confidence set for $\boldsymbol{\psi}_j$. If the proportion $\hat{p}_j$ is small, say less than 0.05, then the law at stage j is likely not exceptional and inferences for the blip parameters based on Wald confidence intervals for earlier stages can be trusted. Otherwise, the law is likely exceptional and confidence intervals for blip parameters in earlier stages are not reliable. Moodie & Richardson (2010) summarize the practical guidelines by pointing out that exceptional laws in the j^{th} stage only affect inferences in the previous stages and that the laws are likely exceptionals if at any stage the null hypothesis of no treatment effect is not rejected.

2.2.2 The Bootstrap

For exceptional laws, Wald confidence intervals based on the asymptotic variance (2.7) do not have the correct coverage. The bootstrap is often proposed as a solution to construct confidence intervals when estimating the variance of an estimator is complicated or impos-

sible (Efron, 1992a). Unfortunately, the bootstrap also fails to provide reliable confidence intervals for the parameters at exceptional laws (Shao, 1994). We review the bootstrap algorithm and explain why it fails in non-regular settings.

Consider the problem of constructing a confidence interval for a unidimensional parameter θ for which we have an estimator $\hat{\theta}_n$ based on a sample of n observations. The goal of inference is to learn about the distribution $F_n(t) = P(\sqrt{n}(\hat{\theta}_n - \theta) \leq t)$ in order to find the quantiles $t_{\alpha/2}$ and $t_{1-\alpha/2}$ that allow constructing a $(1-\alpha)\times 100\%$ confidence interval for θ as $C_n = \left[\hat{\theta}_n - \frac{t_{1-\alpha/2}}{\sqrt{n}}, \hat{\theta}_n - \frac{t_{\alpha/2}}{\sqrt{n}}\right]$. The confidence interval C_n has the correct coverage $P(C_n \subset \theta) = 1-\alpha$ as $n \to \infty$. When $F_n(t)$ does not have a well-defined form, the bootstrap can be used to approximate it. For this, it suffices to draw B samples of size n with replacement from the original sample, compute the estimator $\hat{\theta}_n^{(b)}$ in each b sample, $b = 1, \ldots, B$, and approximate the distribution $F_n(t)$ by $F_n^*(t) = \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\left[\sqrt{n}(\hat{\theta}_n^{(b)} - \hat{\theta}_n) \leq t\right]$. The quantiles $t_{\alpha/2}$ and $t_{1-\alpha/2}$ are approximated with the corresponding bootstrap quantiles $t^*_{\alpha/2}$ and $t^*_{1-\alpha/2}$ of $F_n^*(t)$ and an approximate confidence interval for θ is $C_n^* = \left[\hat{\theta}_n - \frac{t^*_{1-\alpha/2}}{\sqrt{n}}, \hat{\theta}_n - \frac{t^*_{\alpha/2}}{\sqrt{n}}\right]$ which has $P(C_n^* \subset \theta) \approx 1-\alpha$.

The validity of the bootstrap procedure is based on the following key result: $\sup_t |F_n^*(t) - F_n(t)| \to 0 \Rightarrow \mathbb{P}(C_n^* \subset \theta) \to 1-\alpha$ as $n \to \infty$ (Efron, 1992a). In words, this means that, as the sample size n increases, F_n^* becomes a better approximation of F_n which implies that the coverage of C_n^* approaches the desired probability. However, the condition above is not satisfied in some situations, including when the estimator $\hat{\theta}_n$ is non-smooth. Shao (1994) considers constructing a confidence interval for $|\mu|$ with the estimator $|\bar{X}_n|$ defined as a non-differentiable function of a plug-in estimator. When $\mu = 0$, the author shows that the distribution $\sqrt{n}(|\bar{X}_n^{(b)}| - |\bar{X}_n|)$ does not have a limit, thus the bootstrap estimator of the distribution of $|\bar{X}_n|$ is not consistent and the validity of the procedure is not guaranteed. Shao's example translates to our dWOLS parameters, where the blip estimators $\hat{\boldsymbol{\psi}}_{\boldsymbol{j}}$ in all but the last stage are defined with a non-smooth function of plug-in estimators. At exceptional

laws, the bootstrap estimator of the distribution of interest $P(\sqrt{n}(\hat{\boldsymbol{\psi}}_{\boldsymbol{j}} - \boldsymbol{\psi}_{\boldsymbol{j}}) \leq t)$ is not consistent and the derived confidence intervals may not have the nominal coverage.

2.2.3 The m-out-of-n Bootstrap and Other Solutions

Because DTR methods that suffer from non-regularity are widely used, several solutions have been proposed to alleviate its negative impact on inference, that is, unreliable confidence intervals about the blip parameters. We review solutions that could apply to dWOLS.

The m-out-of-n bootstrap has been proposed as an alternative to the standard bootstrap in cases where the bootstrap fails asymptotically, including for constructing confidence intervals with non-regular estimators (Bickel et al., 1997; Bretagnolle, 1983; Shao, 1994). The m-out-of-n bootstrap relies on the same algorithm as the standard bootstrap except that each resample has size $m < n$ where m must be defined as a function of n and must satisfy $m \to \infty$ and $m/n \to 0$ as $n \to \infty$. The estimator $\hat{\theta}_m^{(b)}$ is calculated in each resample of size m and the distribution of interest $F_n(t) = P(\sqrt{n}(\hat{\theta}_n - \theta) \leq t)$ is approximated by the corresponding m-out-of-n bootstrap distribution $F_m^*(t) = \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\left[\sqrt{m}(\hat{\theta}_m^{(b)} - \hat{\theta}_n) \leq t\right]$. The key idea is that, under the asymptotic conditions on m, the distribution $F_m^*(t)$ provides a better approximation for $F_n(t)$ than the distribution $F_n^*(t)$ based on the standard bootstrap with resamples of size n. Chakraborty et al. (2013) present a toy example to help understand why this is the case. Suppose we wish to estimate $|\mu|$ from a sample of n i.i.d. observations X_i drawn from a $N(\mu, 1)$. The MLE for $|\mu|$ is $|\bar{X}_n|$ which is non-regular as $\sqrt{n}(|\bar{X}_n| - |\mu|)$ converges to a standard normal distribution when $\mu \neq 0$ but to a χ_1 distribution when $\mu = 0$. Let $m = m(n)$ define a resample size that depends on the original sample size n. We visualize a random sample of size m drawn from the original n units by defining random variables $W_1, \ldots, W_n$ from a multinomial distribution with m trials and probability of success (i.e. being resampled) $1/n$. Within each bootstrap sample of size m, the bootstrap mean is $\bar{X}_m^{(b)} = \sum_{i=1}^{n} W_i X_i$ where the randomness in the mean now comes from the multinomial

weights, with the X_is held fixed. The sample mean is a regular estimator and it can be shown that, even with resample size $m < n$, the bootstrap distribution $\sqrt{m}(\bar{X}_m^{(b)} - \bar{X}_n)$ can be used to approximate $\sqrt{n}(\bar{X}_n - \mu)$ for $\mu \in \mathbb{R}$ (Bickel & Freedman, 1981). For the non-regular estimator $|\bar{X}_n|$, when $\mu \neq 0$, it can also be shown that $\sqrt{m}(|\bar{X}_m^{(b)}| - |\bar{X}_n|)$ converges to a standard normal distribution as n and m tend to ∞. At the point of non-differentiability $\mu = 0$, the distribution $\sqrt{m}(|\bar{X}_m^{(b)}| - |\bar{X}_n|)$ can be rewritten as $\left|\sqrt{m}(\bar{X}_m^{(b)} - \bar{X}_n) + \sqrt{\frac{m}{n}}\sqrt{n}\bar{X}_n\right| - \left|\sqrt{\frac{m}{n}}\sqrt{n}\bar{X}_n\right|$ which converges to a χ_1 distribution only when the asymptotic conditions on m are satisfied but does not converge when $m = n$.

The asymptotic conditions on m offer little guidance on how to choose m in practice. Bickel & Sakov (2008) suggest that m must not be "too large" or held fixed with respect to n. They propose an adaptive rule for choosing m based on the idea that, when considering a sequence of values for m, the bootstrap distributions should be "close" for m chosen in the "right range" but different when m is too large or fixed. Their adaptive rule requires specifying a metric to measure the distance between the bootstrap distributions from two different choices of m and to choose a sequence $m = f(q, n)$, where f defines positive integers smaller than n and q is a tuning parameter that controls the difference between successive m in the sequence. Chakraborty et al. (2013) adapt the rule proposed by Bickel & Sakov (2008) to the context of Q-learning. They propose to choose m that adapts to the degree of non-regularity in the data. The resample size m is defined as $m = n^{\frac{1+\alpha(1-\hat{p})}{1+\alpha}}$ where $\hat{p}$ is the proportion of individuals with non-unique optimal treatment discussed in Section 2.2.1 and α is a tuning parameter.

Solutions other than the m-out-of-n bootstrap have been proposed to construct valid confidence intervals for the non-regular DTR estimators. Chakraborty et al. (2010) propose hard- and soft-thresholding to reduce the asymptotic bias in Q-learning at exceptional laws. Both approaches involve redefining the pseudo-outcome by replacing the problematic non-smooth function with other functions. Hard-thresholding replaces the non-differentiable term $|\hat{\boldsymbol{\psi}}_{\boldsymbol{2}}^{\boldsymbol{T}}\boldsymbol{h_2}|$

in the Q-learning pseudo-outcome by $|\hat{\boldsymbol{\psi}}_{\mathbf{2}}^{\boldsymbol{T}}\boldsymbol{h_2}|\mathbb{I}\left(\hat{\boldsymbol{\psi}}_{\mathbf{2}}^{\boldsymbol{T}}\boldsymbol{h_2} > \lambda_i\right)$ where $\lambda_i > 0$ is a threshold for the i^{th} individual. This means that individuals with $|\hat{\boldsymbol{\psi}}_{\mathbf{2}}^{\boldsymbol{T}}\boldsymbol{h_2}|$ "close" to zero, which corresponds to the point of non-differentiability of the absolute value function where exceptional laws occur, have the term $|\hat{\boldsymbol{\psi}}_{\mathbf{2}}^{\boldsymbol{T}}\boldsymbol{h_2}|$ in the pseudo-outcome shrunk to zero. A hypothesis test for $\boldsymbol{\psi}_{\mathbf{2}}^{\boldsymbol{T}}\boldsymbol{h_2} = 0$ could be used to determine if the second-stage optimal treatment is likely to be non-unique for individual i and λ_i could be chosen accordingly. Moodie & Richardson (2010) propose a form of hard-thresholding, called zeroing instead of plugging in (ZIPI), in the context of G-estimation. Soft-thresholding also shrinks the pseudo-outcome of some individuals by replacing $|\hat{\boldsymbol{\psi}}_{\mathbf{2}}^{\boldsymbol{T}}\boldsymbol{h_2}|$ with $|\hat{\boldsymbol{\psi}}_{\mathbf{2}}^{\boldsymbol{T}}\boldsymbol{h_2}|\left(1 - \frac{\lambda_i}{|\hat{\boldsymbol{\psi}}_{\mathbf{2}}^{\boldsymbol{T}}\boldsymbol{h_2}|^2}\right)_+$ where $\lambda_i > 0$ is a (different) tuning parameter. While hard-thresholding only shrinks the pseudo-outcome of individuals who may have non-unique optimal treatment a_2, soft-thresholding shrinks the pseudo-outcomes of all individuals but shrinks more importantly that of individuals with $\hat{\boldsymbol{\psi}}_{\mathbf{2}}^{\boldsymbol{T}}\boldsymbol{h_2}$ close to zero. However, both methods still involve a non-differentiable function. Hard- and soft-thresholding are not supported with theoretical results (ZIPI is) and their performance in simulation studies shows conflicting evidence (Chakraborty et al., 2010; Laber et al., 2014b). Laber et al. (2014b) propose constructing adaptive confidence intervals for the blip parameters in Q-learning. They approximate the bounds of the confidence intervals for the non-regular estimators $\hat{\boldsymbol{\psi}}_{\mathbf{1}}$ by separating the contribution of individuals for whom $\boldsymbol{\psi}_{\mathbf{2}}^{\boldsymbol{T}}\boldsymbol{h_2}$ is likely or unlikely close to 0. As in hard-thresholding, this approach involves choosing a threshold to quantify the acceptable distance of $\hat{\boldsymbol{\psi}}_{\mathbf{2}}^{\boldsymbol{T}}\boldsymbol{h_2}$ from 0.

All methods presented in this section have been compared in simulation studies which typically consider three types of data generating mechanisms defining regular, near non-regular and fully non-regular settings in a two-stage DTR (Chakraborty et al., 2013, 2010; Fan et al., 2019; Laber et al., 2014b). Regular simulation scenarios are such that all individuals have a unique optimal treatment in the second stage i.e. $\boldsymbol{\psi}_{\mathbf{2}}^{\boldsymbol{T}}\boldsymbol{h_2}$ is "far" from the point of non-differentiablity for all individuals. In regular scenarios, the laws are not exceptional and inferences for the first-stage blip parameters are reliable. Non-regular simulation scenarios

are such that some individuals (not necessarily all) have a non-unique optimal stage 2 treatment. The laws are then exceptional and alternative methods for inferences in the first stage are necessary. Near non-regular scenarios are such that some or all individuals have $\boldsymbol{\psi_2^T h_2}$ close to but not exactly 0. The laws are not exceptional but the inferences in the first-stage might be affected by how close $\boldsymbol{\psi_2^T h_2}$ is to 0 and by the sample size. Confidence intervals based on hard- and soft-thresholding may under- or over-cover in non-regular or near non-regular scenarios (Fan et al., 2019; Laber et al., 2014b). Adaptive confidence intervals proposed by Laber et al. (2014b) are generally valid but can be conservative (Chakraborty et al., 2013; Fan et al., 2019). Chakraborty et al. (2013) found that their adaptive choice of m using the degree of non-regularity in the data is at least as good as the adaptive rule proposed by Bickel & Sakov (2008) in Q-learning.

2.3 DTR for Censored Data

Often, interest lies in estimating a DTR that optimizes the time until the occurrence of an event. For example, in T2D, a sequence of drug and lifestyle therapies tailored to individual patient characteristics aims to delay the development of diabetic complications, that is, to maximize the time until complications. Estimating an optimal DTR when the outcome of interest is time-to-event poses additional challenges which are discussed in this section. We first review basic concepts and methods in survival analysis (Kalbfleisch & Prentice, 2011). We then describe the main challenges intrinsic to the estimation of an optimal DTR when the outcome is subject to right-censoring and review methods that have been proposed in the literature.

2.3.1 Important Concepts in Survival Analysis

Survival analysis is concerned with the distribution of T, the survival time from a well-defined starting point until the occurrence of an event of interest, which we assumed continuous in this thesis. T may not be observed for all individuals, in which case we say that the individual is censored at time $C < T$. For example, a study comparing cancer treatments considers T defined as the time from initiation of cancer treatment until complete remission. Right-censoring may occur if a patient drops out of the study or is lost to follow-up before achieving complete remission or if a patient has not achieved complete remission by the end of the study period, both cases leading to observe C instead of T. The minimum between the survival and censoring time $Y = \min(T, C)$ is observed. The indicator $\Delta = \mathbb{I}(T < C)$ allows us to distinguish between those who experienced an event and those who were censored.

Several metrics can be used to characterize the distribution of T. The survival function $S(t) = P(T > t)$ gives the probability of surviving beyond time t. It is linked to the cumulative distribution function $F(t) = P(T \leq t)$ through the relationship $S(t) = 1 - F(t)$. The mean survival time is defined as $\mu = \int_0^\infty S(t)\mathrm{dt}$ and quantiles of the distribution of T, such as the median, can also be defined with $S(t)$. The hazard function (hazard rate), defined as $h(t) = \lim_{\Delta t \to \infty} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$, provides information about T where $h(t)\Delta t$ can be viewed as the "probability" that an individual who is still alive at time t experiences an event in the next instant Δt. A related quantity is the cumulative hazard function $H(t)$ defined by $\int_0^t h(u)\mathrm{du}$ and satisfying $S(t) = \exp[-H(t)]$. It is common to assume a parametric model for the survival time such as the Exponential, Weibull and Log-normal distributions. Parametric models are useful because they offer insights into the form of the survival and hazard functions and related quantities. Non-parametric methods to characterize functions of the survival time are also frequently used, for example, the Kaplan-Meier estimator of the survival function and the Nelson-Aalen estimator of the cumulative hazard function.

Beyond merely describing the distribution of T, interest often lies in learning about the effect

of some covariates $\boldsymbol{X}$ on T. There are two well-known approaches to the modeling of covariate effects on survival: the Cox model and the accelerated failure time (AFT) model. The Cox model (Cox, 1972) is a special case of multiplicative hazard rate models. It expresses the conditional hazard rate as the product of a baseline hazard rate $h_0(t)$ and a function of the covariates as $h(t|\boldsymbol{x}) = h_0(t)\exp(\boldsymbol{\beta}^{\boldsymbol{T}}\boldsymbol{x})$. This form of the conditional hazard model implies proportionality of the hazard rates over time of two individuals with different covariate values. Inferences in the Cox regression focus on the hazard ratio parameters $\boldsymbol{\beta}$ which have an interpretation on the hazard scale. Partial likelihood is typically used to estimate the Cox parameters which treats $h_0(t)$ as an infinite dimensional nuisance parameter and has the advantage of not needing to estimate this quantity. If necessary, $h_0(t)$ can be estimated parametrically or non-parametrically, for example, with the Breslow estimator (Breslow, 1974). The AFT model is an alternative to Cox regression. It assumes a linear model for the log-survival time as $\log(T) = \mu + \boldsymbol{\gamma}^{\boldsymbol{T}}\boldsymbol{x} + \sigma W$ with μ an intercept, σ a scale parameter and W an error term. The choice of a distribution for W implies a specific distribution for the survival time, for example, W following a standard normal distribution implies a Log-normal distribution for T. The effect of the covariates is to accelerate (or decelerate) the time to an event by a factor $\exp(-\boldsymbol{\gamma}^{\boldsymbol{T}}\boldsymbol{x})$. The interpretation of the AFT parameters $\boldsymbol{\gamma}$ is made directly on the log-time scale akin to the interpretation of linear regression parameters. Likelihood-based approaches that account for censoring can be used to estimate $\boldsymbol{\gamma}$ in a parametric AFT while rank-based estimation (Wei, 1992) is a common choice for semi-parametric AFT models where the error distribution is left unspecified. Regardless of the modeling choice, assumptions must be made about the relationship between the covariates, survival times and censoring times in order to learn about the effect of $\boldsymbol{X}$ on T. Non-informative censoring assumes that knowing the censoring time of an individual does not provide information about its survival time, meaning that the distributions of T and C provide no information on the value or distribution of the other. This assumption can be relaxed by adding that censoring is non-informative given a set of covariates.

2.3.2 DTRs and Censored Data: Additional Considerations

Finding an optimal DTR when the outcome of interest is survival time subject to right-censoring poses additional conceptual and estimation challenges as compared to the situation with uncensored continuous outcomes (Goldberg & Kosorok, 2012).

The obvious challenge with survival data is to deal with unobserved survival times due to censoring. Individual trajectories may be incomplete because of censoring and it is unclear how to incorporate the data of censored individuals in the estimation procedure for a DTR. Necessarily, some assumptions must be made about the longitudinal relationship between the censoring time, survival time and covariates. Because individuals may experience an event or be censored at any time before the end of the study, a challenge specific to DTRs is that all individuals do not necessarily enter the same number of stages nor do they necessarily spend the same amount of time within each stage. This complicates the recursive estimation procedure underlying regression-based methods of uncensored outcomes in which one starts by finding the optimal decision rule in the last stage and moves backward into previous stages. In particular, it is difficult to conceptualize the pseudo-outcome in stage j, defined as the counterfactual outcome had future treatments been optimal, for individuals who did not enter stages $j+1$ and beyond because future covariates and future treatments received are undefined. The fact that all individuals do not enter the same number of stages also poses a conceptual challenge about how one defines stages over time. The two following stage definitions could be considered, for example.

Example 1. Entrance in a stage of intervention is defined in terms of developing or not a condition. Following transplantation, a sequence of treatments to prevent and treat graft-versus-host disease (GVHD) aims to improve the survival time post-transplantation (Krakow et al., 2017). The first stage of intervention compares preventive treatments for GVHD following transplantation. A patient enters the second stage of intervention if he develops GVHD and salvage treatments are then compared. Thus, entrance in the second stage is

defined in terms of developing or not GVHD and the time from transplantation to GVHD varies across individuals. Similar situations naturally arise in recurring and relapsing diseases where it is not known if and when a patient will enter the second stage of intervention (e.g. Huang et al., 2014).

Example 2. The duration of a stage of intervention is defined as the time elapsed between two fixed treatment decision time points. The STAR*D sequential randomized trial for major depressive disorder compares various treatment options over time for individuals who do not attain a satisfactory response to their current treatment regime (Rush et al., 2004). Participants are assessed at pre-planned clinic visits held every two or three weeks. At each visit, depending on a participant's response to treatment, the participant could either continue on his current treatment regime or be randomized to alternative treatments as dictated by the study protocol. This framework mimics the situation where an individual's condition is assessed at fixed routine clinic visits and the treating clinician makes a treatment decision at each visit as to keep the individual on his current treatment or make any change to the regime.

Beyond censoring and stage definitions, extending optimal DTR methods to accommodate right-censoring requires choosing a criterion of optimality. With uncensored outcomes, the criterion of optimality is the conditional mean (pseudo)-outcome across stages which could also be used with censored data, provided enough information is available about the tail of the distribution of T (Karrison, 1997). Existing DTR methods for survival outcomes have considered other metrics as criteria of optimality. The restricted mean survival time has been widely used (Bai et al., 2017; Goldberg & Kosorok, 2012; Huang et al., 2014; Y. Q. Zhao et al., 2014b) because it can accommodate heavy censoring that may prevent observing enough data about the tail of the survival time distribution (Karrison, 1997). The restricted survival time Y_τ is defined as $Y_\tau = Y$ if $Y < \tau$ and $Y_\tau = \tau$ if $Y \geq \tau$, where $\tau > 0$ is chosen to be smaller than the longest follow-up time. The restricted mean survival time is then

$\mu(\tau) = \int_0^\tau S(t)\mathrm{dt}$. It offers a practical advantage for methods that use inverse probability of censoring weights (IPCW) (Robins & Rotnitzky, 1992) because $P(C > \tau)$, the probability of observing censoring times larger than τ, is guaranteed to be positive due to truncation of the survival and censoring times. Another possible criterion of optimality is the survival probability at a particular time t (Jiang et al., 2017a). One potential problem with the t-year survival probability is that the choice of t is often subjective and that choosing a single value of t complicates the balance of short- and long-term benefits (Jiang et al., 2017b). The median survival time or other quantiles of the survival distribution have also been proposed as criteria of optimality (Jiang et al., 2017b).

Extending existing regression-based methods to accommodate censored data is subject to some modeling constraints. The recursive estimation procedure of regression-based methods requires the construction of pseudo-outcomes across stages. With survival data, this means estimating a relative improvement in (counterfactual) survival time under optimal treatments in future stages. Therefore, one must be able to make predictions on the survival time scale to extend the methods proposed for uncensored continuous outcomes to censored outcomes. As such, modeling the survival time directly may be necessary to allow recovering such predictions. The AFT model has been used (Huang & Ning, 2012; Huang et al., 2014) as it directly models the log-survival time. Because the AFT parameters are defined on the log-survival time scale, their interpretation easily translates into clinical domain knowledge. Relying on Cox regression is a less interesting option as it would lead to decision rules interpretable on the hazard scale, which is less intuitive, and would require specifying a baseline hazard function to construct pseudo-outcomes. Also, the proportional hazard assumption typically made by the Cox model may not be suitable for DTRs which consider that short- and long-term treatment effects are different (Jiang et al., 2017b).

2.3.3 Existing Regression-based Methods

To the best of our knowledge, there exist only two regression-based methods for estimating an optimal DTR with censored data, Q-learning with censored data (Goldberg & Kosorok, 2012) and a method by Huang et al. (2014) for recurrent diseases, which we summarize in this section. We also briefly mention how survival outcomes are handled with G-estimation.

Q-learning with censored data extends Q-learning to accommodate flexible number of stages and incomplete individual trajectories due to censoring. It aims to maximize the expected restricted mean survival time $\mathbb{E}\left[\min\left(\sum_{j=1}^{\bar{J}} R_j, \tau\right)\right]$ where R_j is the reward in stage j defined as the length of the interval between decision time points $j-1$ and j and $\bar{J} \leq J$ is the (random) number of stages for an individual, allowing different individuals to enter different number of stages. Let δ_j be an indicator where $\delta_j = 1$ if no censoring happened before the stage $j+1$ and $\delta_{j-1} = 0 \implies \delta_j = 0$. Because an event or censoring can occur at any time during the follow-up, the individual trajectories are not of equal lengths but rather defined up to stage $\bar{J}$ as $\{\boldsymbol{H_1}, A_1, R_1, \delta_1, ..., \boldsymbol{H_{\bar{J}}}, A_{\bar{J}}, R_{\bar{J}}, \delta_{\bar{J}}\}$ if an event is observed in stage $\bar{J}$ or $\{\boldsymbol{H_1}, A_1, R_1, \delta_1, ..., \boldsymbol{H_{\bar{J}}}, A_{\bar{J}}, \delta_{\bar{J}}\}$ if censoring occurs in stage $\bar{J}$, in which case the censoring time $C < \sum_{j=1}^{\bar{J}} R_j$ is also observed.

The Q-learning algorithm for censored data is similar to the algorithm for uncensored continuous outcomes in the sense that it uses backward induction to find the optimal DTR and requires modeling the outcome, here a survival time, at each step of the recursion. The algorithm differs by incorporating IPCW to accommodate censored data, assuming that censoring time is independent of both the covariates and survival time. It also differs by transforming the observed trajectories before using them in the algorithm. The trajectories of individuals who did not enter the maximal number of stages are imputed as following: for $j > \bar{J}$, set $\boldsymbol{H_j} = \varnothing$, $R_j = 0$ and draw A_j uniformly from $\mathcal{A}_j$. Also, the observed survival or censoring times are replaced by their truncated counterparts, ensuring $\sum_{j=1}^{\bar{J}} R_j < \tau$. At

each stage, Q-learning computes the Q-function

$$\hat{Q}_j(\boldsymbol{h_j}, a_j) = \arg\min_{Q_j} \mathbb{E}_n \left[\left(R_j + \max_{a_{j+1}} \hat{Q}_{j+1}(\boldsymbol{h_{j+1}}, a_{j+1}) - Q_j(\boldsymbol{h_j}, a_j) \right)^2 \frac{\delta_j}{\hat{S}_c(\sum_{k=1}^{j} R_k)} \right]$$

where $\hat{S}_c$ is the Kaplan-Meier estimator of the survival function of the censoring time and $\mathbb{E}_n$ is the empirical expectation. $\hat{Q}_j$ is set to 0 whenever $\boldsymbol{h_j} = \varnothing$, that is, when a failure occurred before stage j. The terms $R_j + \max_{a_{j+1}} \hat{Q}_{j+1}(\boldsymbol{h_{j+1}}, a_{j+1})$ represent the remaining (truncated) survival time from stage j onwards, given that optimal treatments are received in future stages. $Q_j(\boldsymbol{h_j}, a_j)$ is a parametric model for $R_j + \max_{a_{j+1}} \hat{Q}_{j+1}(\boldsymbol{h_{j+1}}, a_{j+1})$. This is akin to the pseudo-outcome with uncensored data defined in Q-learning (see step 2 of the Q-learning algorithm for uncensored data). All individuals contribute to the estimation of $\hat{S}_c$ but $R_j + \max_{a_{j+1}} \hat{Q}_{j+1}(\boldsymbol{h_{j+1}}, a_{j+1}) - Q_j(\boldsymbol{h_j}, a_j)$ is only defined for individuals who were not censored in stage j. The optimal stage j decision rule is then obtain by maximizing $\hat{Q}_j(\boldsymbol{h_j}, a_j)$ with respect to a_j as in the uncensored case. A finite sample bound on the difference between the expected truncated survival times under the true optimal DTR versus under the estimated optimal DTR is derived to evaluate the performance of the algorithm. Q-learning for censored data suffers important limitations. First, the assumption of independent censoring is restrictive. Second, specifying a model for $Q_j(\boldsymbol{h_j}, a_j)$ may be challenging in practice. Third, the method lacks tools to make inference about the decision rules and their parameters. To the best of our knowledge, Q-learning for censored data has never been used in practice.

The method proposed by Huang et al. (2014) applies to DTRs with two stages of intervention in which all individuals do not necessarily reach the second stage. Their work is motivated by an application to the treatment of acute myeloid leukemia which consists of an initial treatment A_1 (first stage) followed by a salvage treatment A_2 (second stage) if the disease recurs or progresses. Following this example, the outcome to maximize is the time to death defined as $T = T_1 + T_2$ where T_1 represents the time until disease recurrence or progression

or death and T_2 is the time from salvage treatment to death, if applicable. As in Q-learning with censored data, the algorithm relies on backward induction, uses the restricted mean survival time as the criterion of optimality and incorporates censoring with IPCW.

Estimates of the decision rule parameters are obtained by solving a series of weighted estimating equations across stages that depend on models for $\log(T_2)$ in the second stage or for the logarithm of $T_1 + T_2^{\text{opt}}$, the overall survival time had the second stage treatment been optimal, in the first stage. First, estimates for the parameters used to construct the optimal salvage treatment decision rule (second stage decision rule) are obtained by solving

$$U_2^{\text{HNW}}(\boldsymbol{\beta_2}) = \sum \eta\delta\omega \begin{bmatrix} \boldsymbol{h_2} \\ a_2\boldsymbol{h_2} \end{bmatrix} \left\{\log(T_2) - \boldsymbol{\beta_{21}^T h_2} - a_2\boldsymbol{\beta_{22}^T h_2}\right\} = 0$$

where η indicates if an individual entered the second stage and ω are IPCW. The model for the survival time in the second stage is an AFT model specified as $\mathbb{E}[\log(T_2)|\boldsymbol{h_2}, a_2; \boldsymbol{\beta_2}] = \boldsymbol{\beta_{21}^T h_2} + a_2\boldsymbol{\beta_{22}^T h_2}$, defining decision rules of the form $\mathbb{I}(\boldsymbol{\beta_{22}^T h_2} > 0)$. A Cox proportional hazard model is used to estimate the hazard function of the censoring time and, using the Breslow estimator, an estimator of the survival function of the censoring time is derived to construct the weights ω. Second, the optimal initial treatment (optimal treatment in the first stage) is estimated by solving estimating equations similar to U_2^{HNW} except that the outcome is a pseudo-survival time defined by adding a positive quantity $|\hat{\boldsymbol{\beta}}_{\boldsymbol{22}}^{\boldsymbol{T}}\boldsymbol{h_2}|$ to the survival time of individuals who entered the second stage but did not receive their optimal treatment. This is similar to how dWOLS defines a pseudo-outcome with the added consideration that not all individuals enter the second stage. The consistency and asymptotic normality of the decision rule estimators are established (see also Huang & Ning, 2012). The method by Huang et al. (2014) improves on Q-learning by allowing the censoring times to depend on baseline covariates. However, the method is not robust to misspecification of the model for the log-survival time.

Despite the problem of non-regularity not being mentioned by the authors of the two methods, Q-learning and the method by Huang et al. (2014) also yield non-regular estimators because of plug-in quantities in the backward estimation procedure. The same negative consequences as those discussed for uncensored continuous outcomes apply (c.f. Section 2.2.1), i.e. confidence intervals for the decision rule parameters may not be reliable when derived with asymptotic variance formulae or the standard bootstrap.

G-estimation with survival outcomes has been described to address the problem of estimating the causal effect of a time-dependent treatment in the presence of time-varying confounding using a class of causal models called structural nested failure time models (SNFTM) (Hernán et al., 2005; Robins, 1998; Robins et al., 1992). In its simplest form, SNFTM assumes that an individual's survival time under no treatment is expanded or contracted by the factor $\exp(-\psi)$ were he continuously exposed to some treatment $\boldsymbol{a}$ i.e. $T_i^{\boldsymbol{a}} = T_i^0 \exp(-\psi)$ where T_i^0 and $T_i^{\boldsymbol{a}}$ are counterfactual survival times under no treatment or continuous treatment over a given follow-up time. The observed survival time T_i is linked to the counterfactual survival times $T_{i,\psi}$ through the relationship $T_{i,\psi} = \int_0^{T_i} \exp(\psi \times A_{i,t})\mathrm{dt}$ where $A_{i,t}$ indicates if an individual receives the treatment at time t. This form specifies an AFT model for the survival time. G-estimation is used to find $\hat{\psi}$ such that $A_{i,t}$ is independent of $T_{i,\psi}$ given (time-varying) confounders at each time t. For example, this can be done by specifying logistic regressions $\mathrm{logit}(A_{i,t}) \sim \alpha t_{i,\psi} + \boldsymbol{\beta h_{i,t}}$, where $\boldsymbol{h_{i,t}}$ are the history at time t, and performing a grid search about ψ to find the value $\hat{\psi}$ which lead to $\hat{\alpha} = 0$. The method also handles multi-dimensional $\boldsymbol{\psi}$, for example, by allowing treatment-covariate interactions.

The counterfactual survival time $T_{i,\psi}$ can only be calculated if individual i experienced an event during the follow-up period. While censoring due to loss to follow-up or competing risks can be handled via a weighting argument as described in Q-learning for censored data and in the method by Huang et al. (2014), SNFTM accommodates administrative censoring, that is, censoring due to reaching the end of the follow-up period in calendar time without

having experienced an event, using artificial censoring (Robins, 1998). Let C_i be the maximal possible follow-up time for individual i, known at the beginning of the study. Assuming that C_i is independent of the counterfactual survival time $T_{i,\psi}$, the indicator $\Delta_{i,\psi}$ of whether the event would have been observed had the individual been continuously treated or untreated can be computed for the individuals who experienced an event as $\Delta_{i,\psi} = \mathbb{I}\left(T_{i,\psi} < C_{i,\psi}\right)$ where $C_{i,\psi} = C_i$ if $\psi \geq 0$ and $C_{i,\psi} = C_i \times \exp(\psi)$ if $\psi < 0$. This indicator is zero for all individuals who are censored and is also zero for individuals who experienced an event but for whom an event would not have been observed had they received another treatment. Estimates of ψ are obtained by G-estimation, now using the fact that $A_{i,t}$ should independent of $\Delta_{i,\psi}$ given (time-varying) confounders at each time t. Artificial censoring with G-estimation has been criticized because it results in loss of information due to censoring individuals who actually experienced an event (Joffe et al., 2012).

2.3.4 Other Existing Methods

Although value search estimation is not the focus of this thesis, we present an overview of methods proposed in the literature for estimating an optimal DTR with censored data.

Extensions of existing value search methods have been proposed to accommodate censored data in single-stage settings (e.g. Cui et al., 2017; Geng et al., 2015; Zhu et al., 2017). For example, Y. Q. Zhao et al. (2014b) and Bai et al. (2017) extend outcome weighted learning by respectively adding one and two augmentation terms to account for censoring, the latter method yielding a more efficient estimator. Methods suitable for multi-stage settings have also been proposed. Y. Zhao et al. (2011) use SVM within the Q-learning framework to fit nonlinear Q-functions. The censored individuals contribute to the SVM procedure via a particular choice of a loss function. The proposed method lacks tools for inference about the resulting optimal DTR or the value of the optimal regime. Jiang et al. (2017a) describe an inverse propensity score weighted Kaplan-Meier estimator to maximize

the t-year survival probability. Conceptually, their estimator can be extended to more than two stages but it may become less reliable. It could also incorporate an augmentation term with a posited model for the survival time distribution to yield a doubly-robust estimator but its formulation would be too complicated. Jiang et al. (2017b) extend the method by Jiang et al. (2017a) to accommodate any user-specified function of the survival function such as the restricted mean survival time or median survival time. Assuming independent censoring, their proposed estimator is consistent whenever the propensity score or the regression model for the survival time is correctly specified, assuming that the survival function of the censoring time can be consistently estimated with the Kaplan-Meier estimator. The authors derive the asymptotic distribution for the value of the regime. Hager et al. (2018) extend the method by Bai et al. (2017) to a DTR with two stages of intervention by using a backward induction procedure. They add augmentation terms to the IPW estimator to capture back information from individuals who are censored in the first and second stages. Their proposed estimator is doubly-robust as it consistently estimates the value of the regime whenever both the propensity score and censoring hazard models are correctly specified or if the survival time hazard model is correctly specified.

2.4 DTRs for the Treatment of Type 2 Diabetes

We conclude our review of the literature by giving an overview of the current guidelines for the treatment of T2D, focusing on aspects that would benefit from the development of DTR methods for survival outcomes.

T2D is a chronic disease characterized by an elevated blood sugar level which can lead to severe complications if untreated. One of the main therapy goals in the management of T2D is lowering, or maintaining an optimal, glycemic level, as measured by glycated hemoglobin (HbA1c). Ultimately, the aims of controlling glycemia are to avoid unstable blood glucose

levels over time and to prevent or delay the development of diabetic complications without significantly altering the patient's quality of life. Treatment strategies for T2D include lifestyle and drug therapies as well as considerations for the joint management of comorbidities (Garber et al., 2019). At diagnosis of T2D, interventions designed to improve the patient's lifestyle habits are recommended before embarking pharmacotherapy. Metformin is the preferred and most cost-effective first-line oral treatment. When metformin in monotherapy fails to achieve therapeutic goals, it is recommended to add a second or even a third oral agent to the current regime before eventually transitioning to injectable therapy. The whole process of decision-making used to determine appropriate treatment strategies following metformin for a specific patient is complex and, to some extent, subjective to the treating clinician and patient's values but it remains widely accepted that the choice of second-line and subsequent agents is best tailored to individual patients (Garber et al., 2019; Inzucchi et al., 2015).

In the absence of comprehensive comparative-effectiveness trials that take into account the dynamic nature of the treatment of T2D, personalized recommendations on the best agent to be combined with metformin cannot easily be made. Thus, there remain uncertainties in the choice of agent to add to metformin and in the sequence of therapies that follows. This uncertainty is reflected in practice as illustrated in a large observational study from the Observational Health Data Sciences and Informatics collaboration (Hripcsak et al., 2016). Using electronic health records and administrative claims data on 250 million patients, the authors found that metformin is indeed favored as the first-line medication for T2D but that second- and third-line add-on treatments are much more variable.

Existing second-line therapies and their effect on the risk of complications or mortality in T2D have been widely studied (recently Kosiborod et al., 2018; Kuo et al., 2019; Nyström et al., 2017; Yu et al., 2015) but the therapy comparisons do not account for the fact that patients will likely make multiple changes to their treatment regime following the initial add-

on to metformin. The estimation of optimal treatment rules that are tailored to patients' characteristics has also been considered (e.g. Fu et al., 2016; Wang et al., 2018) but, again, never beyond a single treatment decision. Others have considered the treatment of T2D from a dynamic perspective (Kreif et al., 2018; Neugebauer et al., 2016). For example, Neugebauer et al. (2013) use marginal structural models to determine when treatment should be intensified (add any drug to the current regime) based on HbA1c levels to decrease all-cause mortality and complications. However, their study and other similar studies fail to provide guidance on how to choose the drug to be added to intensify treatment.

2.5 Summary

The literature review introduced important methods for the estimation of an optimal DTR. We introduced key DTR concepts and methods for uncensored continuous outcomes, including dWOLS. We explained how the inferences with dWOLS may be affected by non-regularity of the estimators in multi-stage DTRs. Methods have been proposed to alleviate the negative impact of non-regularity, offering promising avenues for dWOLS. We presented challenges pertaining to the extension of optimal DTR methods to censored data and summarized the few methods that have been proposed for that purpose. Finally, we gave an overview of how the management of T2D in clinical practice could benefit from methodological advancements in precision medicine.

Chapter 3

Non-regular Inference for Dynamic Weighted Ordinary Least Squares: Understanding the Impact of Solid Food Intake in Infancy on Childhood Weight

Preamble to Manuscript 1. The ideas for this project came when dWOLS was a relatively new method for estimating optimal DTRs with continuous uncensored outcomes. Despite the practical and theoretical advantages of dWOLS, it lacked tools for constructing confidence intervals about the decision rules parameters in situations where the non-regularity of the estimators may negatively affect the inferences. This project started as a course assignment tackling a computationally intensive problem centered around the m-out-of-n bootstrap. A basic simulation study was carried out and initial implementation of the m-out-of-n bootstrap was done in `R`. It then evolved into the more substantive project presented in this chapter.

The original contributions are (i) proposing an empirical measure of non-regularity in conjunction with dWOLS to choose m adaptively, and (ii) deriving a class of data-generating mechanisms for exceptional laws, that is, where inferences with non-regular estimators may be incorrect. The manuscript presented in this chapter was published in *Biostatistics* in 2017. The m-out-of-n bootstrap has been integrated in the `R` package `DTRreg` and released on CRAN.

Non-regular Inference for Dynamic Weighted Ordinary Least Squares: Understanding the Impact of Solid Food Intake in Infancy on Childhood Weight

Gabrielle Simoneau[1], Erica EM Moodie[1], Robert W Platt[1], Bibhas Chakraborty[2]

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

[2]*Duke-NUS Medical School, National University of Singapore, Singapore*

Abstract

A dynamic treatment regime (DTR) is a set of decision rules to be applied across multiple stages of treatments. The decisions are tailored to individuals, by inputing an individual's observed characteristics and outputting a treatment decision at each stage for that individual. Dynamic weighted ordinary least squares (dWOLS) is a theoretically robust and easily implementable method for estimating an optimal DTR. As many related DTR methods, the dWOLS treatment effects estimators can be non-regular when true treatment effects are zero or very small, which results in invalid Wald-type or standard bootstrap confidence intervals. Inspired by an analysis of the effect of diet in infancy on measures of weight and body size in later childhood – a setting where the exposure is distant in time and whose effect is likely to be small – we investigate the use of the m-out-of-n bootstrap with dWOLS as method of analysis for valid inferences of optimal DTR. We provide an extensive simulation study to compare the performance of different choices of resample size m in situations where the treatment effects are likely to be non-regular. We illustrate the methodology using data from the PROmotion of Breastfeeding Intervention Trial to study the effect of solid food intake in infancy on long-term health outcomes.

3.1 Introduction

Personalized medicine is an approach to health care in which treatment decisions are tailored to evolving patient-level information. This approach is especially relevant in the chronic-care environment where the patient's health condition is changing over time and treatments must correspondingly be altered. The statistical study of personalized medicine is known as dynamic treatment regimes (DTR) or sometimes adaptive treatment strategies. In the typical DTR setting, individuals are followed through multiple stages of clinical intervention and the statistical goal is to perform estimation and inference on the sequence of decision rules, one at each stage of intervention, which uses the individual's characteristics as inputs and yields a recommended treatment decision. Of particular interest is the identification of an optimal DTR, that is, the sequence of treatment decisions that yields the best expected outcome for a population of ("similar") individuals.

Methods for estimating optimal DTRs have been widely considered in the last decade (Murphy, 2003; Robins, 2004; B. Zhang et al., 2013; Y. Q. Zhao et al., 2015). Notably, Q-learning (Sutton & Barto, 1998; Watkins, 1989) offers a relatively easy framework via ordinary least squares (OLS) or other regression or prediction methods, but lacks robustness to model mis-specification. Alternatively, G-estimation (Robins, 2004) offers robustness against mis-specification of the outcome model at the cost of a greater investment both in terms of understanding the underlying theory and implementation. The strengths of the two methods have been united under the recently proposed dynamic weighted OLS (dWOLS) approach (Wallace & Moodie, 2015): it possesses G-estimation's double-robustness while relying on a simple estimation framework based on sequences of weighted OLS regressions.

Despite its practical and theoretical advantages, dWOLS, like Q-learning, G-estimation, and other regression-based DTR estimation approaches, can yield estimators with non-regular limiting distributions (Robins, 2004). In a DTR context, non-regularity occurs because the estimation of stage-specific treatment effects involves a non-smooth maximization operation,

and the asymptotic distribution of the resulting estimator does not converge uniformly over the parameter space (see Section 3.2.2 for further details). In the case of dWOLS, this occurs when the true stage-specific treatment effect is small relative to the sample size (or zero) so that the two treatment options are nearly (or precisely) equally optimal. In randomized controlled trials or large-scale observational trials of treatment sequences, the dWOLS estimators are likely to have non-regular limiting distributions as the expected size of the treatment effect at each stage is usually small. A significant negative consequence of the non-regularity is that typical confidence interval calculations for the DTR parameters perform poorly in terms of coverage (Chakraborty et al., 2010; Moodie & Richardson, 2010; Robins, 2004).

The m-out-of-n bootstrap has been proposed as a tool to produce valid confidence intervals in cases where the standard bootstrap fails (Bickel et al., 1997; Bretagnolle, 1983; Shao, 1994). It proceeds similarly to the standard bootstrap, except that the resamples are of size $m < n$. The performance of the m-out-of-n bootstrap is not guaranteed in non-regular situations (Andrews & Guggenberger, 2010) as it relies on heuristic arguments and is highly dependent on the choice of m, for which there exists no finite-sample guidance. It is thus necessary to evaluate the performance of the procedure in every new application. The m-out-of-n bootstrap, applied with a class of resample size that adapts to non-regularity, has been found to produce valid confidence intervals in the DTR framework for Q-learning (Chakraborty et al., 2013), and so offers a promising avenue to explore in the dWOLS context.

Consider the effect of an infant's diet during the first year of life on future health outcomes. Current World Health Organization (WHO) and United Nations Children's Fund (UNICEF) recommendations on infant diet include: (i) exclusive breastfeeding for the first 6 months (mo) of life and (ii) introduction of nutritionally-adequate and safe complementary (solid) food at 6 mo together with continued breastfeeding up to 2 years of age or beyond ("Infant and young child feeding", 2016). Despite these recommendations, recent studies conducted in the

United States (Clayton et al., 2013) and in Australia (Newby & Davies, 2015) showed that the introduction of solid food before 4 mo, and thus non-exclusive breastfeeding, is prevalent. With childhood obesity becoming more and more of a burden in developed countries, a recent meta-analysis (Daniels et al., 2015) indicated that early introduction of solid food (before 4 mo) results in an increased risk of childhood obesity, and that this association between infant diet and childhood obesity is even stronger for infants that were not breastfed. This suggests that the decision to introduce solid food in an infant's diet may be motivated by evolving health characteristics of the infant.

The PROmotion of Breastfeeding Intervention Trial (PROBIT) (Kramer et al., 2001) has collected longitudinal measurements on infant-feeding habits for over 17,000 infant-mother pairs. Following the conflict between WHO recommendations and the practices noted in the literature on infant-feeding patterns, we aim to investigate the effect of solid food intake in the infant's diet between 3 and 9 mo on long-term health outcomes in the DTR framework, taking advantage of the richness of the PROBIT dataset and using the newly developed dWOLS as method of analysis. More precisely, we examine whether there is evidence that there is an optimal decision of when to start solid food (between 3–6 mo or 6–9 mo) on the child's body mass index (BMI), waist circumference and triceps skinfold thickness at 6.5 years of age, and whether this decision should depend on infant characteristics. As the effect of solid food intake between 3–6 mo or 6–9 mo is likely to have a small effect on BMI, waist circumference, and triceps skinfold thickness measured at 6.5 years, inferences using dWOLS will possibly yield effect estimators with non-regular limiting distributions.

In this article, we demonstrate the construction of valid confidence intervals with the m-out-of-n bootstrap with dWOLS as method of analysis, motivated by a case study based on important health data to investigate whether the decision of introducing solid food into an infant's diet should be tailored to reduce weight-related outcomes in childhood. In Section 3.2, we review dWOLS and the m-out-of-n bootstrap. In Section 3.3, we evaluate the

performance of the m-out-of-n bootstrap for inference concerning dWOLS estimators via a simulation study. In Section 3.4, we present details and results of the PROBIT data analysis.

3.2 Methods

3.2.1 Notation and Important Concepts

Data needed to estimate a DTR consist of n longitudinal trajectories of measured covariates and treatment received at each of a fixed number of stages of intervention. For simplicity, we focus on DTRs with two stages of clinical intervention. Individuals' data are given by the trajectories $(\boldsymbol{X_1}, A_1, \boldsymbol{X_2}, A_2, Y)$ of patient characteristics $\boldsymbol{X_j}$ and treatment A_j grouped into stages of intervention, denoted by a subscript j $(j = 1, 2)$. Let $\boldsymbol{X_j}$ be a matrix of covariates measured prior to the j^{th} treatment, A_j denote the treatment received at the j^{th} stage, and Y be the observed outcome measured at the end of the j^{th} stage. We assume the outcome is continuous and defined such that larger values are preferred. We consider the case where treatment is binary, coded as 0 or 1. Denote an individual's history by $\boldsymbol{H_j}$, a shorthand representing the information available prior to making a treatment decision at the j^{th} stage, including previous treatments. We thus define $\boldsymbol{H_1} = \boldsymbol{X_1}$ and $\boldsymbol{H_2} = (\boldsymbol{X_1}, A_1, \boldsymbol{X_2})$. An optimal DTR consists of a set of decision rules that maximizes the expected final outcome $\mathbb{E}(Y^{a_1,a_2})$, where Y^{a_1,a_2} indicates a potentially unobserved, or counterfactual outcome Y under treatment regime (a_1, a_2). At each stage, the decision rule is a function that inputs the history $\boldsymbol{H_j}$ and outputs one of the two available treatments.

An important concept in DTR inference is the *blip*, or contrast, function. In the context of a two-stage DTR with two possible treatments at each stage, the blip function at the first stage is defined as $\gamma_1(\boldsymbol{h_1}, a_1) = \mathbb{E}(Y^{a_1,a_2^{\text{opt}}} - Y^{0,a_2^{\text{opt}}} | \boldsymbol{H_1} = \boldsymbol{h_1})$, and at the second as $\gamma_2(\boldsymbol{h_2}, a_2) = \mathbb{E}(Y^{a_1,a_2} - Y^{a_1,0} | \boldsymbol{H_2} = \boldsymbol{h_2})$. It is interpreted as the difference between the

expected outcome of an individual who received treatment a_j at stage j and the expected outcome of the same individual had he received some reference treatment a_j=0 at stage j, assuming that the individual goes on to receive optimal subsequent treatment thereafter. The (potentially) counterfactual outcome could be modeled as

$$\mathbb{E}(Y^{a_1,a_2}|\boldsymbol{H_1}=\boldsymbol{h_1},\boldsymbol{H_2}=\boldsymbol{h_2})=\sum_{j=1}^{2}\left\{f_j(\boldsymbol{h_{j\beta}};\boldsymbol{\beta_j})+\gamma_j(\boldsymbol{h_{j\psi}},a_j;\boldsymbol{\psi}_j)\right\}.$$

In this model, the expected outcome is separated into two components: a *treatment-free* function f_j, independent of the stage j treatment, and the blip function γ_j, where possibly different subsets of the history vector $\boldsymbol{h_j}$, respectively $\boldsymbol{h_{j\beta}}$ and $\boldsymbol{h_{j\psi}}$, are used for each of these two models. Although the stage j treatment-free function does not depend on the stage j treatment, it may depend on previous treatments. Of interest is to identify the form and parameters of γ_j as it is only through the blip function that the stage j optimal treatment is estimated. The blip function is constrained so that $\gamma_j(\boldsymbol{h_{j\psi}},0;\boldsymbol{\psi_j})=0$. At each stage, the optimal treatment decision is that which maximizes γ_j, given by "prescribe treatment option 1 if $\gamma_j(\boldsymbol{h_{j\psi}},1;\boldsymbol{\psi_j})>0$, prescribe option 0 otherwise." In a slight abuse of notation, we will simply write $\boldsymbol{h_j}$ in place of both $\boldsymbol{h_{j\beta}}$ and $\boldsymbol{h_{j\psi}}$ in much of what follows, with the understanding that not all components of the vector need appear in all models.

Two assumptions are necessary for the estimation of a DTR: the stable unit treatment value assumption (SUTVA) (Rubin, 1980) and no unmeasured confounding (Robins, 1997). SUTVA demands that there is no interference between individuals leading to different outcomes depending on other individuals' treatment allocation. No unmeasured confounding requires that the treatment allocation at stage j is independent of future (counterfactual) outcomes and covariates given the observed j^{th} stage history.

3.2.2 Dynamic Weighted OLS

Inference for DTRs in regression-based approaches such as dWOLS focuses on the *blip parameters* $\boldsymbol{\psi_j}$. As other standard statistical methods for estimating an optimal DTR, dWOLS is based on a recursive backward estimation procedure (Murphy, 2003; Robins, 2004), which performs a sequence of weighted OLS regressions. For a DTR with J stages of intervention, dWOLS first estimates the effect of treatment (the blip parameters) at the final stage and moves backward into previous stages. At each stage, the estimation of the blip parameters is based on a weighted OLS regression of a pseudo-outcome $\tilde{y}_j$ (see Equation (3.1) below) on (a subset of) the individual history $\boldsymbol{h_j}$ available prior to the treatment decision. At the last stage, the pseudo-outcome is identical to the final outcome y. At each previous stage, it defines an expected outcome had the future treatment decisions been optimal and is estimated by adding what was lost from receiving suboptimal future treatments to the observed outcome. At each stage j of clinical intervention, the dWOLS algorithm takes the following steps:

1. Define the pseudo-outcome

$$\tilde{y}_j = y + \sum_{k=j+1}^{J} \left\{ \gamma_j(\boldsymbol{h_j}, a_j^{\text{opt}}) - \gamma_j(\boldsymbol{h_j}, a_j) \right\}; \tag{3.1}$$

2. Propose a *treatment* model $\mathbb{E}(A_j|\boldsymbol{H_j} = \boldsymbol{h_j}; \boldsymbol{\omega_j})$ and use the data to estimate $\boldsymbol{\omega_j}$;

3. Choose a *weight* function $w_j(a_j, \boldsymbol{h_j}; \boldsymbol{\omega_j})$ which satisfies

$$\pi(\boldsymbol{h_j}) w_j(1, \boldsymbol{h_j}; \boldsymbol{\omega_j}) = [1 - \pi(\boldsymbol{h_j})] w_j(0, \boldsymbol{h_j}; \boldsymbol{\omega_j})$$

   where $\pi(\boldsymbol{x_j}) = P(A_j = 1|\boldsymbol{x_j})$ and use the estimates $\hat{\boldsymbol{\omega}}_{\boldsymbol{j}}$ to obtain estimated weights $\hat{w}_j$;

4. Propose a model for $\mathbb{E}(\tilde{Y}_j^{a_j}|\boldsymbol{H_j} = \boldsymbol{h_j})$ by specifying a treatment-free model $f_j(\boldsymbol{h_j}; \boldsymbol{\beta_j})$

and a blip function $\gamma_j(\boldsymbol{h_j}, a_j; \boldsymbol{\psi_j})$, and carry out a weighted OLS regression of $\tilde{y}_j$ on $(\boldsymbol{h_j}, a_j\boldsymbol{h_j})$ with weights $\hat{w}_j$. Use the resulting estimates $\hat{\boldsymbol{\psi}}_{\boldsymbol{j}}$ to construct the next pseudo-outcome, if necessary.

Typically, the blip model and the treatment-free model are linear in the parameters (though they need not be), such that the optimal treatment at stage j, a_j^{opt}, is determined by the decision rule $\mathbb{I}(\hat{\boldsymbol{\psi}}_{\boldsymbol{j}}^{\boldsymbol{T}}\boldsymbol{h_j} > 0)$ which chooses the treatment (A_j=1) as the optimal decision when $\hat{\boldsymbol{\psi}}_{\boldsymbol{j}}^{\boldsymbol{T}}\boldsymbol{h_j}$ is positive.

The resulting DTR parameters have the double-robustness property, that is, the blip parameters are consistently estimated if either the treatment model $\mathbb{E}(A_j|\boldsymbol{H_j} = \boldsymbol{h_j}; \boldsymbol{\omega_j})$ or the treatment-free model f_j is correctly specified, assuming the blip function γ_j is correctly modeled. In other words, the dWOLS estimators are consistent when either the treatment model or the outcome model is correctly specified, with the additional requirement that mis-specification of the outcome model is only with respect to terms in the model that do not involve the treatment. The double-robustness property can also be exploited to provide insights on whether neither or at least one model is correctly specified (Wallace et al., 2016). The treatment models can be assessed using standard diagnostic plots whereas the treatment-free and blip models can be assessed using diagnostic plots introduced in Rich et al. (2010), all implemented in the `DTRreg` package in `R` (Wallace et al., 2014). The choice of weights may affect efficiency of the estimators; following previous work (Wallace & Moodie, 2015), we consider the weights $w_j(a_j, \boldsymbol{h_j}) = |a_j - \mathbb{E}(A_j|\boldsymbol{H_j} = \boldsymbol{h_j})|$.

The dWOLS method suffers from nonstandard limit theory (Robins, 2004) because of a non-smooth maximization operation in the definition of the pseudo-outcomes. For a two-stage

DTR, the first stage pseudo-outcome is defined as

$$\begin{aligned}\tilde{Y}_1 &= Y + \left\{\gamma_2(\boldsymbol{h_2}, a_2^{\text{opt}}) - \gamma_j(\boldsymbol{h_2}, a_2)\right\} \\ &= Y + \left(\max_{a_2} \hat{\boldsymbol{\psi}}_{\boldsymbol{2}}^{\boldsymbol{T}} a_2 \boldsymbol{h_2} - \hat{\boldsymbol{\psi}}_{\boldsymbol{2}}^{\boldsymbol{T}} a_2 \boldsymbol{h_2}\right) \\ &= Y + \hat{\boldsymbol{\psi}}_{\boldsymbol{2}}^{\boldsymbol{T}} \boldsymbol{h_2} \left(\mathbb{I}(\hat{\boldsymbol{\psi}}_{\boldsymbol{2}}^{\boldsymbol{T}} \boldsymbol{h_2} > 0) - a_2\right). \end{aligned} \tag{3.2}$$

The first stage pseudo-outcome is thus a non-smooth function of $\hat{\boldsymbol{\psi}}_{\boldsymbol{2}}$ because the function $\mathbb{I}(\hat{\boldsymbol{\psi}}_{\boldsymbol{2}}^{\boldsymbol{T}} \boldsymbol{h_2} > 0)$ is non-differentiable at $\hat{\boldsymbol{\psi}}_{\boldsymbol{2}}^{\boldsymbol{T}} \boldsymbol{h_2} = 0$ (point of non-differentiability). Since the first stage blip parameters $\boldsymbol{\psi_1}$ are estimated via a weighted OLS regression of the pseudo-outcome $\tilde{y}_1$ on $(\boldsymbol{h_1}, a_1 \boldsymbol{h_1})$, the resulting estimates $\hat{\boldsymbol{\psi}}_{\boldsymbol{1}}$ are in turn a non-smooth function of $\hat{\boldsymbol{\psi}}_{\boldsymbol{2}}$. The DTR estimator of $\boldsymbol{\psi_1}$ is thus non-regular. As a result, the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\psi}}_{\boldsymbol{1}} - \boldsymbol{\psi_1})$ is not uniformly normal. More precisely, the asymptotic distribution is normal if the second stage blip parameters $\boldsymbol{\psi_2}$ are "far" from the point of non-differentiability, i.e. if the probability of generating a history $\boldsymbol{h_2}$ such that $\boldsymbol{\psi}_{\boldsymbol{2}}^{\boldsymbol{T}} \boldsymbol{h_2} = 0$ is zero. Similarly, the asymptotic distribution is non-normal if $\boldsymbol{\psi_2}$ is "near" the point of non-differentiability, i.e. the probability of generating a history $\boldsymbol{h_2}$ such that $\boldsymbol{\psi}_{\boldsymbol{2}}^{\boldsymbol{T}} \boldsymbol{h_2} = 0$ or indeed very near zero is "large." Define $p := P(\boldsymbol{H_2} : \boldsymbol{\psi}_{\boldsymbol{2}}^{\boldsymbol{T}} \boldsymbol{h_2} = 0)$ to be a measure of non-regularity in the data. Specifically, the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\psi}}_{\boldsymbol{1}} - \boldsymbol{\psi_1})$ is normal if p=0, but is non-normal if $p > 0$. Non-regularity can be defined in terms of the optimal second stage treatment estimated by dWOLS. Referring to Equation (3.2), the pseudo-outcome of an individual with history $\boldsymbol{h_2}$ is independent of the optimal second stage treatment decision $\mathbb{I}(\hat{\boldsymbol{\psi}}_{\boldsymbol{2}}^{\boldsymbol{T}} \boldsymbol{h_2} > 0)$ when $\hat{\boldsymbol{\psi}}_{\boldsymbol{2}}^{\boldsymbol{T}} \boldsymbol{h_2} = 0$. Consequently, the expected outcome is the same for both treatments, and the optimal treatment is not unique.

A practical consequence of this non-regularity in the estimation of the first stage treatment effect parameter $\boldsymbol{\psi_1}$ lies in the construction of valid confidence intervals. Wald-type confidence intervals for $\boldsymbol{\psi_1}$ exhibit poor coverage rates (Robins, 2004). Standard bootstrap confidence intervals can also perform badly (Chakraborty et al., 2013, 2010; Shao, 1994), since the re-

quirement that the statistic of interest is smooth such that it can be approximated linearly is not met in non-regular settings.

3.2.3 The m-out-of-n Bootstrap

We propose the use of the m-out-of-n bootstrap in non-regular situations, following the developments in Chakraborty et al. (2013). The m-out-of-n bootstrap is a general bootstrap-type method proposed to rectify the inconsistency of the bootstrap estimator for non-smooth statistics (Chakraborty et al., 2010; Shao, 1994).

The m-out-of-n bootstrap proceeds similarly to the standard bootstrap, except that the resample size m is smaller than the total sample size n. Let θ be the statistic of interest and $\hat{\theta}_n$ be its estimate in the original sample. A total of B bootstrap resamples of size m are drawn, and the statistic of interest is estimated in each b resample, denoted as $\hat{\theta}_m^{(b)}$. The quantiles of the distribution $\sqrt{m}(\hat{\theta}_m^{(b)} - \hat{\theta}_n)$ are then used to construct a confidence interval for θ. Despite its well-developed theoretical framework, there is some lack of practical guidance on the use of the m-out-of-n bootstrap. Specifically, the conditions on the resample size m are entirely asymptotic and cannot be translated into finite samples properties. We follow Chakraborty et al. (2013) in considering both fixed and adaptive choices for m; in particular, we consider an approach to selecting m that reflects the degree of non-regularity in the underlying data generating process.

Adaptive Choice of m

A class of resample sizes m that adapts to non-regularity in the data is introduced in Chakraborty et al. (2013) and is defined as a function of the sample size n and of the regularity measure p. Provided p can be estimated from the data, a simple definition of $\hat{m}$ given by Chakraborty et al. (2013) is $\hat{m} := n^{\frac{1+\alpha(1-\hat{p})}{1+\alpha}}$, where $\alpha > 0$ is a tuning parameter and

$\hat{p}$ is an estimate of the degree of non-regularity in the data (see below). For a fixed n and α, in a regular scenario where the second stage treatment effect is large, p=0 and so too should be $\hat{p}$, so that the resample size would equal n, yielding a standard bootstrap procedure. In a non-regular situation where the second stage treatment effect is small or inexistent, $\hat{p}$ would be close to 1, and the resample size would be as small as dictated by α. Once $\hat{m}$ is chosen, the construction of $(1-\eta)\times 100\%$ confidence interval for $\boldsymbol{\psi_1}$ is straightforward, where η is a fixed significance level. Let $\hat{\boldsymbol{\psi}}_{\mathbf{1}}$ denote the estimate of the first stage treatment effect using all n individuals and $\hat{\boldsymbol{\psi}}_{1,\hat{m}}^{(b)}$ denote its bootstrap estimate based on resamples of size $\hat{m}$. We find the lower and upper $\eta/2\times 100$ percentiles of $\sqrt{\hat{m}}(\hat{\boldsymbol{\psi}}_{1,\hat{m}}^{(b)}-\hat{\boldsymbol{\psi}}_{\mathbf{1}})$, respectively denoted as $\hat{\boldsymbol{l}}$ and $\hat{\boldsymbol{u}}$. A confidence set for $\boldsymbol{\psi_1}$ is then given by $(\hat{\boldsymbol{\psi}}_{\mathbf{1}}-\hat{\boldsymbol{u}}/\hat{m}, \hat{\boldsymbol{\psi}}_{\mathbf{1}}-\hat{\boldsymbol{l}}/\hat{m})$.

As the true generative model is unknown, the measure of non-regularity p used to define $\hat{m}$ must be estimated from the data. An intuitive estimator of p, implemented in the `DTRreg` package in `R`, considers the proportion of individuals for which the optimal second stage treatment is non-unique. Recall that non-regularity occurs when, for an individual with history $\boldsymbol{h_2}$, the two possible treatments lead to the same expected outcome, resulting in the two treatments being optimal. Using the data, the estimated blip parameters and their corresponding standard bootstrap confidence sets, the idea is to identify individuals for whom, when considering all blip parameters within their respective confidence sets, both treatments are optimal in the sense that the expected outcomes under the two treatment choices are not significantly different.

The tuning parameter α controls the smallest possible resample size $\hat{m}$. The choice of α may be justified by practical consideration, by bounding the smallest possible resample size, or it may be tuned in a data-driven way. Chakraborty et al. (2013) proposed a double bootstrap algorithm for choosing α adaptively. The double bootstrap procedure works as a cross-validation tool for choosing the tuning parameter α such that the coverage of the m-out-of-n bootstrap for the parameter of interest, say ψ, is close to the desired nominal rate. The

statistic of interest is estimated in the original sample as $\hat{\psi}$. The double bootstrap algorithm sets α to a small value, say 0.025, and draws B_1 first-level resamples with replacement of size n from the original data. For each of the B_1 resample, (i) the statistics of interest $\hat{\psi}^{(b_1)}$ is estimated, (ii) the resample size $\hat{m}^{(b_1)}$ is calculated as a function of $\hat{p}$ and of the current value of α, and (iii) B_2 second-level bootstrap resamples of size $\hat{m}^{(b_1)}$ are drawn. The statistic of interest is estimated in each second-level bootstrap resample as $\hat{\psi}^{(b_1,b_2)}$, and B_1 confidence intervals are constructed for ψ from the distribution of $\sqrt{m^{(b_1)}}(\hat{\psi}^{(b_1,b_2)} - \hat{\psi}^{(b_1)})$. The coverage of the m-out-of-n bootstrap with the current value of α is then estimated by counting how many of the B_1 confidence intervals constructed this way include $\hat{\psi}$, the estimate from the original sample. If the desired nominal rate is reached or exceeded, the current value of α is chosen to choose a resample size m. Otherwise, α is incremented by a small value, say 0.025, and the procedure is repeated until it yields the targeted nominal rate. Further details can be found in the Supplemental Material B.1.

3.3 Simulation

We perform a simulation study to compare the performance of the m-out-of-n bootstrap to the standard n-out-of-n bootstrap in terms of inference for dWOLS estimators. We consider nine scenarios with two stages of treatment, each with two possible treatments and one observed covariate. The generative models can be summarized in terms of: (i) $X_j \in \{-1, 1\}$, $A_j \in \{0, 1\}$ for $j = 1, 2$; (ii) $P(A_j = 1) = P(A_j = 0) = 0.5$ for $j = 1, 2$; (iii) $X_1 \sim 2 \times \text{Bernoulli}(0.5) - 1$, $X_2|X_1, A_1 \sim 2 \times \text{Bernoulli}(\text{expit}\{\delta_1 X_1 + \delta_2(2A_1 - 1)\}) - 1$ where $\text{expit}(x) = \exp(x)/(1 + \exp(x))$; (iv) $Y_1 \equiv 0$, $Y_2 = \lambda_1 + \lambda_2 X_1 + \lambda_3 A_1 + \lambda_4 X_1 A_1 + \lambda_5 A_2 + \lambda_6 X_2 A_2 + \lambda_7 A_1 A_2 + \varepsilon$ with $\varepsilon \sim N(0, 1)$. Individual histories are given by $\boldsymbol{H_{2\beta}} = (1, X_1, A_1, X_1 A_1)^T$, $\boldsymbol{H_{2\psi}} = (1, X_2, A_1)^T$ and $\boldsymbol{H_{1\beta}} = \boldsymbol{H_{1\psi}} = (1, X_1)^T$. We use the following

treatment-free and blip models specification:

$$
\begin{aligned}
f_2(\boldsymbol{h_{2\beta}};\boldsymbol{\beta_2}) &= \beta_{20} + \beta_{21}X_1 + \beta_{22}A_1 + \beta_{23}X_1A_1 \\
\gamma_2(\boldsymbol{h_{2\psi}};\boldsymbol{\psi_2}) &= A_2\left(\psi_{20} + \psi_{21}X_2 + \psi_{22}A_1\right) \\
f_1(\boldsymbol{h_{1\beta}};\boldsymbol{\beta_1}) &= \beta_{10} + \beta_{11}X_1 \\
\gamma_1(\boldsymbol{h_{1\psi}};\boldsymbol{\psi_2}) &= A_1\left(\psi_{10} + \psi_{11}X_1\right).
\end{aligned}
$$

These generative models have been chosen to consider different degrees of non-regularity (Chakraborty et al., 2010) which can be defined in terms of (i) the probability of generating an individual history such that $\lambda_5 A_2 + \lambda_6 X_2 A_2 + \lambda_7 A_1 A_2 = 0$ and (ii) the standardized effect size $\mathbb{E}[(\lambda_5 + \lambda_6 X_1 + \lambda_7 A_1)/\sqrt{\mathbb{V}\text{ar}(\lambda_5 + \lambda_6 X_1 + \lambda_7 A_1)}]$. Details on how those parameters influence non-regularity are given in the Supplemental Material B.2. Table 3.1 summarizes the parameter settings and the scenario type ("regular", "near non-regular", "non-regular").

Table 3.1: Parameter settings for nine classes of generative model, classified as "non-regular", "near non-regular", and "regular." The models are defined with (i) $X_1 \sim 2\times\text{Bernoulli}(0.5)-1$, (ii) $X_2|X_1, A_1 \sim 2 \times \text{Bernoulli}(\text{expit}\{\delta_1 X_1 + \delta_2(2A_1 - 1)\}) - 1$, and (iii) $Y_2 = \lambda_1 + \lambda_2 X_1 + \lambda_3 A_1 + \lambda_4 X_1 A_1 + \lambda_5 A_2 + \lambda_6 X_2 A_2 + \lambda_7 A_1 A_2 + \varepsilon$.

Scenario	$\boldsymbol{\lambda}$	$\boldsymbol{\delta}$	Type
1	$(0, 0, 0, 0, 0, 0, 0)^T$	$(0.5, 0.5)^T$	non-regular
2	$(0, 0, 0, 0, 0.01, 0, 0)^T$	$(0.5, 0.5)^T$	near non-regular
3	$(0, 0, -0.5, 0, 0.5, 0, -0.5)^T$	$(0.5, 0.5)^T$	non-regular
4	$(0, 0, -0.5, 0, 0.99, 0, -0.98)^T$	$(0.5, 0.5)^T$	near non-regular
5	$(0, 0, -0.5, 0, 1, 0.5, -0.5)^T$	$(1.0, 0)^T$	non-regular
6	$(0, 0, -0.5, 0, 0.25, 0.5, 0.5)^T$	$(0.1, 0.1)^T$	regular
7	$(0, 0, -0.25, 0, 0.75, 0.5, 0.5)^T$	$(0.1, 0.1)^T$	regular
8	$(0, 0, 0, 0, 1, 0, -1)^T$	$(0, 0)^T$	non-regular
9	$(0, 0, 0, 0, 0.25, 0, -0.24)^T$	$(0.5, 0.5)^T$	near non-regular

We compare the performance of the following methods to construct a confidence interval for the main effect of the first stage treatment ψ_{10} using dWOLS as method of analysis: (i) the standard n-out-of-n bootstrap; (ii) the m-out-of-n bootstrap using α=0.05; (iii) the m-out-of-n bootstrap using α=0.1, and (iv) the m-out-of-n bootstrap using double bootstrap to

choose α in a data-driven way. For each scenario, we use 1000 simulated datasets with a fixed sample size n=300. We fix the number of bootstrap replicated to B=1000. For the double bootstrap procedure, we use B_1=B_2=500 replications.

We report simulation results in terms of average resample size $\hat{m}$, coverage rates and average width of the confidence intervals for ψ_{10}. Table 3.2 shows the average resample sizes $\hat{m}$ obtained from the four methods to construct a confidence interval for ψ_{10}. As expected, for each scenario, the average resample sizes are smaller for the m-out-of-n bootstrap with fixed α=0.1 compared to the m-out-of-n bootstrap with fixed α=0.05 as larger values of the tuning parameter α yield smaller possible resample size $\hat{m}$. On average, the m-out-of-n bootstrap with adaptive choice of $\hat{m}$ chooses the tuning parameter α around 0.05, and so the corresponding resample sizes and other performance metrics are similar to those for the m-out-of-n bootstrap with fixed α=0.05. Figure 3.1 shows the corresponding coverage rates and average confidence interval width for the four methods. As anticipated, confidence intervals constructed via the standard bootstrap (solid lines) have the smallest width, on average. However, the corresponding coverage rates are below the nominal rate of 0.95, with the coverage rates being significantly different than the nominal coverage rate of 0.95 (marked in

Table 3.2: Average bootstrap resample size over the 1000 simulated data sets. The columns represent the nine different scenarios with the degree of non-regularity denoted by NR = non-regular, NNR = near non-regular, R = regular. The rows represent the four different methods of constructing CIs: (i) the regular n-out-of-n bootstrap (nn), (ii) the m-out-of-n bootstrap with fixed α=0.05 ($\text{mn}_{0.05}$), (iii) the m-out-of-n bootstrap with fixed α=0.1 ($\text{mn}_{0.1}$), and (iv) the m-out-of-n bootstrap with adaptive choice of α using double bootstrap. The average values of the tuning parameter α for the last method are presented in the last row.

CI	Method	Sc.1 *NR*	Sc.2 *NNR*	Sc.3 *NR*	Sc.4 *NNR*	Sc.5 *NR*	Sc.6 *R*	Sc.7 *R*	Sc.8 *NR*	Sc.9 *NNR*
(i)	nn	300	300	300	300	300	300	300	300	300
(ii)	$\text{mn}_{0.05}$	228.83	228.89	242.14	261.56	254.63	258.45	262.80	261.69	231.38
(iii)	$\text{mn}_{0.1}$	178.97	178.96	200.64	231.09	217.93	225.84	233.32	230.97	182.97
(iv)	$\text{mn}_{\hat{\alpha}}$	235.26	236.39	236.99	264.11	254.62	259.45	261.94	264.62	228.93
	$\hat{\alpha}$	(0.046)	(0.045)	(0.056)	(0.047)	(0.050)	(0.051)	(0.052)	(0.047)	(0.054)

bold) in most examples. The m-out-of-n bootstrap with fixed α=0.1 (dotted lines) yields the most conservative and wider confidence intervals with coverage rates larger than 0.95 in most examples. With fixed α=0.05 (dashed lines), the m-out-of-n bootstrap reduces conservatism in most examples, but exhibits under-coverage in two scenarios (5 and 9). Interestingly, all three m-out-of-n bootstrap approaches outperform the regular bootstrap in the two regular scenarios (6 and 7). We also considered B=5,000 and 10,000; results were stable (results not shown).

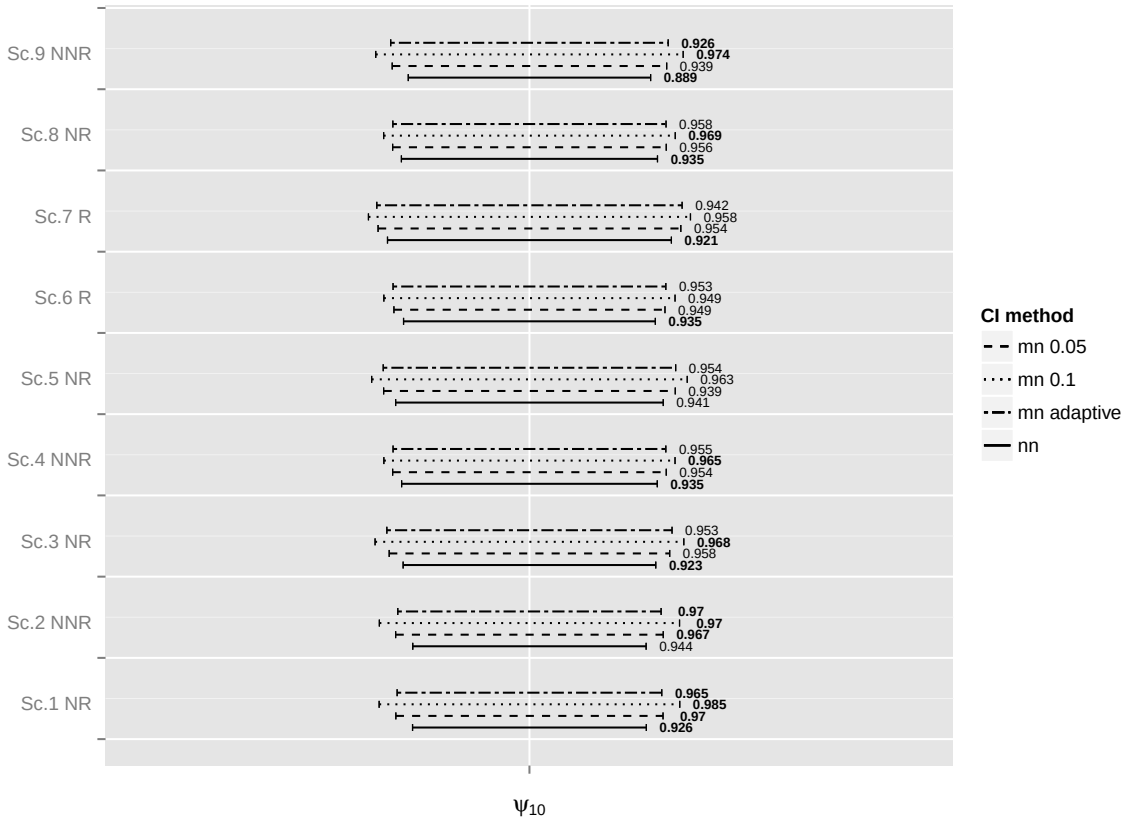

Figure 3.1: Monte Carlo estimates of the mean width of 95% confidence intervals for the main effect of treatment (ψ_{10}) for nine scenarios (y-axis) with corresponding degree of non-regularity denoted by NR = non-regular, NNR = near non-regular, R = regular. Four different methods for constructing CIs are presented: the regular n-out-of-n bootstrap (nn, solid lines), the m-out-of-n bootstrap with fixed α=0.05 (mn 0.05, dashed lines), the m-out-of-n bootstrap with fixed α=0.1 (mn 0.1, dotted lines), and the m-out-of-n bootstrap with adaptive choice of α using double bootstrap (mn adaptive, two-dash lines). Coverage rates are indicated on the right of the CIs, with coverages significantly different than 0.95 at significance level 0.05 marked in **bold**. The actual value of ψ_{10} varies across scenarios.

3.4 The PROBIT

The PROBIT study, a cluster-randomized trial, was initiated in the mid-1990s to investigate the effect of the promotion of exclusive breastfeeding on infant health outcomes (Kramer et al., 2001). Thirty-two hospitals in Belarus were paired according to hospital-level characteristics, and within each pair, one hospital was randomized to the experimental intervention while the other hospital was encouraged to continue with its usual practice in terms of breastfeeding education. Sociodemographic and clinical information on the mother and her infant was recorded at enrolment, and detailed information about infant health and feeding habits were collected at routine visits held at 1, 2, 3, 6, 9 and 12 mo. The PROBIT cohort was further followed, and health and development outcomes were collected when the children in the cohort were approximately 6.5 years old.

We used the PROBIT data to assess whether the introduction of solid food should be tailored to infant characteristics, considering three long-term health outcomes measured at 6.5 years of age: BMI, waist circumference, and triceps skinfold thickness. Note that the introduction of solid food was *not* randomized and thus the available data are "observational" (Moodie et al., 2009). Previous analyses (Kramer et al., 2009; Moodie et al., 2009; Wallace & Moodie, 2015), though not conducted in a DTR framework, have suggested that there may be an effect of the infant diet on long-term health outcomes, but that this effect is likely to be small. DTR parameters are thus likely to have non-regular limiting distributions.

We removed infants who were introduced to cereals or other type of solid food before the age of 3 mo as we believe they represent a different population from the one we aim to study. We defined the two stages of binary intervention as whether or not the infant received cereals between 3–6 mo (A_1) and between 6–9 mo (A_2). For example, an infant who was not feed with cereals between 3 and 6 mo, but received cereals between 6 and 9 mo would have A_1=0 and A_2=1. We consider three separate DTR analyses maximizing each outcome transformed to the negative scale assuming smaller BMI, waist circumference, and triceps skinfold thickness

values are preferred among children who are not underweight. At each stage, we considered infant weight at the start of the interval as the only tailoring variable, defining the blip functions as $\gamma_1(\boldsymbol{h_1}, a_1; \boldsymbol{\psi_1}) = a_1(\psi_{10} + \psi_{11} \times \text{infant weight at 3 mo})$ and $\gamma_2(\boldsymbol{h_2}, a_2; \boldsymbol{\psi_2}) = a_2(\psi_{20} + \psi_{21} \times \text{infant weight at 6 mo})$. We defined the treatment-free model linearly at each stage, depending on the following covariates: infant weight at the start of the interval, infant sex, mother's alcohol consumption and smoking status during the interval, mother's BMI, father's BMI, infant symptom indicators (rash, gastrointestinal and respiratory tract infection, other illnesses), and hospitalization of the infant. We include a cluster indicator in the treatment-free model to account for pairing of hospitals at randomization. We defines the treatment model at each stage as the probability of an infant receiving cereals at that stage, which depends on a wide variety of variables measured prior to the intervention: randomization arm, infant sex, infant weight at the start of the interval, mother's age, parity, mother's alcohol consumption and smoking status during the interval, gestational age, symptom indicators, hospitalization.

For each health outcome separately, we carried out two kind of analyses: a naive complete-case analysis and an analysis using inverse probability of censoring weights (IPCW) (Robins et al., 1995). An infant was considered as censored at a given stage if he has a missing value for the intervention at that stage. For the complete-case analysis, we removed an infant from the analysis at a given stage and in all previous stages if he has missing values in at least one covariate involved in the calculation of the treatment-free, treatment or blip model at that stage. For the IPCW analysis, the probability of censoring at each stage was modeled with a logistic regression using the full covariate history up to that point, except parental BMI since almost all infants who had missing value(s) in the outcome(s) also had missing values for parental BMI. To account for the clustered nature of the PROBIT data (clustering of children within hospital pairs), we used a stratified standard bootstrap (Bickel & Freedman, 1984), and a proportionally allocated, stratified m-out-of-n bootstrap (Müller & Welsh, 2009) to draw resamples to construct the confidence intervals, where stratification

is with respect to hospital pairs.

Table 3.3 presents a subset of the infant-mother characteristics observed at baseline and through the course of the study. Out of the 17,046 infant-mother pairs initially included in the PROBIT trial, 4,031 infants were excluded because they were introduced to solid food before 3 mo, and 28 infants were removed from the analysis because their BMI at 6.5 years old fell below the WHO threshold of severe thinness ("Growth reference 5-19 years", 2007), leaving a sample size of 12,987 infants. Missing values occurred more frequently in variables measured at the 6.5 years old follow-up visit, where more than 27% (n=3,582) of the infants had missing values in at least one of: BMI, waist circumference, triceps skinfold thickness, mother's BMI, or father's BMI.

Table 3.3: Baseline characteristics, stage-specific measurements, and measured outcomes for infant-mother pairs included in the PROBIT data analysis.

Characteristics† (*n=12,987*)		Missing values
Baseline		
Sex (% female)	48.1 (6,247)	0
Gestational age (in weeks)	39.4 (1.0)	0
Mother's age at birth	24.9 (0.4)	0
Mother's BMI	24.5 (4.3)	2,701
Father's BMI	25.7 (3.3)	3,582
Stage-specific		
Infant weight at 3 mo (in kg)	6.1 (0.7)	388
Infant weight at 6 mo (in kg)	8.1 (0.8)	526
Outcomes		
BMI (in kg/m^2)	15.6 (1.7)	2,601
Waist circumference (in cm)	54.5 (4.3)	2,597
Triceps skinfold (in mm)	10.0 (3.9)	2,599

† Reported as mean (standard deviation) or % (n)

The blip parameters estimated via the complete-case analysis are shown in Table 3.4. No significant effect of the the introduction of cereals or its interaction with infant weight was observed for any of the three outcomes. The non-significant effects of the second stage parameters suggest that the first stage estimators have non-regular distributions, and that the corresponding standard bootstrap confidence intervals may have poor coverage. As expected,

all m-out-of-n bootstrap confidence intervals are larger compared to the corresponding standard bootstrap confidence interval. To use these results to identify an optimal DTR with respect to a given outcome, the sign of $\hat{\psi}_{j0} + \hat{\psi}_{j1} \times \text{infant weight}_j$ should be calculated for each j stage, for $j = 1, 2$. For example, for an optimal value of BMI at 6.5 years old, only infants weighting more than $\frac{0.40}{0.06} = 6.67$ kilograms (kg) at 3 months should be introduced to solid food between 3 and 6 months (if we were to ignore the non-significance of the findings). Results with the IPCW analysis are very similar to the complete-case analysis (see the Supplemental Material B.3). Diagnostic plots did not raise any concerns regarding any of the fitted models (see the Supplemental Material B.4).

Table 3.4: Estimates of the blip parameters $(\psi_{10}, \psi_{11}, \psi_{20}, \psi_{21})$ in the PROBIT data analysis with three outcomes using the complete-case observations along with 95% confidence intervals calculated with standard bootstrap (nn), m-out-of-n bootstrap with α=0.05 ($\text{mn}_{0.05}$), m-out-of-n bootstrap with α=0.1 ($\text{mn}_{0.1}$) and m-out-of-n bootstrap with adaptive choice of α ($\text{mn}_{\hat{\alpha}}$).

Complete-case analysis					
		95% Confidence Interval			
Estimates		nn	$\text{mn}_{0.05}$	$\text{mn}_{0.1}$	$\text{mn}_{\hat{\alpha}}$
BMI (n_1†=8,910, n_2‡=9,144, $\hat{\alpha}$††=0.07)					
$\hat{\psi}_{10}$	-0.40	(-1.22; 0.42)	(-1.40; 0.60)	(-1.72; 0.93)	(-1.49; 0.69)
$\hat{\psi}_{11}$	0.06	(-0.10; 0.22)	(-0.12; 0.25)	(-0.16; 0.28)	(-0.14; 0.26)
$\hat{\psi}_{20}$	-0.55	(-1.84; 0.74)	(-2.19; 1.10)	(-2.56; 1.46)	(-2.39; 1.30)
$\hat{\psi}_{21}$	0.06	(-0.12; 0.24)	(-0.16; 0.28)	(-0.19; 0.31)	(-0.18; 0.30)
Waist Circumference (n_1=8,913, n_2=9,147, $\hat{\alpha}$=0.08)					
$\hat{\psi}_{10}$	0.37	(-1.77; 2.52)	(-2.37; 3.11)	(-2.71; 3.46)	(-2.51; 3.26)
$\hat{\psi}_{11}$	-0.09	(-0.45; 0.27)	(-0.54; 0.36)	(-0.59; 0.42)	(-0.56; 0.38)
$\hat{\psi}_{20}$	-1.46	(-4.87; 1.95)	(-5.86; 2.94)	(-6.83; 3.92)	(-6.39; 3.47)
$\hat{\psi}_{21}$	0.22	(-0.22; 0.65)	(-0.33; 0.77)	(-0.45; 0.89)	(-0.40; 0.83)
Triceps Skinfold Thickness (n_1=8,911, n_2=9,145, $\hat{\alpha}$=0.08)					
$\hat{\psi}_{10}$	-1.38	(-3.11; 0.35)	(-3.61; 0.85)	(-3.96; 1.20)	(-3.90; 1.14)
$\hat{\psi}_{11}$	0.22	(-0.07; 0.51)	(-0.15; 0.58)	(-0.20; 0.64)	(-0.20; 0.63)
$\hat{\psi}_{20}$	-0.40	(-3.61; 2.81)	(-4.53; 3.73)	(-5.51; 4.71)	(-5.23; 4.42)
$\hat{\psi}_{21}$	0.05	(-0.35; 0.47)	(-0.46; 0.58)	(-0.58; 0.70)	(-0.55; 0.67)

† Sample size at first stage, ‡ Sample size at second stage, †† Adaptive α
Using double bootstrap

3.5 Discussion

Inferential methods for DTR are typically either robust but theoretically involved (and sometimes also hard to implement), or practically accessible at the cost of relying on stronger modelling assumptions. The dWOLS approach has reunited the robustness of G-estimation and the accessibility of Q-learning into a theoretically accessible, easy to implement, robust and checkable framework. We have further enriched dWOLS with tools for valid inference in any situations, notably in non-regular situations where the effect of a treatment effect is weak or null for all or some subjects. Specifically, when the effect of an intervention is small at a given stage, the dWOLS estimators in previous stages are likely to have non-regular distributions, and standard bootstrap confidence intervals may exhibit poor coverage. With simulations, we showed that constructing confidence intervals with the m-out-of-n bootstrap procedure provides an accurate solution for remedying the standard bootstrap inconsistency.

Small stage-specific effects are anticipated in experimental randomized controlled trials, or in large-scale observation studies of treatment sequences. Our application to the PROBIT dataset was an example of a real situation where DTR estimators are likely to have non-regular distributions. We investigated whether the timing of the introduction of solid food into infant diets affected three long-term health outcomes. We found that solid food intake between 6 and 9 mo had a small, non-significant effect on BMI, waist circumference or triceps skinfold thickness, such that the main effect estimator of solid food intake between 3 and 6 mo was likely to have a non-regular distribution.

Our analysis is subject to some limitations. In particular, in the PROBIT data, it was not known precisely within each interval when solid food was introduced. This may introduce some partial misclassification in the exposure, with children breastfed for very little or nearly all of an interval classified in the same exposure group. Moreover, infant growth during the first year of life is likely a highly dynamic process, in which infant-feeding habits and infant

general health characteristics may affect one another in close succession, so that the length of the intervals considered in our analysis may be too coarse to capture the dynamic nature of infant growth and maternal decisions about continued breastfeeding. Finally, there may be unmeasured confounders affecting the conclusion of our analyses, e.g. maternal education regarding or attitude towards nutrition (Kramer et al., 2002).

The m-out-of-n bootstrap has been used widely in the literature to correct for the inconsistency of the standard bootstrap procedure in some situations, such as in non-regular cases. In the broad class of problems in which a statistic has a discontinuity in its limiting distribution, the m-out-of-n bootstrap does not always work in the sense that the resulting confidence intervals may dramatically exceed the nominal level (Andrews & Guggenberger, 2010). Moreover, theoretical justifications for the choice of m are merely based on asymptotic conditions on m as a function of n, and it is necessary to rely on data-driven methods for the selection of m (Bickel et al., 1997). As the validity of the m-out-of-n bootstrap is not guaranteed in all settings, and as the choice of m may have a considerable impact on the performance of the procedure, it is necessary to evaluate to validate the m-out-of-n bootstrap as a tool for making valid inference in each new application as we did in this work with dWOLS.

We proposed a methodology, the m-out-of-n bootstrap, for constructing valid confidence intervals for the DTR parameters with dWOLS as method of analysis. As expected, the class of resample sizes introduced in Chakraborty et al. (2013) has proven to be particularly well fitted for dWOLS. The adaptive class of resample sizes presented in this work is defined as a function of an estimate of the non-regularity in the data and a user-specified tuning parameter α. As for other methods that require selecting a tuning parameter, general recommendations on how to choose α are dangerous, and cross-validation should be used to select the best value α in every application. The double bootstrap algorithm, implemented in the `DTRreg R` package, provides a convenient cross-validation tool for choosing the tuning parameter α in

a data-driven adaptive way, and we strongly encourage future users to rely on this approach for choosing α.

Software

In an effort to promote reproducible research, scripts in the form of R code are available at https://github.com/gabriellesimoneau/Rcode-Biostatistics. As the PROBIT dataset cannot be shared, an example dataset has been created and sample output provided to show how to reproduce the analyses in Section 3.4. For questions, comments or remarks about the shared code, contact the corresponding author (gabrielle.simoneau@mail.mcgill.ca).

Acknowledgements

We thank Michael Kramer and the PROBIT study investigators for sharing the PROBIT dataset. *Conflict of Interest*: None declared.

Funding

The Promotion of Breastfeeding Intervention Trial was supported by grants from the Thrasher Research Fund; the National Health Research and Development Program (Health Canada); UNICEF; the European Regional Office of WHO; and the CIHR. This work was supported by a doctoral research grant from the Fonds de Recherche du Québec Nature et technologies (FRQNT) [199803].

Appendix B – Supplemental Materials

Contains the following sections:

B.1 **Adaptive Choice of m** – details on the adaptive procedure for choosing the resample size $\hat{m}$.

B.2 **Details of the Data Generating Process used in the Simulation Study** – details and calculation examples for regular and non-regular simulation scenarios.

B.3 **PROBIT: The IPCW Analysis** – results with the IPCW analysis of the PROBIT application.

B.4 **PROBIT: Diagnostic Plots** – diagnostic plots for the PROBIT application.

Chapter 4

Estimating Optimal Dynamic Treatment Regimes With Survival Outcomes

Preamble to Manuscript 2. While the work presented in the previous manuscript built on one specific gap in dWOLS, namely the lack of inferential tools in non-regular scenarios, the idea for this second project was to further generalize dWOLS to accommodate censored outcomes. Extending dWOLS to time-to-event data yielded a new method called dynamic weighted survival modeling (DWSurv), introduced in this manuscript. The original contributions are (i) proposing an extended class of balancing weights that further account for the censoring mechanism, (ii) proposing an algorithm to derive doubly-robust estimators of the treatment effect and its interactions with tailoring variables in a general K-stage DTR, (iii) establishing the consistency and double-robustness of the estimators, (iv) deriving the asymptotic distribution of the estimators, and (v) designing simulation studies for two-stage DTRs. The manuscript presented in this chapter was accepted for publication in the *Journal of the American Statistical Association* and was published online at the time of submitting this thesis. DWSurv is implemented in `R` as part of the `DTRreg` package and is available on CRAN.

Estimating Optimal Dynamic Treatment Regimes With Survival Outcomes

Gabrielle Simoneau[1], Erica EM Moodie[1], Jagtar S Nijjar[2], Robert W Platt[1], the Scottish Early Rheumatoid Arthritis Inception Cohort Investigators

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

[2]*Department of Medicine, University of Cambridge, Cambridge, United Kingdom*

Abstract

The statistical study of precision medicine is concerned with dynamic treatment regimes (DTRs) in which treatment decisions are tailored to patient-level information. Individuals are followed through multiple stages of clinical intervention, and the goal is to perform inferences on the sequence of personalized treatment decision rules to be applied in practice. Of interest is the identification of an optimal DTR, that is, the sequence of treatment decisions that yields the best expected outcome. Statistical methods for identifying optimal DTRs from observational data are theoretically complex and not easily implementable by researchers, especially when the outcome of interest is survival time. We propose a doubly robust, easy to implement method for estimating optimal DTRs with survival endpoints subject to right-censoring which requires solving a series of weighted generalized estimating equations. We provide a proof of consistency that relies on the balancing property of the weights and derive a formula for the asymptotic variance of the resulting estimators. We illustrate our novel approach with an application to the treatment of rheumatoid arthritis using observational data from the Scottish Early Rheumatoid Arthritis Inception Cohort. Our method, called dynamic weighted survival modeling, has been implemented in the `DTRreg` `R` package. Supplementary materials for this article are available online.

4.1 Introduction

Precision medicine is an approach to health care in which treatment decisions are tailored to patient-level information. This approach is especially relevant in the chronic-care environment where a patient's health condition is changing over time, and treatments must correspondingly be altered. One such example of a chronic condition is rheumatoid arthritis (RA). Patients typically experience recurrent episodes of high disease activity followed by periods of remission during their lifetime. An important clinical question is to determine which sequence of treatments minimizes a patient's time to remission based on observed characteristics at the time he/she enters an episode of disease flare-up.

Such adaptive sequences of treatments are called dynamic treatment regimes (DTRs). In the typical DTR setting, individuals are followed through multiple stages of clinical intervention and the statistical goal is to perform estimation and inference on the sequence of treatment decision rules, one at each stage, which uses the individual's characteristics as inputs and yields a recommended treatment. Of particular interest is the identification of an optimal DTR, that is, the sequence of treatment decisions that yields the best expected outcome. In the RA example, an optimal DTR would yield the shortest time to remission in subgroups of individuals with similar characteristics.

Guidelines from the American College of Rheumatology (ACR) recommend an adaptive sequence of treatments to achieve remission in patients with RA (Singh et al., 2016). Recommendations were designed by a group of highly specialized clinicians and epidemiologists based on the balance of relative benefits and harms of the treatment options, and the quality of evidence from the literature. Yet, the evidence derived from single-stage clinical trials or observational studies that compared first- and second-line treatments separately does not account for the dynamic nature of the treatment of RA. This might have profound implications on the identification of optimal regimes; focusing on the efficacy of the first-line treatment may unknowingly set patients on regimes that preclude more effective later-stage treatments,

thus the importance of using adequate statistical methods.

In observational data, the fact that the first-line treatment affects the outcome but also subsequent patient characteristics and treatments, which are in turn predictors of the outcome, leads to a complex interplay between treatments and outcomes, making the discovery of an optimal DTR a challenging task. When the optimization criterion is based on maximizing time-to-event, the possibility of right-censoring complicates the estimation procedure because patients do not necessarily enter all stages of clinical intervention if they experience an event or are censored before reaching the end of the study.

There is a substantial statistical literature on methods to identify an optimal DTR from observational data (e.g. Murphy, 2003; Robins, 2004; B. Zhang et al., 2013; Y. Q. Zhao et al., 2015) yet only a few methods have been proposed to accommodate time-to-event endpoints subject to right-censoring. Q-learning has been extended for censored data to find treatment regimes that lead to longer survival time (Goldberg & Kosorok, 2012). The method requires positing parametric models for the survival time at each stage of intervention and making predictions under optimal treatment assignments. Its approach to modeling is simple and intuitive, using inverse probability of censoring weights (IPCW) to account for censoring. However, it lacks robustness to model misspecification, it is currently only implemented in MatLab and it assumes that censoring is independent of observed trajectories. A related method uses accelerated failure time (AFT) to model the survival time at each stage of intervention (Huang et al., 2014) – a model that could equally be adopted in Q-learning – however, this approach requires making predictions only for individuals who did not receive their optimal treatment. Although the method does not make the restrictive assumption of random (covariate-independent) censoring, it is singly-robust and not implemented in a statistical software package. G-estimation is an alternative regression-based approach to uncover optimal DTR with survival outcomes but it remains unpopular given its theoretical complexity and computational burden (Hernán & Robins, 2010; Robins & Greenland, 1994).

A number of value search or classification-based methods have been proposed with survival endpoints (Bai et al., 2017; Hager et al., 2018; Jiang et al., 2017a; Y. Q. Zhao et al., 2014b) but they provide a less intuitive approach to both modeling and model-checking and often lack tools for valid inference on the decision rules themselves.

This article provides a new, theoretically robust approach to estimation of an optimal DTR with survival outcomes subject to right-censoring. Our method extends the dynamic weighted ordinary least squares (dWOLS) approach (Wallace & Moodie, 2015) to time-to-event data and borrows from the singly robust framework set up by Huang et al. (2014). DWOLS is an easily implementable statistical method that unites the double-robustness of G-estimation and the simplicity of Q-learning when the outcome is continuous and uncensored. It accounts for nonrandomized treatment assignments with a broad class of weights. Our extension to time-to-event data incorporates a flexible number of stages of intervention, as in Q-learning, thus allowing individuals to experience an event or be censored before the end of the follow-up, and it allows censoring to depend on time-varying individual trajectories. The balancing weights introduced in dWOLS and extensively discussed by Li et al. (2018) are extended to incorporate IPCW. Our method, named dynamic weighted survival modeling (DWSurv), is doubly robust, easy to understand by non-statisticians, and has tools for inference and model-checking. We have implemented DWSurv in the `DTRreg` package in `R` (Wallace et al., 2017a).

We introduce our methodology in Section 4.2. Section 4.3 evaluates the performance of DWSurv and a related method in extensive simulation studies. Section 4.4 illustrates DWSurv in an application to the treatment of RA using the Scottish Early Rheumatoid Arthritis (SERA) Inception Cohort.

4.2 Methodology

With the outcome of interest being survival time, an optimal DTR aims to identify the sequence of decision rules that leads to longer survival time. Similar notations and principles apply when the goal is to minimize the time to an event, for example, minimize time to remission.

4.2.1 Notation and Assumptions

Unless specified otherwise, upper cases, lower cases and bold respectively denote random variables, realizations of random variables and vectors. Data needed to estimate a DTR consist of longitudinal trajectories of covariates and treatments across a maximum of J stages of clinical intervention. Let individuals be identified with a subscript $i = 1, ..., n$ (often dropped for clarity) and stages be denoted by a second subscript $j = 1, ..., J$. Let η_j be a random variable which takes value 1 if a participant entered stage j and 0 otherwise, where all should have at least η_1=1. Let $A_j \in (0, 1)$ denote the treatment received at the beginning of stage j. Let T_j be the survival time within stage j, with T_j set to zero when η_j=0. The outcome is the overall survival time defined as the sum of the stage-specific survival times $T = \sum_{j=1}^{J} \eta_j T_j$. We define the counterfactual survival time $T^{\mathbf{a}} = \sum_{j=1}^{J} \eta_j T_j^{\boldsymbol{a_j}}$ where $T_j^{\boldsymbol{a_j}}$ denotes the potential survival time within stage j if, possibly contrary to fact, an individual were given treatments $\boldsymbol{a_j} = (a_1, ..., a_j) \in (0, 1)^j$, with $\mathbf{a} = \mathbf{a_J}$. The censoring time is denoted by C. Let $Y = \min(T, C)$ and let $\Delta = \mathbb{I}(T \leq C)$ denote the failure indicator. Let $\boldsymbol{X_j}$ be a vector of covariates measured prior to the jth treatment. Denote an individual's history by $\boldsymbol{H_j}$ taking values $\boldsymbol{h_j} \in \mathcal{H}_j$, the sample space for $\boldsymbol{H_j}$, which represents a shorthand for the information available prior to making the jth treatment decision, including (functions of) previous treatment assignments, covariates, and survival times. The observed data are given by the individual trajectories $(\eta_{i1}, \boldsymbol{X_{i1}}, A_{i1}, T_{i1}, ..., \eta_{iJ}, \boldsymbol{X_{iJ}}, A_{iJ}, T_{iJ}, \Delta_i)$, where A_{ij} and

$\boldsymbol{X_{ij}}$ are missing when η_{ij}=0. A DTR consists of a set of decision rules $\{d(\boldsymbol{h_1}), ..., d(\boldsymbol{h_J})\} \in \mathcal{D}$ where $\mathcal{D}$ denotes the class of all possible treatment regimes. At each stage j, the decision rule is a function $d(\boldsymbol{h_j}) : \mathcal{H}_j \rightarrow (0, 1)$ that inputs the observed history $\boldsymbol{h_j}$ and outputs a treatment. An optimal DTR is the set of decision rules $\{d^{\text{opt}}(\boldsymbol{h_1}), ..., d^{\text{opt}}(\boldsymbol{h_J})\}$ that maximizes the overall expected survival time $\mathbb{E}(T^{\mathbf{a}})$.

To identify an optimal DTR, we rely on the axiom of consistency to link counterfactuals to observed data. We also make the following assumptions:

1. Stable unit treatment value (Rubin, 1980) – it requires that an individual's outcome is not influenced by other individuals' treatment allocation;

2. Sequential ignorability (Robins, 2000b) – it is the no unmeasured confounder assumption extended to longitudinal settings, which further requires that the treatment assignment at a given stage cannot depend on future covariates. It is expressed as $\left\{\sum_{k\geq l}^{J} T_k^{\boldsymbol{a_k}} : l = j, \ldots, J\right\} \perp\!\!\!\perp A_j | \boldsymbol{H_j}, \eta_1, \ldots, \eta_j$;

3. Coarsening at random (Gill et al., 1997) – it assumes that, at the beginning of each stage, the probability of censoring onwards is independent of future outcomes, given accrued information. It is expressed as $\left\{\sum_{k\geq l}^{J} T_k^{\boldsymbol{a_k}} : l = j, \ldots, J\right\} \perp\!\!\!\perp \Delta | \boldsymbol{H_j}, \eta_1, \ldots, \eta_j$.

4.2.2 Definition of Optimal Dynamic Treatment Regimes

For simplicity, we define an optimal DTR with up to two stages of intervention and denote the optimal stage 1 and stage 2 decision rules, $d^{\text{opt}}(\boldsymbol{h_1})$ and $d^{\text{opt}}(\boldsymbol{h_2})$, with the shorthand a_1^{opt} and a_2^{opt}, respectively. It is straightforward to extend the following derivations and results to two stages or more.

Like Q-learning, our approach relies on backward induction to estimate a sequence of treatments that maximizes survival time. As a first step, the optimal stage 2 decision rule a_2^{opt} is estimated by considering the effect of the stage 2 treatment and its interactions with tai-

loring variables on a function $f(\cdot)$ of the expected counterfactual survival time from stage 2 onwards, $\mathbb{E}[f(T_2^{a_1,a_2})]$. Only individuals who entered the second stage contribute to the estimation of the stage 2 decision rule. In a second step, the optimal stage 1 decision rule is estimated by considering the effect of the stage 1 treatment and its interactions with tailoring variables on a function of the expected counterfactual overall survival time had the second stage treatment been optimal, $\mathbb{E}[f(T^{a_1,a_2^{\text{opt}}})]$. With the aim of maximizing survival time, this step requires to construct counterfactual survival times under a_2^{opt} by adding a positive quantity to the observed survival times of the individuals who did not receive their optimal stage 2 treatment. The first stage treatment comparison is then "fair" as it is with respect to an overall survival time that incorporates the effect of the stage 2 treatment, taken to be optimal for everybody.

Formally, starting with the second stage of intervention, we define the second stage treatments comparison through the *blip* function

$$\gamma_2(a_2, \boldsymbol{h_2}) = \mathbb{E}[\log(T_2^{a_1,a_2}) - \log(T_2^{a_1,0})|\eta_2 = 1, \boldsymbol{H_2} = \boldsymbol{h_2}]$$

where we consider $f(x) = \log(x)$, stretching the domain to $\mathbb{R}$. The stage 2 blip function is interpreted as the difference between the expected log-survival time within the second stage of an individual who received some treatment a_2 at stage 2 and the expected log-survival time within stage 2 of the same individual had he received some reference treatment a_2=0, conditional on reaching the second stage and on his observed history $\boldsymbol{h_2}$. The blip function needs to satisfy $\gamma_2(0, \boldsymbol{h_2}) = 0$. The optimal stage 2 treatment is that which maximizes the blip $a_2^{\text{opt}} = \arg\max_{a_2}\gamma_2(a_2, \boldsymbol{h_2})$. Note that a_2^{opt} is a function of the individual histories such that it may depend on the first and second stage covariates as well as on the first stage treatment.

Next, consider the optimization of the first stage treatment. The comparison of the stage 1 treatments is based on the hypothetical situation in which each individual who entered stage

2 would have received their optimal stage 2 treatment a_2^{opt}. Let $T^{a_1,a_2^{\text{opt}}}$ denote the pseudo-overall survival time, hereafter referred to as pseudo-outcome, had an individual received his optimal stage 2 treatment. It is defined as $T^{a_1,a_2^{\text{opt}}} = T_1^{a_1} + \eta_2 T_2^{a_1,a_2^{\text{opt}}}$ to make explicit that individuals who did not enter the second stage have their pseudo-outcome equal to their overall survival time, which is equal to the time spent in the first stage. We define the first stage treatments comparison through the blip function

$$\gamma_1(a_1, \boldsymbol{h_1}) = \mathbb{E}[\log(T^{a_1,a_2^{\text{opt}}}) - \log(T^{0,a_2^{\text{opt}}})|\boldsymbol{H_1} = \boldsymbol{h_1}].$$

The stage 1 blip function is also interpreted as a difference of expected log-transformed counterfactual outcomes and constrained to $\gamma_1(0, \boldsymbol{h_1}) = 0$. The optimal stage 1 treatment is that which maximizes the blip $a_1^{\text{opt}} = \arg\max_{a_1} \gamma_1(a_1, \boldsymbol{h_1})$.

4.2.3 Accelerated Failure Time Specification

We operationalize the previous optimization procedure by specifying (semi-)parametric models for the blip functions. AFT models are a natural choice for clinical decision-making, as the modeling is performed and treatment strategies are compared on the scale of interest: the expected survival time. We assume an AFT model for $\log(T_2^{a_1,a_2})$ as

$$\log(T_2^{a_1,a_2}) = f_2(\boldsymbol{h_{2\beta}}; \boldsymbol{\beta_2}) + a_2 g_2(\boldsymbol{h_{2\psi}}; \boldsymbol{\psi_2}) + \epsilon_2$$

where the errors ϵ_2 are independent and identically distributed (i.i.d.) across subjects although more flexible forms such as splines could be incorporated into the model. The distribution of the errors is left unspecified with $\mathbb{E}(\epsilon_2) = 0$. Note that if the errors are not centered at zero, any deviation is absorbed in the intercept. The model for $\log(T_2^{a_1,a_2})$ is separated in two parts: a stage 2 treatment-free component $f_2(\boldsymbol{h_{2\beta}}; \boldsymbol{\beta_2})$ for any function f_2 which depends on (a subset of) the stage 2 history $\boldsymbol{h_2}$ but not on the stage 2 treatment, and

a stage 2 treatment effect component $a_2 g_2(\boldsymbol{h_{2\psi}}; \boldsymbol{\psi_2})$ for any function g_2 which depends on (a potentially different subset of) the stage 2 history and specifically includes the main effect of treatment a_2 and its interactions with tailoring variables. A typical choice of parametrization is a linear function

$$\log(T_2^{a_1,a_2}) = \boldsymbol{\beta_2^T h_{2\beta}} + a_2 \boldsymbol{\psi_2^T h_{2\psi}} + \epsilon_2. \tag{4.1}$$

Under assumptions 1-2, the previous parametrization implies a specific form of the stage 2 blip function given by

$$\begin{aligned}
\gamma_2(a_2, \boldsymbol{h_2}; \boldsymbol{\psi_2}) &= \mathbb{E}[\log(T_2^{a_1,a_2}) - \log(T_2^{a_1,0}) | \eta_2 = 1, \boldsymbol{H_2} = \boldsymbol{h_2}; \boldsymbol{\psi_2}] \\
&= \boldsymbol{\beta_2^T h_{2\beta}} + a_2 \boldsymbol{\psi_2^T h_{2\psi}} - (\boldsymbol{\beta_2^T h_{2\beta}} + 0 \times \boldsymbol{\psi_2^T h_{2\psi}}) \\
&= a_2 \boldsymbol{\psi_2^T h_{2\psi}}.
\end{aligned}$$

We identify the optimal stage 2 treatment for each individual who entered the second stage by $a_2^{\text{opt}} = \mathbb{I}(\boldsymbol{\psi_2^T h_{2\psi}} > 0)$.

Now consider the construction of the counterfactual survival time under optimal stage 2 treatment $\tilde{T}^{a_1, a_2^{\text{opt}}}$ as

$$\tilde{T}(\boldsymbol{\psi_2}) := T_1 + \eta_2 \left(T_2 \times \exp\{\boldsymbol{\psi_2^T h_{2\psi}}[a_2^{\text{opt}} - a_2]\} \right)$$

emphasizing its dependency on the parameters in (4.1). An individual who received his optimal stage 2 treatment has the term inside the $\exp(\cdot)$ equal to zero and his pseudo-outcome equal to his observed survival time, that is, $\tilde{T}(\boldsymbol{\psi_2}) = T$. An individual who did not receive his optimal treatment has the term inside the $\exp(\cdot)$ greater than zero and his pseudo-outcome larger than his observed outcome, that is, $\tilde{T}(\boldsymbol{\psi_2}) > T$. An individual who did not enter the second stage has his pseudo-outcome equal to his observed outcome.

The optimization of the first stage treatment proceeds in a similar manner but using the

counterfactual survival time under optimal stage 2 treatment $\tilde{T}^{a_1,a_2^{\text{opt}}}$ as criterion of optimality. We assume an AFT model for the pseudo-outcome, for example, as in (4.1) as

$$\log(\tilde{T}^{a_1,a_2^{\text{opt}}}) = \boldsymbol{\beta_1^T h_{1\beta}} + \boldsymbol{\psi_1^T} a_1 \boldsymbol{h_{1\psi}} + \epsilon_1 \tag{4.2}$$

where the errors ϵ_1 are i.i.d. with distribution left unspecified. The model is also separated in two parts: a stage 1 treatment-free model $\boldsymbol{\beta_1^T h_{1\beta}}$ and a stage 1 treatment effect model $\boldsymbol{\psi_1^T} a_1 \boldsymbol{h_{1\psi}}$. As above, this parametrization yields a similar specific form for the stage 1 blip function and optimal decision rule.

This demonstrates that positing an AFT model can be viewed as considering a restricted class of regimes D_ψ whose elements are indexed by the parameters $\boldsymbol{\psi} = (\boldsymbol{\psi_1}, \boldsymbol{\psi_2})$ involved in the decision rules. The form of the decision rules resulting from the proposed linear parametrization $a_j^{\text{opt}} = \mathbb{I}(\boldsymbol{\psi_j^T h_{j\psi}} > 0)$ is motivated by interpretability and feasibility in practice. For instance, decision rules involving cut-offs are natural and easy to implement in clinical practice.

4.2.4 Estimation and Inference

Interest lies in the estimation of the parameters $\boldsymbol{\psi}$ involved in the decision rules. The estimation procedure needs to account for right-censoring. For this, we use IPCW methods (Robins & Rotnitzky, 1992) with weights proportional to $\mathbb{P}(\Delta = 1|\boldsymbol{H_j}, A_j, \eta_j = 1)$. The IPCW create a pseudo-population with the same size and same distribution of baseline covariates of the original study population with $\eta_j = 1$ by replacing the censored individuals by copies of the uncensored individuals with similar treatments and covariates (Hernán & Robins, 2010). The estimation procedure also accounts for nonrandomized treatment assignments, also via a weighting argument. A positivity assumption is required: at each stage, $P(A_j = a_j|\boldsymbol{H_j}, \eta_j = 1) > 0$ for all treatment options a_j and $\mathbb{P}(\Delta = 1|\boldsymbol{H_j}, A_j, \eta_j = 1) > 0$.

The following algorithm details the estimation procedure:

1. Specify two parametric models for the probability of treatment and the probability of censoring within stage 2 respectively denoted by $\mathbb{P}(A_2 = 1|\boldsymbol{H_2}, \eta_2 = 1; \boldsymbol{\alpha_2})$ and $\mathbb{P}(\Delta = 0|\boldsymbol{H_2}, A_2, \eta_2 = 1; \boldsymbol{\lambda_2})$.

2. Specify weights $w_2(\delta, a_2, \boldsymbol{h_2}; \hat{\boldsymbol{\alpha}}_{\boldsymbol{2}}, \hat{\boldsymbol{\lambda}}_{\boldsymbol{2}})$ and estimate the stage 2 parameters $(\boldsymbol{\beta_2}, \boldsymbol{\psi_2})$ by solving the following weighted generalized estimating equations (GEE)

$$U_2(\boldsymbol{\psi_2}, \boldsymbol{\beta_2}) = \sum_{i=1}^{n} \delta_i \eta_{i2} \hat{w}_{i2} \begin{pmatrix} \boldsymbol{h_{i2\beta}} \\ a_{i2}\boldsymbol{h_{i2\psi}} \end{pmatrix} \left(\log(T_{i2}) - \boldsymbol{\beta_2^T h_{i2\beta}} - a_{i2}\boldsymbol{\psi_2^T h_{i2\psi}}\right) = 0. \quad (4.3)$$

   Note that (4.3) uses outcomes only for those individuals for whom δ=1.

3. Construct stage 1 pseudo-outcome as

$$\tilde{T}(\hat{\boldsymbol{\psi}}_{\boldsymbol{2}}) := T_1 + \eta_2 \left(T_2 \times \exp\{\hat{\boldsymbol{\psi}}_{\boldsymbol{2}}^{\boldsymbol{T}} \boldsymbol{h_{2\psi}}[\mathbb{I}(\hat{\boldsymbol{\psi}}_{\boldsymbol{2}}^{\boldsymbol{T}} \boldsymbol{h_{2\psi}} > 0) - a_2]\}\right).$$

4. Specify two parametric models for the probability of treatment within stage 1 and the probability of censoring from stage 1 onwards, respectively, denoted by $\mathbb{P}(A_1 = 1|\boldsymbol{H_1}; \boldsymbol{\alpha_1})$ and $\mathbb{P}(\Delta = 0|\boldsymbol{H_1}, A_1; \boldsymbol{\lambda_1})$.

5. Specify weights $w_1(\delta, a_1, \boldsymbol{h_1}; \hat{\boldsymbol{\alpha}}_{\boldsymbol{1}}, \hat{\boldsymbol{\lambda}}_{\boldsymbol{1}})$ and estimate the stage 1 parameters $(\boldsymbol{\beta_1}, \boldsymbol{\psi_1})$ by solving the following weighted GEE

$$U_1(\boldsymbol{\psi_1}, \boldsymbol{\beta_1}; \hat{\psi}_2) = \sum_{i=1}^{n} \delta_i \hat{w}_{i1} \begin{pmatrix} \boldsymbol{h_{i1\beta}} \\ a_{i1}\boldsymbol{h_{i1\psi}} \end{pmatrix} \left(\log\{\tilde{T}_i(\hat{\psi}_2)\} - \boldsymbol{\beta_1^T h_{i1\beta}} - a_{i1}\boldsymbol{\psi_1^T h_{i1\psi}}\right) = 0. \quad (4.4)$$

The form of the weights w_1 and w_2 must satisfy the balancing property stated in the theorem below (proof in Supplemental Material C.1).

Theorem 1 (balancing property). Under assumptions 1–3, solving the weighted GEE (4.3) and

(4.4) will yield consistent estimate of $\boldsymbol{\psi}$ if the weights satisfy the balancing property

$$\begin{aligned}[1-g(0,\boldsymbol{h_j})][1-\pi(\boldsymbol{h_j})]w_j(0,0,\boldsymbol{h_j}) &= g(0,\boldsymbol{h_j})[1-\pi(\boldsymbol{h_j})]w_j(0,1,\boldsymbol{h_j}) \\ = [1-g(1,\boldsymbol{h_j})]\pi(\boldsymbol{h_j})w_j(1,0,\boldsymbol{h_j}) &= g(1,\boldsymbol{h_j})\pi(\boldsymbol{h_j})w_j(1,1,\boldsymbol{h_j}) \end{aligned} \tag{4.5}$$

where $\pi(\boldsymbol{h_j}) = \mathbb{P}(A_j = 1|\boldsymbol{H_j} = \boldsymbol{h_j}, \eta_j = 1)$ and $g(a_j, \boldsymbol{h_j}) = \mathbb{P}(\Delta = 1|\boldsymbol{H_j} = \boldsymbol{h_j}, A_j = a_j, \eta_j = 1)$, for $j = 1, 2$.

The balancing property defines an entire family of weights. For example, the overlap weights

$$w_j(\delta, a_j, \boldsymbol{h_j}) = \frac{|a_j - \mathbb{P}(A_j = 1|\eta_j = 1, \boldsymbol{h_j})|}{\mathbb{P}(\Delta = \delta|\eta_j = 1, a_j, \boldsymbol{h_j})}$$

satisfy (4.5), extending a form of weights previously introduced in the context of uncensored outcomes (Li et al., 2018; Wallace & Moodie, 2015). They place more emphasis on individuals with treatment probability 1/2, thus defining a target population of substantive clinical interest, that is, the individuals whose characteristics could appear in any treatment group with equal probability (Li et al., 2018). IPCW further gives importance to individuals who were less likely to experience an event.

The consistency and asymptotic normality of the blip estimators $\hat{\boldsymbol{\psi}}_\mathbf{1}$ and $\hat{\boldsymbol{\psi}}_\mathbf{2}$ can be established under standard regularity conditions and the additional assumption that optimal stage 2 treatments are unique for all subjects (Moodie & Richardson, 2010) (see Supplemental Material C.2). The estimation procedure offers the double-robustness property. At each stage of optimization j, solving the corresponding weighted GEE yields unbiased estimators of the parameter $\boldsymbol{\psi_j}$ when either or both the treatment-free model $\boldsymbol{\beta_j^T h_j}$ or the weighting models, which include the treatment model $\mathbb{P}(A_j = 1|\boldsymbol{H_j}, \eta_j = 1; \boldsymbol{\alpha_j})$ and the censoring model $\mathbb{P}(\Delta = 0|\boldsymbol{H_j}, A_j, \eta_j = 1; \boldsymbol{\lambda_j})$, are correctly specified provided that the form of the decision rule $a_j\boldsymbol{\psi_j^T h_{j\psi}}$ is correct. We derive a formula for the asymptotic variance of the estimators $\hat{\boldsymbol{\psi}}_\mathbf{1}$ and $\hat{\boldsymbol{\psi}}_\mathbf{2}$ by performing a first-order Taylor expansion of the GEE about the

limiting distributions of the nuisance parameters (Moodie, 2009; Robins, 2004).

Note that the construction of the pseudo-outcome might create impossible pseudo-observation times if the pseudo-outcome exceeds the maximum possible follow-up time (see examples in Hernán et al. (2005)). In the context of G-estimation, artificial censoring has been proposed as a solution to correct for this, as such "impossible" pseudo-outcomes are not only unsatisfying but lead to violations of the assumption of independent censoring and observation times. However, this artificial censoring has also been viewed as a major drawback of G-estimation (Joffe, 2001; Joffe et al., 2012), and has not been adopted in related approaches such as that of Huang et al. (2014); we follow these authors and do not implement artificial censoring.

4.3 Simulations

We conducted a simulation study to compare the performance of DWSurv to the method by Huang et al. (2014), hereafter referred to as the method by HNW. Unlike DWSurv, this approach models the censoring time distribution rather than the censoring probability. Also, it is not robust to model misspecification, but instead requires correct specification of the event-time models. The simulation study aimed to (i) evaluate the accuracy, precision and associated inference of the blip estimators, (ii) evaluate the ability to identify the true optimal DTR, and (iii) compare the distribution of the survival time under treatment assignment following the true optimal DTR, the estimated DTR by the two methods and any fixed treatment strategy.

We simulated data from an observational study with two stages of intervention. Denote $\text{expit}(v) = \exp(v)/(1+\exp(v))$ defined for $v \in \mathbb{R}$ and $\text{logit}(u) = \text{expit}^{-1}(u)$ defined for $u \in (0,1)$. For individual i, the first stage treatment was assigned through a Bernoulli distribution with $P(A_{i1} = 1) = \text{expit}(-1 + 2X_{i1})$ where X_{i1} was a baseline continuous

covariate generated from a Uniform$(0.1, 1.29)$. Similarly, the assignment of the second stage treatment was based on a Bernoulli distribution with $P(A_{i2} = 1) = \text{expit}(2.8 - 2X_{i2})$ where X_{i2} was a continuous covariate measured at the beginning of the second stage generated from a Uniform$(0.9, 2)$. We generated Δ_i, the censoring indicator, and η_{i2}, the indicator of whether an individual entered the second stage, independently from Bernoulli distributions with probability 0.70 and 0.80, respectively.

For those who experienced an event and entered the second stage ($\Delta_i = \eta_{i2} = 1$), we used the AFT model in (4.1) to generate the survival time within the second stage as

$$T_{i2} = \exp(4 + 1.1X_{i2} - 0.2X_{i2}^3 - 0.1X_{i1} + A_{i2}(-0.9 + 0.6X_{i2}) + \epsilon_{i2})$$

where ϵ_{i2} had a Normal distribution centered at zero with variance 0.09. The true optimal stage 2 treatment A_{i2}^{opt}, given by $\mathbb{I}(-0.9 + 0.6X_{i2} > 0)$, was used to calculate the stage 2 survival time had everybody received their optimal stage 2 treatment as

$$T_{i2}^{\text{opt}} = \exp\{\log(T_{i2}) + (A_{i2}^{\text{opt}} - A_{i2})(-0.9 + 0.6X_{i2})\}.$$

For all individuals with $\Delta_i = 1$, the (counterfactual) overall survival time under optimal stage 2 treatment was generated from the AFT model in (4.2) as

$$\tilde{T}_i = \exp(6.3 + 1.5X_{i1} - 0.8X_{i1}^4 + A_{i1}(0.1 + 0.1X_{i1}) + \epsilon_{i1})$$

where ϵ_{i1} was also normally distributed with mean zero and variance 0.09. For individuals who did not enter the second stage, $\tilde{T}_i$ was their observed survival time, that is, $T_i = \tilde{T}_i$. For individuals who entered the second stage, their observed survival time was derived as $T_i = T_{i1} + T_{i2}$ where $T_{i1} = \tilde{T}_i - T_{i2}^{\text{opt}}$. For those who did not experience an event, we generated the censoring times from an Exponential distribution with rate $1/300$.

This data generating mechanism yielded 30% independent censoring satisfying the assump-

tions on the censoring mechanism made by both DWSurv and the method by HNW. We also considered data-generating mechanisms with 60% censoring and with the censoring time and probability dependent on baseline covariates. Another data-generating mechanism considered the probability of censoring dependent on time-varying covariates to assess the consistency and double-robustness of DWSurv in more complex situations. Details on these alternatives data generating mechanisms are given in the Supplemental Material C.3.

To compare the accuracy and precision of the methods, we considered four simulation scenarios. Scenario 1 assumed that all models (treatment-free, treatment, censoring) were correctly specified. Scenario 2 misspecified the weight models (treatment and censoring) and correctly specified the treatment-free model. Scenario 3 had the treatment-free model misspecified but the weight models correctly specified. Scenario 4 incorrectly specified all models. To compare the distribution of the survival time under different treatment assignment schemes, we estimated the optimal decision rules with DWSurv and the method by HNW from one simulated dataset and then generated larger datasets (n=10,000) with treatment assignment following the optimal decision rules estimated by the two methods. We considered three sample sizes (n=500, 1000 and 10,000) and simulated 1000 datasets[1].

Figure 4.1 shows the distribution of the blip estimates under four scenarios, with sample size n=1000. As expected, when the treatment-free model was correctly specified (scenarios 1 & 2), both methods were unbiased. Our method yielded unbiased estimators when the weight models were correctly specified, even if the treatment-free model was not (scenario 3), and biased estimators only when all models were misspecified. The method by HNW was not robust to weight model misspecification. The precision of the estimators was comparable between the two methods. Results were similar with smaller and larger sample sizes, with a higher proportion of censoring and with censoring dependent on baseline covariates.

[1]In the event that a data set contained at least one observation with negative $T_{i1} < 0$, the whole dataset was discarded and a new dataset was generated.

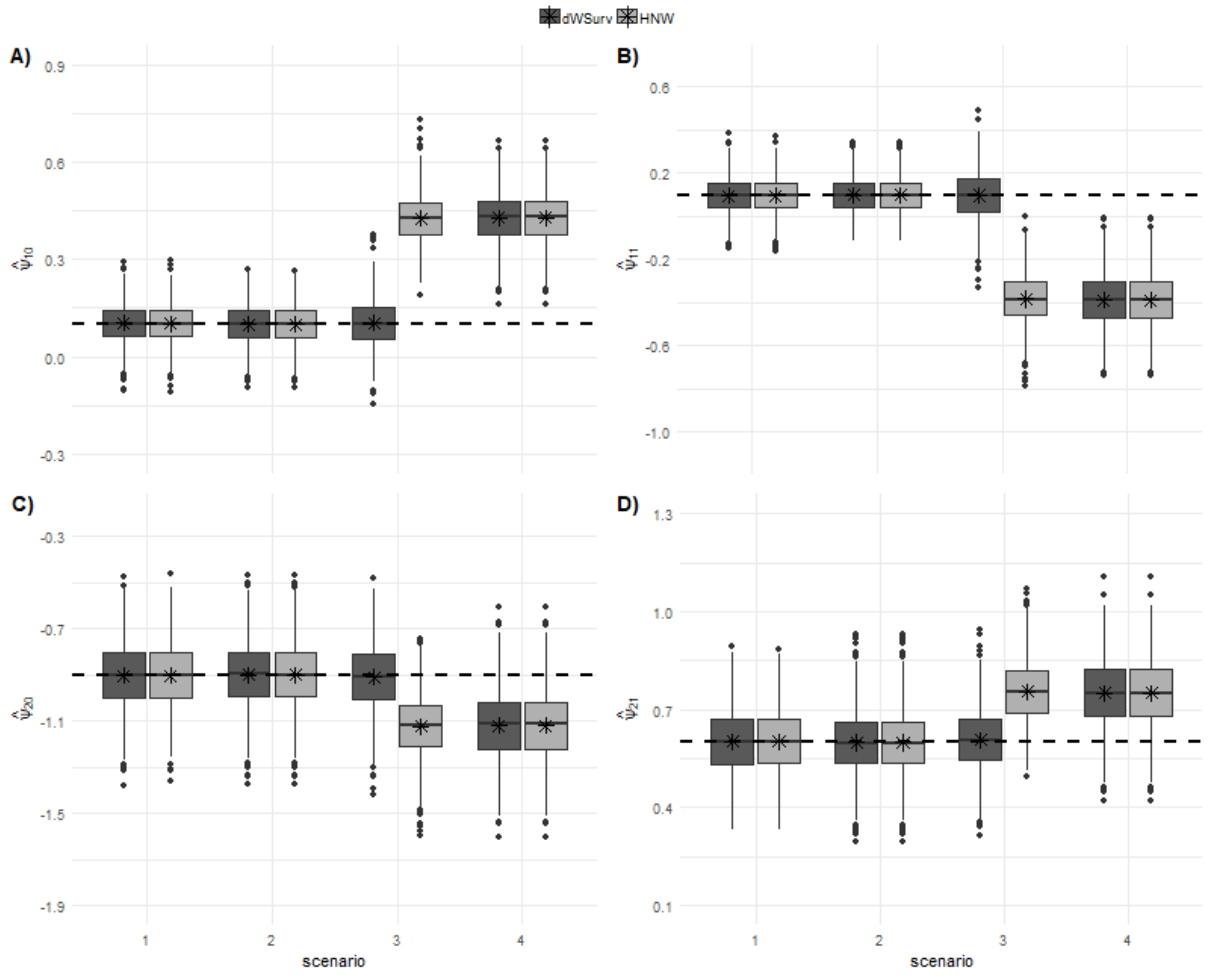

Figure 4.1: Distribution of the blip parameter estimates in the first stage (upper row) and second stage (lower row) with DWSurv (dark grey) and the method by HNW (light grey) with sample size n=1000 across four scenarios: (i) all the models correctly specified, (ii) weight models misspecified but treatment-free model correctly specified, (iii) treatment-free model misspecified but weight models correctly specified, and (iv) all models incorrectly specified. The data were simulated with 30% censoring, with both the probability of censoring and the censoring times independent of the survival times.

With DWSurv, the proportion of individuals with the optimal DTR correctly identified was high. With the smallest sample size (n=500) and 30% independent censoring, DWSurv identified the true optimal DTR for an average of 94% of the individuals over the 1000 simulated datasets (range: 63%–100%). In general, larger sample sizes or lower censoring percentages yielded higher proportions of individuals with their true optimal DTR correctly identified. The method by HNW yielded similar results.

The distribution of the survival times under the optimal treatment decision rules estimated by DWSurv was comparable to that of the survival time under the true optimal DTR. With an initial sample size of n=500, the median survival time across 10,000 observations simulated according to the DTR estimated by DWSurv was 7.07, which was equal to the median survival time under the true optimal DTR. Because the true optimal treatment is A_1=1 for all individuals and the survival time within the second stage contributed to, on average, only 22% of the overall survival time (median: 11%), both fixed treatment strategies with A_1=1 also yielded a survival times distribution comparable to the distribution of the survival times generated under the true optimal DTR. The two fixed treatment strategies with A_1=0 yielded shorter survival times with median 6.90 in both cases. Detailed results are shown in the Supplemental Material C.3.

4.4 An Application to Rheumatoid Arthritis

We applied our proposed method to a cohort of patients with RA. Current treatment recommendations from the ACR targeted remission time in patients with early RA. At symptom onset, it was recommended that traditional disease-modifying antirheumatic drug (DMARD) monotherapy be initiated. At a follow-up visit, if disease activity remained moderate or high despite DMARD monotherapy, it was recommended to use what we will refer to as DMARD combination therapy and which was defined as traditional double or triple DMARD therapies or adding a tumor necrosis factor (TNF) inhibitor or a non-TNF biologic agent to the current regime. We aimed to estimate similar rules in a DTR framework. We used data from the SERA Inception Cohort (Dale et al., 2016), an ongoing cohort of patients with a diagnosis of undifferentiated arthritis (UA) or RA. Patients attended a baseline visit and follow-up visits every 6 months for 3 years, providing data on disease activity, demographics and drug prescription at each visit. The Supplemental Material C.4 details key covariates, outcomes, drug categories, and inclusion criteria.

The primary analysis aimed to estimate two treatment decision rules that minimized time to remission. The first stage of clinical intervention started at the baseline visit and compared DMARD monotherapy (A_1=1) to DMARD combination therapy (A_1=0). The second stage started at the first follow-up visit and compared staying on the same therapy (A_2=1) or making an "acceptable" change to the current regime (A_2=0). Acceptable changes were switching, adding or discontinuing a drug such that the resulting regime was either a DMARD monotherapy or DMARD combination therapy as described in Table C.33 of the Supplemental Material C.4. The outcome was time to remission measured from baseline until the Disease Activity Score 28 (DAS28) erythrocyte sedimentation rate (ESR) was lower or equal to 2.6, or the DAS28 C-reactive protein level (CRP) was lower or equal to 2.3 if DAS28-ESR was not measured. Patients were censored if they were lost to follow-up, made an unauthorized drug change or made any kind of treatment change after the first follow-up visit. Details on the implementation are available in the Supplemental Material C.4.

In a secondary analysis, the same methodology was applied to identify a sequence of treatments that minimized the time until a 35% decrease in DAS28 score from baseline was observed. This alternative endpoint was chosen to increase the number of events, as the primary analysis was subject to an unusually high proportion of censoring. For both analyses, we conducted a complete-case analysis. In a sensitivity analysis, we compared the results to single imputation analysis using chained equations (White et al., 2011).

Table C.34 in the Supplemental Material C.4 summarizes the patients' characteristics at baseline and at the first follow-up visit. A total of 496 patients met our inclusion criteria among which 488 have complete data. The median follow-up time was 307 days. Approximately 66% of the patients were on a DMARD monotherapy in the first stage and 141 patients achieved remission by the end of the first stage. A little less than half of the patients (48%) reached the second stage of intervention, among which 70% remained on the same treatment and only 11% were in remission before the end of the follow-up period.

Table 4.1: Inference for a two-stage DTR in the rheumatoid arthritis application.

	Time to remission			35% decrease in DAS28		
Parameter	Est.	SE	95% CI	Est.	SE	95% CI
	Baseline (n=488)			Baseline (n=482)		
ψ_{10}	0.06	0.08	(-0.08, 0.21)	0.09	0.10	(-0.10, 0.29)
ψ_{11} (DA_1)	-0.10	0.08	(-0.26, 0.06)	-0.14	0.10	(-0.34, 0.06)
	1st follow-up (n=236)			1st follow-up (n=144)		
ψ_{20}	-0.08	0.12	(-0.33, 0.16)	-0.03	0.33	(-0.68, 0.62)
ψ_{21} (DA_2)	0.08	0.41	(-0.72, 0.89)	-0.22	0.30	(-0.81, 0.38)
ψ_{22} (A_1)	0.28	0.32	(-0.34, 0.90)	0.14	0.21	(-0.28, 0.55)

DA_j: disease activity at the beginning of stage j, Est.: estimates, SE: standard error, CI: confidence interval

Table 4.1 shows the parameter estimates involved in the construction of the two treatment decision rules along with measures of uncertainty in the primary and secondary analyses. At baseline, the treatment decision rule took the form $a_1^{\text{opt}} = \mathbb{I}(\hat{\psi}_{10}+\hat{\psi}_{11}\text{DA}_1 < 0)$ where DA_j was 1 if disease activity was moderate or high at the beginning of stage j, 0 otherwise. Although the effects are not significant, the decision rule recommends initiating DMARD monotherapy if disease activity is moderate or high at baseline. At the first follow-up visit, the treatment decision rule $a_2^{\text{opt}} = \mathbb{I}(\hat{\psi}_{20} + \hat{\psi}_{21}\text{DA}_2 + \hat{\psi}_{22}A_1 < 0)$ recommends changing the current regime if the patient was on DMARD monotherapy in the first stage and staying on the same regime otherwise. In the secondary analysis, although more events were observed overall (243 events as opposed to 167 in the primary analysis), fewer patients entered the second stage (n=144) among which a larger proportion (29%) experienced an event. The decision rule in the second stage is different than in the primary analysis. There was inconsistent evidence that tailoring treatment based on disease activity or previous treatment was warranted. Conclusions remained unchanged with single imputation analyses (results not shown). We emphasize that the analyses presented in this paper were not meant to disprove the current treatment recommendations but rather aimed to showcase the usefulness of DWSurv in answering a clinical question.

4.5 Discussion and Conclusion

We proposed dynamic weighted survival modeling to estimate an optimal DTR from observational data when the outcome is survival time subject to right-censoring. At each stage of intervention, our method requires solving weighted GEE with mean structure corresponding to a semi-parametric AFT model and weights that depend on models for the probability of being censored and the probability of treatment. With weights satisfying the balancing property and under standard causal assumptions, the procedure is doubly robust as it yields consistent estimators of the effect of treatment and its interactions with tailoring variables at each stage if only a subset of the nuisance models is correctly specified. A broad class of balancing weights are defined by combining IPCW with a function of the probability of treatment. DWSurv is equipped with tools for inference including formulas for the asymptotic variance of the blip estimators and an approach to model-checking (Wallace et al., 2016, 2017b). Its implementation by other statisticians or epidemiologists is straightforward with the `DWSurv` function in the `DTRreg R` package.

The definition of a stage of intervention was intentionally left vague throughout the notations and derivations, allowing for stages to be defined with respect to the clinical problem under study. For example, our application to RA considered the time between equally-spaced follow-up visits as a stage. This definition is likely relevant for many chronic diseases where the patient's condition is monitored routinely. Entering a stage of intervention could also be viewed as a covariate whose value depends on previous treatments. For example, the treatment of cancer often includes an initial treatment followed by a salvage treatment if cancer recurs. In this case, cancer recurrence defines the beginning of a second stage. Regardless of the definition of a stage, the estimated optimal decision rule in stage j applies to the target population of individuals who reach this stage and require this treatment.

Our simulation study considered complex data dependencies, especially with respect to the censoring mechanism. The data-generating mechanism presented in this article was designed

to be fair to the compared methods with respect to the assumptions made on the censoring mechanism. Alternative data generating mechanisms were also designed to showcase the performance of DWSurv in more realistic and complex situations where both the probability of treatment and of censoring depended on time-varying covariates. The simulation study validated the double-robustness of the blip estimators in a multistage setting when the probability of censoring depended on baseline or time-varying covariates. In particular, when censoring only depended on baseline information, the proposed method showed equally good performance as the method by HNW and outperformed it when all confounders were not included in the treatment-free model.

The double-robustness of the blip estimators is an attractive theoretical property of the DWSurv algorithm. More than merely providing additional protection against model misspecification, the double-robustness property allows taking advantage of clinical knowledge on the mechanisms of treatment assignment and censoring through the specification of models that are deemed easier to inform than the outcome model. The double-robustness property can also be exploited for model-checking purposes (Wallace et al., 2016).

A key requirement for DWSurv is that the survival time needs to be modeled directly in order to estimate the pseudo-outcome for a subset of individuals. The Cox model does not provide a natural framework as it models the hazard function rather than the survival time. It thus requires additional steps and modeling assumptions to translate the estimated hazards into a pseudo-outcome. AFT models provide an interesting alternative to Cox models as they are concerned with the survival time directly. It has been argued that subject-matter knowledge such as biological mechanisms is easier to translate into interpretable parameters of AFT models than into those of Cox models (Hernán & Robins, 2010). Note that any parametric survival models can be used in DWSurv if a specific distribution is deemed more appropriate in a particular setting.

DWSurv relies on the mean survival time to estimate the decision rules. However, when

administrative censoring is heavy, the tail of the survival distribution may be ill-determined and the restricted mean survival time, defined as the mean of the survival time up to $\tau > 0$, could then be used as an alternative outcome (Karrison, 1997). An additional data manipulation step is required to define the restricted survival time Y_τ as $y_\tau = y$ if $y < \tau$ and $y_\tau = \tau$ if $y \geq \tau$, where τ is chosen to be smaller than the longest follow-up time, and an additional step in the DWSurv algorithm would also be required when constructing the pseudo-outcomes to ensure that the resulting counterfactual overall survival time does not exceed τ.

As noted in Section 4.2, we followed Huang et al. (2014) and did not implement artificial censoring. While this could lead to violations of the assumption of independence between observation and censoring times (conditional on covariates), the found performance of our estimators was excellent. There may be settings where this does not hold but any gain in performance due to artificial censoring should be weighed against the additional complexity of the approach relative to the simplicity of the current implementation of DWSurv.

A limitation of our proposed method concerns the estimation of standard errors for the blip estimators. The asymptotic variance formulas we have derived are useful when all models are correctly specified. Although the formulas showed good performance in finite samples, model misspecification may have a significant impact on their performance. Moreover, under specific longitudinal distributions of the data, all but the last stage blip estimators may be nonregular in the sense that their asymptotic distribution does not converge uniformly over the parameter space (Robins, 2004). Future work will look into the performance of alternative standard error formulations in the presence of nonregularity.

We illustrated our new method with an application to the treatment of RA using observational data from the SERA Inception Cohort. We aimed to mimic the treatment decision rules recommended in the 2015 ACR guidelines. We found inconsistent effects of tailoring treatment to disease activity or previous treatment received on the two outcome definitions.

Our analysis was subject to some limitations. First, the use of corticosteroids is a potential confounder which was not accounted for due to the complexity of summarizing, in a low dimensional but meaningful way, drugs that are administered in different routes and at varying frequencies and doses. Second, although we restricted our analysis to patients who had had a diagnosis less than a year prior to entering the cohort, the baseline visit did not necessarily coincide with the recommended timing of the first treatment decision in the ACR guidelines. Patients might have been on different treatments before baseline, leading to a form of exposure misclassification. Finally, despite remission being considered as the treatment goal in RA, there is no widely used definition of remission (Felson et al., 2011). We used the DAS28-ESR as suggested in the ACR guidelines. As approximately 40% of the patients had missing DAS28-ESR score at baseline, we used DAS28-CRP scores to define remission and disease activity level when DAS28-ESR was not available. However, the DAS28 scores based on ESR and CRP are not directly comparable (Sengul et al., 2015; Son et al., 2016) and the DAS28-CRP has not been fully validated although we appreciate that it is widely used (Inoue et al., 2007; Kuriya et al., 2017). We acknowledge that defining levels of disease activity and remission with different thresholds of the DAS28-CRP might have led to different results but emphasize that our analysis merely aimed to illustrate the application of DWSurv to a real clinical problem.

The method that we have proposed in this article, DWSurv, allows estimating an optimal DTR when the outcome of interest is survival time subject to right-censoring. It is theoretically attractive, offering double-robustness and valid asymptotic variance formulas for regular settings, and readily applicable in `R`. Finally, DWSurv eases knowledge translation in the field of precision medicine as it yields parameters that are easily interpretable by clinical collaborators analogous to risk scores and decision rules that can be applied in practice.

Software

In an effort to promote reproducible research, scripts in the form of R code are available at https://github.com/gabriellesimoneau/Rcode-JASA2019. For questions, comments or remarks about the code, contact the corresponding author

Funding

This work was supported by a doctoral scholarship from the Fonds de recherche du Québec – Nature et technologies (Ref 199803) and by awards (INF-GU-168) from (i) the Translational Medicine Research Collaboration, a consortium made up of the Universities of Aberdeen, Dundee, Edinburgh and Glasgow, the four associated NHS Health Boards (Grampian, Tayside, Lothian and Greater Glasgow & Clyde), and Pfizer and (ii) from the Chief Scientific Office (Ref ETM-40).

Appendix C – Supplemental Materials

Contains the following sections:

C.1 **Consistency and Double-robustness** – formal proof of consistency and double-robustness of the blip estimators.

C.2 **Details on the Asymptotic Variance Formulae** – steps to derive the asymptotic variance of the treatment-free and blip estimators.

C.3 **Details on the Simulation Study** – details on alternative data generating mechanisms and additional simulation results.

C.4 **SERA Data Analysis** – additional information about the SERA data application.

Chapter 5

Finite Sample Variance Estimation for Optimal Dynamic Treatment Regimes of Survival Outcomes

Preamble to Manuscript 3. We expected that DWSurv would also suffer from non-regular inference because it relies on the same type of substitution estimators as in dWOLS. Therefore, this project initially aimed to study non-regularity in DWSurv and explore the usefulness of previously proposed tools to correct for its negative impact on inferences. However, preliminary simulation studies suggested that the asymptotic variance formulae of the DWSurv blip estimators led to conservative confidence intervals in regular settings. We realized that it was important to study and compare different approaches to construct confidence intervals for the DWSurv parameters in general, with non-regular situations as special cases. The original contributions in this manuscript are (i) proposing parametric and non-parametric bootstraps for censored data to be used with DWSurv, and (ii) evaluating the performance of the asymptotic variance formulae derived for DWSurv estimators (this has only been done once with G-estimation and uncensored continuous outcomes). This

manuscript has been submitted to *Statistics in Medicine*. All inferential tools presented in this manuscript are implemented in `R` as part of the `DTRreg` package and available on CRAN.

Finite Sample Variance Estimation for Optimal Dynamic Treatment Regimes of Survival Outcomes

Gabrielle Simoneau[1], Erica EM Moodie[1], Robert W Platt[1]

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

Abstract

Dynamic weighted survival modeling is an accessible and doubly-robust statistical method to estimate an optimal dynamic treatment regime when the outcome is survival time. However, it is unclear how different methods to construct confidence intervals for the decision rule parameters compare in practice when the true specification of the nuisance models is unknown or in non-regular situations. Via simulations, we compare two asymptotic variances based on sandwich estimation, adjusting or not for the estimation of the nuisance parameters, to three bootstrap methods. We find that the bootstrap approaches perform consistently well at the cost of longer computational times. The asymptotic variance with adjustments generally yields conservative confidence intervals. The asymptotic variance without adjustments yields nominal coverages for large sample sizes. We recommend using the asymptotic variance with adjustments in small samples and the bootstrap if computationally feasible. Caution should be taken when non-regularity may be an issue.

5.1 Introduction

A dynamic treatment regime (DTR) is a set of treatment decision rules that are tailored to individual characteristics, a form of precision medicine. DTRs are particularly relevant in the chronic care environment where a patient's condition is changing over time and treatments must correspondingly be altered. Across multiple stages of clinical intervention, the decision rules input current patient characteristics and output a recommended treatment. Of interest is usually to identify an optimal DTR, that is, the sequence of treatment rules that maximizes or minimizes a certain outcome. Statistical methods for optimal DTR have been proposed over the last decades (e.g. Murphy, 2003; Robins, 2004; Wallace & Moodie, 2015; B. Zhang et al., 2013; Y. Q. Zhao et al., 2015) yet only a few have been extended to accommodate time-to-event outcomes (Bai et al., 2017; Goldberg & Kosorok, 2012; Hager et al., 2018; Huang et al., 2014; Jiang et al., 2017a). With survival time as the criterion of optimality, the estimation procedure is complicated by the possibility of censoring and by the fact that not all individuals may have the same number of decision points. Dynamic weighted survival modeling (DWSurv) is a method for estimating optimal DTRs with censored survival outcomes (Simoneau et al., 2018). It is robust, easy to understand by practitioners and readily applicable in `R` with the `dWSurv` function in the `DTRreg` package (Wallace et al., 2017a). It requires solving a series of weighted generalized estimating equations (GEE) with mean structure satisfying a semi-parametric accelerated failure time (AFT) model and weights that depend on nuisance models for the probability of treatment and censoring. Under standard regularity conditions, DWSurv estimators are asymptotically normal and confidence intervals with nominal coverage can be constructed for large n.

The asymptotic variance of the DWSurv estimators accounts for the estimation of the parameters in the nuisance models through a series of adjustments. These adjustments are found in the first-order Taylor expansion of the estimating equations about the limiting values of the nuisance parameters. An application of the delta method then allows to approximate the

asymptotic variance of the DWSurv estimators. A "naive" version of the variance formula which ignores the adjustments could also be considered in practice. However, it is unclear how both formulations of the asymptotic variance perform in finite samples, for example, as the number of nuisance parameters increases (Moodie, 2006). Moreover, even if the estimators remain consistent under incorrect specification of one of the nuisance models, it is unclear if model misspecification affects the quality of the asymptotic approximation (Robins, 2004). With G-estimation, where asymptotic derivations are similar to that of DWSurv, it has been noted that model misspecification may slow the convergence of the confidence interval coverage to the nominal level (Moodie, 2009). Also, as in other regression-based approaches for optimal DTR, the DWSurv estimators can be non-regular under certain data generating mechanisms because the estimation procedure relies on a non-smooth maximization operation (Moodie & Richardson, 2010; Robins, 2004). This implies that the distribution of the DWSurv estimators is not uniformly normal over the parameter space; a practical negative consequence of this is that confidence intervals based on the asymptotic variance can perform poorly in terms of coverage. Alternative methods to construct confidence intervals for DWSurv parameters can be used. The standard non-parametric bootstrap, which resamples observations with replacement from the original sample, has been adapted to censored data (Efron, 1981). Parametric bootstraps for survival times have been proposed (Efron & Tibshirani, 1986; Hjort, 1985) and reviewed in the context of Cox regressions (Burr, 1994). However, as with the asymptotic approximation, bootstrap methods can also perform poorly in non-regular situations (Chakraborty et al., 2010; Simoneau et al., 2017), and the impact of model misspecification on the quality of inferences remains unclear.

The objective of this study is to compare the performance of five methods to construct confidence intervals for DWSurv parameters. The comparison via simulations focuses on situations frequently encountered in practice and for which the choice of a method to construct confidence intervals is not simple, for example, when the correct model specifications are unknown and when non-regular inferences can occur. Section 5.2 introduces DWSurv and

formalizes non-regularity. Section 5.3 presents the form of the asymptotic variance formula with and without adjustments for the plug-in nuisance estimators. We also review the bootstrap and discuss specific considerations for its application to DTR. Section 5.4 presents the simulation studies. Section 5.5 discusses the results and derives general recommendations on how to choose the best methods to construct confidence intervals depending on the situation encountered in practice.

5.2 Dynamic Treatment Regimes

Consider an individual followed through multiple stages of intervention. Each stage j is associated with two possible treatments $a_j \in \{0, 1\}$ and a decision rule $d(\boldsymbol{h_j}) : \mathcal{H}_j \to (0, 1)$ that inputs the individual's characteristics, or histories, $\boldsymbol{h_j}$ measured at the beginning of that stage and outputs a recommended treatment. The collection of decision rules across all stages of intervention $\{d(\boldsymbol{h_1}), ..., d(\boldsymbol{h_J})\}$ is a DTR and of interest is to identify and estimate an optimal DTR $\{d^{\text{opt}}(\boldsymbol{h_1}), ..., d^{\text{opt}}(\boldsymbol{h_J})\}$, that is, the collection of decision rules that optimizes an outcome usually measured at the end of the study. With time-to-event data, an optimal DTR typically aims to maximize the survival time from the first stage of intervention until a negative event (e.g. death) but may also aim to minimize the survival time until a positive event (e.g. remission from a disease). The outcome is thus the overall survival time defined as the sum of the stage-specific survival times $T = \sum_{j=1}^{J} \eta_j T_j$ where J is the total possible number of stages, T_j is the survival time within stage j and η_j indicates whether an individual reached intervention j. Right-censoring can occur when the outcome for some individuals is unobserved. The censoring time is denoted with C and the observed time Y is the minimum between T and C, with $\Delta = \mathbb{I}(T \leq C)$ indicating if an event was observed. The following details on the DWSurv algorithm are given for a two-stage DTR and apply directly to DTRs with more than two stages.

Under standard causal assumptions (Simoneau et al., 2018), DWSurv estimates the decision rule parameters by assuming semi-parametric AFT models for the survival times across stages. The estimation of an optimal DTR relies on the principle of backward induction in which estimation starts in the last stage and moves backward into previous stages. In a two-stage DTR, estimation for the second stage decision rule is carried out by specifying a semi-parametric AFT model for the survival time in the second stage with mean

$$\mathbb{E}[\log(T_2)] = \boldsymbol{\beta_2^T h_{2\beta}} + \boldsymbol{\psi_2^T} a_2 \boldsymbol{h_{2\psi}} \tag{5.1}$$

where the distribution of the errors is left unspecified except that it should have zero expectation. The mean model has term $\boldsymbol{\beta_2^T h_{2\beta}}$ that does not depend on the treatment A_2, which is called the treatment-free component, and a term $\boldsymbol{\psi_2^T} a_2 \boldsymbol{h_{2\psi}}$ that depends on the treatment, called the blip component. The optimal stage 2 decision rule is that which maximizes the expectation given in (5.1) and is expressed as $d^{\text{opt}}(\boldsymbol{h_2}) = \mathbb{I}(\boldsymbol{\psi_2^T h_{2\psi}} > 0)$ which only depends on the blip parameters $\boldsymbol{\psi_2}$. If the expression inside the indicator is positive, then receiving treatment $A_2 = 1$ increases the survival time and the optimal treatment for individuals with history $\boldsymbol{h_{2\psi}}$ is $A_2 = 1$. Estimation for the first stage decision rule is based on optimizing the overall survival time had everybody received their optimal stage 2 treatment, namely T_2^{opt}. This (counterfactual) overall survival time is called the pseudo-survival time $\tilde{T}$ and is defined as $\tilde{T} = T_1 + \eta_2 T_2^{\text{opt}}$. This allows for a fair comparison of the first stage treatments on the overall survival time by ruling out the differences between individuals that are attributable to the second stage treatment. A model for the mean pseudo-survival time $\tilde{T}$ is specified as

$$\mathbb{E}[\log(\tilde{T})] = \boldsymbol{\beta_1^T h_{1\beta}} + \boldsymbol{\psi_1^T} a_1 \boldsymbol{h_{1\psi}} \tag{5.2}$$

where the treatment-free and blip components can be identified as in the second stage. Based on this specification of the blip, the optimal stage 1 decision rule is $d^{\text{opt}}(\boldsymbol{h_1}) = \mathbb{I}(\boldsymbol{\psi_1^T h_{1\psi}} > 0)$. Therefore, the optimal DTR is: recommend treatment $A_1 = 1$ if $\boldsymbol{\psi_1^T h_{1\psi}} > 0$. If the individual

reaches the second stage of intervention, recommend $A_2 = 1$ if $\boldsymbol{\psi_2^T h_{2\psi}} > 0$.

DWSurv estimates the parameters of interest, $\boldsymbol{\psi_1}$ and $\boldsymbol{\psi_2}$, by solving a series of weighted GEE. Inverse probability of censoring is used to account for potential informative censoring. The estimation procedure allows for non-randomized treatment assignments, also via a weighting argument. A broad family of weights is available (Simoneau et al., 2018), for example weights of the form $|a_j - P(A_j = 1|\boldsymbol{H_j})|/P(\Delta = \delta|\boldsymbol{H_j})$ or inverse probability of censoring and treatment weights could be used. The following algorithm details the estimation procedure for a DTR with up to two stages:

1. Specify two parametric models for the probability of treatment and the probability of censoring within stage 2 respectively denoted by $\mathbb{P}(A_2 = 1|\eta_2 = 1, \boldsymbol{H_2}; \boldsymbol{\alpha_2})$ and $\mathbb{P}(\Delta = 0|\eta_2 = 1, \boldsymbol{H_2}, A_2; \boldsymbol{\lambda_2})$.

2. Specify weights $w_2(\delta, a_2, \boldsymbol{h_2}; \hat{\boldsymbol{\alpha}}_{\boldsymbol{2}}, \hat{\boldsymbol{\lambda}}_{\boldsymbol{2}})$ using the models specified in step 1 and estimate the stage 2 parameters $(\boldsymbol{\beta_2}, \boldsymbol{\psi_2})$ by solving the following weighted GEE using only individuals with $\delta = 1$

$$U_2(\boldsymbol{\psi_2}, \boldsymbol{\beta_2}) = \sum_{i=1}^{n} \delta_i \eta_{2i} \hat{w}_{2i} \begin{pmatrix} \boldsymbol{h_{2\beta i}} \\ a_{2i}\boldsymbol{h_{2\psi i}} \end{pmatrix} \left(\log(T_{2i}) - \boldsymbol{\beta_2^T h_{2\beta i}} - \boldsymbol{\psi_2^T} a_{2i} \boldsymbol{h_{2\psi i}}\right) = 0. \quad (5.3)$$

3. Construct the pseudo-survival time as

$$\tilde{T}(\hat{\boldsymbol{\psi}}_{\boldsymbol{2}}) := T_1 + \eta_2 \left(T_2 \times \exp\{\hat{\boldsymbol{\psi}}_{\boldsymbol{2}}^{\boldsymbol{T}} \boldsymbol{h_{2\psi}} [\mathbb{I}(\hat{\boldsymbol{\psi}}_{\boldsymbol{2}}^{\boldsymbol{T}} \boldsymbol{h_{2\psi}} > 0) - a_2]\}\right). \quad (5.4)$$

4. Specify two parametric models for the probability of treatment within stage 1 and the probability of censoring from stage 1 onwards, respectively denoted by $\mathbb{P}(A_1 = 1|\boldsymbol{H_1}; \boldsymbol{\alpha_1})$ and $\mathbb{P}(\Delta = 0|\boldsymbol{H_1}, A_1; \boldsymbol{\lambda_1})$.

5. Specify weights $w_1(\delta, a_1, \boldsymbol{h_1}; \hat{\boldsymbol{\alpha}}_{\boldsymbol{1}}, \hat{\boldsymbol{\lambda}}_{\boldsymbol{1}})$ using the models specified in step 4 and estimate

the stage 1 parameters $(\boldsymbol{\beta_1}, \boldsymbol{\psi_1})$ by solving

$$U_1(\boldsymbol{\psi_1}, \boldsymbol{\beta_1}; \hat{\boldsymbol{\psi}}_{\mathbf{2}}) = \sum_{i=1}^{n} \delta_i \hat{w}_{1i} \begin{pmatrix} \boldsymbol{h_{1\beta i}} \\ a_{1i}\boldsymbol{h_{1\psi i}} \end{pmatrix} \left(\log\{\tilde{T}_i(\hat{\boldsymbol{\psi}}_{\mathbf{2}})\} - \boldsymbol{\beta_1^T h_{1\beta i}} - \boldsymbol{\psi_1^T} a_{1i}\boldsymbol{h_{1\psi i}} \right) = 0. \quad (5.5)$$

A one-stage DTR would consist of steps 1 and 2 only while a DTR with three stages or more repeats steps 3-5 as necessary. The estimation procedure offers the double-robustness property by which, at each stage j, the procedure yields consistent estimators of $\boldsymbol{\psi_j}$ when either or both the treatment-free model and/or the treatment and censoring models included in the weights are correctly specified, provided that the form of the blip is correct.

DWSurv suffers from nonstandard limit theory (so-called non-regularity) because the construction of the pseudo-survival time involves a non-smooth maximization operation as per the indicator function in (5.4). As a result, the asymptotic distribution of the first stage blip parameters $\boldsymbol{\psi_1}$ is not uniformly normal. More precisely, the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\psi}}_{\mathbf{1}} - \boldsymbol{\psi_1})$ is non-normal if the expression inside the indicator is close to the point of non-differentiability, which corresponds to data generating mechanisms in which the effect of the stage 2 treatment is null or small.

5.3 Measures of Uncertainty

5.3.1 Asymptotic Variance Formulations

In a one-stage DTR, it has been shown that the asymptotic variance of the estimators $(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\beta}})$ must adjust for the plug-in estimates of the nuisance parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$ in $U(\boldsymbol{\psi}, \boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}})$ (Moodie, 2009; Robins, 2004). This is done by performing a first-order Taylor expansion of the estimating function about the limiting values of $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\lambda}}$, say $\boldsymbol{\alpha_0}$ and $\boldsymbol{\lambda_0}$. An implementable version, which does not depend on the unknown values $\boldsymbol{\alpha_0}$ and $\boldsymbol{\lambda_0}$, is

derived by performing another Taylor expansion, given rise to

$$U_{\text{adj}}(\boldsymbol{\psi}, \boldsymbol{\beta}) \approx U - \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\alpha}} U\right] \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\alpha}} \boldsymbol{s}_{\boldsymbol{\alpha}}\right]^{-1} \boldsymbol{s}_{\boldsymbol{\alpha}} - \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\lambda}} U\right] \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\lambda}} \boldsymbol{s}_{\boldsymbol{\lambda}}\right]^{-1} \boldsymbol{s}_{\boldsymbol{\lambda}}$$

where $\boldsymbol{s}_{\boldsymbol{\alpha}}$ and $\boldsymbol{s}_{\boldsymbol{\lambda}}$ are the score functions for the treatment and censoring models. The variance-covariance matrix of the estimators $(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\beta}})$ is

$$\mathbb{V}\text{ar}(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\beta}}) = \mathbb{E}\left[\left\{\mathbb{E}\left[\frac{\partial}{\partial(\boldsymbol{\psi}, \boldsymbol{\beta})} U_{\text{adj}}\right]^{-1} U_{\text{adj}}\right\}^{\otimes 2}\right] \tag{5.6}$$

where $\mathbb{E}[X^{\otimes 2}] = \mathbb{E}[XX^T]$. For a DTR with two stages or more, in all but the last stage, an additional adjustment term for the plug-in estimators $(\hat{\boldsymbol{\psi}}_{\boldsymbol{j+1}}, ..., \hat{\boldsymbol{\psi}}_{\boldsymbol{J}})$ appearing in the pseudo-survival times is added to U_{adj} (Moodie, 2009). Although the variance (5.6) is asymptotically correct, its finite sample performance may depend on the quality of the approximation of U by U_{adj}, which may, in turn, depend on the dimension of the nuisance parameters and the underlying data generating mechanism (Moodie, 2006). As the quality of the adjusted asymptotic variance may be affected by how well one can estimate the nuisance parameters, it is of interest to compare it to the naive asymptotic variance which does not make adjustments, given by

$$\mathbb{V}\text{ar}_{\text{naive}}(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\beta}}) = \mathbb{E}\left[\left\{\mathbb{E}\left[\frac{\partial}{\partial(\boldsymbol{\psi}, \boldsymbol{\beta})} U\right]^{-1} U\right\}^{\otimes 2}\right]. \tag{5.7}$$

Although asymptotically incorrect, its performance in finite samples is unknown and may be acceptable in some cases. Specifically, it may counterbalance the potential negative impact of an inaccurate approximation of U by U_{adj} in practice which may be too conservative in its adjustment for the estimation of the nuisance parameters.

5.3.2 Bootstrap

We review the standard non-parametric bootstrap by considering the case of uncensored outcomes in a one-stage DTR. For simplicity, suppose that the parameter of interest ψ is unidimensional, which corresponds to assuming no effect of patient characteristics on the treatment decision. The goal of the inference is to characterize the distribution of the estimator $\hat{\psi}$ which depends on n observations of the form (t_i, x_i, a_i) sampled from an unknown population distribution F. With F unknown, the sampling distribution of $\hat{\psi}$ is also unknown. The idea behind the non-parametric bootstrap is to approximate the distribution of $\hat{\psi}$ by sampling from the empirical distribution $\hat{F}$ that puts probability mass $1/n$ on each observed data point (t_i, x_i, a_i). A sample from $\hat{F}$ is called a bootstrap sample $s^\star = \{(t_1^\star, x_1^\star, a_1^\star), ..., (t_n^\star, x_n^\star, a_n^\star)\}$ and it corresponds to a random sample of size n drawn with replacement from the original sample $s = \{(t_1, x_1, a_n), ..., (t_n, x_n, a_n)\}$. The bootstrap algorithm proceeds in three steps:

1. draw a large number of independent bootstrap samples, say $s_1^\star, ..., s_B^\star$;
2. evaluate the statistic $\hat{\psi}$ for the b-th bootstrap sample, say $\hat{\psi}^{(b)}$, for $b = 1, ..., B$; and
3. approximate the distribution of the estimator $\sqrt{n}(\hat{\psi}_n - \hat{\psi})$ with the bootstrap analogue $\sqrt{n}(\hat{\psi}^{(b)} - \hat{\psi}_n)$, where $\hat{\psi}_n$ is the estimate obtained using the actual sample s.

A $(1-\alpha) \times 100\%$ confidence interval for ψ may be constructed with the $\alpha/2$ and $(1-\alpha/2)$ percentiles of the distribution of $\hat{\psi}^{(b)}$, $b = 1, ..., B$. Other methods have been proposed to derive confidence intervals from bootstrap samples (Burr, 1994; Efron, 1992b; Efron & Tibshirani, 1986).

The bootstrap could alternatively be carried out parametrically by conditioning on (X, A), which is ancillary for the parameter of interest ψ, and resampling survival times T from F_T, its parametric distribution. Suppose that the survival times are assumed to follow a semi-parametric AFT model as $\log(T) = x\beta + a\psi + \epsilon$, with $\mathbb{E}(\epsilon) = 0$. Estimators $\hat{\beta}_n$ and $\hat{\psi}_n$ are

obtained with the DWSurv algorithm using the actual sample and residuals are calculated as $\hat{\epsilon}_i = \log(t_i) - x_i\hat{\beta}_n - a_i\hat{\psi}_n$. Then, by drawing n times with replacement from the empirical distribution of the residuals, say $\epsilon_1^\star, ..., \epsilon_n^\star$, and generating the responses conditional on the observed covariates (x_i, a_i) from the AFT parametric mean model as $t_i^\star = x_i\hat{\beta}_n + a_i\hat{\psi}_n + \epsilon_i^\star$, a bootstrap sample is obtained as $s^\star = \{(t_1^\star, x_1, a_1), ..., (t_n^\star, x_n, a_n)\}$. Therefore, step 1 of the bootstrap algorithm samples from $\hat{F}_T$ instead of $\hat{F}$. Note that if a specific distribution for the error term ϵ is deemed reasonable, residuals could be drawn from that parametric distribution instead of the empirical distribution.

With censored data, the observations are in the form $(y_i, \delta_i, x_i, a_i)$. A natural extension of the non-parametric bootstrap consists of resampling the quadruplets $(y_i, \delta_i, x_i, a_i)$ to obtain bootstrap samples $s^\star = \{(y_1^\star, \delta_1^\star, x_1^\star, a_1^\star), ..., (y_n^\star, \delta_n^\star, x_n^\star, a_n^\star)\}$. Alternatively, one could take advantage of the structure of the censored data to generate bootstrap samples. We borrow from Efron & Tibshirani (1986) who discuss a similar idea (see Section 5 therein). A typical data point $(Y_i, \delta_i, X_i, A_i)$ can be thought to be generated in the following way: the event indicator is selected from a Bernoulli distribution $\delta_i|X_i, A_i \sim \text{Ber}(p_\delta)$. If the event is observed, $Y_i = T_i$ is generated from a survival distribution $T_i|X_i, A_i \sim F_T$. If the event is not observed, $Y_i = C_i$ is generated from a censoring distribution F_C. The observed sample was generated from the unknown probability mechanism (p_δ, F_T, F_C) and an obvious choice to draw samples from that mechanism is to replace (p_δ, F_T, F_C) by $(\hat{p}_\delta, \hat{F}_T, \hat{F}_C)$. The DWSurv algorithm assumes a parametric model for the probability of observing the event, $P(\Delta = 0|X, A; \lambda)$, such that $\hat{p}_\delta$ could naturally be taken as $P(\Delta = 0|X, A; \hat{\lambda})$. Also, the DWSurv algorithm does not make use of the actual censoring times in the estimation procedure and hence specifying $\hat{F}_C$ and resampling censoring times is not necessary to construct bootstrap samples. The parametric bootstrap modifies step 1 as following: for each observation i, generate an event indicator $\delta_i^\star$ from $\hat{p}_\delta$; if $\delta_i^\star = 1$, generate $y_i^\star$ as $t_i^\star = x_i\hat{\beta}_n + a_i\hat{\psi}_n + \epsilon_i^\star$ from $\hat{F}_T$ as in the uncensored case; if $\delta_i^\star = 0$, assign any value to $y_i^\star$. The resulting bootstrap sample is $s^\star = \{(y_1^\star, \delta_1^\star, x_1, a_1), ..., (y_n^\star, \delta_n^\star, x_n, a_n)\}$ and steps 2 and 3 of the bootstrap algorithm

are applied. Compared to uncensored data, this parametric bootstrap makes an additional parametric assumption on the probability of observing the event.

The non-parametric bootstrap can be directly applied to a DTR with two stages or more, where individuals with their complete trajectories are drawn with replacement to construct bootstrap samples in which the DWSurv algorithm is applied to estimate the parameters across all stages. For a DTR with two stages or more, the parametric bootstrap needs to break down the estimation into as many one-stage DTRs as the total number of stages. For example, in a two-stage DTR, estimation and inference for the second stage blip parameter ψ_2 are made by considering only the individuals who entered the second stage. Conditional on their current history (H_2, A_2), bootstrap resamples of the form $s^\star = \{(y_{21}^\star, \delta_1^\star, h_{21}, a_{21}), ..., (y_{2n}^\star, \delta_n^\star, h_{2n}, a_{2n})\}$ are drawn, where only the stage 2 survival time is resampled. The event indicator $\delta^\star$ is selected from a Bernoulli distribution with probability $\hat{p}_{\delta_2}$ corresponding to the fitted values of the censoring model in the second stage specified in step 1 of the DWSurv algorithm. For those with an observed event $\delta^\star = 1$, $y_{2i}^\star$ is generated from the assumed AFT model for the stage 2 survival time by drawing from the residuals. From the B bootstrap resamples constructed this way, a one-stage DTR is fitted to each resample with $y_{2i}^\star$ as the outcome and confidence intervals are derived for ψ_2. Estimation and inference for the first stage parameters ψ_1 are carried out separately. Conditional on the stage 1 history (H_1, A_1), bootstrap resamples $s^\star = \{(\tilde{y}_1^\star, \delta_1^\star, h_{11}, a_{11}), ..., (\tilde{y}_n^\star, \delta_n^\star, h_{1n}, a_{1n})\}$ are drawn, where now the pseudo-survival times are resampled. The event indicator is generated with probability $\hat{p}_{\delta_1}$ corresponding to the fitted values of the censoring model in the first stage specified in step 4 of the DWSurv algorithm and $\tilde{y}_i^\star$ is generated from the AFT model assumed for the pseudo-survival times by drawing from the residuals. A one-stage DTR is then fitted to each bootstrap resample with $\tilde{y}_i^\star$ as the outcome and confidence intervals are derived for ψ_1.

5.4 Simulation Study

The simulation study compares the performance of the five following methods to construct 95% confidence intervals for the DWSurv blip parameters in one- and two-stage DTRs:

1. the adjusted asymptotic variance given by (5.6);
2. the naive asymptotic variance without adjustments given by (5.7);
3. the non-parametric bootstrap;
4. the parametric bootstrap which resamples residuals from their empirical distribution (also referred to as parametric bootstrap 1);
5. the parametric bootstrap which assumes a Log-normal distribution for the survival times and thus resamples residuals from a Normal distribution (also referred to as parametric bootstrap 2).

The performance of the methods is characterized in terms of coverage probability, interval width, and computational time.

5.4.1 Data Generating Mechanisms

The simulated data mimic two observational studies, one with only one stage of intervention and the other with up to two stages, both with 30% censoring that depends on individual characteristics at baseline. Denote $\text{expit}(v) = \exp(v)/(1+\exp(v))$ defined for $v \in \mathbb{R}$. The first stage treatment is assigned with $P(A_1 = 1|X_{11}) = \text{expit}(-1+2X_{11})$ where X_{11} is a continuous covariate measured at baseline generated from a Uniform [0.1,1.29]. An additional binary baseline covariate X_{12} is generated from a Bernoulli distribution with $P(X_{12} = 1) = 0.4$. When applicable, the second stage treatment is assigned with $P(A_2 = 1|X_{21}) = \text{expit}(2.8 - 2X_{21})$ where X_{21} is a continuous covariate measured at the beginning of the second stage generated from a Uniform [0.9,2]. An additional binary covariate measured at the beginning

of the second stage X_{22} is generated from a Bernoulli distribution with probability p_2, whose value will be defined later to yield different non-regular scenarios. The censoring indicator is generated with $P(\Delta = 1|X_{12}) = \text{expit}(0.1 + 3X_{12})$ and the indicator η_2 of whether an individual enters the second stage is taken as 1 for all individuals such that the event or censoring only can occur in the second stage.

For the one-stage DTR, individuals who experience an event have their survival time T generated from an AFT model as

$$\log(T) = \beta_{10} + \beta_{11}X_{11} + \beta_{12}X_{12} + \beta_{13}X_{11}^4 + A_1(\psi_{10} + \psi_{11}X_{11}) + \epsilon$$

where $\epsilon \sim N(0, 0.09)$ and $(\psi_{10}, \psi_{11})^T = (0.1, 0.1)^T$, defining Log-normal survival times. Two sets of treatment-free parameters are considered: a linear relationship between X_{11} and $\log(T)$ sets $\beta_{13} = 0$ and $(\beta_{10}, \beta_{11}, \beta_{12})^T = (4.7, 1.5, -0.8)^T$, and a nonlinear relationship sets $(\beta_{10}, \beta_{11}, \beta_{12}, \beta_{13})^T = (4.7, 3, -0.9, 0.05)^T$. Results from the simulations with a nonlinear relationship are presented in the Supplemental Materials D.2–D.4.

For the two-stage DTR, the stage 2 survival time T_2 is generated for the individuals who experienced an event as

$$\log(T_2) = \beta_{20} + \beta_{21}X_{21} + \beta_{22}X_{22} + \beta_{23}X_{21}^3 + \beta_{24}X_{11} + A_2(\psi_{20} + \psi_{21}X_{22}) + \epsilon_2$$

where $\epsilon_2 \sim N(0, 0.09)$ and $(\beta_{20}, \beta_{21}, \beta_{22}, \beta_{23}, \beta_{24})^T = (4, 1.1, 0.01, -0.2, 0.1)^T$. The value of the blip parameters (ψ_{20}, ψ_{21}) is chosen later to yield different non-regular scenarios. The pseudo-survival time under optimal stage 2 treatment is generated as

$$\log(\tilde{T}) = \beta_{10} + \beta_{11}X_{11} + \beta_{12}X_{12} + A_1(\psi_{10} + \psi_{11}X_{11}) + \epsilon_1$$

where $\epsilon_1 \sim N(0, 0.09)$ and $(\beta_{10}, \beta_{11}, \beta_{12}, \psi_{10}, \psi_{11})^T = (6.3, 0.5, -0.01, 0.1, 0.1)^T$. The observed overall survival time T depends on whether an individual receives his optimal stage

2 treatment $A_2^{\text{opt}} = \mathbb{I}(\psi_{20} + \psi_{21}X_{22} > 0)$. An individual who receives his optimal stage 2 treatment has $T = \tilde{T}$. Individuals who do not receive their optimal stage 2 treatment have $T = \tilde{T} - T_2^{\text{opt}}$ where

$$T_2^{\text{opt}} = \exp\{\log(T_2) + (\psi_{20} + \psi_{21}X_{22})(A_2^{\text{opt}} - A_2)\}.$$

In both the one- and two-stage DTRs, individuals who are censored have their censoring time generated from an Exponential distribution with rate 1/300. Sample sizes between 100 and 10,000 for the one-stage DTR and between 300 and 10,000 for the two-stage DTR are considered and the number of bootstrap resamples is 1000. Each simulation study is based on 1000 simulated data sets.

5.4.2 Unknown Error Distribution

In a one-stage DTR, we compare the five methods across various sample sizes when the survival times are generated from a Log-normal distribution with a linear relationship between X_{11} and $\log(T)$, which corresponds to the data generating mechanism described in Section 5.4.1. In this case, the expectation of the errors is zero as required by DWSurv. We also perform the comparison with survival times generated from a Weibull distribution obtained by defining $\epsilon = \log(\epsilon^{\dagger})$ with $\epsilon^{\dagger}$ following a Weibull distribution with shape 4 and scale 1. This defines a left-skewed error distribution with mean -0.1, thus violating the requirement on the expectation of the errors. In this case, the estimators are still unbiased as the systematic deviation from zero is absorbed by the intercept of the model and the impact on inferences is investigated in the simulations.

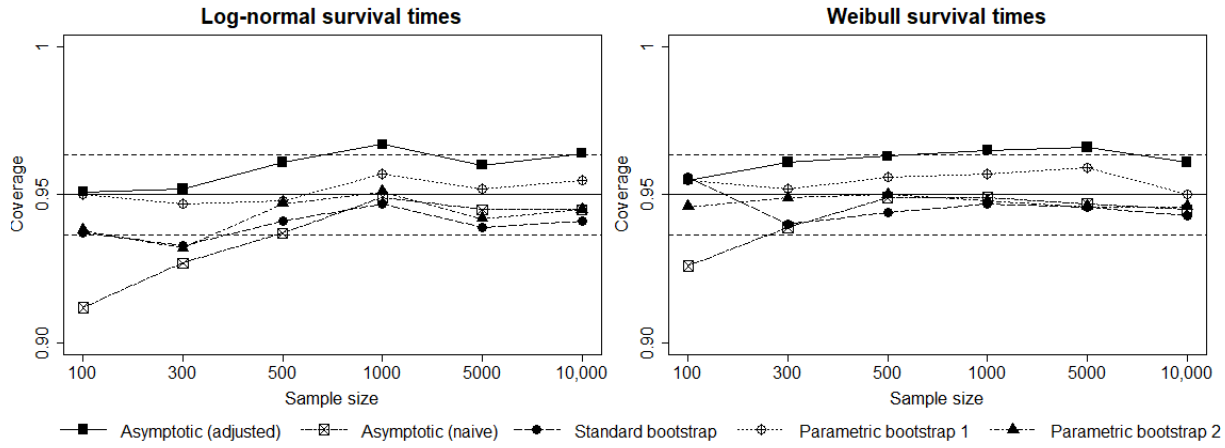

Figure 5.1: Coverage of 95% confidence intervals for ψ_{11} in a one-stage DTR derived with five methods across 1000 simulated data sets with sample sizes ranging from 100 to 10,000 with Log-normal or Weibull survival times. The horizontal dashed lines represent the bounds for testing the null hypothesis that the coverage equals the nominal rate.

Figure 5.1 shows the coverage of 95% confidence intervals for the blip parameter ψ_{11} constructed with the five methods across various sample sizes. As expected, the coverages approach the nominal level as the sample size increases. While the naive asymptotic variance yields lower coverage than the other methods for the smallest sample size of n=100, the adjusted asymptotic variance exhibits over-coverage for all sample sizes except n=100. Results are comparable regardless of the true error distribution. Specifically, the parametric bootstrap 2, which mistakenly resamples the survival times from a Log-normal distribution when the true distribution is Weibull, performs comparably to the other methods. Results for ψ_{10} or with data simulated with a nonlinear association between X_{11} and $\log(T)$ are similar (see Supplemental Material D.2).

5.4.3 Model Misspecification

This set of simulations considers a one-stage DTR with true Log-normal survival times generated with a linear relationship between $\log(T)$ and the baseline covariate X_{11}. The five methods for inferences on $\boldsymbol{\psi_1}$ are compared when the treatment-free, treatment or cen-

soring models are alternatively misspecified as following: the true treatment-free model $\beta_{10}+\beta_{11}X_{11}+\beta_{12}X_{12}$ is misspecified by omitting X_{12}, a variable associated with the survival time and the probability of censoring; the treatment and censoring models are both fitted as a function of an intercept only while their distribution truly depends on X_{11} and X_{12}, respectively. Additional misspecifications are considered, such as generating survival times with a true nonlinear treatment-free model as $\beta_{10}+\beta_{11}X_{11}+\beta_{12}X_{12}+\beta_{13}X_{11}^4$ and misspecifying the treatment-free model by omitting the nonlinear component X_{11}^4. Additional results are presented in the Supplemental Material D.3.

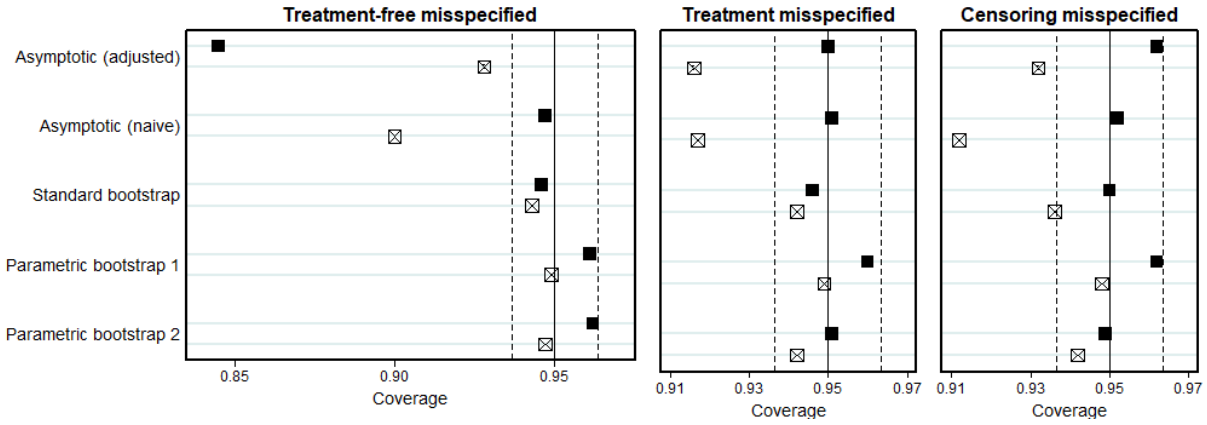

Figure 5.2: Coverage of 95% confidence intervals for ψ_{11} in a one-stage DTR derived with five methods across 1000 simulated data sets for sample sizes n=100 (☒) and n=1000 (■) under misspecification of the treatment-free, treatment or censoring model. The dashed lines represent the bounds for testing the null hypothesis that the coverage equals the nominal rate.

Figure 5.2 shows the coverage of 95% confidence intervals for the blip parameter ψ_{11} under misspecification of the treatment-free, treatment or censoring model for sample sizes n=100 and 1000. The non-parametric and parametric bootstraps perform well in terms of coverage regardless of which model is misspecified, for small and larger sample sizes. However, other ways of misspecifying the treatment-free model lead to poorer performance of the three bootstrap approaches, with coverage just above 92% across all sample sizes (see Supplemental Material D.3). The two asymptotic variances yield low coverage with n=100 when either one of the models is misspecified but yield nominal coverage with a larger sample size when

the treatment or censoring model is misspecified. However, misspecifying the treatment-free model yields even lower coverage for the confidence intervals constructed with the adjusted asymptotic variance with n=1000 despite the corresponding mean confidence interval width of 3.15 being much larger than that with the other methods (0.48 to 0.51). A closer look into the inferences with the two asymptotic variances shows that the distribution of the asymptotic standard errors has a heavy right tail driven by large outliers, explaining the large mean width of the confidence intervals, but also contains standard error estimates smaller than that estimated with the other methods, explaining the poor coverage (see Supplemental Material D.5).

5.4.4 Non-regularity

The impact of non-regularity on inferences is studied with a two-stage DTR with Log-normal survival times. Table 5.1 presents the parameters of eight data generating mechanisms characterizing regular and non-regular scenarios, which have already been used to study the impact of non-regularity on inferences for DTR parameters with continuous outcomes (Chakraborty et al., 2013; Simoneau et al., 2017). Sample sizes starting from n=300 are considered in a two-stage DTR to ensure enough individuals reach the second stage. The degree of non-regularity p is the probability of generating histories $\boldsymbol{h_2} = (a_1, x_{11}, x_{12}, x_{21}, x_{22})$ such that the stage 2 treatment effect is null or small. Therefore, p is equal to $P(\psi_{20} + \psi_{21}X_{22} = 0)$ and depends on the stage 2 blip parameters $\boldsymbol{\psi_2} = (\psi_{21}, \psi_{21})$ and on X_{22} through $p_2 = P(X_{22} = 1)$. In practice, p can be estimated by the proportion of individuals for whom the optimal stage 2 treatment is non-unique. Altogether, the parameters p, p_2 and $\boldsymbol{\psi_2}$ define a non-regular scenario if $p > 0$, a near non-regular scenario if $p = 0$ but the linear combination $\psi_{20} + \psi_{21}X_{22}$ is "close" to zero for all or some of the individuals, and a regular scenario otherwise.

Figure 5.3 shows the coverage of 95% confidence intervals for the blip parameter ψ_{11} across

Table 5.1: Description of the eight regular to non-regular simulation scenarios.

	Scenario	p	$\boldsymbol{\psi_2}$	p_2	Description
1	Non-regular	1	$(0, 0)$	0.3	No effect of stage 2 treatment.
2	Near non-regular	0	$(0.01, 0)$	0.3	Weak stage 2 treatment effect.
3	Non-regular	1/2	$(-0.5, 0.5)$	0.5	No effect of stage 2 treatment for half of the subjects, large effect for the other half.
4	Near non-regular	0	$(-0.5, 0.49)$	0.5	Weak stage 2 treatment effect for half of the subjects, large effect for the other half.
5	Non-regular	1/2	$(-0.2, 0.2)$	0.5	No effect of stage 2 treatment for half of the subjects, moderate effect for the other half.
6	Near non-regular	0	$(-0.2, 0.19)$	0.5	Weak stage 2 treatment effect for half of the subjects, moderate effect for the other half.
7	Regular	0	$(-0.9, 0.6)$	0.5	Large effect of stage 2 treatment for half of the subject, moderate effect for the other half.
8	Regular	0	$(0.2, -0.7)$	0.5	Idem to scenario 7, with smaller effects.

eight regular to non-regular scenarios with sample sizes n=300 and 1000. Across all scenarios, the adjusted asymptotic variance always yields confidence intervals with coverage higher than the nominal rate, which goes in line with what is observed in the previous simulations (c.f. Section 5.4.2). The coverages with the four other methods are similar within each scenario and are typically closer to the nominal rate for larger sample sizes. There is no discernable pattern allowing to differentiate between the performance of the five methods across regular, near non-regular or non-regular scenarios. All methods are comparable when focusing on inferences for ψ_{10} (see Supplemental Material D.4).

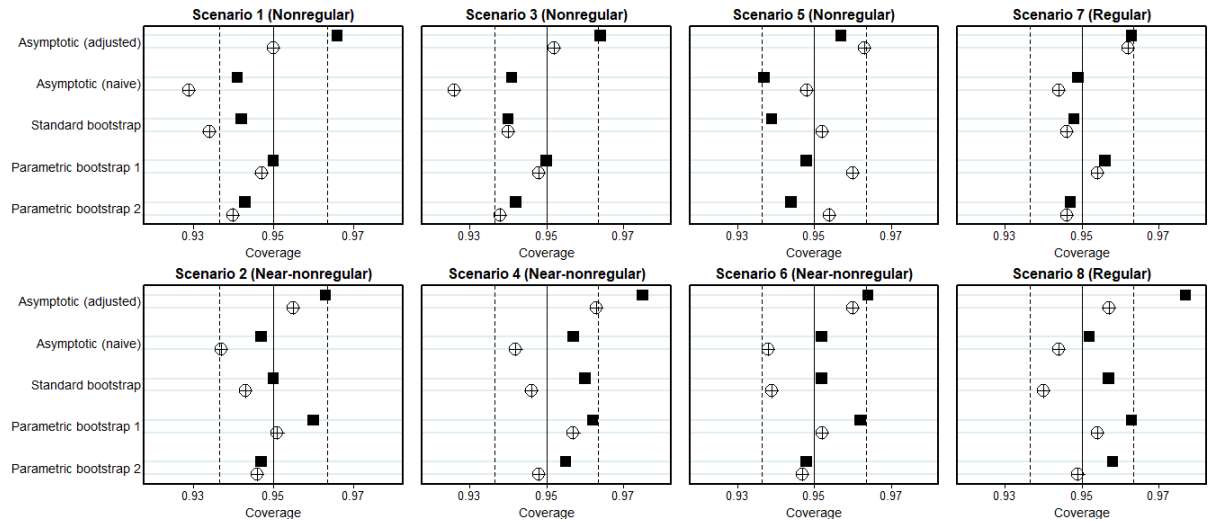

Figure 5.3: Coverage of 95% confidence intervals for ψ_{11} in a two-stage DTR derived with five methods across 1000 simulated data sets with different degrees of non-regularity for sample sizes n=300 ($\oplus$) and n=1000 (■). The dashed lines represent the bounds for testing the null hypothesis that the coverage equals the nominal rate.

5.5 Discussion and Conclusion

Characterizing the uncertainty of DWSurv or, more generally, DTR estimators is a challenging task in practice. Although asymptotic variance formulae are derived for the DWSurv estimators, it is not clear how this variance performs in finite samples or in any situation likely to occur in practice. Our simulation study focuses on the performance of five methods to construct 95% confidence intervals for DWSurv parameters: the asymptotic variance with adjustments for the estimation of the nuisance parameters, the asymptotic variance without adjustments, the standard non-parametric bootstrap, and two parametric bootstraps resampling from the distribution of the survival times. A discussion of the simulation study results and of practical considerations follows, allowing to derive well-informed recommendations for users of DTR statistical methods, in particular, of DWSurv.

The first aspect of the simulation study aims to assess how robust the methods are to the violation of the requirement for zero expectation of the true, yet unknown in practice,

error distribution. Except for the adjusted asymptotic variance, the compared methods yield confidence intervals with nominal coverage probabilities for Log-normal and Weibull survival times despite the fact that the error distribution of the latter is skewed with expectation smaller than zero. The adjusted asymptotic variance generally yields conservative confidence intervals, even for larger sample size. This over-coverage, which has already been observed previously with DWSurv (Simoneau et al., 2018), was one of the motivations for investigating the performance of the naive asymptotic variance, which performs adequately for moderate to large sample sizes in a one-stage DTR.

The robustness of the methods to misspecification of the nuisance models is evaluated in the second series of simulations. Only the adjusted asymptotic variance exhibits poor coverage despite larger average confidence widths, specifically in the case where the treatment-free model is misspecified. The coverages and confidence interval widths obtained with the two parametric bootstraps are not affected by misspecification of the nuisance models. As the parametric bootstrap relies on the fitted probabilities of censoring to resample right-censored survival times, it is surprising that misspecification of the censoring model does not affect the performance of the method. Omitting a covariate strongly associated with the censoring mechanism and the outcome may lead to less desirable performance.

In the last set of simulations, all methods yield confidence intervals with good coverage in the first stage of a two-stage DTR where non-regularity may affect inferences. The regular, near non-regular and non-regular scenarios in our simulation study have been used before to showcase the negative impact of non-regularity for constructing confidence intervals in DTR regression-based methods. It is unclear why those same scenarios do not yield similar conclusions with DWSurv. One possible explanation is that, with survival times, not all individuals necessarily enter the second stage due to censoring or experiencing the event before. For them, a pseudo-survival time is not computed, eliminating their contribution to non-regular estimation, and thus perhaps attenuating the negative impact of non-regularity.

We do note that, in some scenarios, coverages are near the lower bound for testing the null hypothesis that the coverage is different than the nominal rate of 0.95 but this problem disappears with the larger sample sizes of n=10,000. The m-out-of-n bootstrap (Chakraborty et al., 2013) has been proposed as an alternative to the standard non-parametric bootstrap for the inference of DTR parameters in non-regular scenarios, which could provide an interesting alternative for DWSurv.

The simulation study considers several data generating mechanisms, each with variations in the choice of the parameters. The parameters are not chosen to mimic one particular data set or application but rather to encompass general simple DTRs. We chose to consider a moderate, plausible amount of censoring across all data generating mechanisms. The performance of the asymptotic variance was previously evaluated in a simulation study with censoring up to 60% and showed comparable results to those depicted here. Note that as the DWSurv algorithm does not use the censored individuals except in the construction of the weights, increasing the proportion of censoring has an impact similar to that of decreasing the sample size. The applicability of the standard non-parametric bootstrap may be compromised with an increasing proportion of censoring because the censored individuals may be overrepresented in some resamples and the estimation of the parameters in those samples may be unstable, especially through matrix inversion operations in the GEE in DWSurv. The parametric bootstrap would not suffer from this problem as it controls the proportion of censored observations in the resamples by sampling the event indicator δ.

In practice, computational resources may be limited. While the computational cost of the two asymptotic variances simply increases with the sample size n, the computational cost of the bootstrap procedures also increases with the total number B of bootstrap samples (see Supplemental Material D.1). Moreover, the computational cost of the parametric bootstrap, which needs to be applied as many times as the maximal number of stages in the DTR, also increases with the number of stages.

The results from the simulation study, practical considerations, and insights from work done by others permit some recommendations for the construction of confidence intervals for decision rule parameters in DWSurv. Although the adjusted asymptotic variance formula is justified on theoretical grounds, it overestimates the variance of the estimators in some cases, leading to conservative confidence intervals as compared to alternative methods. Moreover, it is not robust to misspecification of the treatment-free model, leading to the more serious problem of confidence intervals with poor coverage yet large width, on average. We thus recommend using the asymptotic formula only if one is confident about the correct specification of the treatment-free model, for example, if enough covariates are available to the user and the sample size is large. We recommend using the naive asymptotic variance to benchmark confidence intervals obtained with the adjusted asymptotic variance only when the number of observed events is above 700, which may correspond to a sample size of n=1000 with 30% censoring. The bootstrap approaches consistently yield confidence intervals with nominal coverages when the underlying error distribution does not have zero expectation or when one of the nuisance models is misspecified, although under-coverage is observed under certain misspecifications of the treatment-free model. The bootstrap is thus a viable option when computationally feasible. Although the expected negative effects of non-regularity on inferences were not observed in our simulation study, caution should be taken when the first stage estimators may be non-regular, that is, when the treatment effect in the second stage is thought to be null or small as any method may lead to under-coverage.

Funding

This work was supported by a doctoral research grant from the Fonds de recherche du Québec – Nature et technologies [199803].

Appendix D – Supplemental Materials

Contains the following sections:

D.1 **Computational Times** – additional simulation results about computational time of the five methods.

D.2 **Additional Simulation Results: Unknown Error Distribution** – simulation results from other data generating mechanisms to complement Section 5.4.2.

D.3 **Additional Simulation Results: Model Misspecification** – simulation results from other data generating mechanisms to complement Section 5.4.3.

D.4 **Additional Simulation Results: Non-regularity** – simulation results from other data generating mechanisms to complement Section 5.4.4.

D.5 **Details on the Performance of the Asymptotic Variance** – additional descriptive statistics about the performance of the adjusted asymptotic variance in Section 5.4.3

Chapter 6

Optimal Dynamic Treatment Regimes with Survival Outcomes: An Application to the Treatment of Type 2 Diabetes using a Large Observational Database

Preamble to Manuscript 4. The idea of this case study arose from an actual clinical question about the management of type 2 diabetes: once metformin in monotherapy fails to achieve adequate glycemic control, what treatment(s) should be recommended next in order to delay the occurrence of diabetic complications? This project highlighted the usefulness of DWSurv to answer a real clinical question but also pointed out the difficulty of applying DWSurv when patients follow heterogeneous treatment pathways in practice. The original contributions in this manuscript are (i) providing an individualized treatment rule that can be used in clinical practice about which of sulfonylurea or dipeptidyl peptidase-4 inhibitors should be added to metformin once metformin in monotherapy fails to achieve the therapeutic goals, and (ii) providing detailed summary statistics about treatment path-

ways followed by patients with T2D in the CPRD database. At the time of submitting this thesis, this manuscript was under a second round of review in the *American Journal of Epidemiology*.

Optimal Dynamic Treatment Regimes with Survival Outcomes: An Application to the Treatment of Type 2 Diabetes using a Large Observational Database

Gabrielle Simoneau[1], Erica EM Moodie[1], Laurent Azoulay[1], Robert W Platt[1]

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

Abstract

Sequences of treatments that adapt to the patient's changing condition over time are often needed for the management of chronic diseases. A dynamic treatment regime (DTR) consists of personalized treatment rules to be applied through the course of a disease that input the patient's characteristics at the time of decision-making and output a recommended treatment. An optimal DTR is the sequence of treatments that yields the best clinical outcome for patients sharing similar characteristics. Methods to estimate optimal DTRs, which must disentangle short- and long-term treatment effects, can be theoretically involved and hard to explain to clinicians, especially when the outcome to optimize is a survival time subject to right-censoring. In this paper, we decipher dynamic weighted survival modeling, a method to estimate DTRs with survival outcomes, and illustrate how it can be used to answer an important clinical question about the treatment of type 2 diabetes using data from the Clinical Practice Research Datalink, a large primary care database. We identify an individualized treatment rule about which add-on treatment to recommend when metformin in monotherapy does not achieve the therapeutic goals but fail to answer more complex questions given the heterogeneity of treatment pathways observed in practice.

6.1 Introduction

The treatment of chronic and recurring diseases often consists of a sequence of therapies that adapt to a patient's evolving condition over time. The clinician decides which treatment to recommend next based on the patient's characteristics (value of a biomarker, comorbidities, patient's response to previous treatments) observed at the time of decision-making. In such a situation, making the "best" treatment decision does not simply involve answering the question "What is the best treatment for a specific patient at this time to prevent a disease-related outcome?" but rather addresses the more complex question "What is the best sequence of treatments for a specific patient?" Methods for estimating an optimal dynamic treatment regime (DTR) are concerned with the latter question.

A DTR is a sequence of treatment rules, one for each decision time point, that inputs current patient's characteristics, including information about previous treatments, and outputs a recommended treatment. Of interest is to identify an optimal DTR, that is, the sequence of treatment rules that yields the best expected outcome for individuals sharing similar characteristics. Because short- and long-term treatment effects may be hard to disentangle when multiple treatments are taken successively, simple statistical methods that optimize each treatment decision separately rather than optimizing the sequence of treatments jointly may fail to identify the DTR that indeed leads to an optimal outcome.

We consider the treatment of type 2 diabetes (T2D) which typically consists of a sequence of lifestyle and drug therapies that aims to delay major diabetic complications and death. Our interest is to identify a set of personalized treatment rules that maximizes the time until such negative events occur. When the outcome to maximize is a survival time subject to right-censoring, dynamic weighted survival modeling (DWSurv) offers a theoretically robust and interpretable framework to identify an optimal DTR (Simoneau et al., 2018).

The aims of this article are to offer an accessible overview of DWSurv and to illustrate how

the method can be used to help discover optimal DTRs for T2D using large administrative databases. We study the question "Is sulfonylurea or dipeptidyl peptidase-4 inhibitors (DPP-4i) the best add-on treatment to metformin for maximizing the time until a cardiovascular event or death?" and describe extensions of this to a multi-stage DTR.

6.2 Illustrative Example: Type 2 Diabetes

Guidelines on the management of T2D recommend a sequence of lifestyle and drug therapies to lower, and maintain an optimal, glycemic level in order to reduce the risk of diabetic complications (Garber et al., 2019; McGuire et al., 2016). Metformin is the recommended first-line oral agent. When metformin in monotherapy fails to achieve adequate glycemic control, treatment guidelines recommend to add a second and even a third oral agent before eventually transitioning to injectable therapy such as insulin. However, there remain uncertainties on which sequence of treatments should be followed when treatment with metformin fails, although it is widely recognized that the choice of therapies must be individualized (Garber et al., 2019; Inzucchi et al., 2015). In the absence of comparative-effectiveness trials that take into account the dynamic nature of the treatment of T2D, recommendations cannot easily be made.

6.3 Methods: Dynamic Weighted Survival Modeling

Compared to other methods that have been proposed to estimate multi-stage optimal DTRs with survival outcomes from observational data (Goldberg & Kosorok, 2012; Hager et al., 2018; Huang et al., 2014; Jiang et al., 2017b), DWSurv is appealing because of its accessibility and its theoretical robustness. It requires specifying parametric models for quantities that can be informed by clinical knowledge (e.g. treatment assignment and censoring mech-

anisms). It is doubly-robust and is equipped with tools for model checking and inferences about the estimated decision rules.

6.3.1 Notation and Assumptions

We first consider a single-stage DTR and assume that observational data from n individuals are available. Individual data are $(\boldsymbol{X_1}, A_1, Y_1, \delta)$ where $\boldsymbol{X_1}$ are covariates observed at the beginning of the first stage, prior to the administration of treatment A_1, a binary option coded as $\{0, 1\}$. Y_1 is a survival or censoring time observed at the end of the follow-up. We use T_1 and C to respectively denote the survival and censoring times and the indicator δ determines if Y corresponds to a survival time (1) or a censoring time (0). A single-stage DTR is defined as an individualized treatment rule $d_1(\boldsymbol{h_1})$ to be applied at the beginning of the first stage, after the patient's chacteristics $\boldsymbol{h_1}$ (history) are observed, i.e. $\boldsymbol{H_1} = \boldsymbol{X_1}$. The decision rule $d_1(\boldsymbol{h_1})$ is a function of the history that returns a recommended treatment a_1=0 or a_1=1.

DWSurv adopts the counterfactual outcomes framework. It defines $T_1^{a_1}$, the survival time if, possibly contrary to the fact, treatment a_1 is received. An optimal single-stage DTR is the treatment rule $d_1^{\text{opt}}(\boldsymbol{h_1})$ that maximizes the average counterfactual survival time $T_1^{a_1}$ conditional on individual characteristics. Four assumptions (general enough to apply to DTRs with more than one stage) are necessary for the estimation of an optimal DTR with DWSurv: (i) an individual's survival time is not influenced by others' treatment allocations, (ii) there are no unmeasured confounders at each stage, (iii) conditional on the observed history, censoring is non-informative at each stage, and (iv) an individual has a positive probability of receiving either treatment at each stage and a positive probability of experiencing an event.

6.3.2 Estimation

DWSurv requires modeling the treatment assignment and censoring mechanisms as well as the survival time. A semi-parametric accelerated failure time (AFT) model for the average counterfactual survival time, referred to as the outcome model, is specified as

$$\mathbb{E}[\log(T_1^{a_1})|\boldsymbol{H_1} = \boldsymbol{h_1}, A_1 = a_1; \boldsymbol{\beta_1}, \boldsymbol{\psi_1}] = \boldsymbol{\beta_1^T h_{1\beta}} + a_1 \boldsymbol{\psi_1^T h_{1\psi}} \tag{6.1}$$

where $\boldsymbol{h_{1\psi}}$, called tailoring variables, and $\boldsymbol{h_{1\beta}}$ are two (possibly different) subsets of the history $\boldsymbol{h_1}$. The outcome model is separated into two components: the treatment-free model $\boldsymbol{\beta_1^T h_{1\beta}}$ which does not depend on the stage 1 treatment and the blip component $a_1 \boldsymbol{\psi_1^T h_{1\psi}}$ which depends on a_1. The optimal first stage treatment is that which maximizes (6.1) with respect to a_1, here implying decision rules of the form $d_1^{\text{opt}} = \mathbb{I}(\boldsymbol{\psi_1^T h_{1\psi}} > 0)$.

Th estimation of the parameters $(\boldsymbol{\beta_1}, \boldsymbol{\psi_1})$ must account for censored individuals for whom T_1 is not observed and must also eliminate any confounding between the treatment assignment and the survival time. This is achieved via a weighting argument by using balancing weights (Li et al., 2018; Simoneau et al., 2018; Wallace & Moodie, 2015). For example, the overlap weights

$$w_1(\delta, a_1, \boldsymbol{h_1}) = \frac{|a_1 - \mathbb{E}(A_1|\boldsymbol{h_1})|}{\mathbb{P}(\Delta = \delta_i|\boldsymbol{h_1}, A_1)} \tag{6.2}$$

define a target population in clinical equipoise, that is, individuals whose characteristics make them almost equally likely to receive any treatment option. Models for the probability of treatment $\mathbb{E}(A_1|\boldsymbol{h_1})$ and the probability of experiencing an event $\mathbb{E}(\Delta|\boldsymbol{h_1}, a_1)$ are proposed and estimated, for example with logistic regressions, and the weights are computed.

Given the outcome model (6.1) and the estimated weights $\hat{w}_1$, estimators $(\hat{\boldsymbol{\beta}}_{\boldsymbol{1}}, \hat{\boldsymbol{\psi}}_{\boldsymbol{1}})$ are obtained by solving weighted estimating equations implemented in an unpublished version of the `DTRreg` package in `R`. The estimated optimal first stage treatment is $a_1^{\text{opt}} = \mathbb{I}(\hat{\boldsymbol{\psi}}_{\boldsymbol{1}}^{\boldsymbol{T}} \boldsymbol{h_{1\psi}} > 0)$.

6.3.3 Double-robustness

DWSurv yields doubly-robust estimators of the parameters $\boldsymbol{\psi_1}$ used to construct the decision rule. All other parameters ($\boldsymbol{\beta_1}$ and the parameters in the treatment and censoring models) are nuisance quantities. The double-robustness property means that $\hat{\boldsymbol{\psi}}_{\mathbf{1}}$ are unbiased estimators of $\boldsymbol{\psi_1}$ when the treatment and censoring models are correctly specified or when the treatment-free component of the outcome model is correctly specified or when all three models are correctly specified. The obvious advantage of doubly-robustness is that it provides protection against misspecification of some models. Another advantage is that some models may be easier to inform from a clinical perspective.

6.3.4 Inferences and Model Checking

Confidence intervals can be calculated for the parameters $\boldsymbol{\psi_1}$ using asymptotic variance formulae or parametric and non-parametric bootstraps. Residual plots can be used with DWSurv to assess model specifications (Rich et al., 2010), with residuals calculated on the logarithmic scale for individuals who experienced an event as $\log(T_1) - \mathbb{E}[\log(T_1^{a_1})|\boldsymbol{h_1}, A_1; \hat{\boldsymbol{\beta}}_{\mathbf{1}}, \hat{\boldsymbol{\psi}}_{\mathbf{1}}]$. The double-robustness property can also be exploited to provide reassurance that some models are correctly specified (Wallace et al., 2016).

6.4 The Data

The objective of the illustrative example is to estimate an individualized treatment rule that recommends the best agent to add to metformin between sulfonylurea and DPP-4i in order to maximize the time until a cardiovascular event or death. We used data from the Clinical Practice Research Datalink (CPRD), a large primary care database in the United Kingdom (UK).

6.4.1 Study Population and Definitions

We assembled a base cohort of patients aged 40 or older with a first-ever prescription of metformin in monotherapy between April 1, 1997, and March 31, 2018, and at least one year of history in the CPRD prior to metformin. Women with a history of polycystic ovarian syndrome and gestational diabetes (other known indications for metformin) were excluded. The study cohort was composed of all patients who added sulfonylurea or DPP-4i to metformin, with study entry defined as the date when the first add-on was recorded. Patients were considered to have added one of the drugs if a new prescription was recorded within 30 days after a prescription of metformin (Yu et al., 2015). Details on the CPRD database and the study cohort are available in the Supplemental Material E.1.

Our analysis compared the addition of sulfonylurea ($A_1 = 0$) and DPP-4i ($A_1 = 1$) to metformin. The outcome is the time from study entry until the occurrence of a cardiovascular event (stroke, myocardial infarction or peripheral vascular disease) or death. Patients were censored when they made any change to their treatment regime (adding or switching drugs) or when they were lost to follow-up (end of study date or transferred out of the practice). The following covariates were recorded at study entry: age, sex, years on metformin, socio-economic status (SES), smoking status, body mass index (BMI), glycated hemoglobin (HbA1c), and comorbidities. Covariates were defined with the presence or absence of medical, diagnosis, or prescription codes any time before the study entry (see Supplemental Material E.2 for details).

6.4.2 Model Development and Fitting

Covariates selection for all models was made *a priori*. A parametric model for the outcome was specified as a linear combination of all available covariates, with the following tailoring variables: glycemic control (good, HbA1c $\leq 7\%$; borderline, $7\% <$ HbA1c $\leq 10\%$;

bad, HbA1c $>$ 10%), history of severe hypoglycemia (yes/no) and BMI. The probabilities of treatment and censoring were modeled with logistic regressions using all covariates as predictors. Overlap weights were used and the distributions of the fitted treatment probability, fitted censoring probability and estimated weights were inspected to determine if truncation or trimming was necessary. Model specifications were checked with residual and double-robustness plots. A complete-case analysis (using singly imputed values of SES) was conducted. Several sensitivity analyses were performed, including restricting study entry to after January 1st, 2007 (date when DPP-4i were first approved in the UK) and using more stringent time-windows to record covariates before study entry. Details are given in the Supplemental Materials E.3 and E.4.

6.5 Results

6.5.1 Cohort Description

The study cohort consisted of 36,911 patients among whom 28,370 patients (77%) added sulfonylurea to metformin and the remaining 8,541 patients (33%) added DPP-4i. Table 6.1 presents the characteristics of the patients at study entry. A total of 2,551 events (7%) were recorded and the median time to an event was 25 months. More events were observed in the metformin-sulfonylurea group (8%, 2,293 events) than in the metformin-DPP-4i group (3%, 258 events).

Figure 6.1 shows the treatment and response trajectories by treatment group. The inner rings show the distributions of patients according to the type of events they experienced or to why they were censored. In both groups, about half of the patients who were censored were lost to follow-up before making any treatment change. The other half were censored because of a change to their treatment regime. Referring to Figure 6.1 A, 21% of the patients on metformin-sulfonylurea combination therapy added a drug to their regime, 4% replaced

Table 6.1: Characteristics of type 2 diabetes patients at the time of first add-on to metformin, United Kingdom, 1997-2018.

Characteristics	All patients n=36,911	Added sulfonylurea n=28,370	Added DPP-4i n=8,541
Age, mean (SD)	62.4 (11.0)	62.5 (11.0)	61.9 (10.7)
Male, n (%)	22,195 (60.1)	17,031 (60.0)	5,164 (60.5)
SES, n (%)			
1^{st} quintile	3,813 (10.3)	2,991 (10.5)	822 (9.6)
2^{nd} quintile	4,577 (12.4)	3,691 (13.0)	886 (10.4)
3^{rd} quintile	4,465 (12.1)	3,615 (12.7)	850 (10.0)
4^{th} quintile	4,534 (12.3)	3,689 (13.0)	845 (9.9)
5^{th} quintile	4,005 (10.9)	3,145 (11.1)	860 (10.1)
Unknown	15,517 (42.0)	11,239 (39.6)	4,278 (50.1)
Smoking status, n (%)			
Never	16,938 (45.9)	12,915 (45.5)	4,023 (47.1)
Current	6,931 (18.8)	5,650 (19.9)	1,281 (15.0)
Ex	13,042 (35.3)	9,805 (34.6)	3,237 (37.9)
Body mass index, n (%)			
≤ 25 kg/m^2	3,411 (9.2)	2,910 (10.3)	501 (5.9)
25 to 30 kg/m^2	11,358 (30.8)	9,178 (32.4)	2,180 (25.5)
30 to 40 kg/m^2	17,755 (48.1)	13,214 (46.6)	4,541 (53.2)
>40 kg/m^2	3,910 (10.6)	2,647 (9.3)	1,263 (14.8)
Unknown	477 (1.3)	421 (1.5)	56 (0.7)
Years of metformin, mean (SD)	2.5 (2.4)	2.2 (2.3)	3.2 (2.7)
HbA1c, n (%)			
$\leq 7\%$	1,684 (4.6)	1,342 (4.7)	342 (4.0)
7% to 10%	25,759 (69.8)	18,944 (66.8)	6,815 (79.8)
$>10\%$	8,221 (22.3)	6,904 (24.3)	1,317 (15.4)
Unknown	1,247 (3.4)	1,180 (4.2)	67 (0.8)
Alcohol misuse, n (%)	2,143 (5.8)	1,572 (5.5)	571 (6.7)
Renal disease, n (%)	3,251 (8.8)	2,398 (8.5)	853 (10.0)
Dyslipidemia, n (%)	29,009 (78.6)	21,692 (76.5)	7,317 (85.7)
Hypertension, n (%)	28,276 (76.6)	21,635 (76.3)	6,641 (77.8)
Severe hypoglycemia, n (%)	301 (0.8)	209 (0.7)	92 (1.1)

DPP-4i: dipeptidyl peptidase-4 inhibitors, HbA1c: glycated hemoglobin, SD: standard deviation, SES: socio-economic status

metformin by another agent, 12% replaced sulfonylurea by another agent and 9% stopped their current regime to start a completely new regime. The outer rings provide more details on the type of treatment changes, showing the distribution of the chosen agent to add to the current regime categorized by drug classes. The choice of a second add-on is variable in both groups. Thiazolidinedione and DPP-4i are the preferred add-ons for metformin-sulfonylurea users while sulfonylurea is the preferred add-on for metformin-DPP-4i users.

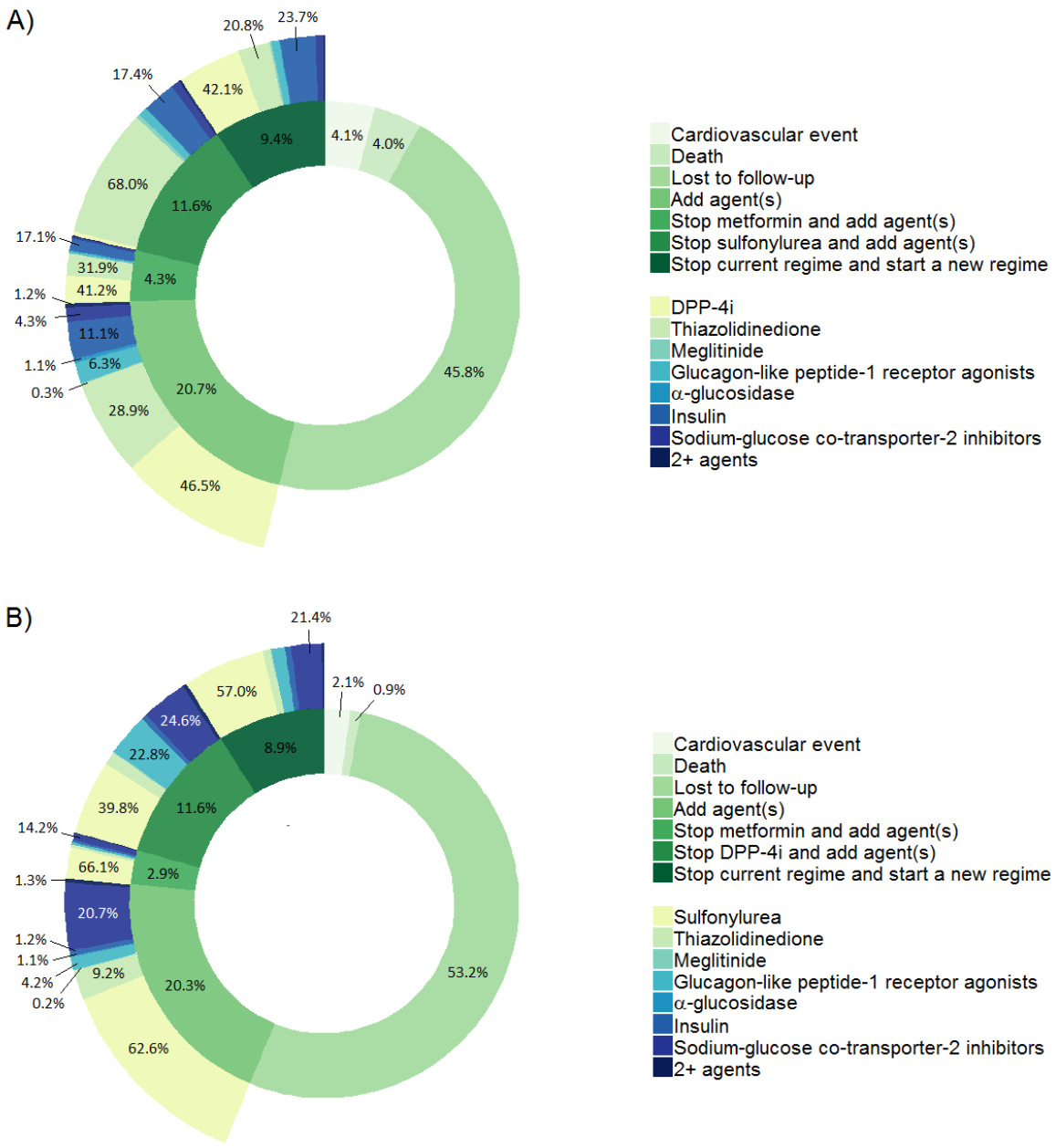

Figure 6.1: Treatment and response trajectories for metformin-sulfonylurea users (A) and metformin-DPP-4i users (B). The inner ring shows the distribution of patients according to the type of events and the reason for censoring. The outer ring shows the distribution of the drug that was added to the current regime categorized by drug classes.

6.5.2 Estimated Treatment Rule

Table 6.2 presents estimates of the parameters in the treatment rule along with measures of uncertainty using 35,287 patients. The estimated decision rule recommends DPP-4i if the linear combination

$$-0.99 - 0.87 \times \mathbb{I}(7\% < \text{HbA1c} \leq 10\%) - 1.09 \times \mathbb{I}(\text{HbA1c} > 10\%) + 1.69 \times \mathbb{I}(\text{Hypoglycemia}) + 0.04 \times \text{BMI}$$

is positive and sulfonylurea otherwise. The resulting rule is depicted in Figure 6.2. Patients with a history of hypoglycemia are recommended to add DPP-4i, regardless of other characteristics. In the absence of a history of hypoglycemia, patients with a high BMI are recommended to add DPP-4i and the worse the glycemic control, the higher the BMI must be to recommend DPP-4i over sulfonylurea. Among patients included in the analysis, the estimated rule recommends adding sulfonylurea for 32,642 patients (93%) and DPP-4i for 2,645 patients (7%), with 25,753 patients (73%) having actually received their optimal treatment.

Table 6.2: Treatment rule parameters estimates based on 35,287 patients.

Tailoring variable	$\hat{\boldsymbol{\psi}}_{\mathbf{1}}$ (SE)	95% CI
Intercept	-0.99 (0.76)	-2.48, 0.51
HbA1c (ref: $\leq 7\%$)		
7% to 10%	-0.87 (0.45)	-1.75, 0.01
$> 10\%$	-1.09 (0.52)	-2.12, -0.07
Hypoglycemia	1.69 (0.71)	0.31, 3.07
BMI	0.04 (0.02)	0.005, 0.08

BMI: body mass index, CI: confidence interval, HbA1c: glycated hemoglobin, SE: standard error

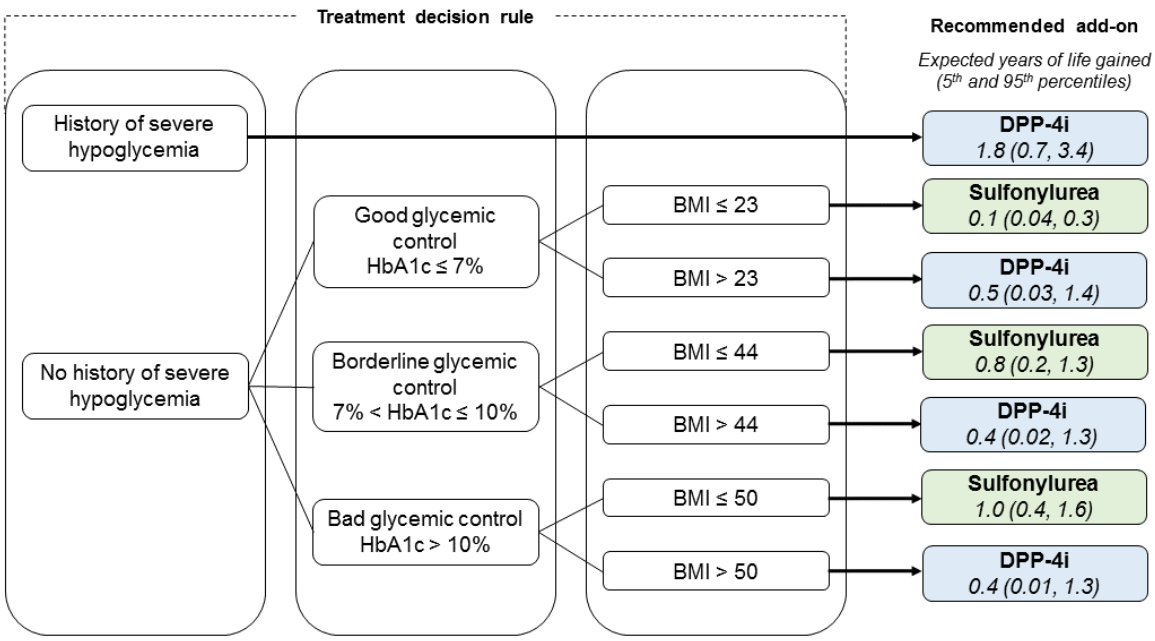

Figure 6.2: Estimated individualized treatment rule using history of severe hypoglycemia, glycemic control and BMI. The recommended add-on treatments are shown on the right along with the estimated expected years of life gained from receiving the optimal add-on versus the other treatment option and the estimated 5^{th} and 95^{th} percentiles of that distribution.

6.6 Extension Beyond One Stage

We introduce additional notation and theory needed to estimate a two-stage DTR with DWSurv and discuss how this extension could apply to the T2D illustrative example.

6.6.1 Estimation of a Two-stage DTR

Data needed to estimate an optimal two-stage DTR are also in the form of trajectories $(\boldsymbol{X_1}, A_1, Y_1, \eta_2, \boldsymbol{X_2}, A_2, Y_2, \delta)$ grouped into stages denoted with the subscript j, where $\boldsymbol{X_j}$ represents covariates measured at the beginning of stage j, prior to the administration of treatment A_j. The variable η_2 indicates if an individual entered stage 2. Y_1 and Y_2 are stage-specific survival or censoring time and the overall survival or censoring time Y is $Y_1 + \eta_2 Y_2$. An individual who does not enter the second stage because he experiences the

event or is censored in the first stage has missing stage 2 treatment, covariates and outcome. We define $T_2^{a_1,a_2}$, the survival time in the second stage if, possibly contrary to the fact, the treatment regime (a_1, a_2) is followed. A two-stage DTR is the sequence of two decision rules $d_1(\boldsymbol{h_1})$ and $d_2(\boldsymbol{h_2})$ to be applied at the beginning of stages 1 and 2 respectively, where $\boldsymbol{H_2} = (\boldsymbol{X_1}, A_1, Y_1, \boldsymbol{X_2})$. The optimal DTR $\{d_1^{\text{opt}}(\boldsymbol{h_1}), d_2^{\text{opt}}(\boldsymbol{h_2})\}$ is that which maximizes the average counterfactual survival time T^{a_1,a_2} conditional on individual characteristics accrued over time.

Backward induction is used to extend the estimation procedure described in a single-stage setting to two stages. It implies identifying the optimal second stage treatment and then optimizing the first stage treatment, thus working backward in time. The optimal second stage treatment maximizes the average counterfactual survival time in the second stage $T_2^{a_1,a_2}$. An outcome model $\mathbb{E}[\log(T_2^{a_1,a_2})|\boldsymbol{h_2}, a_2, \eta_2 = 1; \boldsymbol{\beta_2}, \boldsymbol{\psi_2}]$ akin to (6.1) but with the stage 2 quantities is specified. Models for the probability of treatment A_2 and the probability of experiencing an event are proposed and estimated using only individuals who entered the second stage and weights w_2 such as (6.2) are computed accordingly. Weighted estimating equations are solved and the optimal second stage treatment is derived as $a_2^{\text{opt}} = \mathbb{I}(\hat{\boldsymbol{\psi}}_{\boldsymbol{2}}^{\boldsymbol{T}} \boldsymbol{h_{2\psi}} > 0)$. Once this initial step is completed, DWSurv proceeds with the optimization of the first stage treatment using all individuals.

Because the first stage treatment affects the overall survival time (not only the survival time in the first stage) which is also affected by the second stage treatment, the optimization in the first stage uses a pseudo-survival time $\tilde{T}$ defined as the overall survival time had all individuals, possibly contrary to the fact, received their optimal stage 2 treatment. For individuals who did not enter the second stage, this pseudo-survival time is equal to the observed survival time in the first stage i.e. $\tilde{T} = T_1$. Among individuals who entered stage 2, some may actually have received their optimal treatment ($a_2^{\text{opt}} = a_2$), in which case $\tilde{T}$ is equal to their observed survival time $T_1 + T_2$. For individuals who entered the second stage

but have not received their optimal treatment ($a_2^{\text{opt}} \neq a_2$), $\tilde{T}$ is not observed and must be estimated by adding what was lost from receiving a suboptimal stage 2 treatment, i.e. $\tilde{T} := T_1 + T_2 \times \exp\{\hat{\boldsymbol{\psi}}_{\mathbf{2}}^{\boldsymbol{T}} \boldsymbol{h}_{\mathbf{2}\boldsymbol{\psi}}(a_2^{\text{opt}} - a_2)\}$. Once $\tilde{T}$ is estimated for individuals who experienced an event, similar steps as presented above are followed: an outcome model for the pseudo-survival time $\mathbb{E}[\log(\tilde{T}^{a_1,a_2})|\boldsymbol{h}_{\mathbf{1}}, a_1; \boldsymbol{\beta}_{\mathbf{1}}, \boldsymbol{\psi}_{\mathbf{1}}]$ is specified, weights w_1 are computed based on models for the probability of treatment a_1 and the probability of experiencing an event and the first stage optimal treatment is derived as $a_1^{\text{opt}} = \mathbb{I}(\hat{\boldsymbol{\psi}}_{\mathbf{1}}^{\boldsymbol{T}} \boldsymbol{h}_{\mathbf{1}\boldsymbol{\psi}} > 0)$. The resulting optimal DTR would be: at the beginning of the first stage, recommend a_1^{opt}; if the patient enters the second stage, recommend a_2^{opt}.

6.6.2 T2D Treatment Pathways Beyond One Stage

We explored treatment and response pathways about the following two-stage DTR: the first stage compares adding sulfonylurea or DPP-4i to metformin, as described previously, and the second stage compares adding DPP-4i or insulin if patients were on metformin-sulfonylurea combination therapy in the first stage or adding glucagon-like peptide-1 receptor agonists (GLP-1) or insulin if patients were previously on metformin-DPP-4i. The outcome of interest remains the time until the occurrence of a cardiovascular event or death.

Figure 6.3 summarizes treatment and response pathways of this two-stage DTR using the study cohort described before. Among metformin-sulfonylurea users, 80% of the patients (n=22,723) were censored in the first stage because they made an unacceptable treatment change or were lost to follow-up before making any change to their treatment (see also Figure 6.1). Only 12% of metformin-sulfonylurea users (n=3,354) entered the second stage. The number of events is small in the second stage, 90 (3%) among users who added DPP-4i to their metformin-sulfonylurea regime and 94 (14%) among users who added insulin. Among metformin-DPP-4i users in the first stage, fewer patients entered the second stage (n=377, 4%). Fewer events were observed in the second stage in that group, four events (1%) among

users who added GLP-1 and only one event (1%) among those who added insulin.

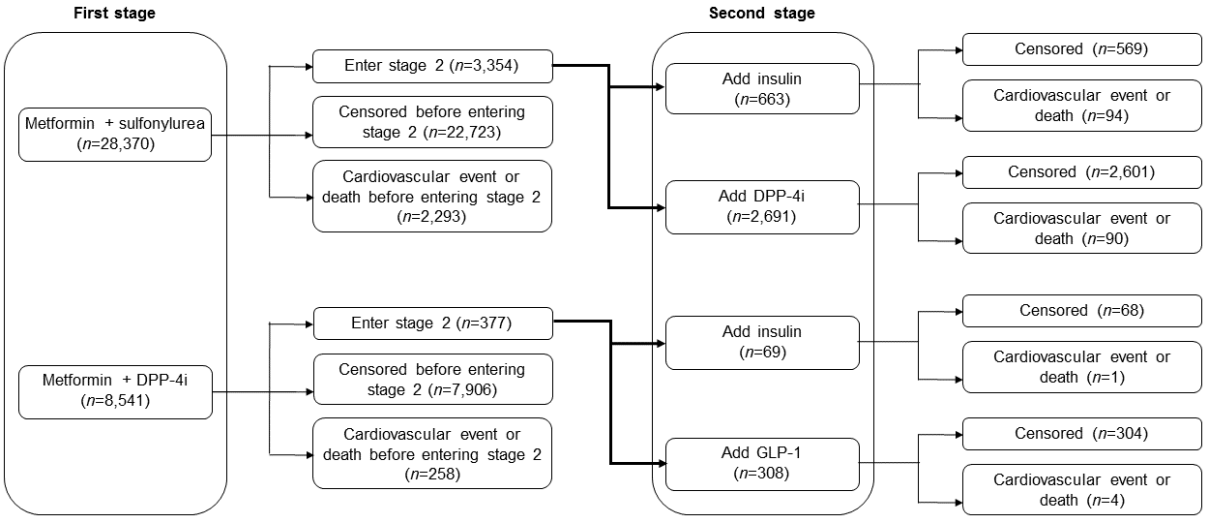

Figure 6.3: Treatment and response pathways exploring a two-stage DTR that compares adding sulfonylurea or DPP-4i to metformin in the first stage and further adding DPP-4i or insulin (if patients were on metformin-sulfonylurea in the first stage) or adding GLP-1 or insulin (if patients were on metformin-DPP-4i in the first stage) in the second stage.

6.7 Discussion

This illustrative example served as a proof-of-concept about the importance of focusing on personalized treatment rules and the difficulty of studying multi-stage DTRs in the real world. We presented a method to estimate an optimal DTR with survival outcomes called DWSurv which takes into account the heterogeneity in response to treatment across patients and thus allows answering the question "What is the best sequence of treatments for this specific patient?" This differs from standard comparative-effectiveness studies which answer the question "What is the best treatment for the average patient?", ignoring that patients may respond differently to the treatment. DWSurv estimates treatment rules that are tailored to evolving patient's characteristics. The decision rules can be used in clinical practice for recommending the best treatment for a specific patient. We illustrated the usefulness of DWSurv in an application to the management of T2D. Using a large observational database,

we discovered an individualized treatment rule for deciding which of sulfonylurea or DPP-4i should be added to metformin once metformin alone fails to achieve the therapeutic goals. We found that extending this rule to treatment decisions beyond the first add-on treatment to metformin was challenging given that, in practice, patients follow heterogeneous treatment pathways.

The illustrative example focused on an important clinical question in the management of T2D: what to do when metformin in monotherapy does not achieve the therapeutic goals. We derived an individualized treatment rule that inputs the patient's glycemic control, history of severe hypoglycemia and BMI at the time of decision-making and outputs the best add-on treatment to metformin between sulfonylurea or DPP-4i. The recommended add-on is that which maximizes the time until a cardiovascular event or death. The decision rule estimated with DWSurv favors DPP-4i for patients with higher BMI and a history of severe hypoglycemia while it favors sulfonylurea for patients with borderline or bad glycemic controls. The estimated rule is in line with known medication profiles (Garber et al., 2019). The rule recommends DPP-4i when a patient has a history of severe hypoglycemia, a decision that makes sense given that sulfonylurea is associated with a higher risk of hypoglycemia and DPP-4i is not. In the absence of a history of hypoglycemia, patients with a suboptimal glycemic control will tend to be recommended to add sulfonylurea, which is known to be more aggressive in achieving good glycemic control. However, patients with a higher BMI will tend to be recommended DPP-4i, which is neutral on weight change while sulfonylurea is associated with weight gain. The estimated decision rule provides additional tools to the existing guidelines by proposing thresholds about BMI and glycemic control to identify the best add-on treatment for a specific patient, thus facilitating decision-making by clinicians.

The estimated individualized treatment rule for choosing an add-on treatment to metformin is to be applied at the time when the treatment should be intensified but does not dictate when the current regime should be changed. Therefore, DWSurv addresses the question

"Given that the current regime does not work, what treatment decision should be made next?" rather than "When should a treatment change should be made?" Individualized targets for deciding when the current regime does not achieve the therapeutic goals have been considered previously (e.g. Neugebauer et al., 2013).

The results of the analysis rely on several assumptions. First, we assumed that patients were continuously exposed to metformin before study entry and continuously exposed to metformin-sulfonylurea or metformin-DDP-4i from study entry until the event or censoring. This assumed that a patient would not be withdrawn from pharmacotherapy once metformin was started. Second, covariates were measured at study entry but, in the absence of a record on the date of study entry, the value of the most recent record before study entry was used instead. This strategy assumed that the value of the covariate remained constant since the last time it was recorded before study entry. Third, we assumed that all confounders were measured. Although the double-robustness property offered additional protection against misspecification of some models, the validity of the estimated treatment rule and of related quantities (e.g. the expected life gained shown in Figure 6.2) may be compromised if important confounders were missing. In fact, the diagnostic plots suggested that the treatment-free model may be misspecified (see Supplemental Material E.3).

There is an increasing interest in statistical methods that estimate optimal DTRs with observational data, with DWSurv being one of the few that can handle survival outcomes. Observational data needed to apply such methods are also increasingly available. Large primary care databases such as the CPRD in the UK reflect how the treatment of T2D is managed in practice and offer insights about treatment effectiveness that would not be captured in controlled trials. However, this example on T2D highlighted a tradeoff between feasibility and clinical relevance in that clinical questions may not always be easy to answer even when the appropriate data and statistical methods are available. Using the CPRD, we could not answer our two-stage DTR question because treatment pathways in practice

were too heterogeneous to capture enough events across stages. More events could have been observed had we considered other or more general treatment comparisons, for example, comparing treatment switches versus treatment add-ons across stages, regardless of the drug classes or the number of agents added or switched to. This question could be answered with DWSurv but the practical implication of the resulting decision rules would be modest in that it would not help clinicians to decide which treatment to actually recommend to their patient.

The management of T2D served as an illustration for the practical feasibility of estimating optimal DTRs from observational data. Other chronic conditions could be studied from that perspective. Hripcsak et al. (2016) describe the heterogeneity of treatment pathways for T2D, hypertension and depression using an international data network regrouping over 250 million patients. However, T2D showed less variability in treatment pathways as compared to hypertension and depression, mainly owing to the fact that metformin is a well-accepted first-line treatment choice. We expect that the study of DTR would be even more challenging for hypertension and depression, both diseases being treated with a variety of drugs even in the first stage of the sequence of treatments.

Ethics

The study protocol was approved by the Independent Scientific Advisory Committee of the CPRD (protocol number 18_169) and by the Faculty of Medicine Institutional Review Board at McGill University, Montréal, Québec, Canada.

Funding

This work was supported by a doctoral research grant from the Fonds de recherche du Québec – Nature et technologies [199803].

Appendix E – Supplemental Materials

Contains the following sections:

E.1 **CPRD and Study Cohort** – details on the CPRD database and on the assembling of the study cohort.

E.2 **Covariates Definitions** – definitions of covariates using medical, diagnosis and prescription codes.

E.3 **Implementation** – details on the implementation of the single-stage DTR in `R` and model-checking plots.

E.4 **Sensitivity Analyses** – results of six sensitivity analyses.

Chapter 7

Conclusion

7.1 Summary

The work presented in this thesis introduces consistent and interpretable methods for the study of precision medicine and demonstrates the usefulness of such tools in data applications motivated by real clinical problems. Chapter 3 proposes using the m-out-of-n bootstrap to construct confidence intervals for the blip parameters in dWOLS in situations where non-regularity is likely to affect the inferences. We recommend choosing the resample size m in a way that adapts to the degree of non-regularity in the data using the double bootstrap algorithm. Chapter 4 develops DWSurv, a method for estimating optimal DTRs with censored outcomes. DWSurv requires specifying semi-parametric AFT models for the survival time across stages and yields doubly-robust estimators by relying on balancing weights constructed with models for the treatment assignment and censoring mechanisms. Chapter 5 explores the finite sample performance of two asymptotic formulae and three bootstrap approaches to construct confidence intervals for the decision rule parameters in DWSurv. This work provides guidance for using DWSurv in practice. Chapter 6 demonstrates how DWSurv can be used to fill knowledge gaps in the treatment of T2D. Using data from a

large primary care database, we estimate an individualized treatment rule for whether sulfonylurea or DPP-4i should be added to metformin to delay the occurrence of cardiovascular events or death. We highlight how extending this rule to multiple stages of clinical intervention can be challenging given that treatment pathways followed by patients in practice are heterogeneous. All methods and inferential tools presented in this thesis are or will be freely available in the `DTRreg` package in the hope of facilitating the uptake of the proposed methods in applied sciences.

The simulation studies allow showcasing our theoretical developments or supporting heuristic justifications (of the m-out-of-n bootstrap). The first manuscript considers nine simulation scenarios characterized by the extent to which the non-regularity of the blip estimators is likely to affect the inferences. The simulation studies highlight that choosing m in a data-adaptive manner can be computationally intensive depending on the sample size. The second and third manuscripts use novel simulation scenarios for two-stage DTRs with censored data that showcase the double-robustness of DWSurv and evaluate the performance of competing methods to construct confidence intervals for its parameters. The simulation scenarios are complex enough to handle treatment and censoring mechanisms that depend on time-varying characteristics across stages.

All but the third manuscript include an illustration of the methods in a data application. The first manuscript estimates two decision rules about the best timing for introducing solid food in an infant's diet to optimize metabolic outcomes measured during childhood. In a situation where the outcome is measured a long time (six years) after the treatments are received (solid food intake at 3 and 6 months), non-regularity likely affects the inferences thus the importance of considering the m-out-of-n bootstrap in this data application. The second manuscript applies DWSurv to observational data from patients with rheumatoid arthritis to validate existing guidelines about the management of episodes of disease activity. We identify two decision rules that do not entirely coincide with recommendations in guidelines but

recognize that the conclusions of this analysis are only as convincing as the assumptions they rely on. Specifically, assuming that the baseline visit in the data corresponds to the beginning of an episode of disease activity likely does not hold. The case study on the treatment of T2D presented in the last manuscript provides insights into the practical challenges of applying DWSurv to large observational databases. Despite having data for over 35,000 patients and observing over 2,000 events, few patients follow the treatment pathways under study and not enough patients experience an event beyond the first stage, preventing the application of DWSurv to estimate a two-stage DTR. We identify a sensible single-stage treatment rule which recommends adding DPP-4i to metformin if the patient has a history of severe hypoglycemia or if the patient's BMI is high and recommends adding sulfonylurea otherwise.

7.2 Future Work

Together, DWOLS and DWSurv provide a complete framework for estimating optimal DTRs with censored or uncensored outcomes. Indications on how dWOLS can handle continuous or non-binary treatments have already been given (Wallace & Moodie, 2015), which could be used with DWSurv and implemented in the corresponding function in the `DTRreg` package. Beyond theoretical indications, applying DWSurv and dWOLS to problems with multiple treatment arms or continuous treatments (e.g. comparing doses) would be useful to promote widespread usage of the methods. For instance, the illustrative example about the treatment of T2D could compare multiple treatment arms corresponding to the multiple drug classes, thus improving the relevance of the resulting treatment rules for guiding clinical decision-making.

The design of the simulation study for multi-stage DTRs with survival endpoints used in Chapters 4 and 5 is subject to some limitations. The simulation study suits the purpose

of demonstrating important properties of DWSurv at the cost of providing little intuition on how the data generating mechanisms mimic the way the data arise in practice. Also, the parameters of the data generating mechanisms require fine-tuning to yield acceptable survival times whose distribution also satisfy (or not) certain modeling assumptions across stages. Designing simulation studies that translate into real data examples and limit the extent to which the parameters of the data generating mechanism must be tuned is an important avenue for future work on DTR methods of survival outcomes.

7.3 Concluding Remarks

DWOLS and DWSurv provide useful tools to answer open clinical questions about the treatment of chronic or recurring diseases. With the increasing availability of large registry and claims databases, such methods gain from being robust enough to handle non-experimental data and from exploiting frameworks that are easy to understand and accessible to scientists who are most likely to use the rules. More widespread usage of dWOLS and DWSurv can improve clinical practice about how treatment decisions are made as well as improve the statistical understanding of complex longitudinal data.

Appendix A

Ethics Approvals

Ethics approvals for the work presented in Chapter 6 were obtained from the Independent Scientific Advisory Committee of the CPRD (protocol number 18_169) and from the Faculty of Medicine Institutional Review Board at McGill University. Approvals are shown on the next three pages.

ISAC EVALUATION OF PROTOCOLS FOR RESEARCH INVOLVING CPRD DATA

FEEDBACK TO APPLICANTS

CONFIDENTIAL			***by e-mail***
PROTOCOL NO:	18_169R		
PROTOCOL TITLE:	Towards an accessible methodology in precision medicine: methods for time-to-event data and non-regular inferences with an application to treatment of type 2 diabetes		
APPLICANT:	Samy Suissa, Director, Centre for Clinical Epidemiology, Jewish General Hospital, samy.suissa@mcgill.ca		
APPROVED ☒	**APPROVED WITH COMMENTS (resubmission not required)** ☐	**REVISION/ RESUBMISSION REQUESTED** ☐	**REJECTED** ☐
INSTRUCTIONS: ***Protocols with an outcome of 'Approved' or 'Approved with comments' do not require resubmission to the ISAC.*** **REVIEWER COMMENTS:** **APPLICANT FEEDBACK:**			
DATE OF ISAC FEEDBACK:	28/09/18		
DATE OF APPLICANT FEEDBACK:			

For protocols approved from 01 April 2014 onwards, applicants are required to include the ISAC protocol in their journal submission with a statement in the manuscript indicating that it had been approved by the ISAC (with the reference number) and made available to the journal reviewers. If the protocol was subject to any amendments, the last amended version should be the one submitted.

***** Please refer to the ISAC advice about protocol amendments provided below*****

McGill

Faculty of Medicine
3655 Promenade Sir William Osler #633
Montreal, QC H3G 1Y6

Faculté de médecine
3655, Promenade Sir William Osler #633
Montréal, QC H3G 1Y6

Fax/Télécopieur: (514) 398-3870
Tél/Tel: (514) 398-3124

CERTIFICATION OF ETHICAL ACCEPTABILITY FOR RESEARCH INVOLVING HUMAN SUBJECTS

The Faculty of Medicine Institutional Review Board (IRB) is a registered University IRB working under the published guidelines of the Tri-Council Policy Statement, in compliance with the Plan d'action ministériel en éthique de la recherche et en intégrité scientifique (MSSS, 1998), and the Food and Drugs Act (17 June 2001); and acts in accordance with the U.S. Code of Federal Regulations that govern research on human subjects. The IRB working procedures are consistent with internationally accepted principles of Good Clinical Practices.

At a Board meeting on 13 February 2018, the Faculty of Medicine Institutional Review Board, consisting of:

Frances Aboud, PhD	Frank Elgar, PhD
Carolyn Ells, PhD	Catherine Lecompte
Lucille Panet-Raymond, BA	Shahad Salman, LL.M.
Daniel Saumier, PhD	Blossom Shaffer, MBA

Examined the research project **A02-M04-18B** titled: *Towards an accessible methodology in personalized medicine: methods for time-to-event data and non-regular inferences with an application for treatment of type 2 diabetes*

As proposed by: Dr. Erica E. M. Moodie (Applicant) to _____ (Granting Agency, if any)

And consider the experimental procedures to be acceptable on ethical grounds for research involving human subjects.

13 February 2018		
Date	Chair, IRB	Dean of Faculty

Institutional Review Board Assurance Number: FWA 00004545

McGill

Faculty of Medicine
3655 Promenade Sir William Osler #633
Montreal, QC H3G 1Y6

Faculté de médecine
3655, Promenade Sir William Osler #633
Montréal, QC H3G 1Y6

Fax/Télécopieur: (514) 398-3870
Tél/Tel: (514) 398-3124

February 12, 2019

Erica E. M. Moodie
Epidemiology, Biostatistics and Occupational Health
1020 Pine Avenue
Montreal, Quebec H3A 1A2

RE: IRB Study Number A02-M04-18B
Towards an accessible methodology in personalized medicine: methods for time-to-event data and non-regular inferences with an application for treatment of type 2 diabetes

Dear Dr. Moodie,

Thank you for submitting an application for Continuing Ethics Review for the above-referenced study.

The study progress report was reviewed and Full Board re-approval was provided on February 11, 2019. The ethics certification renewal is valid from **February 7, 2019 to February 6, 2020**.

The Investigator is reminded of the requirement to report all IRB approved protocol and consent form modifications to the Research Ethics Offices (REOs) for the participating hospital sites. Please contact the individual hospital REOs for instructions on how to proceed. Research funds may be withheld and / or the study's data may be revoked for failing to comply with this requirement.

Should any modification or unanticipated development occur prior to the next review, please notify the IRB promptly. Regulation does not permit the implementation of study modifications prior to IRB review and approval.

Sincerely,

Carolyn Ells, PhD
Co-Chair
Institutional Review Board

cc: Gabrielle Simoneau
A02-M04-18B

Appendix B

Supplemental Materials for Chapter 3

B.1 Adaptive Choice of m

The class of resample sizes m introduced in Chakraborty et al. (2013) is defined as $\hat{m} := n^{f(p)}$, with $f(p)$ being a function of the regularity measure p that satisfies: (1) $f(p)$ is monotone decreasing in p, takes value between $(0, 1]$, and $f(0) = 1$; (2) $f(p)$ is continuous and has bounded first derivative. It was shown that this definition of $\hat{m}$ satisfies the consistency conditions stated before (Chakraborty et al., 2013). Provided that we can estimate p from the data, a simple definition of $\hat{m}$ (Chakraborty et al., 2013) is given by $\hat{m} := n^{\frac{1+\alpha(1-\hat{p})}{1+\alpha}}$, where $\alpha > 0$ is a tuning parameter and $\hat{p}$ is an estimate of the degree of non-regularity in the data. The tuning parameter α controls the smallest possible resample size $\hat{m}$. For a fixed n and $\hat{p} \in (0, 1)$, $\hat{m}$ can take any value between $n^{\frac{1}{1+\alpha}}$ and n.

We consider an estimate of p based on the proportion of patients for whom the optimal treatment is non-unique. An alternative estimator of p is given by $\hat{p} = P_n \left[\mathbb{I} \left\{ n(\hat{\boldsymbol{\psi}}_{\mathbf{2}}^{\boldsymbol{T}} \boldsymbol{H}_{\mathbf{2}\boldsymbol{\psi}})^2 \leq \tau_n(\boldsymbol{H}_{\mathbf{2}\boldsymbol{\psi}}) \right\} \right]$ (Chakraborty et al., 2010), where $\tau_n(\boldsymbol{H}_{\mathbf{2}\boldsymbol{\psi}})$ is a chosen cutoff. We first notice that the indicator $\mathbb{I} \left\{ n(\hat{\boldsymbol{\psi}}_{\mathbf{2}}^{\boldsymbol{T}} \boldsymbol{h}_{\mathbf{2}\boldsymbol{\psi}})^2 \leq \tau_n(\boldsymbol{h}_{\mathbf{2}\boldsymbol{\psi}}) \right\}$ can be viewed as the acceptance region for testing the null hypothesis ${\boldsymbol{\psi}_{\mathbf{2}}}^T \boldsymbol{h}_{\mathbf{2}\boldsymbol{\psi}} = 0$ for a patient with history

$\boldsymbol{h_{2\psi}}$. Therefore, a sensible choice for $\tau_n(\boldsymbol{H_{2\psi}})$ is $\left(\boldsymbol{h_{2\psi}^T}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{21}}\boldsymbol{h_{2\psi}}\right) \times \chi^2_{1,1-\nu}$, where $\chi^2_{1,1-\nu}$ is the $(1-\nu)\times 100$ percentile of a χ^2 distribution with one degree of freedom and $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{21}}$ is the plug-in estimator of $\sqrt{n}\text{Cov}(\hat{\boldsymbol{\psi}}_{\boldsymbol{2}}, \hat{\boldsymbol{\psi}}_{\boldsymbol{2}})$. For a fixed n and α, the results in terms of coverage and width of confidence intervals are robust to the choice of ν (Chakraborty et al., 2013).

We detail the double bootstrap algorithm proposed in Chakraborty et al. (2013) for choosing α in a data-driven way. Recall that we are interested in producing a confidence interval for the parameter $\boldsymbol{\psi_1}$ which is estimated by $\hat{\boldsymbol{\psi}}_{\boldsymbol{1}}$ from the original data. We provide a review of the double bootstrap algorithm below.

1. Consider a set of candidate values for α e.g. $\{0.025, 0.05, \ldots, 1\}$
2. Fix α to the smallest value in the set of candidate values.
3. Draw B_1 standard n-out-of-n bootstrap samples from the original data and calculate the bootstrap estimates $\hat{\boldsymbol{\psi}}_1^{(b_1)}$, for $b_1 = 1, \ldots, B_1$.
4. For each b_1 bootstrap sample, estimate $\hat{p}$ and use the definition of $\hat{m}$ to estimate $\hat{m}^{(b_1)}$, for $b_1 = 1, \ldots, B_1$.
5. For each b_1 first-level bootstrap sample, draw B_2 $\hat{m}^{(b_1)}$-out-of-n second-level (nested) bootstrap samples and calculate the double bootstrap estimate $\hat{\boldsymbol{\psi}}_{1,\hat{m}}^{(b_1 b_2)}$, for $b_1 = 1, \ldots, B_1$, $b_2 = 1, \ldots, B_2$.
6. For each b_1 first-level bootstrap sample, compute the lower and upper $\eta/2 \times 100$ percentiles of
$$\left\{\sqrt{\hat{m}^{(b1)}}\left(\hat{\boldsymbol{\psi}}_{1,\hat{m}^{(b_1)}}^{(b_1 b_2)} - \hat{\boldsymbol{\psi}}_1^{(b_1)}\right), b_2 = 1, \ldots, B_2\right\},$$
respectively denoted as $\hat{\boldsymbol{l}}_{DB}^{(b_1)}$ and $\hat{\boldsymbol{u}}_{DB}^{(b_1)}$. Construct the double centered percentile bootstrap (Efron & Tibshirani, 1994) confidence interval for each b_1 first-level bootstrap sample as
$$(\hat{\boldsymbol{\psi}}_1^{(b_1)} - \hat{\boldsymbol{u}}_{DB}^{(b_1)}/\sqrt{\hat{m}^{(b_1)}}, \hat{\boldsymbol{\psi}}_1^{(b_1)} - \hat{\boldsymbol{l}}_{DB}^{(b_1)}/\sqrt{\hat{m}^{(b_1)}}),$$

for $b_1 = 1, \ldots, B_1$.

7. Estimate the coverage rate of the double bootstrap confidence interval from all the first-level bootstrap samples as

$$\frac{1}{B_1}\sum_{b_1=1}^{B_1} \mathbb{I}\left\{\hat{\boldsymbol{\psi}}_1^{(b_1)} - \hat{\boldsymbol{u}}_{DB}^{(b_1)}/\sqrt{\hat{m}^{(b_1)}} \leq \hat{\boldsymbol{\psi}}_1 \leq \hat{\boldsymbol{\psi}}_1^{(b_1)} - \hat{\boldsymbol{l}}_{DB}^{(b_1)}/\sqrt{\hat{m}^{(b_1)}}\right\}.$$

   If the above coverage rate is at or above the nominal level $(1-\eta)\times 100\%$, then pick the current values of α as the final value. Otherwise, increment α to the next highest value in the set and repeat steps 3–7.

Further details on the double bootstrap algorithm for choosing α can be found in Chakraborty et al. (2013).

B.2 Details of the Data Generating Process used in the Simulation Study

We adapt the data generating models developed in Chakraborty et al. (2010) to the treatment coded as $A_j \in \{0, 1\}$, $j = 1, 2$. Following the developments in the Web-based Supplementary Materials of Chakraborty et al. (2016) and details in Chakraborty et al. (2010) (c.f. Section 4) and in Laber et al. (2014b), we calculate the true target value ψ_{10} in terms of the parameters of the data generating model λ and δ.

Recall that dynamic treatment regimens with two stages of clinical intervention, with stages indicated by the subscript j, are defined by the following variables: (i) the patient outcome Y_j (continuous) after treatment j; (ii) the j-th treatment decision A_j, where treatments are assumed binary $\{0, 1\}$; (iii) the non-treatment information (covariates) X_j available prior to the j-th treatment, and (iv) the patient history $\boldsymbol{H_j}$ defined as a matrix containing patient history prior to the j-th treatment, including prior treatment(s). The expected outcome

models for two stages of intervention, referred to as Q-functions in Q-learning, are given by:

$$Q_2(\boldsymbol{H_2}, A_2) = \mathbb{E}[Y_2|\boldsymbol{H_2}, A_2] = \boldsymbol{H_{2\beta}^T}\boldsymbol{\beta_2} + \boldsymbol{H_{2\psi}^T}\boldsymbol{\psi_2}A_2$$
$$Q_1(\boldsymbol{H_1}, A_1) = \mathbb{E}[\max_{a_2} Q_2(\boldsymbol{H_2}, a_2)|\boldsymbol{H_1}, A_1] = \boldsymbol{H_{1\beta}^T}\boldsymbol{\beta_1} + \boldsymbol{H_{1\psi}^T}\boldsymbol{\psi_1}A_1$$

with $\boldsymbol{H_{2\beta}} = (1, X_1, A_1, X_1A_1)$, $\boldsymbol{H_{2\psi}} = (1, X_2, A_1)$, $\boldsymbol{H_{1\beta}} = (1, X_1)$ and $\boldsymbol{H_{1\psi}} = (1, X_1)$. The target parameter is the main effect of the stage 1 treatment ψ_{11} found in $Q_1(H_1, A_1)$. Precisely, it is the coefficient in front of A_1 in $\boldsymbol{H_{1\psi}^T}\boldsymbol{\psi_1}A_1 = A_1(\psi_{10} + \psi_{11}X_1)$. Recall that the generative models can be summarized in terms of: (i) $X_j \in \{-1, 1\}$, $A_j \in \{0, 1\}$ for $j = 1, 2$; (ii) $P(A_j = 1) = P(A_j = 0) = 0.5$ for $j = 1, 2$; (iii) $X_1 \sim 2 \times \text{Bernoulli}(0.5) - 1$, $X_2|X_1, A_1 \sim 2 \times \text{Bernoulli}(\text{expit}\{\delta_1 X_1 + \delta_2(2A_1 - 1)\}) - 1$ where $\text{expit}(x) = \exp(x)/(1 + \exp(x))$; (iv) $Y_1 \equiv 0$, $Y_2 = \lambda_1 + \lambda_2 X_1 + \lambda_3 A_1 + \lambda_4 X_1 A_1 + \lambda_5 A_2 + \lambda_6 X_2 A_2 + \lambda_7 A_1 A_2 + \varepsilon$ with $\varepsilon \backsim N(0, 1)$. It follows that

$$\begin{aligned}\max_{a_2} Q_2(\boldsymbol{H_2}, a_2) &= \lambda_1 + \lambda_2 X_1 + \lambda_3 A_1 + \lambda_4 X_1 A_1 + \max_{a_2} a_2(\lambda_5 + \lambda_6 X_2 + \lambda_7 A_1) \\ &= M + (\lambda_5 + \lambda_6 X_2 + \lambda_7 A_1) \times \mathbb{I}[\lambda_5 + \lambda_6 X_2 + \lambda_7 A_1 > 0] \qquad \text{(B.1)}\end{aligned}$$

We can express (B.1) in terms of the four possible values (X_2, A_1) can take as

$$\begin{aligned}\max_{a_2} Q_2(\boldsymbol{H_2}, a_2) = M &+ \frac{1}{2}(1 + X_2)A_1(\lambda_5 + \lambda_6 + \lambda_7)\mathbb{I}[\lambda_5 + \lambda_6 + \lambda_7 > 0] \\ &+ \frac{1}{2}(1 + X_2)(1 - A_1)(\lambda_5 + \lambda_6)\mathbb{I}[\lambda_5 + \lambda_6 > 0] \\ &+ \frac{1}{2}(1 - X_2)A_1(\lambda_5 - \lambda_6 + \lambda_7)\mathbb{I}[\lambda_5 - \lambda_6 + \lambda_7 > 0] \\ &+ \frac{1}{2}(1 - X_2)(1 - A_1)(\lambda_5 - \lambda_6)\mathbb{I}[\lambda_5 - \lambda_6 > 0] \\ = M &+ \frac{1}{2}(1 + X_2)A_1 f_1 \mathbb{I}(f_1 > 0) + \frac{1}{2}(1 + X_2)(1 - A_1)f_2\mathbb{I}(f_2 > 0) \\ &+ \frac{1}{2}(1 - X_2)A_1 f_3 \mathbb{I}(f_3 > 0) + \frac{1}{2}(1 - X_2)(1 - A_1)f_4\mathbb{I}(f_4 > 0)\end{aligned}$$

where $f_1 = \lambda_5 + \lambda_6 + \lambda_7$, $f_2 = \lambda_5 + \lambda_6$, $f_3 = \lambda_5 - \lambda_6 + \lambda_7$ and $f_4 = \lambda_5 - \lambda_6$. We have

$$\begin{aligned}
\mathbb{E}[X_2|X_1, A_1] &= -P(X_2 = -1|X_1, A_1) + P(X_2 = 1|X_1, A_1) \\
&= \frac{-1}{\exp\{\delta_1 X_1 + \delta_2(2A_1 - 1)\} + 1} + \frac{\exp\{\delta_1 X_1 + \delta_2(2A_1 - 1)\}}{\exp\{\delta_1 X_1 + \delta_2(2A_1 - 1)\} + 1} \\
&= \frac{\exp\{\delta_1 X_1 + \delta_2(2A_1 - 1)\} - 1}{\exp\{\delta_1 X_1 + \delta_2(2A_1 - 1)\} + 1} \\
1 + \mathbb{E}[X_2|X_1, A_1] &= \frac{2\exp\{\delta_1 X_1 + \delta_2(2A_1 - 1)\}}{\exp\{\delta_1 X_1 + \delta_2(2A_1 - 1)\} + 1} = 2\text{expit}\{\delta_1 X_1 + \delta_2(2A_1 - 1)\} \\
1 - \mathbb{E}[X_2|X_1, A_1] &= 2(1 - \text{expit}\{\delta_1 X_1 + \delta_2(2A_1 - 1)\})
\end{aligned}$$

An expression for $Q_1(\boldsymbol{H_1}, A_1)$ is given by

$$\begin{aligned}
Q_1(\boldsymbol{H_1}, A_1) &= \mathbb{E}[\max_{a_2} Q_2(\boldsymbol{H_2}, a_2)|\boldsymbol{H_1}, A_1] \\
&= M + \frac{1}{2}(1 + \mathbb{E}[X_2|X_1, A_1])A_1 f_1 \mathbb{I}(f_1 > 0) + \frac{1}{2}(1 + \mathbb{E}[X_2|X_1, A_1])(1 - A_1) f_2 \mathbb{I}(f_2 > 0) \\
&\quad + \frac{1}{2}(1 - \mathbb{E}[X_2|X_1, A_1])A_1 f_3 \mathbb{I}(f_3 > 0) + \frac{1}{2}(1 - \mathbb{E}[X_2|X_1, A_1])(1 - A_1) f_4 \mathbb{I}(f_4 > 0) \\
&= M + A_1 f_3 \mathbb{I}(f_3 > 0) + (1 - A_1)\mathbb{I}(f_4 > 0) + A_1 \text{expit}\{\delta_1 X_1 + \delta_2(2A_1 - 1)\}\{f_1 \mathbb{I}(f_1 > 0) \\
&\quad - f_3 \mathbb{I}(f_3 > 0)\} + (A_1 - 1)\text{expit}\{\delta_1 X_1 + \delta_2(2A_1 - 1)\}\{f_2 \mathbb{I}(f_2 > 0) - f_4 \mathbb{I}(f_4 > 0)\}
\end{aligned}$$

Again, we notice that $\text{expit}\{\delta_1 X_1 + \delta_2(2A_1 - 1)\}$ can be expressed in terms of the four possible values (X_1, A_1) as

$$\begin{aligned}
\text{expit}\{\delta_1 X_1 + \delta_2(2A_1 - 1)\} &= \frac{1}{2}(1 + X_1)A_1 \text{expit}(\delta_1 + \delta_2) + \frac{1}{2}(1 + X_1)(1 - A_1)\text{expit}(\delta_1 - \delta_2) \\
&= \frac{1}{2}(1 - X_1)A_1 \text{expit}(-\delta_1 + \delta_2) + \frac{1}{2}(1 - X_1)(1 - A_1)\text{expit}(-\delta_1 - \delta_2)
\end{aligned}$$

Since $A_1 \in \{0,1\}$, we have $A_1(1-A_1)=0$, $A_1^2=A_1$ and $(1-A_1)^2=(1-A_1)$. We get

$$\begin{aligned}
\text{expit}\{\delta_1 X_1 + \delta_2(2A_1-1)\}A_1 &= \frac{1}{2}(1+X_1)A_1\text{expit}(\delta_1+\delta_2) \\
&\quad + \frac{1}{2}(1-X_1)A_1\text{expit}(-\delta_1+\delta_2) \\
\text{expit}\{\delta_1 X_1 + \delta_2(2A_1-1)\}(1-A_1) &= \frac{1}{2}(1+X_1)(1-A_1)\text{expit}(\delta_1-\delta_2) \\
&\quad + \frac{1}{2}(1-X_1)(1-A_1)\text{expit}(-\delta_1-\delta_2)
\end{aligned}$$

We then have

$$\begin{aligned}
&Q_1(\boldsymbol{H_1}, A_1) = M + A_1 f_3 \mathbb{I}(f_3>0) + (1-A_1)\mathbb{I}(f_4>0) \\
&\quad + \left(\frac{1}{2}(1+X_1)A_1\text{expit}(\delta_1+\delta_2) + \frac{1}{2}(1-X_1)A_1\text{expit}(-\delta_1+\delta_2)\right)\{f_1\mathbb{I}(f_1>0) \\
&\quad - f_3\mathbb{I}(f_3>0)\} + \left(\frac{1}{2}(1+X_1)(1-A_1)\text{expit}(\delta_1-\delta_2) + \frac{1}{2}(1-X_1)(1-A_1)\text{expit}(-\delta_1-\delta_2)\right) \\
&\quad \times \{f_2\mathbb{I}(f_2>0) - f_4\mathbb{I}(f_4>0)\}
\end{aligned}$$

The coefficient in front of A_1 is expressed in terms of the parameters of the data generating model λ and δ and is given by

$$\begin{aligned}
\psi_{10} =& \lambda_3 + f_3\mathbb{I}(f_3>0) - f_4\mathbb{I}(f_4>0) \\
&+ \left(\frac{1}{2}\text{expit}(\delta_1+\delta_2) + \frac{1}{2}\text{expit}(-\delta_1+\delta_2)\right)\{f_1\mathbb{I}(f_1>0) - f_3\mathbb{I}(f_3>0)\} \\
&- \left(\frac{1}{2}\text{expit}(\delta_1-\delta_2) + \frac{1}{2}\text{expit}(-\delta_1-\delta_2)\right)\{f_2\mathbb{I}(f_2>0) - f_4\mathbb{I}(f_4>0)\} \qquad \text{(B.2)}
\end{aligned}$$

B.2.1 Degree of Non-regularity

We classify the simulation scenarios in terms of (1) the probability p of generating an individual history such that $\lambda_5 A_2 + \lambda_6 X_2 A_2 + \lambda_7 A_1 A_2 = 0$, and (2) the standardized effect size $\phi = \mathbb{E}[(\lambda_5+\lambda_6 X_1+\lambda_7 A_1)/\sqrt{\mathbb{V}\text{ar}(\lambda_5+\lambda_6 X_1+\lambda_7 A_1)}]$. Scenarios with $p>0$ are characterized

as "non-regular". Scenarios with $p = 0$ and large effect of the stage 2 treatment are classified as "regular". Scenarios with $p > 0$ but weaker effect of the stage 2 treatment (such that the effect may be hard to detect) are classified as "near non-regular".

The two measures of non-regularity can be calculated in terms of the parameters of the data generating model λ and δ. The standardized effect size ϕ can be calculated in terms of the distribution of the linear combination $(\lambda_5 + \lambda_6 X_2 + \lambda_7 A_1)$ (see Table B.1). We have $\mathbb{E}[\lambda_5 + \lambda_6 X_1 + \lambda_7 A_1] = q_1 f_1 + q_2 f_2 + q_3 f_3 + q_4 f_4$ and $\mathbb{E}[(\lambda_5 + \lambda_6 X_1 + \lambda_7 A_1)^2] = q_1 f_1^2 + q_2 f_2^2 + q_3 f_3^2 + q_4 f_4^2$, which can be used to calculate $\mathbb{V}\text{ar}(\lambda_5 + \lambda_6 X_1 + \lambda_7 A_1)$. Similarly, the probability p can be calculated with $\mathbb{P}(\lambda_5 + \lambda_6 X_1 + \lambda_7 A_1 = 0)$.

Table B.1: Distribution of the linear combination $(\lambda_5 + \lambda_6 X_2 + \lambda_7 A_1)$.

(X_2, A_1) cell	Cell probability	Value of $(\lambda_5 + \lambda_6 X_2 + \lambda_7 A_1)$
(1,1)	$q_1 = \frac{1}{4}(\text{expit}\{\delta_1 + \delta_2\} + \text{expit}\{-\delta_1 + \delta_2\})$	$f_1 = \lambda_5 + \lambda_6 + \lambda_7$
(1,0)	$q_2 = \frac{1}{4}(\text{expit}\{\delta_1\} + \text{expit}\{-\delta_1\})$	$f_2 = \lambda_5 + \lambda_6$
(-1,1)	$q_3 = \frac{1}{4}(\text{expit}\{\delta_1 + \delta_2\} + \text{expit}\{-\delta_1 + \delta_2\})$	$f_3 = \lambda_5 - \lambda_6 + \lambda_7$
(-1,0)	$q_4 = \frac{1}{4}(\text{expit}\{\delta_1\} + \text{expit}\{-\delta_1\})$	$f_4 = \lambda_5 - \lambda_6$

B.2.2 Calculation Examples

We give calculation examples and details for the 9 simulation scenarios described in Table 3.1 of the manuscript.

Scenario 1. Non-regular scenario ($p = 1$, $\phi = 0/0$, $\psi_{10} = 0$). No treatment effect for all subjects in either stage.

Scenario 2. Near non-regular scenario ($p = 0$, $\phi = \infty$, $\psi_{10} = 0$). The main effect of the treatment at stage 2 is very small (λ_5=0.01) for all subjects. This scenario is regular as

p=0, but is very close to the non-regular scenario 1.

Scenario 3. Non-regular scenario ($p = 1/2$, $\phi = 1$, $\psi_{10} = -1$). No effect of the treatment at stage 2 for half of the subjects, and a large effect of the stage 2 treatment for the other half.

Scenario 4. Near non-regular scenario ($p = 0$, $\phi = 1.03$, $\psi_{10} = -1.48$). Weak effect of the treatment at stage 2 for half of the subjects, and a large effect of the stage 2 treatment for the other half. The weak effect may be hard to detect, hence this scenario is close to the non-regular previous scenario.

Scenario 5. Non-regular scenario ($p = 1/4$, $\phi = 1.34$, $\psi_{10} = -1$). Weak effect of the treatment at stage 2 for half of the subjects.

Scenario 6. Regular scenario ($p = 0$, $\phi = 0.93$, $\psi_{10} = -0.08$). Large effect of the treatment at stage 2 for all subjects. This is a regular scenario.

Scenario 7. Regular scenario ($p = 0$, $\phi = 1.90$, $\psi_{10} = 0.30$). Large effect of the treatment at stage 2 for half of the subjects.

Scenario 8. Non-regular scenario ($p = 1/2$, $\phi = 1$, $\psi_{10} = -1$). No effect of the treatment at stage 2 for half of the subjects, and a moderate effect of the stage 2 treatment for the other half.

Scenario 9. Near non-regular scenario ($p = 0$, $\phi = 1.08$, $\psi_{10} = -0.24$). Weak effect of the treatment at stage 2 for half of the subjects and moderate effect of the treatment at stage 2

for the other half. This scenario is very close to the previous non-regular scenario.

For details on the scenarios, see Chakraborty et al. (2010) for the first 6 scenarios, and Laber et al. (2014b) for scenarios 7, 8 and 9, respectively corresponding to examples A, B and C.

B.3 PROBIT: The IPCW Analysis

The blip parameters estimated via the inverse probability of censoring (IPCW) approach, along with 95% confidence intervals constructed in four different ways, are shown in Table B.2. Estimates, confidence intervals and conclusions are similar to the one obtained with the complete-case analysis.

B.4 PROBIT: Diagnostic Plots

Residual analysis is commonly used tool to diagnose model misspecification in standard regression. In the context of optimal DTR, residual analysis can be performed at each stage to simultaneously diagnose model misspecification of the blip and treatment-free models. Details on how to calculate the residuals and the fitted values at each stage are given in Rich et al. (2010). Under correct specification of the two models, residuals should be symmetrically distributed around zero, and show no trend when plotted against the fitted values, or against the covariates included in the blip model.

Residual analysis for the complete-case analysis with BMI as the outcome is shown in Figure B.1. The upper row shows the diagnostic plots for the first stage, and the row below shows the same diagnostic plots, but for the second stage. The left figures show the residuals against the fitted values, and the right figures show the residuals against the infant's weight

Table B.2: Estimates of the blip parameters $\{\psi_{10}, \psi_{11}, \psi_{20}, \psi_{21}\}$ in the PROBIT data analysis with three outcomes using inverse probability of censoring weighting along with 95% confidence intervals calculated with standard bootstrap (nn), m-out-of-n bootstrap with α=0.05 ($\text{mn}_{0.05}$), m-out-of-n bootstrap with α=0.1 ($\text{mn}_{0.1}$) and m-out-of-n bootstrap with adaptive choice of α ($\text{mn}_{\hat{\alpha}}$).

IPCW analysis					
		95% Confidence Interval			
Estimates		nn	$\text{mn}_{0.05}$	$\text{mn}_{0.1}$	$\text{mn}_{\hat{\alpha}}$
BMI (n_1†=8,910, n_2‡=9,144, $\hat{\alpha}$††=0.07)					
$\hat{\psi}_{10}$	-0.40	(-1.23; 0.43)	(-1.42; 0.62)	(-1.66; 0.85)	(-1.55; 0.75)
$\hat{\psi}_{11}$	0.06	(-0.10; 0.22)	(-0.12; 0.25)	(-0.16; 0.29)	(-0.14; 0.27)
$\hat{\psi}_{20}$	-0.46	(-1.76; 0.84)	(-2.12; 1.21)	(-2.54; 1.62)	(-2.28; 1.36)
$\hat{\psi}_{21}$	0.05	(-0.13; 0.24)	(-0.17; 0.28)	(-0.22; 0.33)	(-0.19; 0.30)
Waist Circumference (n_1=8,913, n_2=9,147, $\hat{\alpha}$=0.07)					
$\hat{\psi}_{10}$	0.36	(-1.69; 2.41)	(-2.17; 2.90)	(-2.63; 3.35)	(-2.33; 3.05)
$\hat{\psi}_{11}$	-0.09	(-0.43; 0.26)	(-0.51; 0.34)	(-0.58; 0.41)	(-0.54; 0.36)
$\hat{\psi}_{20}$	-1.38	(-4.78; 2.02)	(-5.65; 2.88)	(-6.89; 4.12)	(-6.07; 3.31)
$\hat{\psi}_{21}$	0.21	(-0.22; 0.64)	(-0.33; 0.74)	(-0.49; 0.90)	(-0.39; 0.80)
Triceps Skinfold Thickness (n_1=8,911, n_2=9,145, $\hat{\alpha}$=0.07)					
$\hat{\psi}_{10}$	-1.39	(-3.10; 0.32)	(-3.61; 0.83)	(-4.19; 1.41)	(-3.85; 1.07)
$\hat{\psi}_{11}$	0.22	(-0.07; 0.50)	(-0.15; 0.59)	(-0.24; 0.68)	(-0.19; 0.63)
$\hat{\psi}_{20}$	-0.27	(-3.73; 3.18)	(-4.39; 3.84)	(-5.35; 4.80)	(-4.97; 4.42)
$\hat{\psi}_{21}$	-0.04	(-0.40; 0.48)	(-0.47; 0.56)	(-0.60; 0.68)	(-0.55; 0.64)

† sample size at first stage, ‡ sample size at second stage, †† adaptive α using double bootstrap

at the beginning of the interval. A loess smoother is drawn through the points (solid red curve), and the dashed blue line shows the zero level. For all diagnostic plots, we notice no weird trend. Figures B.2 and B.3 shows the same diagnostic plots for the complete-case analysis with waist circumference and tricep skinfold thickness as the outcome, respectively. Again, we identify no alarming pattern in the residual plots.

Figures B.4, B.5 and B.6 show the residual analysis for the IPCW analysis with BMI, waist circumference and tricep skinfold thickness as the outcome, respectively. All residual plots show no trend.

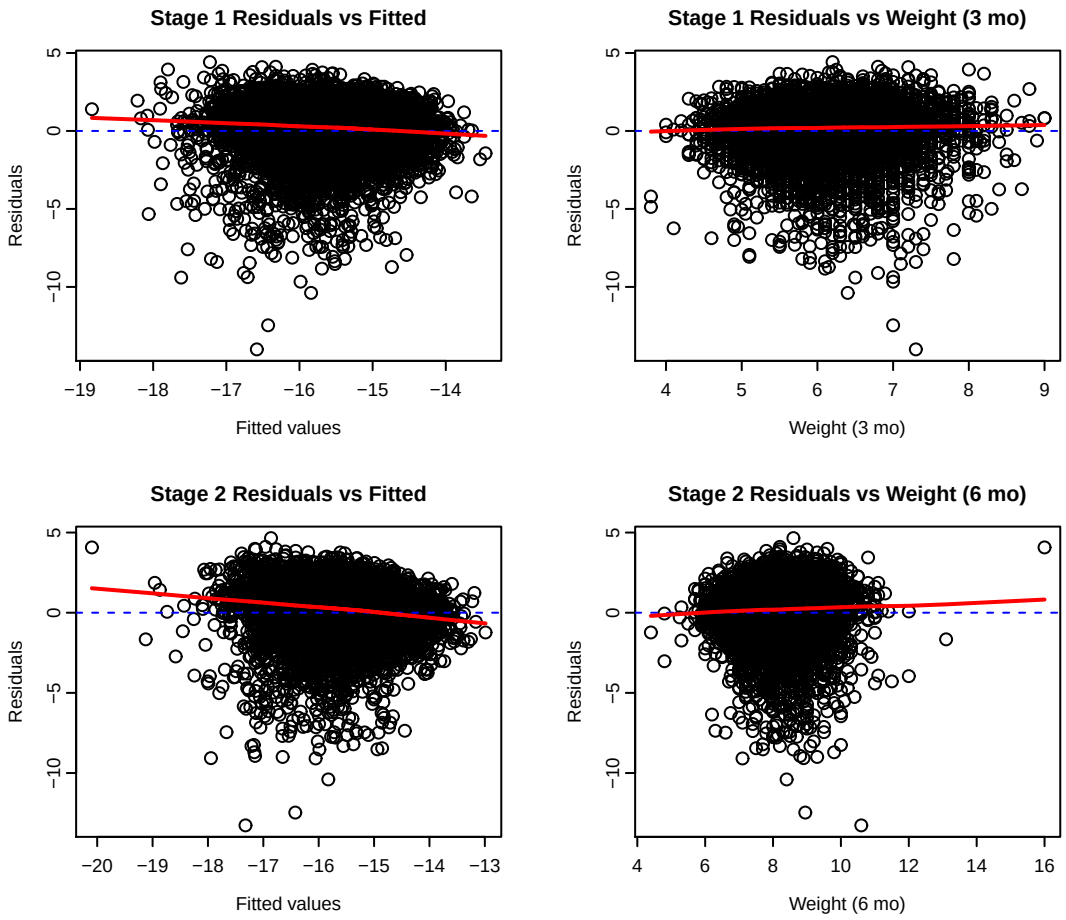

Figure B.1: Plots of residuals at each stage vs. fitted values (left), and vs. infant weight at the start of the interval (right) for the complete-case analysis with BMI as the outcome. A loess smoother is drawn through the points (solid red curve), and the dashed blue horizontal line shows zero level.

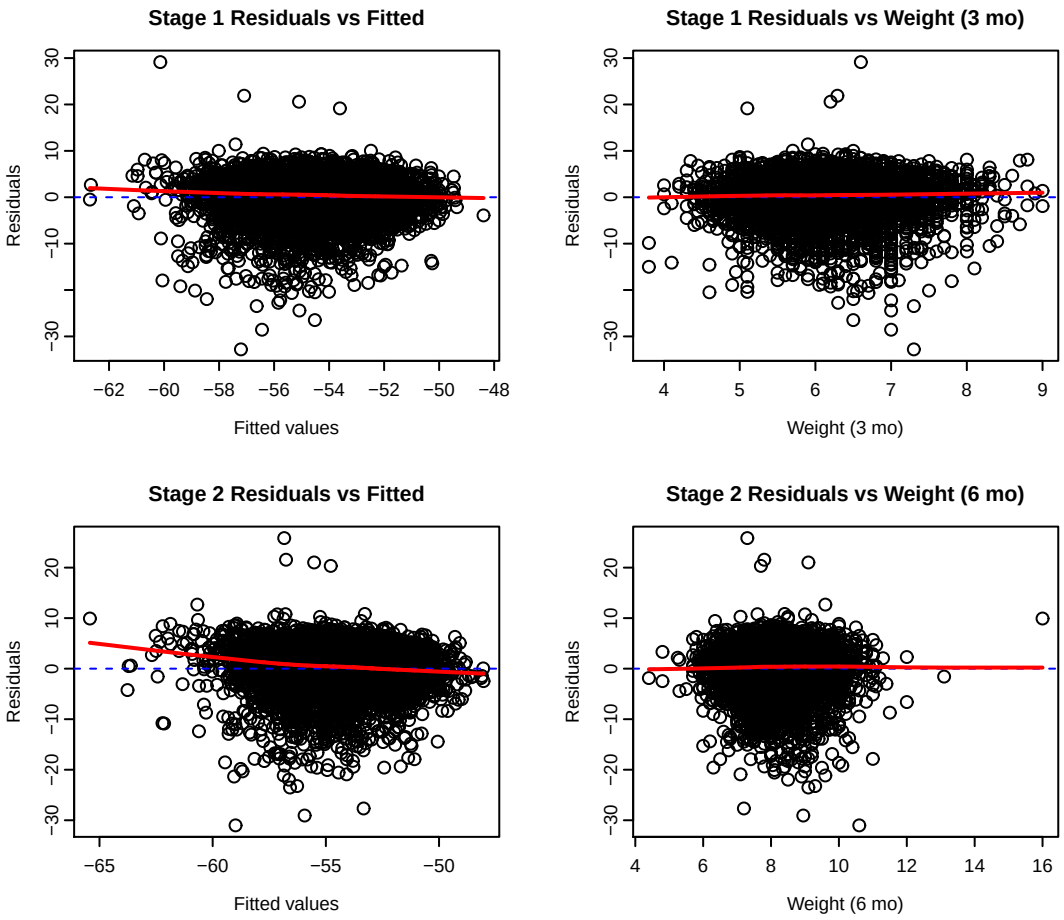

Figure B.2: Plots of residuals at each stage vs. fitted values (left), and vs. infant weight at the start of the interval (right) for the complete-case analysis with waist circumference as the outcome. A loess smoother is drawn through the points (solid red curve), and the dashed blue horizontal line shows zero level.

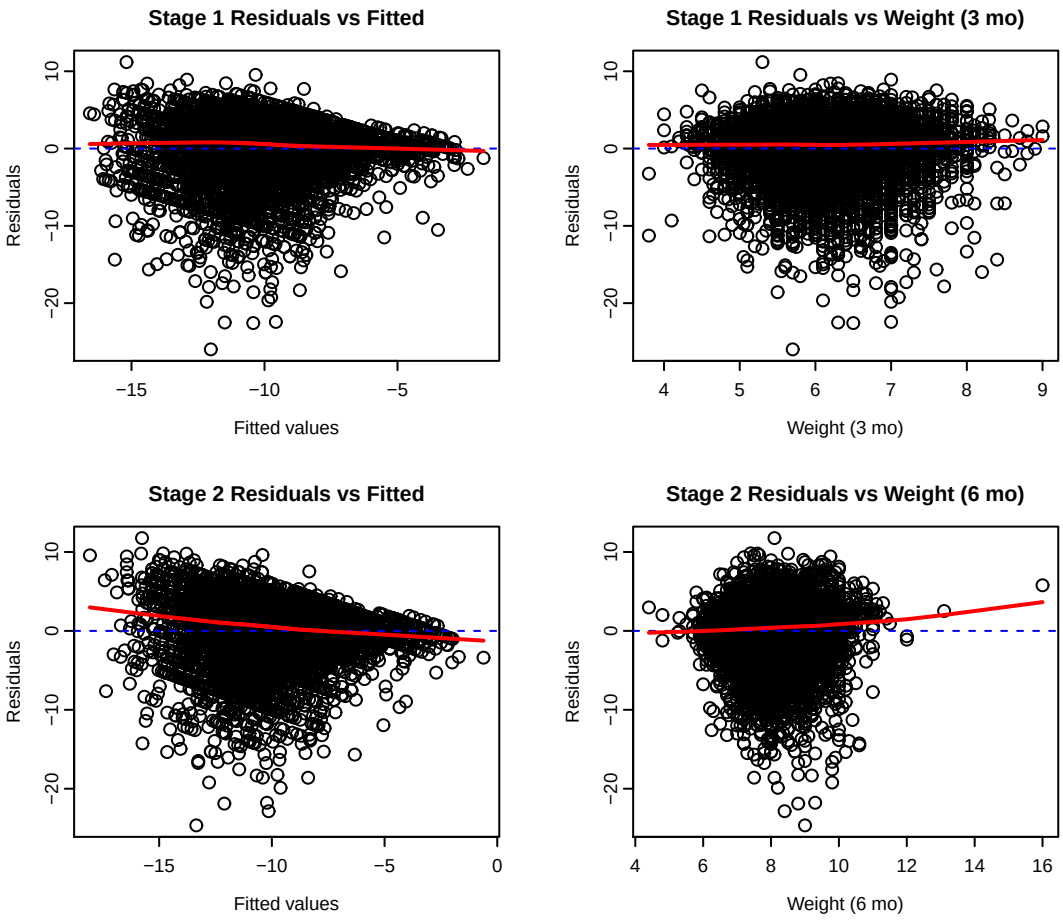

Figure B.3: Plots of residuals at each stage vs. fitted values (left), and vs. infant weight at the start of the interval (right) for the complete-case analysis with tricep skinfold thickness as the outcome. A loess smoother is drawn through the points (solid red curve), and the dashed blue horizontal line shows zero level.

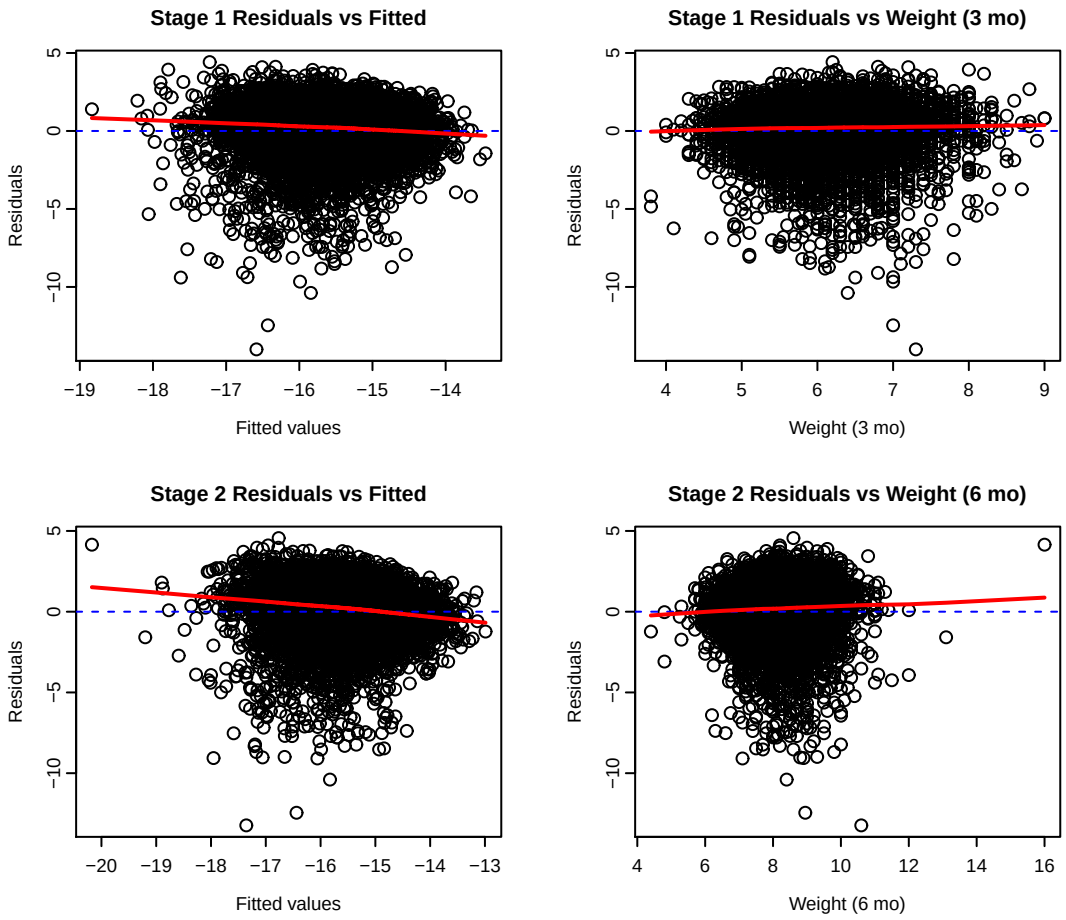

Figure B.4: Plots of residuals at each stage vs. fitted values (left), and vs. infant weight at the start of the interval (right) for the IPCW analysis with BMI as the outcome. A loess smoother is drawn through the points (solid red curve), and the dashed blue horizontal line shows zero level.

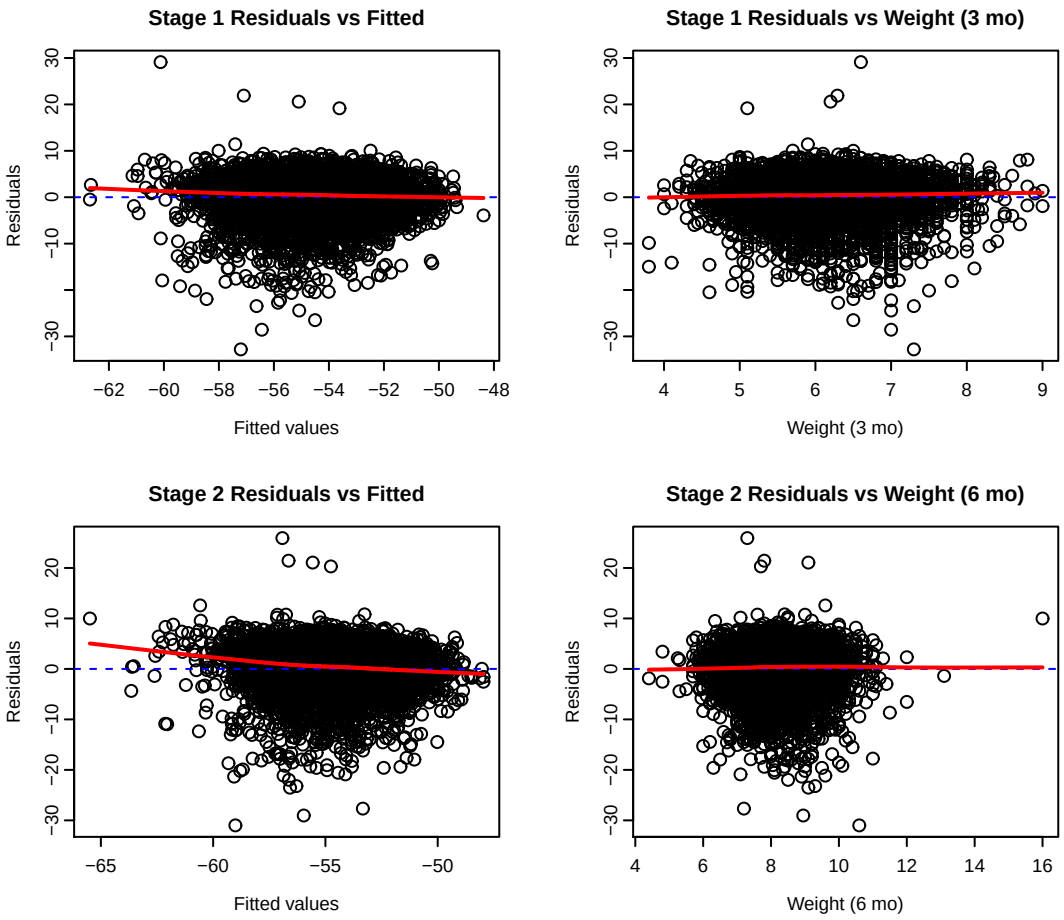

Figure B.5: Plots of residuals at each stage vs. fitted values (left), and vs. infant weight at the start of the interval (right) for the IPCW analysis with waist circumference as the outcome. A loess smoother is drawn through the points (solid red curve), and the dashed blue horizontal line shows zero level.

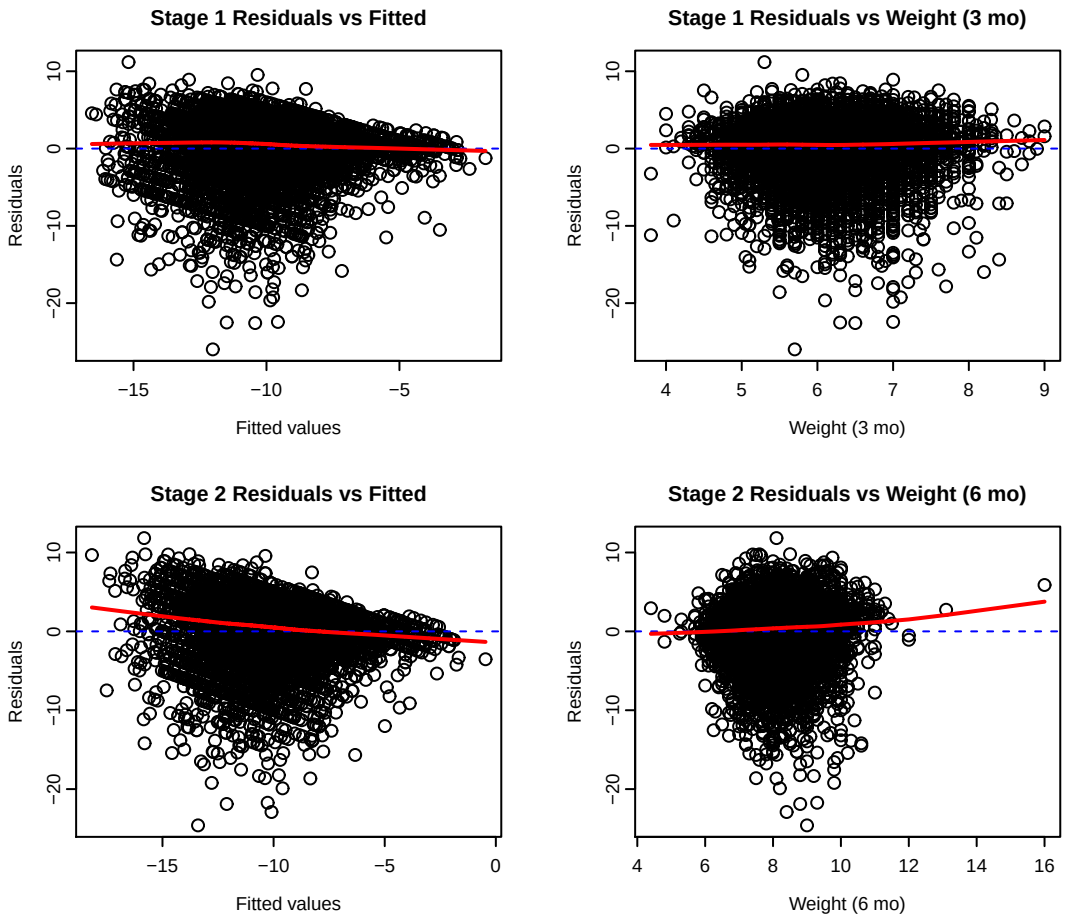

Figure B.6: Plots of residuals at each stage vs. fitted values (left), and vs. infant weight at the start of the interval (right) for the IPCW analysis with tricep skinfold thickness as the outcome. A loess smoother is drawn through the points (solid red curve), and the dashed blue horizontal line shows zero level.

Appendix C

Supplemental Materials for Chapter 4

C.1 Consistency and Double-robustness

We give a proof for the consistency and double-robustness of DWSurv. The proof is concerned with the blip estimators $\hat{\boldsymbol{\psi}}_{\boldsymbol{j}}$. For simplicity, the proof is detailed for a DTR with a single stage of intervention. It is straightforward to extend the reasoning to more than one stage.

In a one-stage DTR, DWSurv estimates the parameters $(\boldsymbol{\beta}, \boldsymbol{\psi})$ by solving the following GEE:

$$U(\boldsymbol{\beta}, \boldsymbol{\psi}) = \sum_{i=1}^{n} \delta_i \hat{w}_i \begin{pmatrix} \boldsymbol{h_{i\beta}} \\ a_i \boldsymbol{h_{i\psi}} \end{pmatrix} \left(\log(T_i) - \boldsymbol{\beta}^T \boldsymbol{h_{i\beta}} - a_i \boldsymbol{\psi}^T \boldsymbol{h_{i\psi}}\right) = 0.$$

We simplify the notations

$$U(\boldsymbol{\beta}, \boldsymbol{\psi}) = \sum_{i=1}^{n} \delta_i \hat{w}_i \begin{pmatrix} \boldsymbol{x_i} \\ a_i \boldsymbol{x_i} \end{pmatrix} \left(\log(T_i) - \boldsymbol{\beta}^T \boldsymbol{x_i} - a_i \boldsymbol{\psi}^T \boldsymbol{x_i}\right) = 0.$$

Note that, for simplicity, we took $\boldsymbol{H_\beta} = \boldsymbol{H_\psi} = \boldsymbol{X}$ but that the following double-robustness proof holds in the general case where $\boldsymbol{H_\beta}$ and $\boldsymbol{H_\psi}$ are different subsets of $\boldsymbol{H}$.

C.1.1 Treatment-free Model Correctly Specified, Weight Models Misspecified

First, let the treatment-free model be correctly specified i.e.

$$\mathbb{E}[\log(T_i)|\boldsymbol{x}, a] = \mathbb{E}\left[\log(T_i) - a_i\boldsymbol{\psi}^T\boldsymbol{x_i}|\boldsymbol{x}, a; \boldsymbol{\beta}\right] = \boldsymbol{\beta}^T\boldsymbol{x_i},$$

assuming the blip model $a_i\boldsymbol{\psi}^T\boldsymbol{x_i}$ is correctly specified. With the expectation taken with respect to $p(T, \Delta|A, \boldsymbol{X})$, we have

$$\begin{aligned}
\mathbb{E}[U(\boldsymbol{\beta}, \boldsymbol{\psi})] &= \sum_{i=1}^{n}\sum_{\delta_i=0}^{1}\left[\int_t \delta_i\hat{w}_i\begin{pmatrix}\boldsymbol{x_i}\\ a_i\boldsymbol{x_i}\end{pmatrix}\left(\log(T_i) - \boldsymbol{\beta}^T\boldsymbol{x_i} - a_i\boldsymbol{\psi}^T\boldsymbol{x_i}\right)p(T_i|\delta_i, a_i, x_i)dT\right]p(\Delta_i|a_i, x_i) \qquad \text{(C.1)}\\
&= \sum_{i=1}^{n}\sum_{\delta_i=0}^{1}\delta_i\hat{w}_i\begin{pmatrix}\boldsymbol{x_i}\\ a_i\boldsymbol{x_i}\end{pmatrix}\left[\int_t \left(\log(T_i) - \boldsymbol{\beta}^T\boldsymbol{x_i} - a_i\boldsymbol{\psi}^T\boldsymbol{x_i}\right)p(T_i|\delta_i, a_i, \boldsymbol{x_i})dT\right]p(\Delta_i|a_i, \boldsymbol{x_i})\\
&= \sum_{i=1}^{n}\sum_{\delta_i=0}^{1}\delta_i\hat{w}_i\begin{pmatrix}\boldsymbol{x_i}\\ a_i\boldsymbol{x_i}\end{pmatrix}p(\Delta_i|a_i, \boldsymbol{x_i})\left[\int_t \left(\log(T_i) - \boldsymbol{\beta}^T\boldsymbol{x_i} - a_i\boldsymbol{\psi}^T\boldsymbol{x_i}\right)p(T_i|\delta_i, a_i, \boldsymbol{x_i})dT\right]\\
&= \sum_{i=1}^{n}\mathbb{E}\left[\delta_i\hat{w}_i\begin{pmatrix}\boldsymbol{x_i}\\ a_i\boldsymbol{x_i}\end{pmatrix}\middle| a_i, \boldsymbol{x_i}\right]\mathbb{E}\left[\log(T_i) - \boldsymbol{\beta}^T\boldsymbol{x_i} - a_i\boldsymbol{\psi}^T\boldsymbol{x_i}|a_i, \boldsymbol{x_i}\right]
\end{aligned}$$

where (C.1) is due to expressing $p(T, \Delta|A, \boldsymbol{X})$ as $p(T|\Delta, A, \boldsymbol{X}) \times p(\Delta|A, \boldsymbol{X}) = p(T|A, \boldsymbol{X}) \times p(\Delta|A, \boldsymbol{X})$ because of the coarsening at random assumption (i.e. $T \perp\!\!\!\perp \Delta|\boldsymbol{X}, A$). When the treatment-free model is correctly specified, the second expectation is

$$\begin{aligned}
\mathbb{E}\left[\log(T_i) - \boldsymbol{\beta}^T\boldsymbol{x_i} - a_i\boldsymbol{\psi}^T\boldsymbol{x_i}|a_i, \boldsymbol{x_i}\right] &= \mathbb{E}\left[\log(T_i) - a_i\boldsymbol{\psi}^T\boldsymbol{x_i} - \boldsymbol{\beta}^T\boldsymbol{x_i}|a_i, \boldsymbol{x_i}\right]\\
&= \mathbb{E}\left[\log(T_i) - a_i\boldsymbol{\psi}^T\boldsymbol{x_i}|a_i, \boldsymbol{x_i}\right] - \mathbb{E}\left[\boldsymbol{\beta}^T\boldsymbol{x_i}|a_i, \boldsymbol{x_i}\right]\\
&= 0
\end{aligned}$$

such that $U(\boldsymbol{\beta}, \boldsymbol{\psi})$ consistently estimates the blip parameters $\boldsymbol{\psi}$ when the treatment-free model is correctly specified.

C.1.2 Treatment-free Model Misspecified, Weight Models Correctly Specified

Second, let the weight models be correctly specified i.e. $\mathbb{P}(A = 1|\boldsymbol{X}; \boldsymbol{\alpha}) = \mathbb{P}(A = 1|\boldsymbol{X})$ and $\mathbb{P}(\Delta = 1|A, \boldsymbol{X}; \boldsymbol{\lambda}) = \mathbb{P}(\Delta = 1|A, \boldsymbol{X})$. We show that finding the root of the estimating function $U(\boldsymbol{\beta}, \boldsymbol{\psi})$ yields consistent estimators of $\boldsymbol{\psi}$ if the weights satisfy the following balancing property.

Theorem (balancing property): Under assumptions 1–3 listed in the article, solving the weighted GEE $U(\boldsymbol{\beta}, \boldsymbol{\psi}) = 0$ yields consistent estimators of $\boldsymbol{\psi}$ if the weights satisfy the balancing property

$$[1-g(0,\boldsymbol{x})][1-\pi(\boldsymbol{x})]w(0,0,\boldsymbol{x}) = g(0,\boldsymbol{x})[1-\pi(\boldsymbol{x})]w(0,1,\boldsymbol{x}) = [1-g(1,\boldsymbol{x})]\pi(\boldsymbol{x})w(1,0,\boldsymbol{x}) = g(1,\boldsymbol{x})\pi(\boldsymbol{x})w(1,1,\boldsymbol{x}) \tag{C.2}$$

where $\pi(\boldsymbol{x}) = \mathbb{P}(A = 1|\boldsymbol{X} = \boldsymbol{x})$ and $g(a, \boldsymbol{x}) = \mathbb{P}(\Delta = 1|A = a, \boldsymbol{X} = \boldsymbol{x})$.

Proof. In a standard linear regression $\log(y) \sim \boldsymbol{\beta}^T\boldsymbol{x} + a\boldsymbol{\psi}^T\boldsymbol{x}$ restricted to observations with $\delta = 1$, the estimators $\boldsymbol{\psi}$ are confounded (and potentially biased) by any lack of independence between the elements of $\boldsymbol{X}$ and (A, Δ). It is therefore sufficient to perform a standard linear regression on a weighted data set $(y_w, \boldsymbol{x}_w, a_w, \delta_w)$ wherein the covariates x_w are independent of exposure (a_w, δ_w) or, equivalently,

$$\mathbb{E}(\boldsymbol{X}_w|A_w = 0, \Delta_w = 1) = \mathbb{E}(\boldsymbol{X}_w|A_w = 1, \Delta_w = 1) = \mathbb{E}(\boldsymbol{X}_w|A_w = 0, \Delta_w = 0) = \mathbb{E}(\boldsymbol{X}_w|A_w = 1, \Delta_w = 0)$$

For this, it suffices to find weight such that

$$\frac{\mathbb{P}(A_w=0,\Delta_w=0|\boldsymbol{X}_w=\boldsymbol{x})}{\mathbb{P}(A_w=0,\Delta_w=0)}=\frac{\mathbb{P}(A_w=0,\Delta_w=1|\boldsymbol{X}_w=\boldsymbol{x})}{\mathbb{P}(A_w=0,\Delta_w=1)}=\frac{\mathbb{P}(A_w=1,\Delta_w=0|\boldsymbol{X}_w=\boldsymbol{x})}{\mathbb{P}(A_w=1,\Delta_w=0)}=\frac{\mathbb{P}(A_w=1,\Delta_w=1|\boldsymbol{X}_w=\boldsymbol{x})}{\mathbb{P}(A_w=1,\Delta_w=1)}. \tag{C.3}$$

To satisfy (C.3), it is sufficient to find weights that ensure the numerators are equal and the denominators are equal. Noticing that $\mathbb{P}(A_w=0,\Delta_w=0|\boldsymbol{X}_w=\boldsymbol{x})+\mathbb{P}(A_w=0,\Delta_w=1|\boldsymbol{X}_w=\boldsymbol{x})+\mathbb{P}(A_w=1,\Delta_w=0|\boldsymbol{X}_w=\boldsymbol{x})+\mathbb{P}(A_w=1,\Delta_w=1|\boldsymbol{X}_w=\boldsymbol{x})$ must equal 1, each numerator must equal 1/4. Let $k=w(0,0,\boldsymbol{x})\mathbb{P}(A=0,\Delta=0|\boldsymbol{X}=\boldsymbol{x})+w(0,1,\boldsymbol{x})\mathbb{P}(A=0,\Delta=1|\boldsymbol{X}=\boldsymbol{x})+w(1,0,\boldsymbol{x})\mathbb{P}(A=1,\Delta=0|\boldsymbol{X}=\boldsymbol{x})+w(1,1,\boldsymbol{x})\mathbb{P}(A=1,\Delta=1|\boldsymbol{X}=\boldsymbol{x})$, then each numerator can be expressed as

$$\mathbb{P}(A_w=a,\Delta_w=\delta|\boldsymbol{X}_w=\boldsymbol{x}_w)=\frac{1}{k}\mathbb{P}(A=a,\Delta=\delta|\boldsymbol{X}=\boldsymbol{x})w(a,\delta,\boldsymbol{x}).$$

The left-hand side must equal 1/4, and the right-hand side will be 1/4 for $(a,\delta)\in(0,1)^2$ if the weights are of the form (C.2). As an example, for $a=1$ and $\delta=1$, we have

$$\begin{aligned}\mathbb{P}(A_w=1,\Delta_w=1|\boldsymbol{X}_w=\boldsymbol{x}_w)&=\frac{1}{k}\mathbb{P}(A=1,\Delta=1|\boldsymbol{X}=\boldsymbol{x})w(1,1,\boldsymbol{x})\\&=\frac{1}{k}\mathbb{P}(\Delta=1|A=1,\boldsymbol{X}=\boldsymbol{x})\mathbb{P}(A=1|\boldsymbol{X}=\boldsymbol{x})w(1,1,\boldsymbol{x})\\&=\frac{1}{k}g(1,\boldsymbol{x})\pi(\boldsymbol{x})w(1,1,\boldsymbol{x}).\end{aligned} \tag{C.4}$$

If the weights satisfy (C.2), then

$$\begin{aligned}w(0,0,\boldsymbol{x})&=\frac{g(1,\boldsymbol{x})\pi(\boldsymbol{x})w(1,1,\boldsymbol{x})}{[1-g(0,\boldsymbol{x})][1-\pi(\boldsymbol{x})]}\\w(0,1,\boldsymbol{x})&=\frac{g(1,\boldsymbol{x})\pi(\boldsymbol{x})w(1,1,\boldsymbol{x})}{g(1,\boldsymbol{x})[1-\pi(\boldsymbol{x})]}\\w(1,0,\boldsymbol{x})&=\frac{g(1,\boldsymbol{x})\pi(\boldsymbol{x})w(1,1,\boldsymbol{x})}{[1-g(1,\boldsymbol{x})]\pi(\boldsymbol{x})}\end{aligned}$$

such that $k=4g(1,\boldsymbol{x})\pi(\boldsymbol{x})w(1,1,\boldsymbol{x})$ and (C.4) is indeed 1/4. This shows that the numerators are equal if the weights are of the specified form.

Next, using $f_{\boldsymbol{X}}(\boldsymbol{x})$ to denote the probability density function of $\boldsymbol{X}$ and writing $\int f_{\boldsymbol{X}}(\boldsymbol{x})kd\boldsymbol{x}$, the denominators in (C.3) can be written as

$$\begin{aligned}
\mathbb{P}(A_w = 0, \Delta_w = 0) &= \frac{1}{l}\int f_{\boldsymbol{X}}(\boldsymbol{x})\mathbb{P}(A = 0, \Delta = 0|\boldsymbol{X} = \boldsymbol{x})w(0, 0, \boldsymbol{x})d\boldsymbol{x} \\
&= \frac{1}{l}\int f_{\boldsymbol{X}}(\boldsymbol{x})[1 - \pi(\boldsymbol{x})][1 - g(0, \boldsymbol{x})]w(0, 0, \boldsymbol{x})d\boldsymbol{x}
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}(A_w = 0, \Delta_w = 1) &= \frac{1}{l}\int f_{\boldsymbol{X}}(\boldsymbol{x})\mathbb{P}(A = 0, \Delta = 1|\boldsymbol{X} = \boldsymbol{x})w(0, 1, \boldsymbol{x})d\boldsymbol{x} \\
&= \frac{1}{l}\int f_{\boldsymbol{X}}(\boldsymbol{x})[1 - \pi(\boldsymbol{x})]g(0, \boldsymbol{x})w(0, 1, \boldsymbol{x})d\boldsymbol{x}
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}(A_w = 1, \Delta_w = 0) &= \frac{1}{l}\int f_{\boldsymbol{X}}(\boldsymbol{x})\mathbb{P}(A = 1, \Delta = 0|\boldsymbol{X} = \boldsymbol{x})w(1, 0, \boldsymbol{x})d\boldsymbol{x} \\
&= \frac{1}{l}\int f_{\boldsymbol{X}}(\boldsymbol{x})\pi(\boldsymbol{x})[1 - g(1, \boldsymbol{x})]w(1, 0, \boldsymbol{x})d\boldsymbol{x}
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}(A_w = 1, \Delta_w = 1) &= \frac{1}{l}\int f_{\boldsymbol{X}}(\boldsymbol{x})\mathbb{P}(A = 1, \Delta = 1|\boldsymbol{X} = \boldsymbol{x})w(1, 1, \boldsymbol{x})d\boldsymbol{x} \\
&= \frac{1}{l}\int f_{\boldsymbol{X}}(\boldsymbol{x})\pi(\boldsymbol{x})g(1, \boldsymbol{x})w(1, 1, \boldsymbol{x})d\boldsymbol{x}
\end{aligned}$$

and again these four expressions will be equal if the weights satisfy (C.2). □

C.2 Details on the Asymptotic Variance Formulae

The derivations of an expression for the asymptotic variance of the estimators $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\psi}})$ are shown below. The details are given for a single-stage DTR in Section C.2.1. Section C.2.2 specifies the additional steps needed to extend the formulae to a multi-stage setting. Section C.2.3 discusses a situation where the asymptotic variance formulae may be inadequate. Section C.2.4 illustrates the performance of the formulae in a simulation study.

For a one-stage DTR, the parameters $(\boldsymbol{\beta}, \boldsymbol{\psi})$ are estimated by solving the following GEE:

$$\boldsymbol{U}(\boldsymbol{\beta}, \boldsymbol{\psi}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = \sum_{i=1}^{n} \delta_i w_i(\boldsymbol{\alpha}, \boldsymbol{\lambda}) \begin{pmatrix} \boldsymbol{h_{i\beta}} \\ a_i \boldsymbol{h_{i\psi}} \end{pmatrix} \left(\log(T_i) - \boldsymbol{\beta}^T \boldsymbol{h_{i\beta}} - a_i \boldsymbol{\psi}^T \boldsymbol{h_{i\psi}} \right) = 0$$

where we emphasize that the GEE also depend on the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$ involved in the estimation of the weights. The following derivations assume that all models (treatment, treatment-free, censoring, blip) are correctly specified.

We simplified the notations to make the derivations easier to read. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\psi})$, let the vector $\boldsymbol{X_{i\theta}} = (\boldsymbol{H_{i\beta}}, a_i \boldsymbol{H_{i\psi}})^T$, let $\boldsymbol{X_{i\lambda}} = (\boldsymbol{H_{i\lambda}}, a_i)^T$ be the covariates, including the treatment, used to construct the censoring model, and let $\boldsymbol{X_{i\alpha}} = \boldsymbol{H_{i\alpha}}$ be the covariates used to construct the treatment model. We consider the GEE with plug-in estimators of the nuisance parameters $(\boldsymbol{\alpha}, \boldsymbol{\lambda})$ as

$$\boldsymbol{U}(\boldsymbol{\theta}; \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}}) = \sum \delta_i w_i(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}}) \boldsymbol{X_{i\theta}} \left(\log(T_i) - \boldsymbol{\theta}^T \boldsymbol{X_{i\theta}} \right) = 0.$$

The following developments follow from Robins (2004), Moodie (2009) and Wallace & Moodie (2015).

C.2.1 Asymptotic Variance Formula

The variance of the estimator $\hat{\boldsymbol{\theta}}$ depends on the variance of $U(\boldsymbol{\theta}; \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}})$, which must adjust for the plug-in estimators of the nuisance parameters. The variance of $U(\boldsymbol{\theta}; \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}})$ can be derived by performing a first-order Taylor expression of the function about the limiting values of $\hat{\boldsymbol{\alpha}}$

and $\hat{\boldsymbol{\lambda}}$, $\boldsymbol{\alpha_0}$ and $\boldsymbol{\lambda_0}$, given by the adjusted estimating functions

$$\begin{aligned}\boldsymbol{U}_{\text{adj}}(\boldsymbol{\theta}) = \boldsymbol{U}(\boldsymbol{\theta}, \boldsymbol{\alpha_0}, \boldsymbol{\lambda_0}) + \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\alpha}} \boldsymbol{U}(\boldsymbol{\theta}, \boldsymbol{\alpha_0}, \boldsymbol{\lambda_0})\right](\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha_0}) \\ + \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\lambda}} \boldsymbol{U}(\boldsymbol{\theta}, \boldsymbol{\alpha_0}, \boldsymbol{\lambda_0})\right](\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda_0}) + o_p(1).\end{aligned} \tag{C.5}$$

The adjusted GEE have variance $\mathbb{E}[\boldsymbol{U}_{\text{adj}}(\boldsymbol{\theta})^{\otimes 2}] = \mathbb{E}[\boldsymbol{U}_{\text{adj}}(\boldsymbol{\theta})\boldsymbol{U}_{\text{adj}}(\boldsymbol{\theta})^T]$. From the delta method, an expression for the asymptotic variance of the estimator $\hat{\boldsymbol{\theta}}$ is given by

$$\mathbb{V}\text{ar}(\hat{\boldsymbol{\theta}}) = \mathbb{E}\left[\left\{\left(\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{U}_{\text{adj}}(\boldsymbol{\theta}, \boldsymbol{\alpha_0}, \boldsymbol{\lambda_0})\right]\right)^{-1} \boldsymbol{U}_{\text{adj}}(\boldsymbol{\theta}, \boldsymbol{\alpha_0}, \boldsymbol{\lambda_0})\right\}^{\otimes 2}\right]. \tag{C.6}$$

The variance of the estimator $\hat{\boldsymbol{\theta}}$ depends on the choice of weights through the derivatives with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$ in (C.5).

Details

Assuming logistic regressions for the treatment and censoring models

$$\begin{aligned}\hat{a}_i &:= \mathbb{P}(A_i = 1 | \boldsymbol{x_{i\alpha}}; \boldsymbol{\alpha_0}) = \frac{1}{1 + \exp(-\boldsymbol{\alpha_0^T x_{i\alpha}})} \\ \hat{d}_i &:= \mathbb{P}(\Delta_i = 1 | \boldsymbol{x_{i\lambda}}; \boldsymbol{\lambda_0}) = \frac{1}{1 + \exp(-\boldsymbol{\lambda_0^T x_{i\lambda}})},\end{aligned}$$

we derive an expression for the variance of $\hat{\boldsymbol{\theta}}$ using weights of the form

$$w_i(a_i, \delta_i, \boldsymbol{x_{i\alpha}}, \boldsymbol{x_{i\lambda}}; \boldsymbol{\alpha}, \boldsymbol{\lambda}) = \frac{|a_i - \hat{a}_i|}{\delta_i \hat{d}_i + (1 - \delta_i)(1 - \hat{d}_i)}.$$

The derivatives of the weight function in $\boldsymbol{U}(\boldsymbol{\theta};\boldsymbol{\alpha},\boldsymbol{\lambda})$ with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$ are

$$\frac{\partial \boldsymbol{w}}{\partial \boldsymbol{\alpha}} = \frac{\partial \boldsymbol{w}}{\partial \hat{\boldsymbol{a}}}\frac{\partial \hat{\boldsymbol{a}}}{\partial \boldsymbol{\alpha}} = \left[\frac{(\mathbf{1}-2\boldsymbol{a})\cdot\hat{\boldsymbol{a}}\cdot(\mathbf{1}-\hat{\boldsymbol{a}})}{\boldsymbol{\delta}\cdot\hat{\boldsymbol{d}}+(\mathbf{1}-\boldsymbol{\delta})\cdot(\mathbf{1}-\hat{\boldsymbol{d}})}\right]^T \boldsymbol{X}_{\boldsymbol{\alpha}}$$
$$\frac{\partial \boldsymbol{w}}{\partial \boldsymbol{\lambda}} = \left[\frac{|\boldsymbol{a}-\hat{\boldsymbol{a}}|\hat{\boldsymbol{d}}(\mathbf{1}-\hat{\boldsymbol{d}})}{[\boldsymbol{\delta}\hat{\boldsymbol{d}}+(\mathbf{1}-\boldsymbol{\delta})(\mathbf{1}-\hat{\boldsymbol{d}})]^2}\right]^T \boldsymbol{X}_{\boldsymbol{\lambda}}$$

where $\mathbf{1}$ is a vector of 1's of length n, $\cdot$ is the element-wise product, and the divisions and the square are also element-wise. The derivatives in (C.5) are

$$\mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\alpha}}\boldsymbol{U}(\boldsymbol{\theta},\boldsymbol{\alpha_0},\boldsymbol{\lambda_0})\right](\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha_0}) = n^{-1}\left[\frac{\boldsymbol{\delta}\cdot(\mathbf{1}-2\boldsymbol{a})\cdot\hat{\boldsymbol{a}}\cdot(\mathbf{1}-\hat{\boldsymbol{a}})}{\boldsymbol{\delta}\cdot\hat{\boldsymbol{d}}+(\mathbf{1}-\boldsymbol{\delta})\cdot(\mathbf{1}-\hat{\boldsymbol{d}})}\right]^T \boldsymbol{X}_{\boldsymbol{\theta}}^{\boldsymbol{T}}(\log(\boldsymbol{T})-\boldsymbol{\theta}^T\boldsymbol{X}_{\boldsymbol{\theta}})\boldsymbol{X}_{\boldsymbol{\alpha}}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha_0}),$$
$$\mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\lambda}}\boldsymbol{U}(\boldsymbol{\theta},\boldsymbol{\alpha_0},\boldsymbol{\lambda_0})\right](\hat{\boldsymbol{\lambda}}-\boldsymbol{\lambda_0}) = n^{-1}\left[\frac{\boldsymbol{\delta}\cdot\hat{\boldsymbol{d}}\cdot(1-\hat{\boldsymbol{d}})\cdot|\boldsymbol{a}-\hat{\boldsymbol{a}}|}{\boldsymbol{\delta}\cdot\hat{\boldsymbol{d}}+(\mathbf{1}-\boldsymbol{\delta})\cdot(\mathbf{1}-\hat{\boldsymbol{d}})}\right]^T \boldsymbol{X}_{\boldsymbol{\theta}}^{\boldsymbol{T}}(\log(\boldsymbol{T})-\boldsymbol{\theta}^T\boldsymbol{X}_{\boldsymbol{\theta}})\boldsymbol{X}_{\boldsymbol{\lambda}}(\hat{\boldsymbol{\lambda}}-\boldsymbol{\lambda_0}).$$

An expression for $U_{\text{adj}}(\boldsymbol{\theta})$ is then

$$\begin{aligned}\boldsymbol{U}_{\text{adj}}(\boldsymbol{\theta}) = &\left[\frac{\boldsymbol{\delta}\cdot|\boldsymbol{a}-\hat{\boldsymbol{a}}|}{\boldsymbol{\delta}\cdot\hat{\boldsymbol{d}}+(\mathbf{1}-\boldsymbol{\delta})\cdot(\mathbf{1}-\hat{\boldsymbol{d}})}\right]^T \boldsymbol{X}_{\boldsymbol{\theta}}^{\boldsymbol{T}}(\log(\boldsymbol{T})-\boldsymbol{\theta}^T\boldsymbol{X}_{\boldsymbol{\theta}})\\ &+ n^{-1}\left[\frac{\boldsymbol{\delta}\cdot(\mathbf{1}-2\boldsymbol{a})\cdot\hat{\boldsymbol{a}}\cdot(\mathbf{1}-\hat{\boldsymbol{a}})}{\boldsymbol{\delta}\cdot\hat{\boldsymbol{d}}+(\mathbf{1}-\boldsymbol{\delta})\cdot(\mathbf{1}-\hat{\boldsymbol{d}})}\right]^T \boldsymbol{X}_{\boldsymbol{\theta}}^{\boldsymbol{T}}(\log(\boldsymbol{T})-\boldsymbol{\theta}^T\boldsymbol{X}_{\boldsymbol{\theta}})\boldsymbol{X}_{\boldsymbol{\alpha}}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha_0})\\ &+ n^{-1}\left[\frac{\boldsymbol{\delta}\cdot\hat{\boldsymbol{d}}\cdot(1-\hat{\boldsymbol{d}})\cdot|\boldsymbol{a}-\hat{\boldsymbol{a}}|}{\boldsymbol{\delta}\cdot\hat{\boldsymbol{d}}+(\mathbf{1}-\boldsymbol{\delta})\cdot(\mathbf{1}-\hat{\boldsymbol{d}})}\right]^T \boldsymbol{X}_{\boldsymbol{\theta}}^{\boldsymbol{T}}(\log(\boldsymbol{T})-\boldsymbol{\theta}^T\boldsymbol{X}_{\boldsymbol{\theta}})\boldsymbol{X}_{\boldsymbol{\lambda}}(\hat{\boldsymbol{\lambda}}-\boldsymbol{\lambda_0})\end{aligned}$$

The derivative of $U_{\text{adj}}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ is given by

$$\begin{aligned}\mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{U}_{\text{adj}}(\boldsymbol{\theta},\boldsymbol{\alpha_0},\boldsymbol{\lambda_0})\right] = n^{-1}\Bigg(&\frac{\boldsymbol{\delta}\cdot|\boldsymbol{a}-\hat{\boldsymbol{a}}|}{\boldsymbol{\delta}\cdot\hat{\boldsymbol{d}}+(\mathbf{1}-\boldsymbol{\delta})\cdot(\mathbf{1}-\hat{\boldsymbol{d}})} + n^{-1}\frac{\boldsymbol{\delta}\cdot(\mathbf{1}-2\boldsymbol{a})\cdot\hat{\boldsymbol{a}}\cdot(\mathbf{1}-\hat{\boldsymbol{a}})}{\boldsymbol{\delta}\cdot\hat{\boldsymbol{d}}+(\mathbf{1}-\boldsymbol{\delta})\cdot(\mathbf{1}-\hat{\boldsymbol{d}})}\boldsymbol{X}_{\boldsymbol{\alpha}}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha_0})\\ &+ n^{-1}\frac{\boldsymbol{\delta}\cdot\hat{\boldsymbol{d}}\cdot(1-\hat{\boldsymbol{d}})\cdot|\boldsymbol{a}-\hat{\boldsymbol{a}}|}{\boldsymbol{\delta}\cdot\hat{\boldsymbol{d}}+(\mathbf{1}-\boldsymbol{\delta})\cdot(\mathbf{1}-\hat{\boldsymbol{d}})}\boldsymbol{X}_{\boldsymbol{\lambda}}(\hat{\boldsymbol{\lambda}}-\boldsymbol{\lambda_0})\Bigg)\boldsymbol{X}_{\boldsymbol{\theta}}^{\boldsymbol{T}}\boldsymbol{X}_{\boldsymbol{\theta}}\end{aligned}$$

Implementation

The previous form of the asymptotic variance of $\hat{\boldsymbol{\theta}}$ cannot be implemented because it depends on the unknown true values $\boldsymbol{\alpha_0}$ and $\boldsymbol{\lambda_0}$. A solution is to use a second Taylor expansion, giving rise to the function

$$\boldsymbol{U}_{\text{adj}} = \boldsymbol{U} - \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\alpha}}\boldsymbol{U}\right]\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\alpha}}\boldsymbol{s_\alpha}\right]^{-1}\boldsymbol{s_\alpha} - \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\lambda}}\boldsymbol{U}\right]\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\lambda}}\boldsymbol{s_\lambda}\right]^{-1}\boldsymbol{s_\lambda}$$

where $\boldsymbol{s_\alpha}$ and $\boldsymbol{s_\lambda}$ denote the score functions of the treatment and censoring models, respectively. Then, the variance of $\hat{\boldsymbol{\theta}}$ is given by

$$\mathbb{V}\text{ar}(\hat{\boldsymbol{\theta}}) = \mathbb{E}\left[\left\{\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{U}_{\text{adj}}\right]^{-1}\boldsymbol{U}_{\text{adj}}\right\}^{\otimes 2}\right].$$

We have

$$\begin{aligned}
&\boldsymbol{s_\alpha} = (\boldsymbol{a} - \hat{\boldsymbol{a}})\boldsymbol{X_\alpha}\\
&\boldsymbol{s_\lambda} = (\boldsymbol{\delta} - \hat{\boldsymbol{d}})\boldsymbol{X_\lambda}\\
&\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\alpha}}\boldsymbol{s_\alpha}\right]^{-1} = \left(\frac{1}{n}\hat{\boldsymbol{a}}\cdot(\boldsymbol{1}-\hat{\boldsymbol{a}})\boldsymbol{X_\alpha^T X_\alpha}\right)^{-1}\\
&\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\lambda}}\boldsymbol{s_\lambda}\right]^{-1} = \left(\frac{1}{n}\hat{\boldsymbol{d}}\cdot(\boldsymbol{1}-\hat{\boldsymbol{d}})\boldsymbol{X_\lambda^T X_\lambda}\right)^{-1}\\
&\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\alpha}}\boldsymbol{U}\right] = -\frac{1}{n}\left(\frac{\boldsymbol{\delta}\cdot(\boldsymbol{1}-2\boldsymbol{a})\cdot\hat{\boldsymbol{a}}\cdot(\boldsymbol{1}-\hat{\boldsymbol{a}})}{\boldsymbol{\delta}\cdot\hat{\boldsymbol{d}}+(\boldsymbol{1}-\boldsymbol{\delta})\cdot(\boldsymbol{1}-\hat{\boldsymbol{d}})}\boldsymbol{X_\alpha X_\theta}[\log(\boldsymbol{T}) - \boldsymbol{\theta}^T\boldsymbol{X_\theta}]\right)\\
&\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\lambda}}\boldsymbol{U}\right] = -\frac{1}{n}\left(\frac{\boldsymbol{\delta}\cdot|\boldsymbol{a}-\hat{\boldsymbol{a}}|\cdot\hat{\boldsymbol{d}}\cdot(\boldsymbol{1}-\hat{\boldsymbol{d}})}{\boldsymbol{\delta}\cdot\hat{\boldsymbol{d}}+(\boldsymbol{1}-\boldsymbol{\delta})\cdot(\boldsymbol{1}-\hat{\boldsymbol{d}})}\boldsymbol{X_\lambda X_\theta}[\log(\boldsymbol{T}) - \boldsymbol{\theta}^T\boldsymbol{X_\theta}]\right)
\end{aligned}$$

such that $\boldsymbol{U}_{\text{adj}}$ is given by

$$\boldsymbol{U}_{\text{adj}} = \frac{\boldsymbol{\delta} \cdot |\boldsymbol{a} - \hat{\boldsymbol{a}}|}{\boldsymbol{\delta} \cdot \hat{\boldsymbol{d}} + (\mathbf{1} - \boldsymbol{\delta}) \cdot (\mathbf{1} - \hat{\boldsymbol{d}})} \boldsymbol{X}_{\boldsymbol{\theta}} [\log(\boldsymbol{T}) - \boldsymbol{\theta}^T \boldsymbol{X}_{\boldsymbol{\theta}}] + \frac{1}{n} \left(\frac{\boldsymbol{\delta} \cdot (\mathbf{1} - 2\boldsymbol{a}) \cdot \hat{\boldsymbol{a}} \cdot (\mathbf{1} - \hat{\boldsymbol{a}})}{\boldsymbol{\delta} \cdot \hat{\boldsymbol{d}} + (\mathbf{1} - \boldsymbol{\delta}) \cdot (\mathbf{1} - \hat{\boldsymbol{d}})} \boldsymbol{X}_{\boldsymbol{\alpha}} \boldsymbol{X}_{\boldsymbol{\theta}} \right.$$
$$\left. [\log(\boldsymbol{T}) - \boldsymbol{\theta}^T \boldsymbol{X}_{\boldsymbol{\theta}}] \right) \left(\frac{1}{n} \hat{\boldsymbol{a}} \cdot (\mathbf{1} - \hat{\boldsymbol{a}}) \boldsymbol{X}_{\boldsymbol{\alpha}}^{\boldsymbol{T}} \boldsymbol{X}_{\boldsymbol{\alpha}} \right)^{-1} (\boldsymbol{a} - \hat{\boldsymbol{a}}) \boldsymbol{X}_{\boldsymbol{\alpha}} + \frac{1}{n} \left(\frac{\boldsymbol{\delta} \cdot |\boldsymbol{a} - \hat{\boldsymbol{a}}| \cdot \hat{\boldsymbol{d}} \cdot (\mathbf{1} - \hat{\boldsymbol{d}})}{\boldsymbol{\delta} \cdot \hat{\boldsymbol{d}} + (\mathbf{1} - \boldsymbol{\delta}) \cdot (\mathbf{1} - \hat{\boldsymbol{d}})} \right.$$
$$\left. \boldsymbol{X}_{\boldsymbol{\lambda}} \boldsymbol{X}_{\boldsymbol{\theta}} [\log(\boldsymbol{T}) - \boldsymbol{\theta}^T \boldsymbol{X}_{\boldsymbol{\theta}}] \right) \left(\frac{1}{n} \hat{\boldsymbol{d}} \cdot (\mathbf{1} - \hat{\boldsymbol{d}}) \boldsymbol{X}_{\boldsymbol{\lambda}}^{\boldsymbol{T}} \boldsymbol{X}_{\boldsymbol{\lambda}} \right)^{-1} (\boldsymbol{\delta} - \hat{\boldsymbol{d}}) \boldsymbol{X}_{\boldsymbol{\lambda}}$$

and its expectation is

$$\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{U}_{\text{adj}} \right] = \frac{1}{n} \left(\frac{\boldsymbol{\delta} \cdot |\boldsymbol{a} - \hat{\boldsymbol{a}}|}{\boldsymbol{\delta} \cdot \hat{\boldsymbol{d}} + (\mathbf{1} - \boldsymbol{\delta}) \cdot (\mathbf{1} - \hat{\boldsymbol{d}})} \boldsymbol{X}_{\boldsymbol{\theta}}^{\boldsymbol{T}} \boldsymbol{X}_{\boldsymbol{\theta}} - \frac{1}{n} \left(\frac{\boldsymbol{\delta} \cdot (\mathbf{1} - 2\boldsymbol{a}) \cdot \hat{\boldsymbol{a}} \cdot (\mathbf{1} - \hat{\boldsymbol{a}})}{\boldsymbol{\delta} \cdot \hat{\boldsymbol{d}} + (\mathbf{1} - \boldsymbol{\delta}) \cdot (\mathbf{1} - \hat{\boldsymbol{d}})} \boldsymbol{X}_{\boldsymbol{\alpha}} \boldsymbol{X}_{\boldsymbol{\theta}}^{\boldsymbol{T}} \boldsymbol{X}_{\boldsymbol{\theta}} \right) \right.$$
$$\left(\frac{1}{n} \hat{\boldsymbol{a}} \cdot (\mathbf{1} - \hat{\boldsymbol{a}}) \boldsymbol{X}_{\boldsymbol{\alpha}}^{\boldsymbol{T}} \boldsymbol{X}_{\boldsymbol{\alpha}} \right)^{-1} (\boldsymbol{a} - \hat{\boldsymbol{a}}) \boldsymbol{X}_{\boldsymbol{\alpha}} - \frac{1}{n} \left(\frac{\boldsymbol{\delta} \cdot |\boldsymbol{a} - \hat{\boldsymbol{a}}| \cdot \hat{\boldsymbol{d}} \cdot (\mathbf{1} - \hat{\boldsymbol{d}})}{\boldsymbol{\delta} \cdot \hat{\boldsymbol{d}} + (\mathbf{1} - \boldsymbol{\delta}) \cdot (\mathbf{1} - \hat{\boldsymbol{d}})} \boldsymbol{X}_{\boldsymbol{\lambda}} \boldsymbol{X}_{\boldsymbol{\theta}}^{\boldsymbol{T}} \boldsymbol{X}_{\boldsymbol{\theta}} \right)$$
$$\left. \left(\frac{1}{n} \hat{\boldsymbol{d}} \cdot (\mathbf{1} - \hat{\boldsymbol{d}}) \boldsymbol{X}_{\boldsymbol{\lambda}}^{\boldsymbol{T}} \boldsymbol{X}_{\boldsymbol{\lambda}} \right)^{-1} (\boldsymbol{\delta} - \hat{\boldsymbol{d}}) \boldsymbol{X}_{\boldsymbol{\lambda}} \right).$$

From these expressions, the asymptotic variance in (C.6) can be calculated.

C.2.2 Extension to More Than One Stage

With more than one stage of intervention, the procedure described above is adapted to account for the fact that the estimating functions in all stages except the last depend on plug-in blip estimators from later stages through the pseudo-outcome. For a DTR with two stages, the variance of the second stage estimators $(\hat{\boldsymbol{\beta}}_{\mathbf{2}}, \hat{\boldsymbol{\psi}}_{\mathbf{2}})$ is obtained following the procedure described above, where the adjusted estimating functions in (C.5) is denoted $\boldsymbol{U}_{2,\text{adj}}(\boldsymbol{\theta}_{\mathbf{2}})$. For the first stage estimators, the estimating functions $\boldsymbol{U}_{1,\text{adj}}(\boldsymbol{\theta}_{\mathbf{1}})$ further depends on the plug-in estimator $\hat{\boldsymbol{\psi}}_{\mathbf{2}}$ through the construction of the pseudo-outcome, which is treated as a nuisance quantity. The same principles as described above apply but an additional term is added to the adjusted estimating functions in the first stage to account for the additional

nuisance parameters as

$$\boldsymbol{U}^{\epsilon}_{1,\text{adj}}(\boldsymbol{\theta_1}) = \boldsymbol{U}_{1,\text{adj}}(\boldsymbol{\theta_1}) - \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\psi_2}}\boldsymbol{U}_1(\boldsymbol{\theta_1}, \boldsymbol{\alpha_{10}}, \boldsymbol{\lambda_{10}}, \boldsymbol{\psi_{20}})\right]\left(\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\psi_2}}\boldsymbol{U}_{2,\text{adj}}(\boldsymbol{\theta_2}, \boldsymbol{\alpha_{20}}, \boldsymbol{\lambda_{20}}, \boldsymbol{\psi_{20}})\right]\right)^{-1}$$
$$\times \boldsymbol{U}_{2,\text{adj}}(\boldsymbol{\theta_2}, \boldsymbol{\alpha_{20}}, \boldsymbol{\lambda_{20}}, \boldsymbol{\psi_{20}}).$$

C.2.3 A Note on Non-regularity

Under specific longitudinal distributions of the data, the first stage blip estimators may be non-regular in the sense that their asymptotic distributions do not converge uniformly over the parameter space (Robins, 2004). Non-regularity occurs because the pseudo-outcome $\tilde{T}$ is a non-smooth function of a plug-in estimator $\hat{\boldsymbol{\psi}}_{\boldsymbol{2}}$ as the function $\mathbb{I}(\boldsymbol{\psi_2^T H_{2\psi}} > 0)$ is not differentiable at $\{\boldsymbol{\psi_2^T H_{2\psi}} = 0\}$ i.e. when the treatment effect in the second stage is small or null. The first stage blip estimators $\hat{\boldsymbol{\psi}}_{\boldsymbol{1}}$ are in turn a non-smooth function of $\hat{\boldsymbol{\psi}}_{\boldsymbol{2}}$ and the asymptotic distribution $\sqrt{n}(\hat{\boldsymbol{\psi}}_{\boldsymbol{1}} - \boldsymbol{\psi_1})$ is not uniformly normal. A significant negative consequence of non-regularity is that typical confidence interval calculations for the blip parameters, including asymptotic and standard bootstrap procedures, perform poorly in terms of coverage (Chakraborty et al., 2010; Moodie & Richardson, 2010; Robins, 2004; Simoneau et al., 2017). The m-out-of-n bootstrap has been proposed to correct for non-regularity with uncensored continuous outcomes and showed good performance with dWOLS (Chakraborty et al., 2010; Simoneau et al., 2017), thus providing a promising solution for DWSurv.

C.2.4 A Simulation Study

We used the data generating mechanisms described in the Supplementary Material "Details on the simulation study". We considered situations where the censoring indicator was independent of the survival times, conditionally independent given baseline covariates and conditionally independent given time-varying covariates.

Tables C.1–C.3 compare the Monte Carlo standard errors and the standard errors estimated by the asymptotic variance formulae in a two-stage DTR. The formulae yielded conservative coverages.

Table C.1: Comparison of Monte Carlo standard errors and standard errors calculated with the asymptotic variance, with 95% Wald confidence interval coverage, in 1000 data sets with independent censoring.

	n=300			n=1,000			n=10,000		
θ	SE^{MC}	ESE	Cov.	SE^{MC}	ESE	Cov.	SE^{MC}	ESE	Cov.
30% independent censoring									
ψ_{10}	0.09	0.09	96.7	0.06	0.06	96.8	0.02	0.02	96.5
ψ_{11}	0.12	0.15	98.5	0.08	0.10	98.5	0.03	0.03	98.5
ψ_{20}	0.20	0.23	96.8	0.14	0.16	96.8	0.04	0.05	97.8
ψ_{21}	0.14	0.16	97.1	0.09	0.11	97.7	0.03	0.04	97.7
60% independent censoring									
ψ_{10}	0.10	0.11	96.3	0.07	0.08	96.4	0.02	0.02	96.8
ψ_{11}	0.14	0.17	98.7	0.10	0.12	98.1	0.03	0.04	98.5
ψ_{20}	0.24	0.27	97.3	0.17	0.19	97.5	0.05	0.06	97.9
ψ_{21}	0.17	0.19	97.0	0.11	0.13	98.1	0.04	0.04	97.6

SE^{MC}, Monte Carlo standard error; ESE, estimated standard errors; Cov., coverage of 95% CI

Table C.2: Comparison of Monte Carlo standard errors and standard errors calculated with the asymptotic variance, with 95% Wald confidence interval coverage, in 1000 data sets with censoring dependent on baseline covariates.

	n=300			n=1,000			n=10,000		
θ	SE^{MC}	ESE	Cov.	SE^{MC}	ESE	Cov.	SE^{MC}	ESE	Cov.
30% independent censoring									
ψ_{10}	0.08	0.09	96.8	0.06	0.07	97.0	0.02	0.02	96.7
ψ_{11}	0.11	0.14	98.3	0.08	0.10	99.1	0.02	0.03	99.1
ψ_{20}	0.18	0.21	97.4	0.13	0.15	97.8	0.04	0.05	97.3
ψ_{21}	0.12	0.14	97.5	0.09	0.10	97.8	0.03	0.03	98.3
60% independent censoring									
ψ_{10}	0.13	0.13	95.0	0.09	0.09	95.1	0.03	0.03	96.9
ψ_{11}	0.15	0.18	97.7	0.11	0.12	97.8	0.03	0.04	98.3
ψ_{20}	0.24	0.27	96.3	0.17	0.19	96.5	0.05	0.06	97.1
ψ_{21}	0.17	0.19	96.1	0.12	0.13	96.5	0.04	0.04	96.9

SE^{MC}, Monte Carlo standard error; ESE, estimated standard errors; Cov., coverage of 95% CI

Table C.3: Comparison of Monte Carlo standard errors and standard errors calculated with the asymptotic variance, with 95% Wald confidence interval coverage, in 1000 data sets with censoring dependent on time-varying covariates.

	n=300			n=1,000			n=10,000		
θ	SE^{MC}	ESE	Cov.	SE^{MC}	ESE	Cov.	SE^{MC}	ESE	Cov.
30% independent censoring									
ψ_{10}	0.08	0.09	95.9	0.06	0.06	96.9	0.02	0.02	98.0
ψ_{11}	0.10	0.13	98.7	0.07	0.09	98.6	0.02	0.03	99.0
ψ_{20}	0.18	0.22	97.4	0.13	0.15	97.9	0.04	0.05	97.7
ψ_{21}	0.12	0.15	97.9	0.09	0.10	98.4	0.03	0.03	97.8
60% independent censoring									
ψ_{10}	0.13	0.13	95.7	0.09	0.09	95.8	0.03	0.03	95.9
ψ_{11}	0.15	0.17	97.1	0.10	0.12	97.9	0.03	0.04	98.6
ψ_{20}	0.26	0.30	96.5	0.18	0.21	97.0	0.06	0.07	98.4
ψ_{21}	0.17	0.20	96.6	0.12	0.14	97.1	0.04	0.04	98.7

SE^{MC}, Monte Carlo standard error; ESE, estimated standard errors; Cov., coverage of 95% CI

C.3 Details on the Simulation Study

We offer a complement to the simulation study presented in the article which compared our method, DWSurv, to the method by HNW. In Section C.3.1, we introduce alternative data generating mechanisms. In Section C.3.2, we illustrate the consistency and double-robustness of DWSurv for all data generating mechanisms. In Sections C.3.3 & C.3.4, we consider additional metrics to evaluate the performance of DWSurv.

C.3.1 Alternative Data Generating Mechanisms

Independent censoring

In the article, we presented results for a data generating mechanism that yielded 30% independent censoring. We also considered data generating mechanisms yielding 60% independent censoring. For this, we used the same steps as described in Section 4.3 of the article but

generated the censoring indicator from a Bernoulli with probability of success 0.4 instead of 0.7 (second paragraph of the section, last sentence).

Censoring dependent on baseline covariates

We considered a data generating mechanism where both the probability of censoring and the censoring time depended on the baseline covariate X_1. We used the same steps as described in Section 4.3 of the article but generated the censoring indicator and the censoring time as a function of X_1. This implied generating the censoring indicator Δ from a Bernoulli with probability of success $\text{expit}(-0.4 + 2X_1)$ (30% censoring) or $\text{expit}(-1.8 + 2X_1)$ (60% censoring). For individuals who were censored ($\Delta = 0$), we generated their censoring time C using the inverse probability method (Bender et al., 2005). It sufficed to generate $U \sim \text{U}(0, 1)$ and make the inverse tranformation $C = S_C^{-1}(u|x_1)$ where $S_c^{-1}(\cdot|x_1)$ is the conditional survival function i.e. $S(t|x_1) = \exp\{-H_0(t)\exp(x_1\beta)\}$ where $H_0(t)$ is the baseline hazard. We used a Weibull baseline hazard $H_0(t) = \lambda t^\rho$ with shape $\rho = 0.9$ and scale $\lambda = 1/300$ and fixed $\beta = 2$.

Censoring dependent on time-varying covariates (DWSurv only)

We considered a data generating mechanism where the probability of censoring depended on time-varying covariates. The proposed data generating mechanism followed the same steps as described in the article, but the censoring indicator was generated such that it depended on the first and second stage covariates. Importantly, the data generating mechanism allowed to recover the correct specification of the censoring models i.e. the probability of censoring given that a patient entered the second stage $P(\Delta = 0|\eta_2 = 1, \boldsymbol{H_2}, A_2)$ and the overall probability of censoring given baseline information $P(\Delta = 0|\boldsymbol{H_1}, A_1)$. First, we describe the strategy that allowed generating a censoring indicator that depended on time-varying covariates. Then, we detail the steps for generating right-censored survival times with 30%

censoring.

Let δ_1 be an indicator in the first stage where $\delta_1 = 1$ if an event was observed in the first stage or if an individual did not experience an event but entered the second stage and $\delta_1 = 0$ if an individual was censored in the first stage. Let δ_2 be an indicator in the second stage where $\delta_2 = 1$ if an event occurred in the second stage and $\delta_2 = 0$ otherwise. An individual who was censored in the first stage will have overall censoring indicator equal to zero i.e. $\Delta = 0$ and $\eta_2 = 0$. An individual who was not censored in the first stage can either (i) experience an event in the first stage and have $\Delta = 1$ and $\eta_2 = 0$ or (ii) reach the second stage without experiencing an event before and have $\eta_2 = 1$. We have

$$
\begin{aligned}
&P(\Delta = 0|X_1, A_1) \\
&\quad = P(\Delta = 0 \text{ and } \delta_1 = 0|X_1, A_1) + P(\Delta = 0 \text{ and } \delta_1 = 1|X_1, A_1) \\
&\quad = P(\Delta = 0|\delta_1 = 0, X_1, A_1)P(\delta_1 = 0|X_1, A_1) + P(\Delta = 0|\delta_1 = 1, X_1, A_1)P(\delta_1 = 1|X_1, A_1) \\
&\quad = P(\delta_1 = 0|X_1, A_1) + P(\Delta = 0|\delta_1 = 1, X_1, A_1)P(\delta_1 = 1|X_1, A_1) \qquad \text{(C.7)}
\end{aligned}
$$

which decomposes the probability of being censored at any point during the follow-up into functions of the events "being censored in the first stage" and "being censored, but not in the first stage". The first part of the second term is

$$
\begin{aligned}
&P(\Delta = 0|\delta_1 = 1, X_1, A_1) \\
&\quad = P(\Delta = 0 \text{ and } \eta_2 = 0|\delta_1 = 1, X_1, A_1) + P(\Delta = 0 \text{ and } \eta_2 = 1|\delta_1 = 1, X_1, A_1) \\
&\quad = P(\Delta = 0|\eta_2 = 1, \delta_1 = 1, X_1, A_1)P(\eta_2 = 1|\delta_1 = 1, X_1, A_1) \\
&\quad = P(\delta_2 = 0|\eta_2 = 1, \delta_1 = 1, X_1, A_1)P(\eta_2 = 1|\delta_1 = 1, X_1, A_1). \qquad \text{(C.8)}
\end{aligned}
$$

Note that $P(\Delta = 0|X_1, A_1)$ and $P(\delta_2 = 0|\eta_2 = 1, \delta_1 = 1, X_1, A_1)$ are the two censoring models that we need to be able to correctly specify to assess the double-robustness property, respectively corresponding to $P(\Delta = 0|\boldsymbol{H_1}, A_1)$ and $P(\Delta = 0|\eta_2 = 1, \boldsymbol{H_2}, A_2)$. Using the

relationships (C.7) and (C.8), we express the probability of not being censored in the first stage $P(\delta_1 = 1|X_1, A_1)$ as a function of these two models as

$$P(\delta_1 = 1|X_1, A_1) = \frac{P(\Delta = 1|X_1, A_1)}{1 - P(\delta_2 = 1|\eta_2 = 1, \delta_1 = 1, X_1, A_1)P(\eta_2 = 1|\delta_1 = 1, X_1, A_1)}. \quad \text{(C.9)}$$

Up to this point, the probability of censoring still only depends on baseline characteristics (X_1, A_1). We want the probability of censoring given that an individual entered the second stage to further depend on second stage characteristics (X_2, A_2). We have

$$\begin{aligned}
&P(\delta_2 = 1|\eta_2 = 1, \delta_1 = 1, X_1, A_1) \\
&\quad = P(\delta_2 = 1 \text{ and } A_2 = 0|\eta_2 = 1, \delta_1 = 1, X_1, A_1) + P(\delta_2 = 1 \text{ and } A_2 = 1|\eta_2 = 1, \delta_1 = 1, X_1, A_1) \\
&\quad = P(\delta_2 = 1|A_2 = 0, \eta_2 = 1, \delta_1 = 1, X_1, A_1)P(A_2 = 0|\eta_2 = 1, \delta_1 = 1, X_1, A_1) \\
&\qquad + P(\delta_2 = 1|A_2 = 1, \eta_2 = 1, \delta_1 = 1, X_1, A_1)P(A_2 = 1|\eta_2 = 1, \delta_1 = 1, X_1, A_1)
\end{aligned}$$

and

$$P(\delta_2 = 1|A_2, \eta_2 = 1, \delta_1 = 1, X_1, A_1) = \mathbb{E}_{X_2}[P(\delta_2 = 1|X_2, A_2, \eta_2 = 1, \delta_1 = 1, X_1, A_1)]$$

such that the probability of censoring given that an individual entered the second stage can further depend on X_2 and A_2. Similarly, we have

$$P(A_2 = j|\eta_2 = 1, \delta_1 = 1, X_1, A_1) = \mathbb{E}_{X_2}[P(A_2 = j|X_2, \eta_2 = 1, \delta_1 = 1, X_1, A_1)]$$

for $j = 1, 2$. Then, $P(\delta_2 = 0|\eta_2 = 1, \delta_1 = 1, X_1, A_1)$ in (C.9) is replaced by

$$\begin{aligned}
&\mathbb{E}_{X_2}[P(\delta_2 = 0|X_2, A_2 = 0, \eta_2 = 1, \delta_1 = 1, X_1, A_1)]\mathbb{E}_{X_2}[P(A_2 = 0|X_2, \eta_2 = 1, \delta_1 = 1, X_1, A_1)] \\
&+ \mathbb{E}_{X_2}[P(\delta_2 = 0|X_2, A_2 = 1, \eta_2 = 1, \delta_1 = 1, X_1, A_1)]\mathbb{E}_{X_2}[P(A_2 = 1|X_2, \eta_2 = 1, \delta_1 = 1, X_1, A_1)]
\end{aligned} \quad \text{(C.10)}$$

The relationships derived above allow to generate a censoring indicator that depends on the baseline and second stage characteristics by specifying models for $P(\delta_2 = 1|X_2, A_2, \eta_2 = 1, \delta_1 = 1, X_1, A_1)$ and $P(\Delta = 1|X_1, A_1)$. We now assume parametric models for all random variables and describe the steps of the corresponding data generating mechanism.

For individual i, the first stage treatment was assigned with a Bernoulli distribution with $P(A_{i1} = 1|X_{i1}) = \text{expit}(-1 + 2X_{i1})$ where X_{i1} was a baseline continuous covariate generated from a Uniform$(0.1, 1.29)$. The assignment of the second stage treatment was also based on a Bernoulli distribution with $P(A_{i2} = 1|X_{i2}) = \text{expit}(2.8 - 2X_{i2})$ where X_{i2} was a continuous covariate measured at the beginning of the second stage generated from a Uniform$(0.9, 2)$. Let the probability of being censored at any point during the follow-up be $P(\Delta_i = 0|X_{i1}) = \text{expit}(-0.2 + 2X_{i1})$. Let the probability of entering the second stage given that the patient was not censored in the first stage be $P(\eta_{i2} = 1) = 0.8$. Let the probability of being censored in the second stage be $P(\delta_{i2} = 0|\eta_{i2} = 1, X_{i2}) = \text{expit}(0.3 - 2X_{i2})$. From (C.9) and (C.10), and given the distribution of X_2, we have that the probability of *not* being censored in the first stage is

$$P(\delta_{i1} = 1) = \text{expit}(-0.2 + 2X_{i1})/(1 - 0.0805 \times 0.8)$$

where $\mathbb{E}_{X_{i2}}[P(\delta_{i2} = 0|\eta_2 = 1, X_{i2})] \approx 0.0805$. Note that taking the expectation involved solving an integral and that this integral was approximated numerically.

We started by generating δ_{i1}. Those who were censored in the first stage had $\eta_{i2} = 0$, δ_{i2} was not applicable and $\Delta_i = 0$. For those who were not censored in the first stage ($\delta_{i1} = 1$), we generated η_{i2}. For those who entered the second stage ($\eta_{i2} = 1$), we generated δ_{i2} and set $\Delta_i = \delta_{i2}$. For those who did not enter the second stage, we set $\Delta_i = 1$. Once the censoring indicator was generated, the remaining steps of the algorithm were similar to what was described in the paper (starting on the 3$^{\text{rd}}$ paragraph of Section 4.3).

The derivations above yielded approximately 30% censoring. We also considered 60% censoring by specifying $P(\Delta_i = 0|X_{i1}) = \text{expit}(-1.8 + 2X_{i1})$, $P(\delta_{i2} = 0|\eta_{i2} = 1, X_{i2}) =$

$\text{expit}(1-2X_{i2})$ and $P(\delta_{i1}=1)=\text{expit}(-1.8+2X_{i1})/(1-0.1466\times 0.8)$ where $\mathbb{E}_{X_{i2}}[P(\delta_{i2}=0|\eta_2=1,X_{i2})]\approx 0.1466$.

C.3.2 Illustration of the Consistency and Double-robustness

In the article, the accuracy and precision of the blip parameter estimators were only presented with the data generating mechanism that yielded 30% independent censoring for a sample size of n=1000. Here, we present more detailed results on accuracy and precision of the blip estimators by considering dependent censoring, a higher proportion of censoring (60%) and different sample sizes (n=500, 1000, 10,000). The distribution of the blip parameter estimators is summarized with figures and tables for all sample sizes, censoring proportions and censoring dependencies. Each blip estimator corresponds to one panel in the figures. Common y-axes were kept across the different schemes to allow for a fair visual comparison. Each figure is followed by a table which describes the distribution of the blip estimators in terms of mean, standard error (SE), bias and root mean squared error (RMSE). The table below helps navigate through the results of this section.

		Censoring dependency		
% censoring	n	Independent	Dependent (baseline)	Dependent (time-varying)
Low (30%)	500	Table C.5 Figure C.1	Table C.10 Figure C.6	Table C.16 Figure C.12
	1000	Table C.4 Figure 4.1 (article)	Table C.11 Figure C.7	Table C.17 Figure C.13
	10,000	Table C.6 Figure C.2	Table C.12 Figure C.8	Table C.18 Figure C.14
High (60%)	500	Table C.7 Figure C.3	Table C.13 Figure C.9	Table C.19 Figure C.15
	1000	Table C.8 Figure C.4	Table C.14 Figure C.10	Table C.20 Figure C.16
	10,000	Table C.9 Figure C.5	Table C.15 Figure C.11	Table C.21 Figure C.17

Larger sample sizes and lower percentages of censoring yielded more precise estimators. The distribution of the blip estimators was comparable for independent and dependent (baseline or time-varying) censoring. The first stage estimators were more precise which was explained by the larger sample size in the first stage as not all individuals reached the second stage.

With independent censoring

In the article, Figure 4.1 illustrates the consistency and double-robustness of DWSurv with 30% independent censoring and n=1000. Table C.4 complements Figure 4.1 in the paper.

Table C.4: Mean, standard error, bias and root mean squared error of the blip estimators with DWSurv and the method by HNW with sample size n=1000 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 30% censoring, with both the probability of censoring and the censoring times independent of the survival times.

		dWSurv				HNW			
	Sc.	Mean	(SE)	Bias	RMSE	Mean	(SE)	Bias	RMSE
$\hat{\psi}_{10}$	1	0.10	(0.06)	3.40×10^{-3}	0.06	0.10	(0.06)	2.30×10^{-3}	0.06
	2	0.10	(0.06)	-1.97×10^{-4}	0.06	0.10	(0.06)	-1.06×10^{-4}	0.06
	3	0.10	(0.08)	3.82×10^{-3}	0.08	0.43	(0.08)	0.33	0.34
	4	0.43	(0.08)	0.33	0.34	0.43	(0.08)	0.33	0.34
$\hat{\psi}_{11}$	1	0.10	(0.08)	-3.44×10^{-3}	0.08	0.10	(0.08)	-3.45×10^{-3}	0.08
	2	0.10	(0.08)	-6.91×10^{-5}	0.08	0.10	(0.08)	-2.02×10^{-4}	0.08
	3	0.10	(0.12)	-2.64×10^{-3}	0.12	-0.38	(0.11)	-0.48	0.50
	4	-0.39	(0.12)	-0.49	0.50	-0.39	(0.12)	-0.49	0.50
$\hat{\psi}_{20}$	1	-0.90	(0.14)	-1.42×10^{-3}	0.14	-0.90	(0.14)	-2.43×10^{-3}	0.14
	2	-0.90	(0.14)	1.75×10^{-3}	0.14	-0.90	(0.14)	1.79×10^{-3}	0.14
	3	-0.91	(0.14)	-9.96×10^{-3}	0.14	-1.12	(0.14)	-0.22	0.26
	4	-1.12	(0.15)	-0.22	0.26	-1.12	(0.15)	-0.22	0.26
$\hat{\psi}_{21}$	1	0.60	(0.10)	1.60×10^{-3}	0.10	0.60	(0.10)	2.25×10^{-3}	0.10
	2	0.60	(0.10)	-1.01×10^{-3}	0.10	0.60	(0.10)	-1.04×10^{-3}	0.10
	3	0.61	(0.10)	7.61×10^{-3}	0.10	0.76	(0.10)	0.16	0.18
	4	0.75	(0.10)	0.15	0.18	0.75	(0.10)	0.15	0.18

True value of the parameters: $\psi_{10} = 0.1$, $\psi_{11} = 0.1$, $\psi_{20} = -0.9$, $\psi_{21} = 0.6$. Sc. = scenario, SE = standard error, RMSE = root mean squared error.

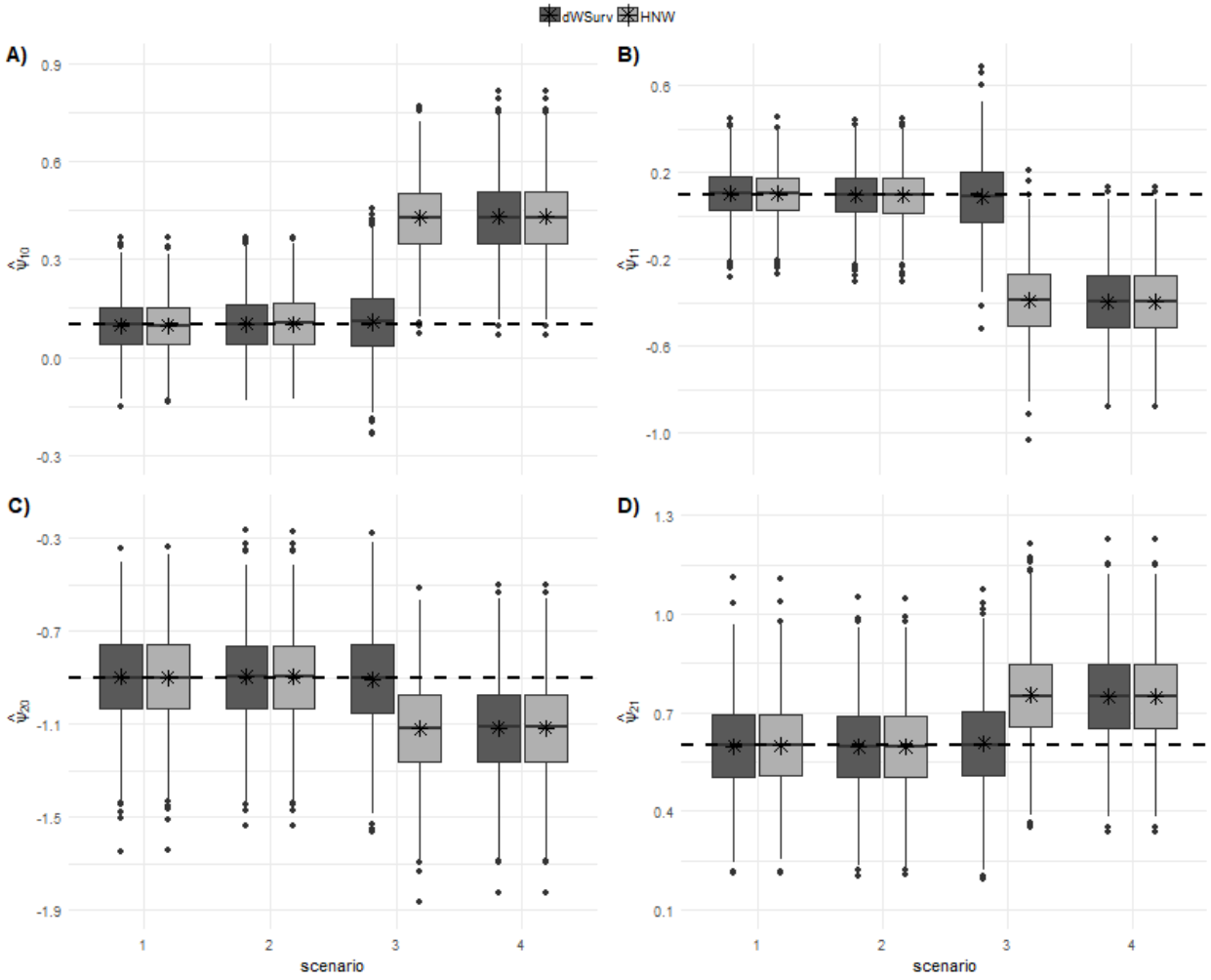

Figure C.1: Distribution of the blip parameter estimates in the first stage (upper row) and second stage (lower row) with DWSurv (dark grey) and the method by HNW (light grey) with sample size n=500 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 30% censoring, with both the probability of censoring and the censoring times independent of the survival times.

Table C.5: Mean, standard error, bias and root mean squared error of the blip estimators with DWSurv and the method by HNW with sample size n=500 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 30% censoring, with both the probability of censoring and the censoring times independent of the survival times.

		dWSurv				HNW			
	Sc.	Mean	(SE)	Bias	RMSE	Mean	(SE)	Bias	RMSE
$\hat{\psi}_{10}$	1	0.10	(0.09)	-2.01×10^{-3}	0.09	0.10	(0.08)	-3.19×10^{-3}	0.08
	2	0.10	(0.09)	2.80×10^{-3}	0.09	0.10	(0.09)	2.86×10^{-3}	0.09
	3	0.11	(0.11)	7.89×10^{-3}	0.11	0.43	(0.12)	0.33	0.35
	4	0.43	(0.12)	0.33	0.35	0.43	(0.12)	0.33	0.35
$\hat{\psi}_{11}$	1	0.10	(0.12)	1.08×10^{-3}	0.12	0.10	(0.11)	1.31×10^{-3}	0.11
	2	0.10	(0.12)	-3.37×10^{-3}	0.12	0.10	(0.12)	-3.45×10^{-3}	0.12
	3	0.09	(0.18)	-0.01	0.18	-0.39	(0.18)	-0.49	0.52
	4	-0.40	(0.17)	-0.50	0.53	-0.40	(0.17)	-0.50	0.53
$\hat{\psi}_{20}$	1	-0.90	(0.20)	2.92×10^{-3}	0.20	-0.90	(0.20)	1.53×10^{-3}	0.20
	2	-0.89	(0.20)	5.45×10^{-3}	0.20	-0.89	(0.20)	5.46×10^{-3}	0.20
	3	-0.91	(0.21)	-7.75×10^{-3}	0.21	-1.12	(0.21)	-0.22	0.30
	4	-1.12	(0.21)	-0.22	0.30	-1.12	(0.21)	-0.22	0.30
$\hat{\psi}_{21}$	1	0.60	(0.14)	-2.06×10^{-3}	0.14	0.60	(0.14)	-1.10×10^{-3}	0.14
	2	0.60	(0.13)	-3.33×10^{-3}	0.13	0.60	(0.13)	-3.33×10^{-3}	0.13
	3	0.61	(0.15)	6.70×10^{-3}	0.15	0.75	(0.14)	0.15	0.21
	4	0.75	(0.14)	0.15	0.20	0.75	(0.14)	0.15	0.20

True value of the parameters: $\psi_{10} = 0.1$, $\psi_{11} = 0.1$, $\psi_{20} = -0.9$, $\psi_{21} = 0.6$. Sc. = scenario, SE = standard error, RMSE = root mean squared error.

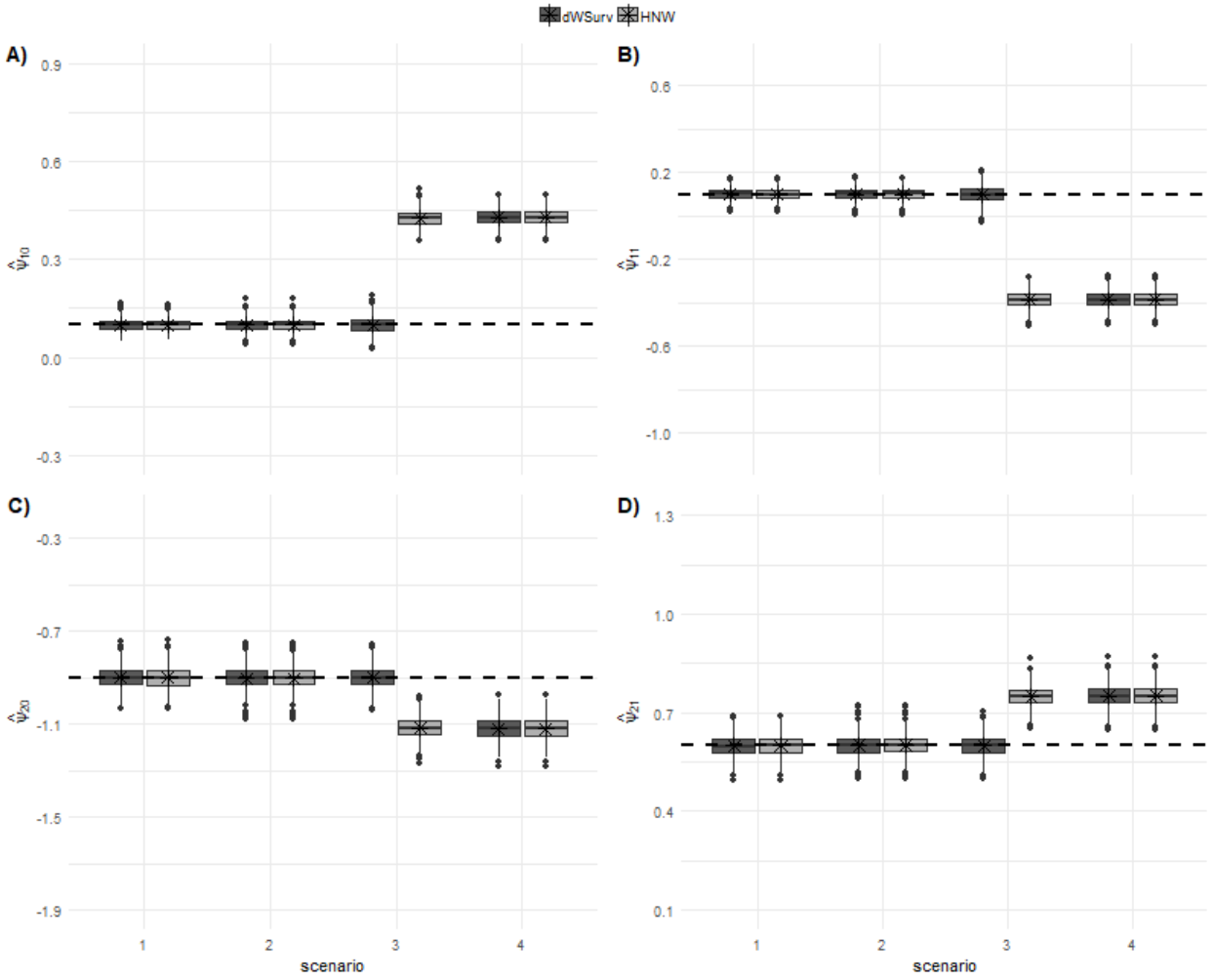

Figure C.2: Distribution of the blip parameter estimates in the first stage (upper row) and second stage (lower row) with DWSurv (dark grey) and the method by HNW (light grey) with sample size n=10,000 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 30% censoring, with both the probability of censoring and the censoring times independent of the survival times.

Table C.6: Mean, standard error, bias and root mean squared error of the blip estimators with DWSurv and the method by HNW with sample size n=10,000 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 30% censoring, with both the probability of censoring and the censoring times independent of the survival times.

		dWSurv				HNW			
	Sc.	Mean	(SE)	Bias	RMSE	Mean	(SE)	Bias	RMSE
$\hat{\psi}_{10}$	1	0.10	(0.02)	-4.99×10^{-4}	0.02	0.10	(0.02)	-1.27×10^{-3}	0.02
	2	0.10	(0.02)	-4.13×10^{-4}	0.02	0.10	(0.02)	-4.17×10^{-4}	0.02
	3	0.10	(0.02)	2.14×10^{-4}	0.02	0.43	(0.03)	0.33	0.33
	4	0.43	(0.02)	0.33	0.33	0.43	(0.02)	0.33	0.33
$\hat{\psi}_{11}$	1	0.10	(0.03)	1.18×10^{-3}	0.03	0.10	(0.02)	7.69×10^{-4}	0.02
	2	0.10	(0.03)	6.70×10^{-4}	0.03	0.10	(0.03)	6.77×10^{-4}	0.03
	3	0.10	(0.04)	-3.90×10^{-4}	0.04	-0.38	(0.04)	-0.48	0.49
	4	-0.39	(0.04)	-0.49	0.49	-0.39	(0.04)	-0.49	0.49
$\hat{\psi}_{20}$	1	-0.90	(0.05)	1.01×10^{-3}	0.05	-0.90	(0.05)	-4.17×10^{-4}	0.05
	2	-0.90	(0.04)	-1.75×10^{-3}	0.04	-0.90	(0.04)	-1.75×10^{-3}	0.04
	3	-0.90	(0.05)	1.19×10^{-3}	0.05	-1.12	(0.04)	-0.22	0.22
	4	-1.12	(0.05)	-0.22	0.22	-1.12	(0.05)	-0.22	0.22
$\hat{\psi}_{21}$	1	0.60	(0.03)	-6.72×10^{-4}	0.03	0.60	(0.03)	2.77×10^{-4}	0.03
	2	0.60	(0.03)	1.03×10^{-3}	0.03	0.60	(0.03)	1.04×10^{-3}	0.03
	3	0.60	(0.03)	-3.95×10^{-4}	0.03	0.75	(0.03)	0.15	0.15
	4	0.75	(0.03)	0.15	0.15	0.75	(0.03)	0.15	0.15

True value of the parameters: $\psi_{10} = 0.1$, $\psi_{11} = 0.1$, $\psi_{20} = -0.9$, $\psi_{21} = 0.6$. Sc. = scenario, SE = standard error, RMSE = root mean squared error.

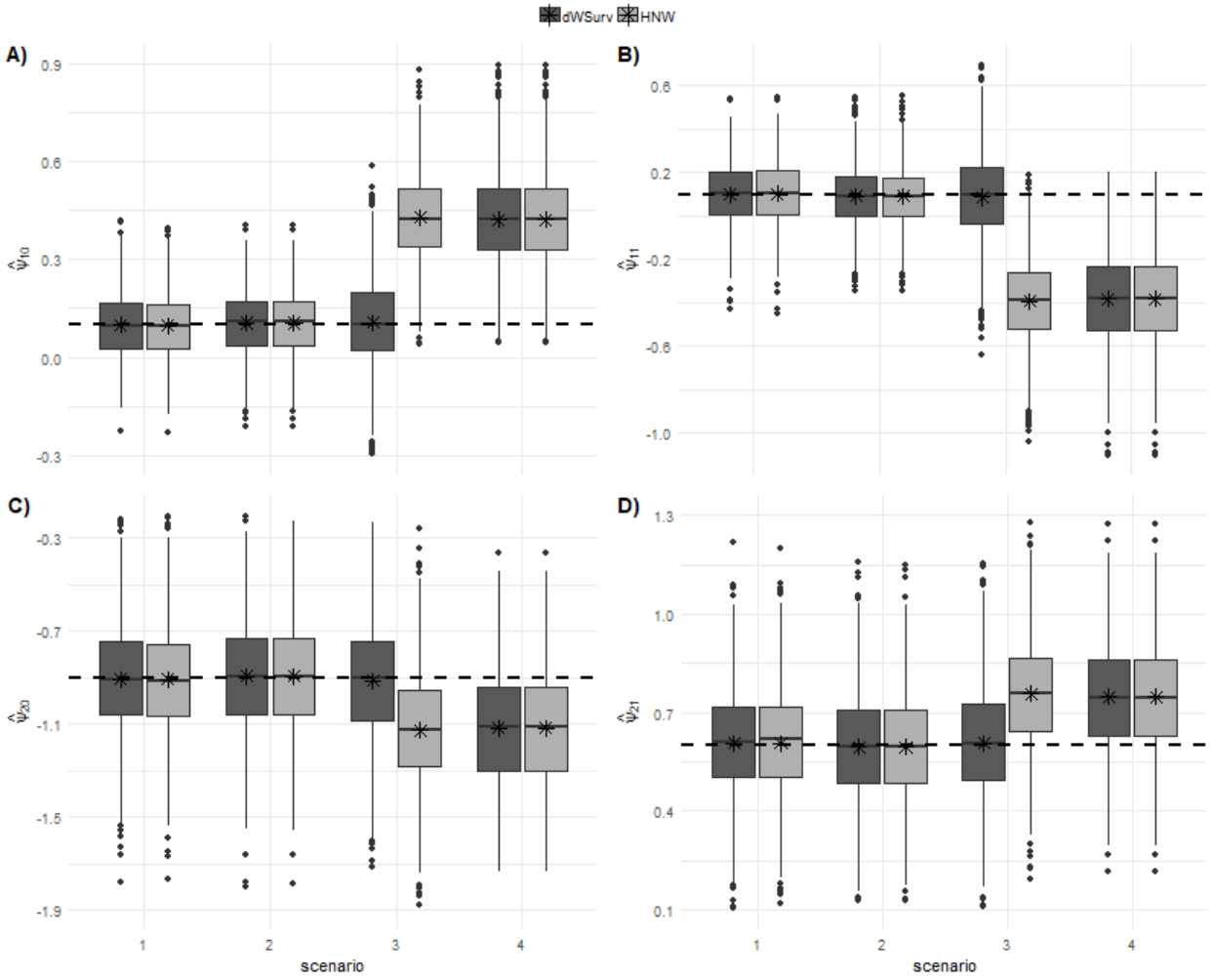

Figure C.3: Distribution of the blip parameter estimates in the first stage (upper row) and second stage (lower row) with DWSurv (dark grey) and the method by HNW (light grey) with sample size n=500 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 60% censoring, with both the probability of censoring and the censoring times independent of the survival times.

Table C.7: Mean, standard error, bias and root mean squared error of the blip estimators with DWSurv and the method by HNW with sample size n=500 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 60% censoring, with both the probability of censoring and the censoring times independent of the survival times.

		dWSurv				HNW			
	Sc.	Mean	(SE)	Bias	RMSE	Mean	(SE)	Bias	RMSE
$\hat{\psi}_{10}$	1	0.10	(0.10)	-2.26×10^{-4}	0.10	0.10	(0.10)	-3.42×10^{-3}	0.10
	2	0.11	(0.10)	5.66×10^{-3}	0.10	0.11	(0.10)	5.61×10^{-3}	0.10
	3	0.11	(0.14)	6.70×10^{-3}	0.14	0.43	(0.14)	0.33	0.36
	4	0.42	(0.14)	0.32	0.35	0.42	(0.14)	0.32	0.35
$\hat{\psi}_{11}$	1	0.10	(0.14)	2.15×10^{-4}	0.14	0.10	(0.14)	2.76×10^{-3}	0.14
	2	0.09	(0.14)	-7.87×10^{-3}	0.14	0.09	(0.14)	-7.77×10^{-3}	0.14
	3	0.09	(0.21)	-0.01	0.21	-0.39	(0.2)	-0.49	0.53
	4	-0.38	(0.21)	-0.48	0.53	-0.38	(0.21)	-0.48	0.53
$\hat{\psi}_{20}$	1	-0.90	(0.24)	-4.10×10^{-3}	0.24	-0.91	(0.24)	-5.87×10^{-3}	0.24
	2	-0.89	(0.25)	9.74×10^{-3}	0.25	-0.89	(0.25)	9.64×10^{-3}	0.25
	3	-0.91	(0.26)	-7.29×10^{-3}	0.26	-1.13	(0.25)	-0.23	0.34
	4	-1.12	(0.26)	-0.22	0.34	-1.12	(0.26)	-0.22	0.34
$\hat{\psi}_{21}$	1	0.60	(0.17)	4.99×10^{-3}	0.17	0.61	(0.17)	6.37×10^{-3}	0.17
	2	0.59	(0.17)	-5.42×10^{-3}	0.17	0.59	(0.17)	-5.37×10^{-3}	0.17
	3	0.61	(0.17)	6.06×10^{-3}	0.17	0.76	(0.17)	0.16	0.23
	4	0.75	(0.17)	0.15	0.23	0.75	(0.17)	0.15	0.23

True value of the parameters: $\psi_{10} = 0.1$, $\psi_{11} = 0.1$, $\psi_{20} = -0.9$, $\psi_{21} = 0.6$. Sc. = scenario, SE = standard error, RMSE = root mean squared error.

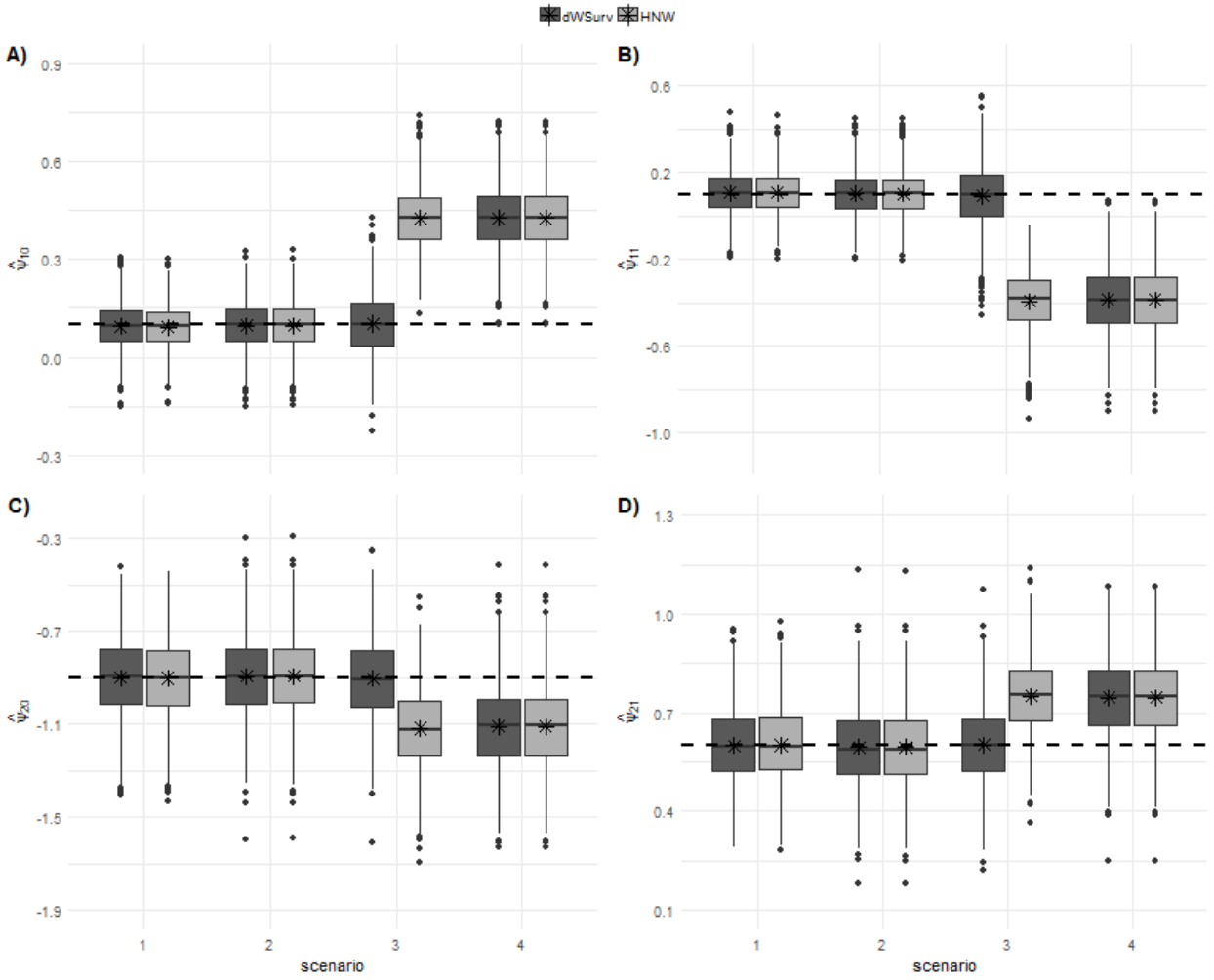

Figure C.4: Distribution of the blip parameter estimates in the first stage (upper row) and second stage (lower row) with DWSurv (dark grey) and the method by HNW (light grey) with sample size n=1000 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 60% censoring, with both the probability of censoring and the censoring times independent of the survival times.

Table C.8: Mean, standard error, bias and root mean squared error of the blip estimators with DWSurv and the method by HNW with sample size n=1000 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 60% censoring, with both the probability of censoring and the censoring times independent of the survival times

		dWSurv				HNW			
	Sc.	Mean	(SE)	Bias	RMSE	Mean	(SE)	Bias	RMSE
$\hat{\psi}_{10}$	1	0.10	(0.07)	-5.00×10^{-3}	0.07	0.09	(0.07)	-6.75×10^{-3}	0.07
	2	0.10	(0.07)	-2.15×10^{-3}	0.07	0.10	(0.07)	-2.12×10^{-3}	0.07
	3	0.10	(0.10)	3.28×10^{-3}	0.10	0.43	(0.09)	0.33	0.34
	4	0.43	(0.10)	0.33	0.34	0.43	(0.10)	0.33	0.34
$\hat{\psi}_{11}$	1	0.11	(0.10)	6.02×10^{-3}	0.10	0.11	(0.09)	6.17×10^{-3}	0.09
	2	0.10	(0.10)	2.35×10^{-3}	0.10	0.10	(0.10)	2.30×10^{-3}	0.10
	3	0.09	(0.15)	-7.75×10^{-3}	0.15	-0.39	(0.14)	-0.49	0.51
	4	-0.39	(0.15)	-0.49	0.51	-0.39	(0.15)	-0.49	0.51
$\hat{\psi}_{20}$	1	-0.90	(0.17)	1.49×10^{-3}	0.17	-0.90	(0.17)	-9.81×10^{-4}	0.17
	2	-0.89	(0.18)	7.46×10^{-3}	0.18	-0.89	(0.18)	7.38×10^{-3}	0.18
	3	-0.90	(0.18)	-2.95×10^{-3}	0.18	-1.12	(0.18)	-0.22	0.28
	4	-1.11	(0.18)	-0.21	0.27	-1.11	(0.18)	-0.21	0.27
$\hat{\psi}_{21}$	1	0.60	(0.11)	-5.90×10^{-4}	0.11	0.60	(0.12)	1.01×10^{-3}	0.12
	2	0.59	(0.12)	-5.44×10^{-3}	0.12	0.59	(0.12)	-5.38×10^{-3}	0.12
	3	0.60	(0.12)	2.27×10^{-3}	0.12	0.75	(0.12)	0.15	0.19
	4	0.75	(0.12)	0.15	0.19	0.75	(0.12)	0.15	0.19

True value of the parameters: $\psi_{10} = 0.1$, $\psi_{11} = 0.1$, $\psi_{20} = -0.9$, $\psi_{21} = 0.6$. Sc. = scenario, SE = standard error, RMSE = root mean squared error.

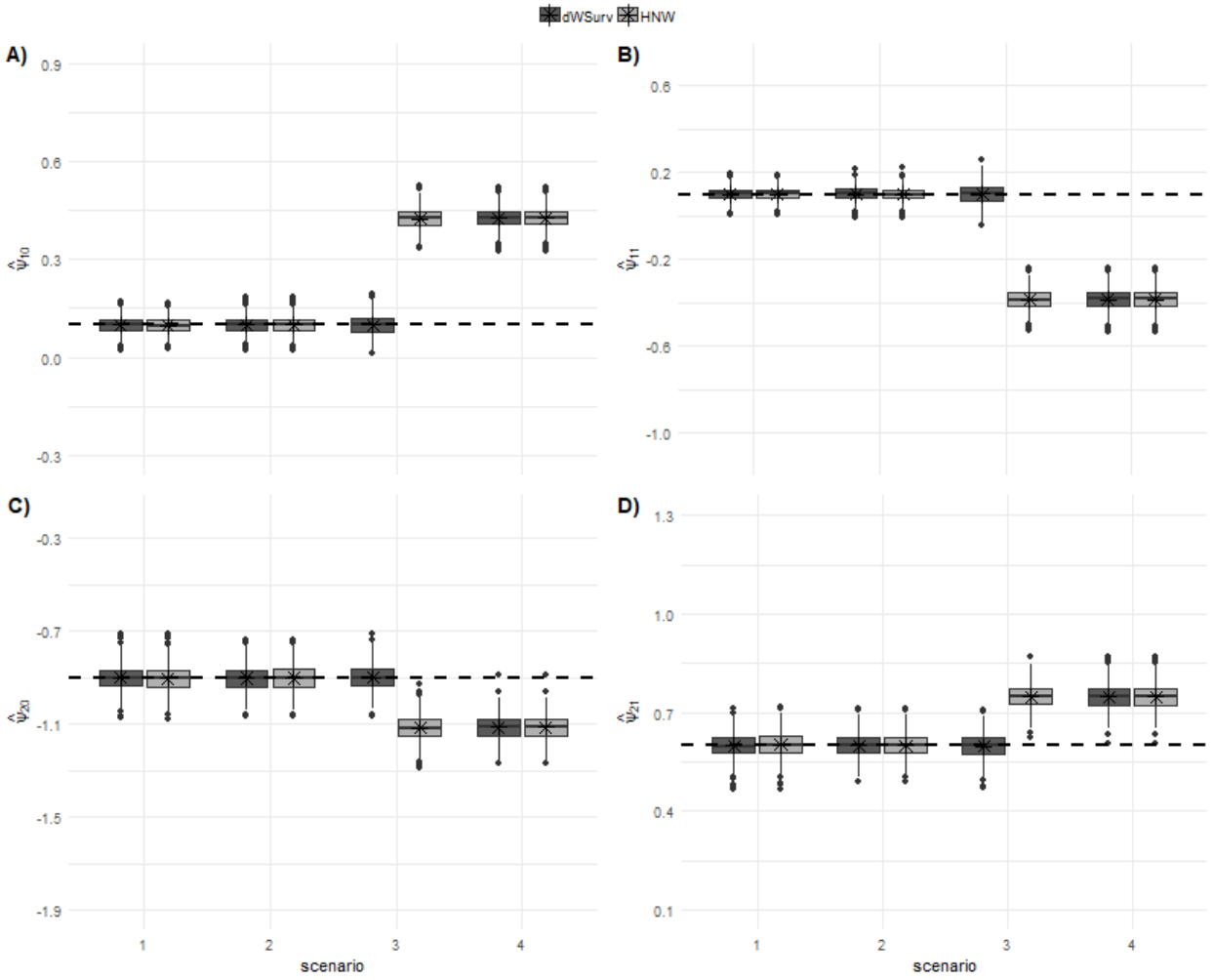

Figure C.5: Distribution of the blip parameter estimates in the first stage (upper row) and second stage (lower row) with DWSurv (dark grey) and the method by HNW (light grey) with sample size n=10,000 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 60% censoring, with both the probability of censoring and the censoring times independent of the survival times.

Table C.9: Mean, standard error, bias and root mean squared error of the blip estimators with DWSurv and the method by HNW with sample size n=10,000 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 60% censoring, with both the probability of censoring and the censoring times independent of the survival times.

		dWSurv				HNW			
	Sc.	Mean	(SE)	Bias	RMSE	Mean	(SE)	Bias	RMSE
$\hat{\psi}_{10}$	1	0.10	(0.02)	-6.25×10^{-4}	0.02	0.10	(0.02)	-2.24×10^{-3}	0.02
	2	0.10	(0.02)	-6.16×10^{-4}	0.02	0.10	(0.02)	-6.15×10^{-4}	0.02
	3	0.10	(0.03)	1.43×10^{-4}	0.03	0.43	(0.03)	0.33	0.33
	4	0.43	(0.03)	0.33	0.33	0.43	(0.03)	0.33	0.33
$\hat{\psi}_{11}$	1	0.10	(0.03)	6.67×10^{-4}	0.03	0.10	(0.03)	3.16×10^{-4}	0.03
	2	0.10	(0.03)	7.97×10^{-4}	0.03	0.10	(0.03)	7.96×10^{-4}	0.03
	3	0.10	(0.05)	1.04×10^{-3}	0.05	-0.38	(0.04)	-0.48	0.49
	4	-0.38	(0.05)	-0.48	0.49	-0.38	(0.05)	-0.48	0.49
$\hat{\psi}_{20}$	1	-0.90	(0.05)	2.58×10^{-4}	0.05	-0.90	(0.05)	-2.82×10^{-3}	0.05
	2	-0.90	(0.05)	-9.52×10^{-4}	0.05	-0.90	(0.05)	-9.43×10^{-4}	0.05
	3	-0.90	(0.06)	2.02×10^{-3}	0.06	-1.12	(0.06)	-0.22	0.22
	4	-1.11	(0.05)	-0.21	0.22	-1.11	(0.05)	-0.21	0.22
$\hat{\psi}_{21}$	1	0.60	(0.04)	-4.55×10^{-4}	0.04	0.60	(0.04)	1.58×10^{-3}	0.04
	2	0.60	(0.04)	9.15×10^{-4}	0.04	0.60	(0.04)	9.08×10^{-4}	0.04
	3	0.60	(0.04)	-1.28×10^{-3}	0.04	0.75	(0.04)	0.15	0.15
	4	0.75	(0.04)	0.15	0.15	0.75	(0.04)	0.15	0.15

True value of the parameters: $\psi_{10} = 0.1$, $\psi_{11} = 0.1$, $\psi_{20} = -0.9$, $\psi_{21} = 0.6$. Sc. = scenario, SE = standard error, RMSE = root mean squared error.

With censoring dependent on baseline covariates

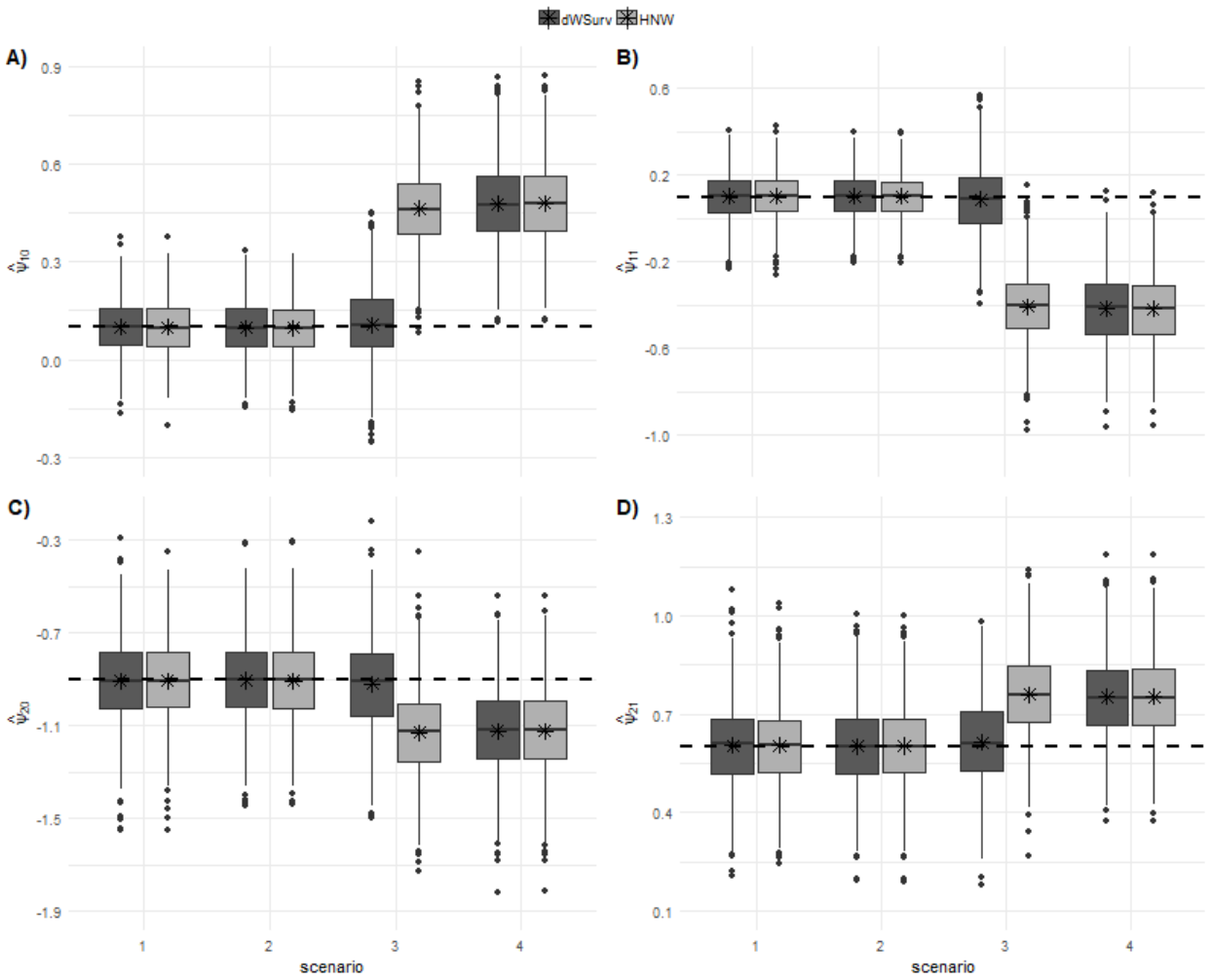

Figure C.6: Distribution of the blip parameter estimates in the first stage (upper row) and second stage (lower row) with DWSurv (dark grey) and the method by HNW (light grey) with sample size n=500 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 30% censoring, with both the probability of censoring and the censoring times conditionnally independent of the survival times given baseline covariates.

Table C.10: Mean, standard error, bias and root mean squared error of the blip estimators with DWSurv and the method by HNW with sample size n=500 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 30% censoring, with both the probability of censoring and the censoring times conditionally independent of the survival times given baseline covariates.

		dWSurv				HNW			
	Sc.	Mean	(SE)	Bias	RMSE	Mean	(SE)	Bias	RMSE
$\hat{\psi}_{10}$	1	0.10	(0.08)	-5.09×10^{-5}	0.08	0.10	(0.08)	-2.39×10^{-5}	0.08
	2	0.10	(0.08)	-2.13×10^{-3}	0.08	0.10	(0.08)	-1.87×10^{-3}	0.08
	3	0.11	(0.11)	7.21×10^{-3}	0.11	0.46	(0.12)	0.36	0.38
	4	0.48	(0.13)	0.38	0.40	0.48	(0.13)	0.38	0.40
$\hat{\psi}_{11}$	1	0.10	(0.11)	1.78×10^{-3}	0.11	0.10	(0.11)	1.72×10^{-3}	0.11
	2	0.10	(0.10)	3.20×10^{-3}	0.10	0.10	(0.10)	3.38×10^{-3}	0.10
	3	0.09	(0.15)	-0.01	0.15	-0.40	(0.16)	-0.50	0.53
	4	-0.42	(0.17)	-0.52	0.54	-0.42	(0.17)	-0.52	0.54
$\hat{\psi}_{20}$	1	-0.91	(0.19)	-5.33×10^{-3}	0.19	-0.91	(0.18)	-5.15×10^{-3}	0.18
	2	-0.90	(0.18)	-4.26×10^{-3}	0.18	-0.90	(0.18)	-4.93×10^{-3}	0.18
	3	-0.92	(0.19)	-0.02	0.19	-1.13	(0.19)	-0.23	0.30
	4	-1.12	(0.18)	-0.22	0.28	-1.12	(0.18)	-0.22	0.28
$\hat{\psi}_{21}$	1	0.61	(0.13)	5.07×10^{-3}	0.13	0.60	(0.12)	4.69×10^{-3}	0.12
	2	0.60	(0.13)	3.15×10^{-3}	0.13	0.60	(0.13)	3.55×10^{-3}	0.13
	3	0.61	(0.13)	0.01	0.13	0.76	(0.13)	0.16	0.20
	4	0.75	(0.12)	0.15	0.19	0.75	(0.12)	0.15	0.19

True value of the parameters: $\psi_{10} = 0.1$, $\psi_{11} = 0.1$, $\psi_{20} = -0.9$, $\psi_{21} = 0.6$. Sc. = scenario, SE = standard error, RMSE = root mean squared error.

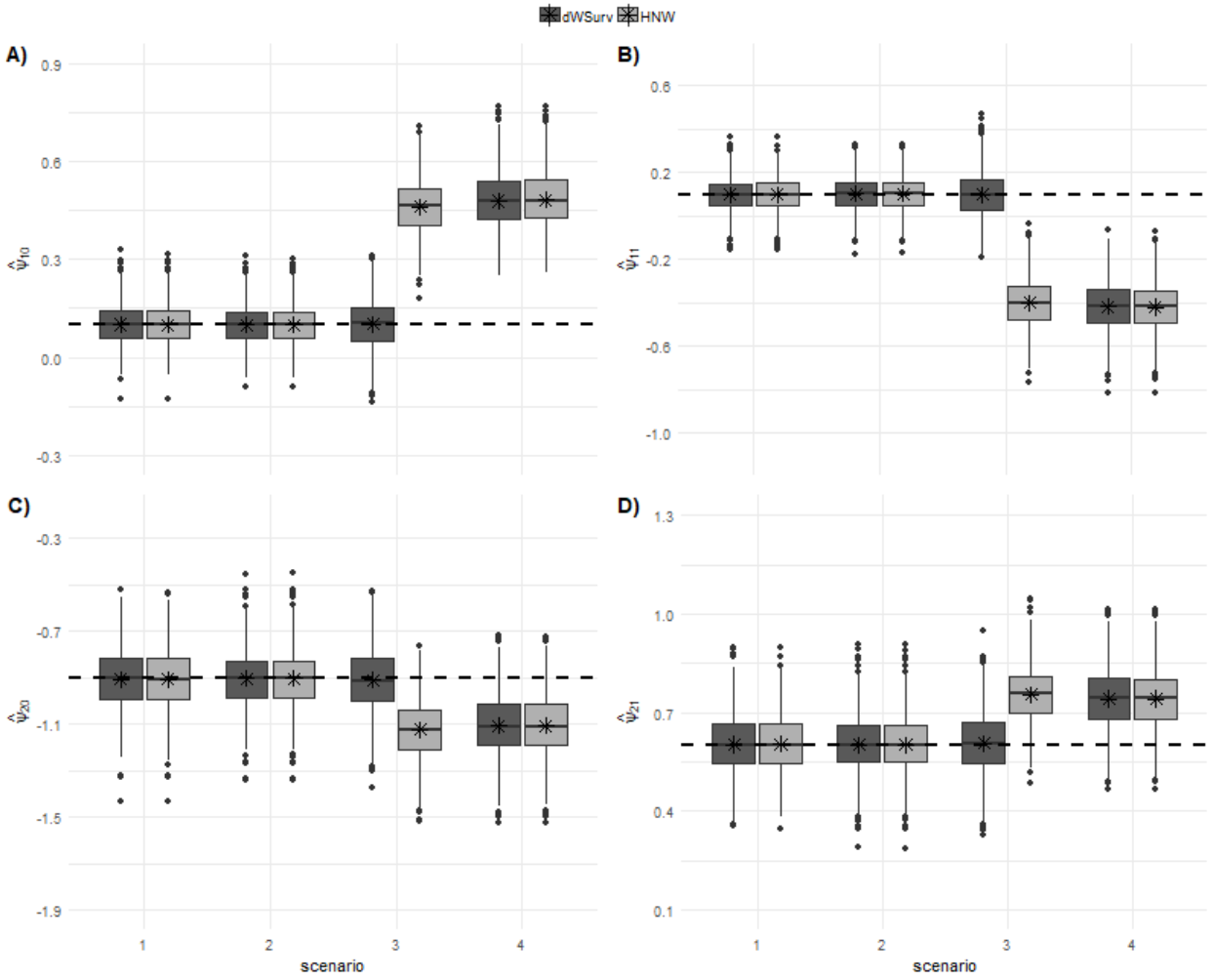

Figure C.7: Distribution of the blip parameter estimates in the first stage (upper row) and second stage (lower row) with DWSurv (dark grey) and the method by HNW (light grey) with sample size n=1000 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 30% censoring, with both the probability of censoring and the censoring times conditionnally independent of the survival times given baseline covariates.

Table C.11: Mean, standard error, bias and root mean squared error of the blip estimators with DWSurv and the method by HNW with sample size n=1000 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 30% censoring, with both the probability of censoring and the censoring times conditionally independent of the survival times given baseline covariates.

		dWSurv				HNW			
	Sc.	Mean	(SE)	Bias	RMSE	Mean	(SE)	Bias	RMSE
$\hat{\psi}_{10}$	1	0.10	(0.06)	7.14×10^{-4}	0.06	0.10	(0.06)	4.02×10^{-4}	0.06
	2	0.10	(0.06)	-1.25×10^{-3}	0.06	0.10	(0.06)	-1.15×10^{-3}	0.06
	3	0.10	(0.08)	1.17×10^{-3}	0.08	0.46	(0.08)	0.36	0.37
	4	0.48	(0.09)	0.38	0.39	0.48	(0.09)	0.38	0.39
$\hat{\psi}_{11}$	1	0.10	(0.08)	-1.86×10^{-3}	0.08	0.10	(0.08)	-1.46×10^{-3}	0.08
	2	0.10	(0.08)	3.29×10^{-3}	0.08	0.10	(0.08)	3.91×10^{-3}	0.08
	3	0.10	(0.11)	-9.37×10^{-5}	0.11	-0.40	(0.11)	-0.50	0.51
	4	-0.42	(0.12)	-0.52	0.53	-0.42	(0.12)	-0.52	0.53
$\hat{\psi}_{20}$	1	-0.91	(0.13)	-5.22×10^{-3}	0.13	-0.91	(0.13)	-5.54×10^{-3}	0.13
	2	-0.90	(0.13)	-3.20×10^{-3}	0.13	-0.90	(0.13)	-3.83×10^{-3}	0.13
	3	-0.91	(0.14)	-8.93×10^{-3}	0.14	-1.12	(0.13)	-0.22	0.26
	4	-1.11	(0.13)	-0.21	0.25	-1.11	(0.13)	-0.21	0.25
$\hat{\psi}_{21}$	1	0.60	(0.09)	3.49×10^{-3}	0.09	0.60	(0.09)	3.76×10^{-3}	0.09
	2	0.60	(0.09)	2.46×10^{-3}	0.09	0.60	(0.09)	2.88×10^{-3}	0.09
	3	0.61	(0.09)	6.90×10^{-3}	0.09	0.75	(0.09)	0.15	0.18
	4	0.74	(0.09)	0.14	0.17	0.74	(0.09)	0.14	0.17

True value of the parameters: $\psi_{10} = 0.1$, $\psi_{11} = 0.1$, $\psi_{20} = -0.9$, $\psi_{21} = 0.6$. Sc. = scenario, SE = standard error, RMSE = root mean squared error.

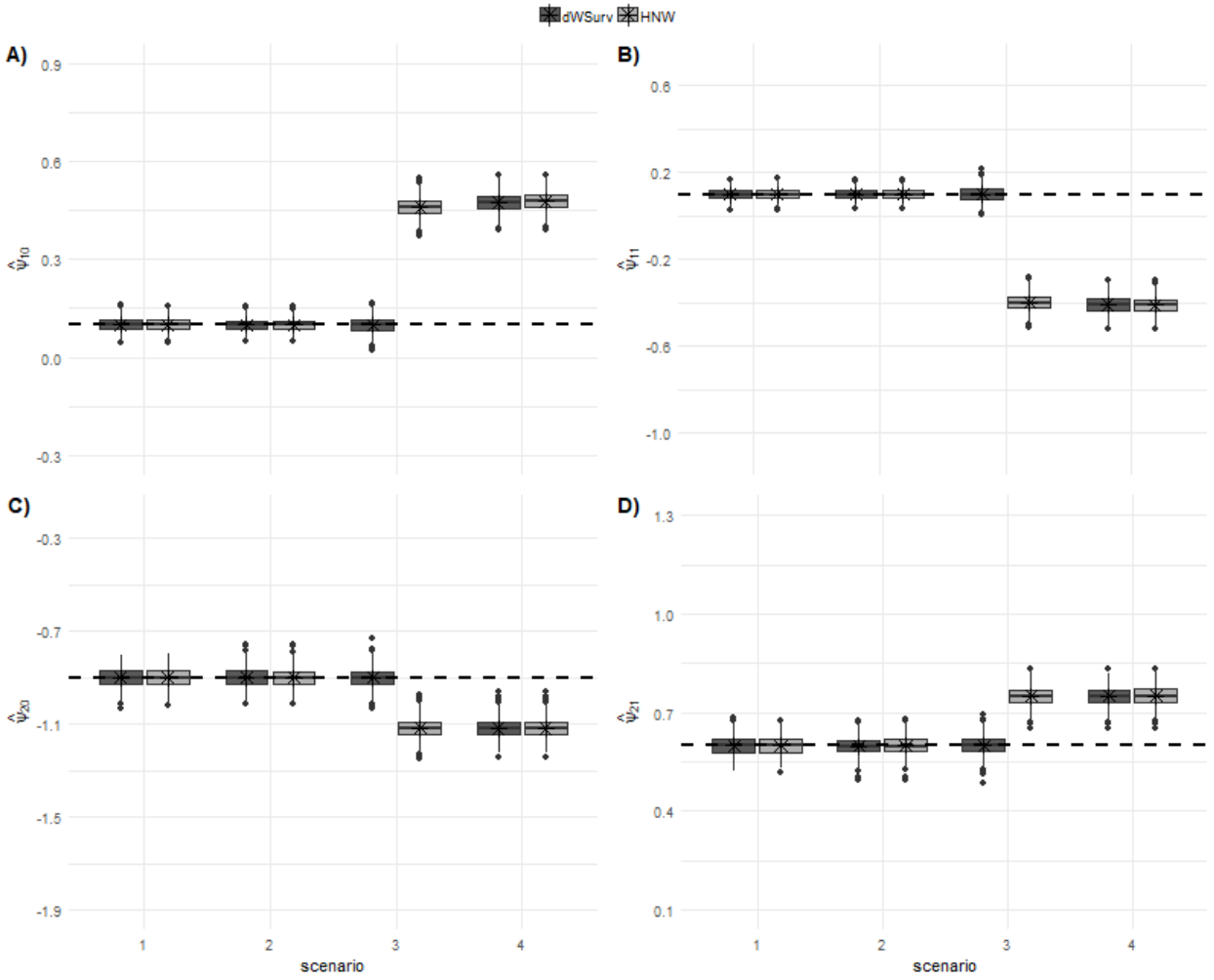

Figure C.8: Distribution of the blip parameter estimates in the first stage (upper row) and second stage (lower row) with DWSurv (dark grey) and the method by HNW (light grey) with sample size n=10,000 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 30% censoring, with both the probability of censoring and the censoring times conditionnally independent of the survival times given baseline covariates.

Table C.12: Mean, standard error, bias and root mean squared error of the blip estimators with DWSurv and the method by HNW with sample size n=10,000 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 30% censoring, with both the probability of censoring and the censoring times conditionally independent of the survival times given baseline covariates.

		dWSurv				HNW			
	Sc.	Mean	(SE)	Bias	RMSE	Mean	(SE)	Bias	RMSE
$\hat{\psi}_{10}$	1	0.10	(0.02)	-4.86×10^{-4}	0.02	0.10	(0.02)	-2.99×10^{-4}	0.02
	2	0.10	(0.02)	1.02×10^{-4}	0.02	0.10	(0.02)	5.97×10^{-4}	0.02
	3	0.10	(0.02)	-7.49×10^{-4}	0.02	0.46	(0.03)	0.36	0.36
	4	0.47	(0.03)	0.37	0.38	0.48	(0.03)	0.38	0.38
$\hat{\psi}_{11}$	1	0.10	(0.02)	2.67×10^{-4}	0.02	0.10	(0.02)	1.33×10^{-5}	0.02
	2	0.10	(0.02)	-2.55×10^{-4}	0.02	0.10	(0.02)	-1.18×10^{-4}	0.02
	3	0.10	(0.03)	6.81×10^{-4}	0.03	-0.40	(0.04)	-0.50	0.50
	4	-0.41	(0.04)	-0.51	0.51	-0.41	(0.04)	-0.51	0.51
$\hat{\psi}_{20}$	1	-0.90	(0.04)	5.27×10^{-4}	0.04	-0.90	(0.04)	5.94×10^{-4}	0.04
	2	-0.90	(0.04)	1.59×10^{-3}	0.04	-0.90	(0.04)	9.25×10^{-4}	0.04
	3	-0.90	(0.04)	-4.98×10^{-4}	0.04	-1.12	(0.04)	-0.22	0.22
	4	-1.12	(0.04)	-0.22	0.22	-1.12	(0.04)	-0.22	0.22
$\hat{\psi}_{21}$	1	0.60	(0.03)	-2.59×10^{-4}	0.03	0.60	(0.03)	-3.42×10^{-4}	0.03
	2	0.60	(0.03)	-1.34×10^{-3}	0.03	0.60	(0.03)	-8.87×10^{-4}	0.03
	3	0.60	(0.03)	5.55×10^{-4}	0.03	0.75	(0.03)	0.15	0.15
	4	0.75	(0.03)	0.15	0.15	0.75	(0.03)	0.15	0.15

True value of the parameters: $\psi_{10} = 0.1$, $\psi_{11} = 0.1$, $\psi_{20} = -0.9$, $\psi_{21} = 0.6$. Sc. = scenario, SE = standard error, RMSE = root mean squared error.

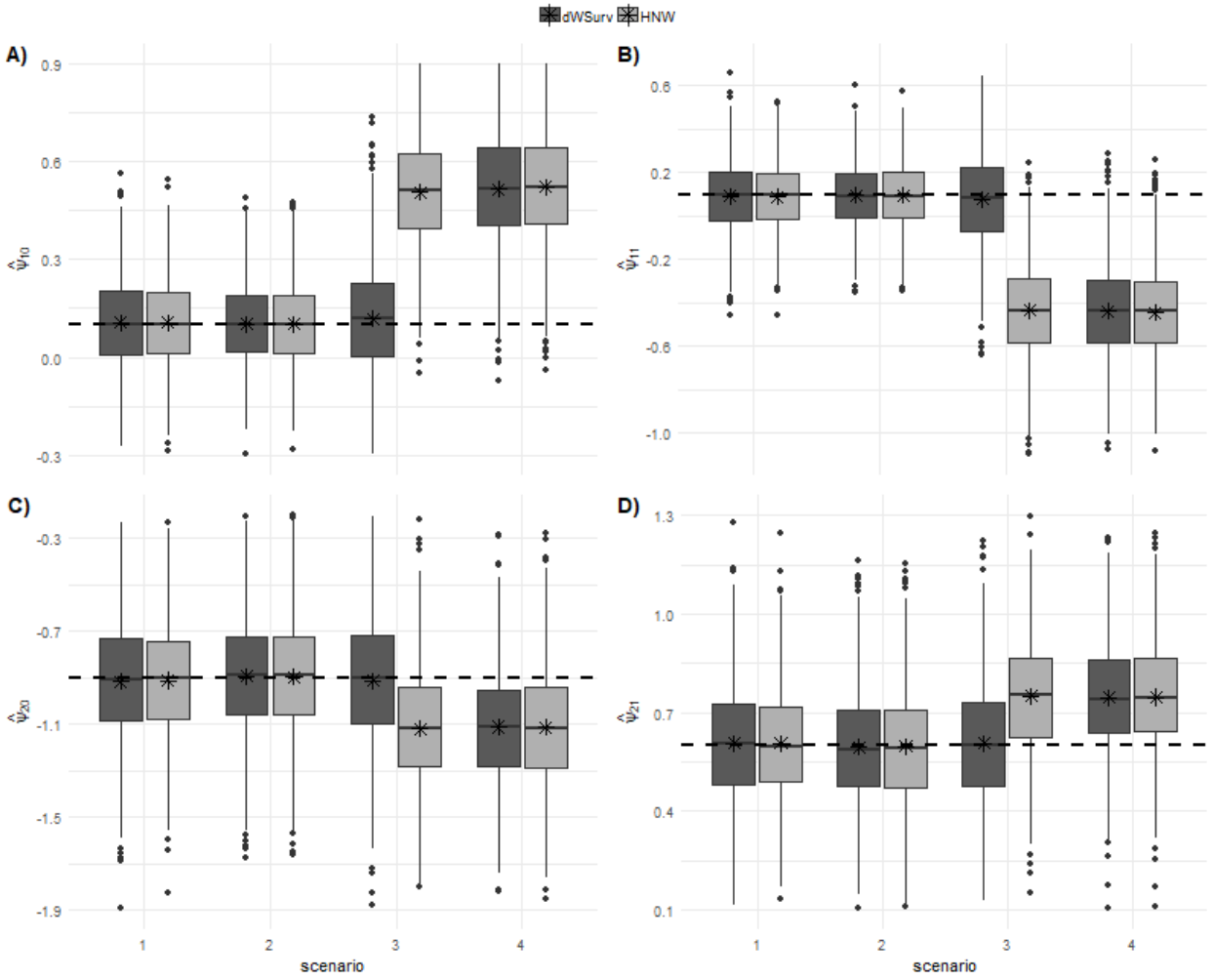

Figure C.9: Distribution of the blip parameter estimates in the first stage (upper row) and second stage (lower row) with DWSurv (dark grey) and the method by HNW (light grey) with sample size n=500 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 60% censoring, with both the probability of censoring and the censoring times conditionnally independent of the survival times given baseline covariates.

Table C.13: Mean, standard error, bias and root mean squared error of the blip estimators with DWSurv and the method by HNW with sample size n=500 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 60% censoring, with both the probability of censoring and the censoring times conditionally independent of the survival times given baseline covariates.

		dWSurv				HNW			
	Sc.	Mean	(SE)	Bias	RMSE	Mean	(SE)	Bias	RMSE
$\hat{\psi}_{10}$	1	0.11	(0.14)	5.14×10^{-3}	0.14	0.11	(0.13)	6.51×10^{-3}	0.13
	2	0.10	(0.13)	1.60×10^{-3}	0.12	0.10	(0.13)	2.46×10^{-3}	0.13
	3	0.12	(0.17)	0.02	0.17	0.52	(0.18)	0.42	0.45
	4	0.53	(0.19)	0.43	0.47	0.54	(0.18)	0.44	0.47
$\hat{\psi}_{11}$	1	0.09	(0.17)	-9.00×10^{-3}	0.17	0.09	(0.16)	-0.01	0.16
	2	0.09	(0.15)	-5.35×10^{-3}	0.15	0.10	(0.15)	-4.59×10^{-3}	0.15
	3	0.08	(0.22)	-0.02	0.22	-0.44	(0.22)	-0.54	0.58
	4	-0.44	(0.23)	-0.54	0.59	-0.45	(0.22)	-0.55	0.59
$\hat{\psi}_{20}$	1	-0.91	(0.27)	-0.01	0.27	-0.91	(0.25)	-9.43×10^{-3}	0.25
	2	-0.89	(0.25)	8.29×10^{-3}	0.25	-0.89	(0.25)	5.92×10^{-3}	0.25
	3	-0.91	(0.28)	-7.72×10^{-3}	0.28	-1.12	(0.26)	-0.22	0.34
	4	-1.11	(0.25)	-0.21	0.33	-1.11	(0.25)	-0.21	0.33
$\hat{\psi}_{21}$	1	0.61	(0.18)	6.10×10^{-3}	0.18	0.61	(0.17)	5.63×10^{-3}	0.17
	2	0.59	(0.17)	-6.23×10^{-3}	0.17	0.60	(0.17)	-4.83×10^{-3}	0.17
	3	0.60	(0.19)	4.05×10^{-3}	0.19	0.75	(0.17)	0.15	0.23
	4	0.75	(0.17)	0.15	0.22	0.75	(0.17)	0.15	0.23

True value of the parameters: $\psi_{10} = 0.1$, $\psi_{11} = 0.1$, $\psi_{20} = -0.9$, $\psi_{21} = 0.6$. Sc. = scenario, SE = standard error, RMSE = root mean squared error.

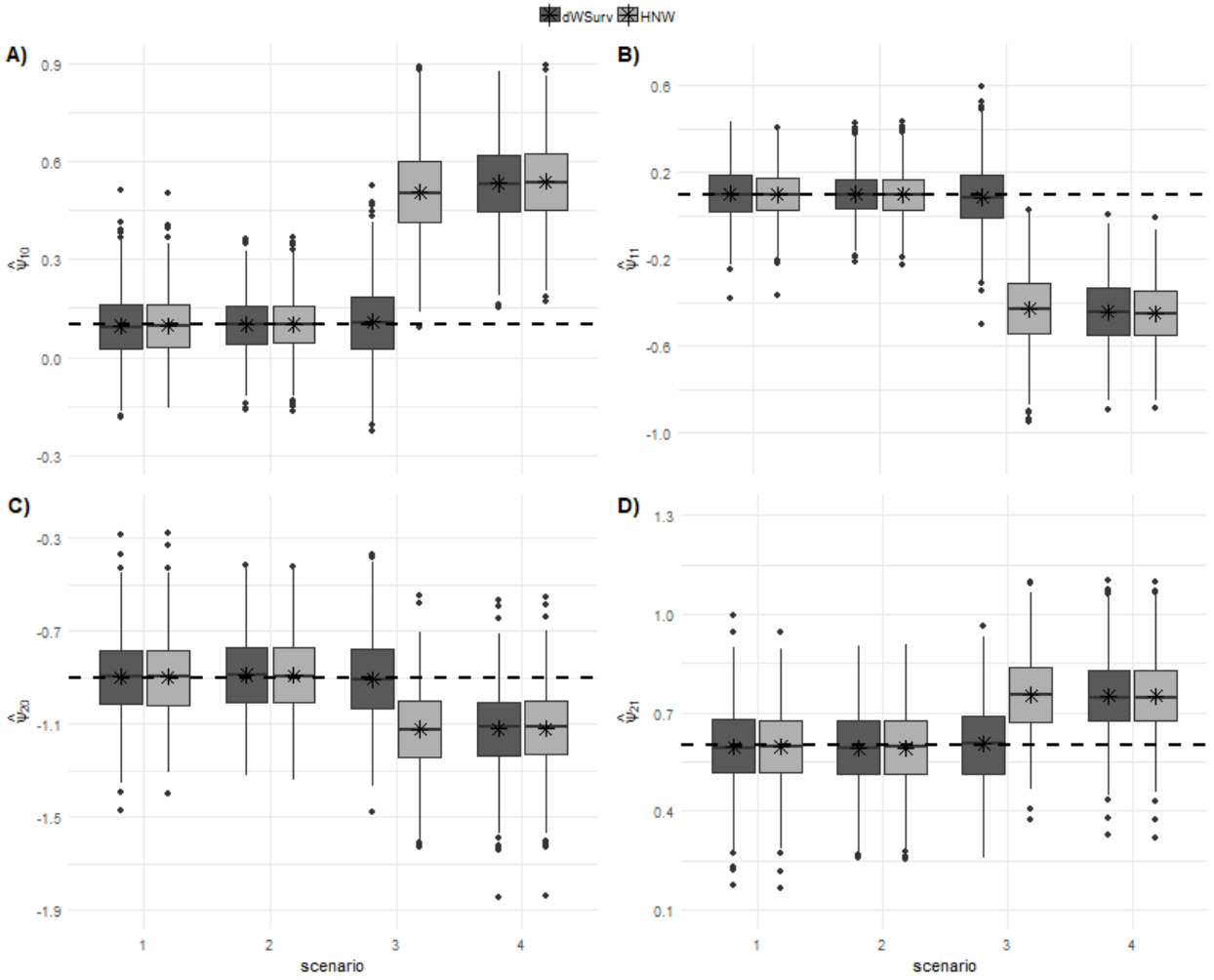

Figure C.10: Distribution of the blip parameter estimates in the first stage (upper row) and second stage (lower row) with DWSurv (dark grey) and the method by HNW (light grey) with sample size n=1000 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 60% censoring, with both the probability of censoring and the censoring times conditionnally independent of the survival times given baseline covariates.

Table C.14: Mean, standard error, bias and root mean squared error of the blip estimators with DWSurv and the method by HNW with sample size n=1000 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 60% censoring, with both the probability of censoring and the censoring times conditionally independent of the survival times given baseline covariates.

		dWSurv				HNW			
	Sc.	Mean	(SE)	Bias	RMSE	Mean	(SE)	Bias	RMSE
$\hat{\psi}_{10}$	1	0.10	(0.10)	-3.92×10^{-3}	0.10	0.10	(0.09)	-2.25×10^{-3}	0.09
	2	0.10	(0.09)	-5.42×10^{-4}	0.09	0.10	(0.09)	1.14×10^{-3}	0.09
	3	0.11	(0.12)	8.85×10^{-3}	0.12	0.51	(0.14)	0.41	0.43
	4	0.53	(0.13)	0.43	0.45	0.54	(0.12)	0.44	0.46
$\hat{\psi}_{11}$	1	0.10	(0.12)	1.98×10^{-3}	0.12	0.10	(0.11)	-3.95×10^{-4}	0.11
	2	0.10	(0.10)	9.63×10^{-4}	0.10	0.10	(0.10)	5.62×10^{-4}	0.10
	3	0.09	(0.15)	-0.01	0.15	-0.43	(0.16)	-0.53	0.55
	4	-0.44	(0.15)	-0.54	0.56	-0.45	(0.15)	-0.55	0.57
$\hat{\psi}_{20}$	1	-0.90	(0.17)	3.44×10^{-3}	0.17	-0.9	(0.17)	3.10×10^{-3}	0.17
	2	-0.89	(0.17)	0.01	0.17	-0.89	(0.17)	0.01	0.17
	3	-0.91	(0.19)	-7.04×10^{-3}	0.19	-1.12	(0.18)	-0.22	0.28
	4	-1.12	(0.17)	-0.22	0.28	-1.12	(0.17)	-0.22	0.28
$\hat{\psi}_{21}$	1	0.60	(0.12)	-3.74×10^{-3}	0.12	0.60	(0.11)	-3.55×10^{-3}	0.11
	2	0.59	(0.12)	-9.26×10^{-3}	0.12	0.59	(0.12)	-8.32×10^{-3}	0.12
	3	0.61	(0.13)	5.89×10^{-3}	0.13	0.75	(0.12)	0.15	0.19
	4	0.75	(0.12)	0.15	0.19	0.75	(0.12)	0.15	0.19

True value of the parameters: $\psi_{10} = 0.1$, $\psi_{11} = 0.1$, $\psi_{20} = -0.9$, $\psi_{21} = 0.6$. Sc. = scenario, SE = standard error, RMSE = root mean squared error.

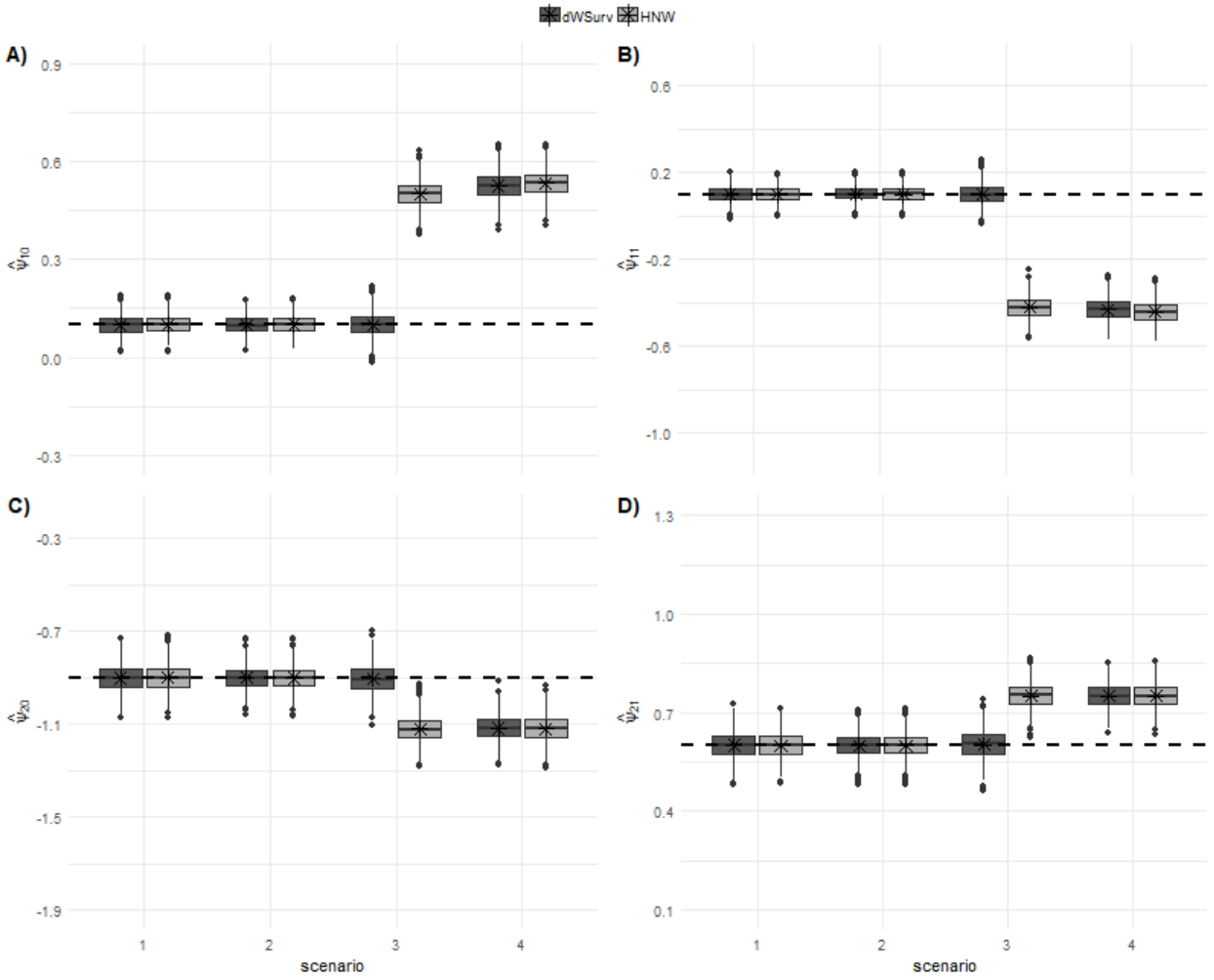

Figure C.11: Distribution of the blip parameter estimates in the first stage (upper row) and second stage (lower row) with DWSurv (dark grey) and the method by HNW (light grey) with sample size n=10,000 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 60% censoring, with both the probability of censoring and the censoring times conditionnally independent of the survival times given baseline covariates.

Table C.15: Mean, standard error, bias and root mean squared error of the blip estimators with DWSurv and the method by HNW with sample size n=10,000 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 60% censoring, with both the probability of censoring and the censoring times conditionally independent of the survival times given baseline covariates.

		dWSurv				HNW			
	Sc.	Mean	(SE)	Bias	RMSE	Mean	(SE)	Bias	RMSE
$\hat{\psi}_{10}$	1	0.10	(0.03)	-7.68×10^{-4}	0.03	0.10	(0.03)	-7.21×10^{-4}	0.03
	2	0.10	(0.03)	-1.23×10^{-3}	0.03	0.10	(0.03)	8.41×10^{-4}	0.03
	3	0.10	(0.04)	5.63×10^{-4}	0.04	0.50	(0.04)	0.40	0.40
	4	0.53	(0.04)	0.43	0.43	0.53	(0.04)	0.43	0.44
$\hat{\psi}_{11}$	1	0.10	(0.03)	7.24×10^{-4}	0.03	0.10	(0.03)	5.81×10^{-4}	0.03
	2	0.10	(0.03)	1.32×10^{-3}	0.03	0.10	(0.03)	7.52×10^{-4}	0.03
	3	0.10	(0.05)	-1.99×10^{-3}	0.05	-0.42	(0.05)	-0.52	0.52
	4	-0.43	(0.05)	-0.53	0.53	-0.44	(0.05)	-0.54	0.54
$\hat{\psi}_{20}$	1	-0.90	(0.06)	-1.39×10^{-3}	0.06	-0.90	(0.05)	-5.89×10^{-4}	0.05
	2	-0.90	(0.05)	7.26×10^{-4}	0.05	-0.90	(0.05)	-7.41×10^{-4}	0.05
	3	-0.90	(0.06)	-3.79×10^{-3}	0.06	-1.12	(0.06)	-0.22	0.23
	4	-1.12	(0.05)	-0.22	0.22	-1.12	(0.05)	-0.22	0.23
$\hat{\psi}_{21}$	1	0.60	(0.04)	9.72×10^{-4}	0.04	0.60	(0.04)	4.42×10^{-4}	0.04
	2	0.60	(0.04)	-6.57×10^{-4}	0.04	0.60	(0.04)	3.06×10^{-4}	0.04
	3	0.60	(0.04)	2.63×10^{-3}	0.04	0.75	(0.04)	0.15	0.16
	4	0.75	(0.04)	0.15	0.15	0.75	(0.04)	0.15	0.16

True value of the parameters: $\psi_{10} = 0.1$, $\psi_{11} = 0.1$, $\psi_{20} = -0.9$, $\psi_{21} = 0.6$. Sc. = scenario, SE = standard error, RMSE = root mean squared error.

With censoring dependent on time-varying covariates

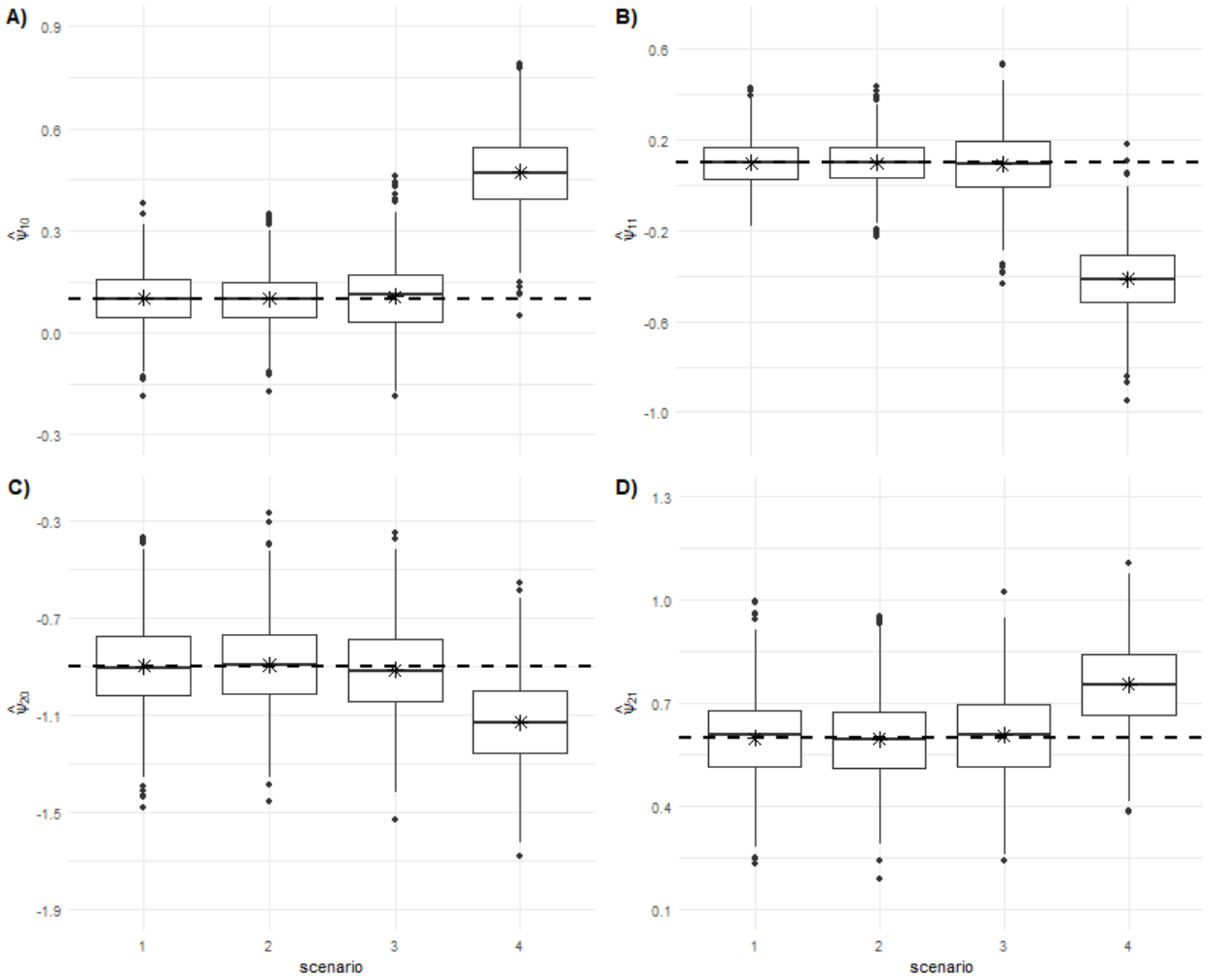

Figure C.12: Distribution of the blip parameter estimates in the first stage (upper row) and second stage (lower row) with DWSurv with sample size n=500 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 30% censoring, with both the probability of censoring dependent on time-varying covariates.

Table C.16: Mean, standard error, bias and root mean squared error of the blip estimators with DWSurv with sample size n=500 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 30% censoring, with the probability of censoring dependent on time-varying covariates.

		dWSurv			
	Sc.	Mean	(SE)	Bias	RMSE
$\hat{\psi}_{10}$	1	0.10	(0.08)	1.90×10^{-3}	0.08
	2	0.10	(0.08)	-5.82×10^{-4}	0.08
	3	0.11	(0.11)	6.87×10^{-3}	0.11
	4	0.47	(0.12)	0.37	0.39
$\hat{\psi}_{11}$	1	0.10	(0.10)	-2.10×10^{-3}	0.10
	2	0.10	(0.10)	-1.04×10^{-3}	0.10
	3	0.09	(0.15)	-8.94×10^{-3}	0.15
	4	-0.41	(0.16)	-0.51	0.54
$\hat{\psi}_{20}$	1	-0.90	(0.18)	2.96×10^{-3}	0.18
	2	-0.89	(0.18)	7.83×10^{-3}	0.18
	3	-0.91	(0.19)	-0.01	0.19
	4	-1.13	(0.18)	-0.23	0.29
$\hat{\psi}_{21}$	1	0.60	(0.12)	-1.54×10^{-3}	0.12
	2	0.60	(0.12)	-3.64×10^{-3}	0.12
	3	0.61	(0.13)	7.01×10^{-3}	0.13
	4	0.75	(0.12)	0.15	0.20

True value of the parameters: $\psi_{10} = 0.1$, $\psi_{11} = 0.1$, $\psi_{20} = -0.9$, $\psi_{21} = 0.6$. Sc. = scenario, SE = standard error, RMSE = root mean squared error.

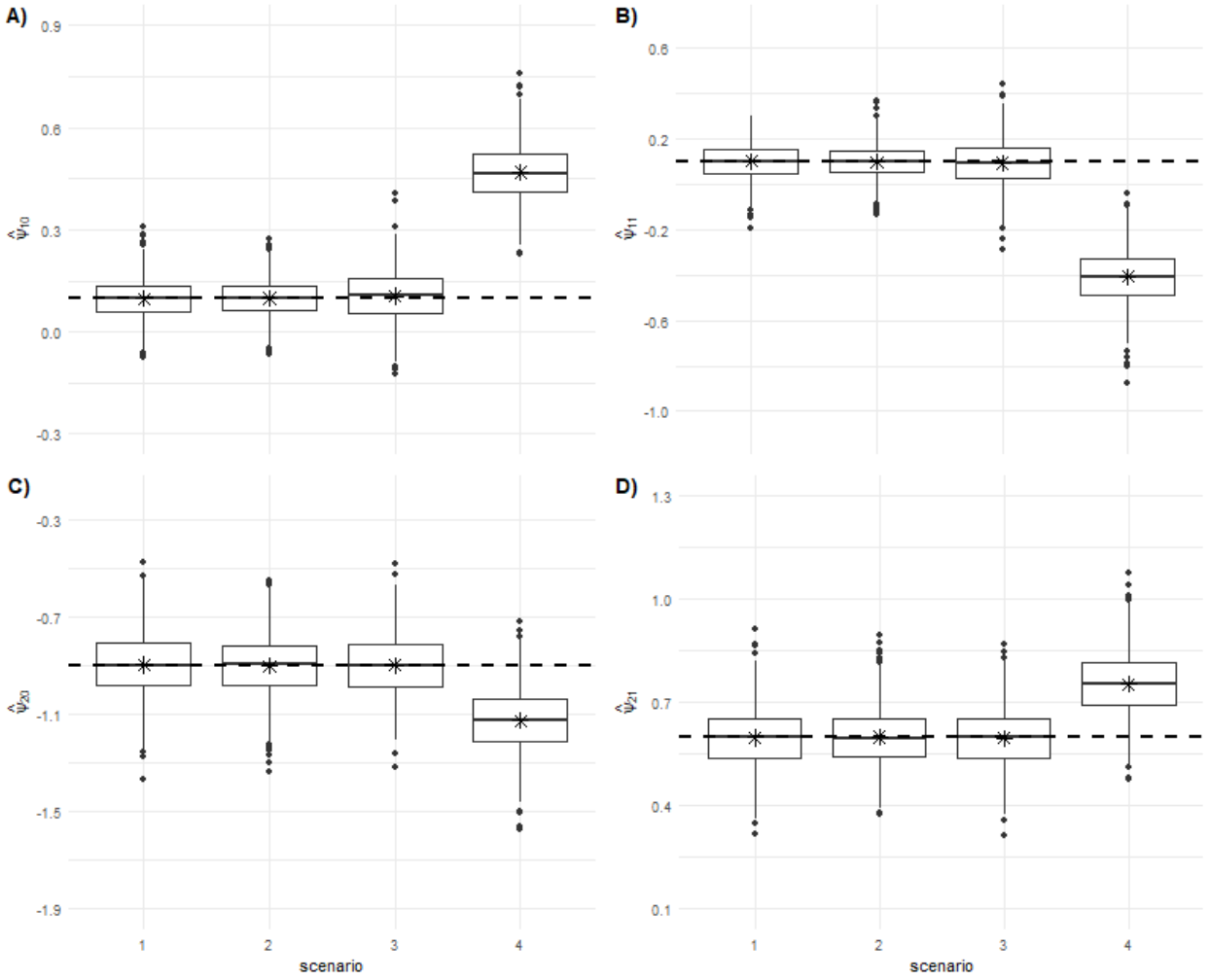

Figure C.13: Distribution of the blip parameter estimates in the first stage (upper row) and second stage (lower row) with DWSurv with sample size n=1000 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 30% censoring, with the probability of censoring dependent on time-varying covariates.

Table C.17: Mean, standard error, bias and root mean squared error of the blip estimators with DWSurv with sample size n=1000 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 30% censoring, with the probability of censoring dependent on time-varying covariates.

		dWSurv			
	Sc.	Mean	(SE)	Bias	RMSE
$\hat{\psi}_{10}$	1	0.10	(0.06)	-2.29×10^{-3}	0.06
	2	0.10	(0.06)	-1.38×10^{-3}	0.06
	3	0.11	(0.07)	5.35×10^{-3}	0.07
	4	0.47	(0.08)	0.37	0.38
$\hat{\psi}_{11}$	1	0.10	(0.08)	1.63×10^{-3}	0.08
	2	0.10	(0.07)	-9.77×10^{-5}	0.07
	3	0.09	(0.10)	-6.36×10^{-3}	0.10
	4	-0.41	(0.11)	-0.51	0.52
$\hat{\psi}_{20}$	1	-0.90	(0.13)	4.84×10^{-3}	0.13
	2	-0.90	(0.12)	-3.30×10^{-4}	0.12
	3	-0.90	(0.13)	3.18×10^{-3}	0.13
	4	-1.12	(0.13)	-0.22	0.26
$\hat{\psi}_{21}$	1	0.60	(0.09)	-2.26×10^{-3}	0.09
	2	0.60	(0.08)	-2.14×10^{-4}	0.08
	3	0.60	(0.09)	-2.58×10^{-3}	0.09
	4	0.75	(0.09)	0.15	0.18

True value of the parameters: $\psi_{10} = 0.1$, $\psi_{11} = 0.1$, $\psi_{20} = -0.9$, $\psi_{21} = 0.6$. Sc. = scenario, SE = standard error, RMSE = root mean squared error.

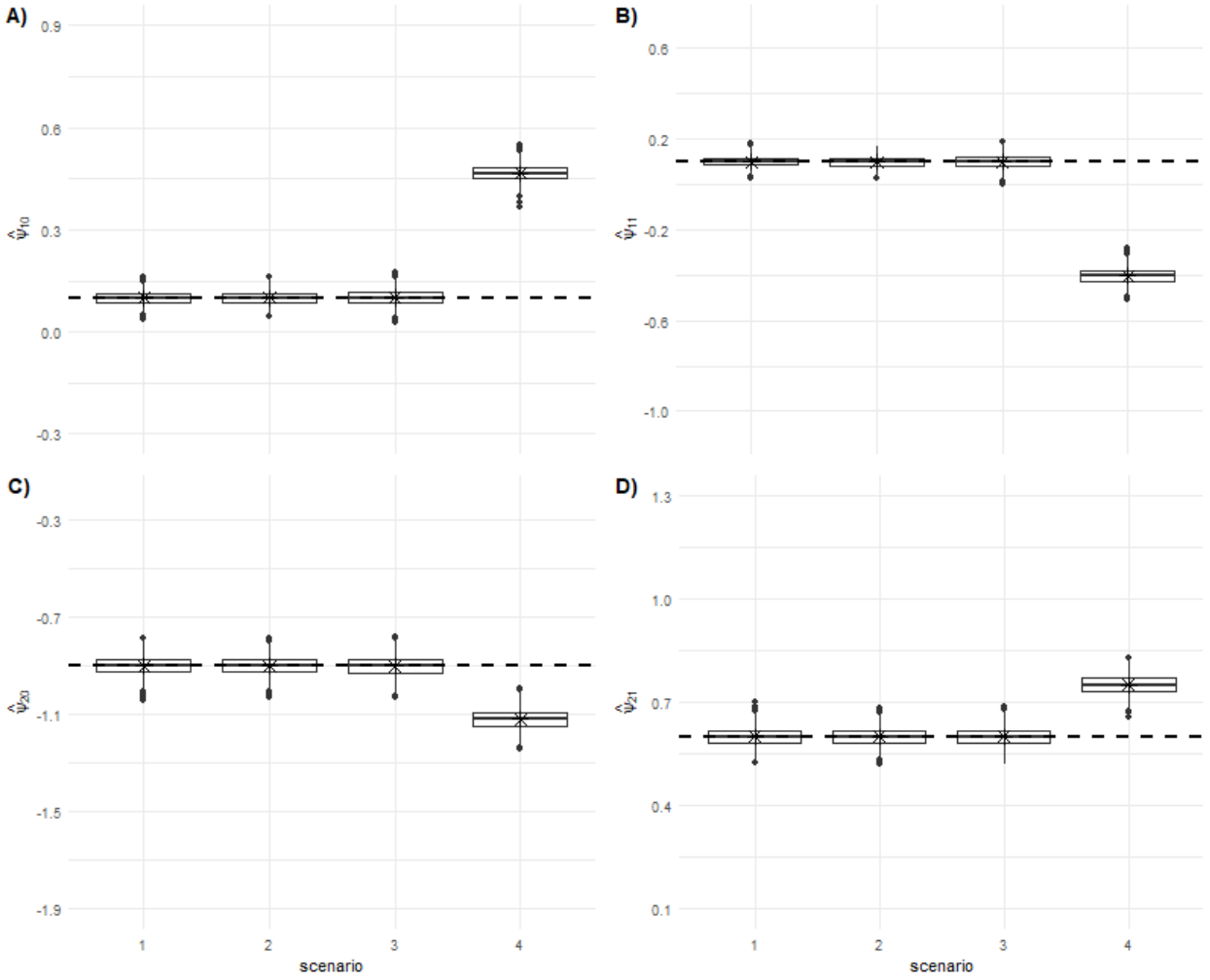

Figure C.14: Distribution of the blip parameter estimates in the first stage (upper row) and second stage (lower row) with DWSurv with sample size n=10,000 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 30% censoring, with the probability of censoring dependent on time-varying covariates.

Table C.18: Mean, standard error, bias and root mean squared error of the blip estimators with DWSurv with sample size n=10,000 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 30% censoring, with the probability of censoring dependent on time-varying covariates.

		dWSurv			
	Sc.	Mean	(SE)	Bias	RMSE
$\hat{\psi}_{10}$	1	0.10	(0.02)	-9.29×10^{-4}	0.02
	2	0.10	(0.02)	-1.84×10^{-4}	0.02
	3	0.10	(0.02)	8.41×10^{-4}	0.02
	4	0.47	(0.03)	0.37	0.37
$\hat{\psi}_{11}$	1	0.10	(0.02)	-1.28×10^{-4}	0.02
	2	0.10	(0.02)	-5.76×10^{-4}	0.02
	3	0.10	(0.03)	-1.99×10^{-3}	0.03
	4	-0.40	(0.03)	-0.50	0.50
$\hat{\psi}_{20}$	1	-0.90	(0.04)	2.09×10^{-4}	0.04
	2	-0.90	(0.04)	-4.08×10^{-4}	0.04
	3	-0.90	(0.04)	-3.45×10^{-4}	0.04
	4	-1.12	(0.04)	-0.22	0.22
$\hat{\psi}_{21}$	1	0.60	(0.03)	5.25×10^{-4}	0.03
	2	0.60	(0.03)	2.76×10^{-4}	0.03
	3	0.60	(0.03)	2.67×10^{-4}	0.03
	4	0.75	(0.03)	0.15	0.15

True value of the parameters: $\psi_{10} = 0.1$, $\psi_{11} = 0.1$, $\psi_{20} = -0.9$, $\psi_{21} = 0.6$. Sc. = scenario, SE = standard error, RMSE = root mean squared error.

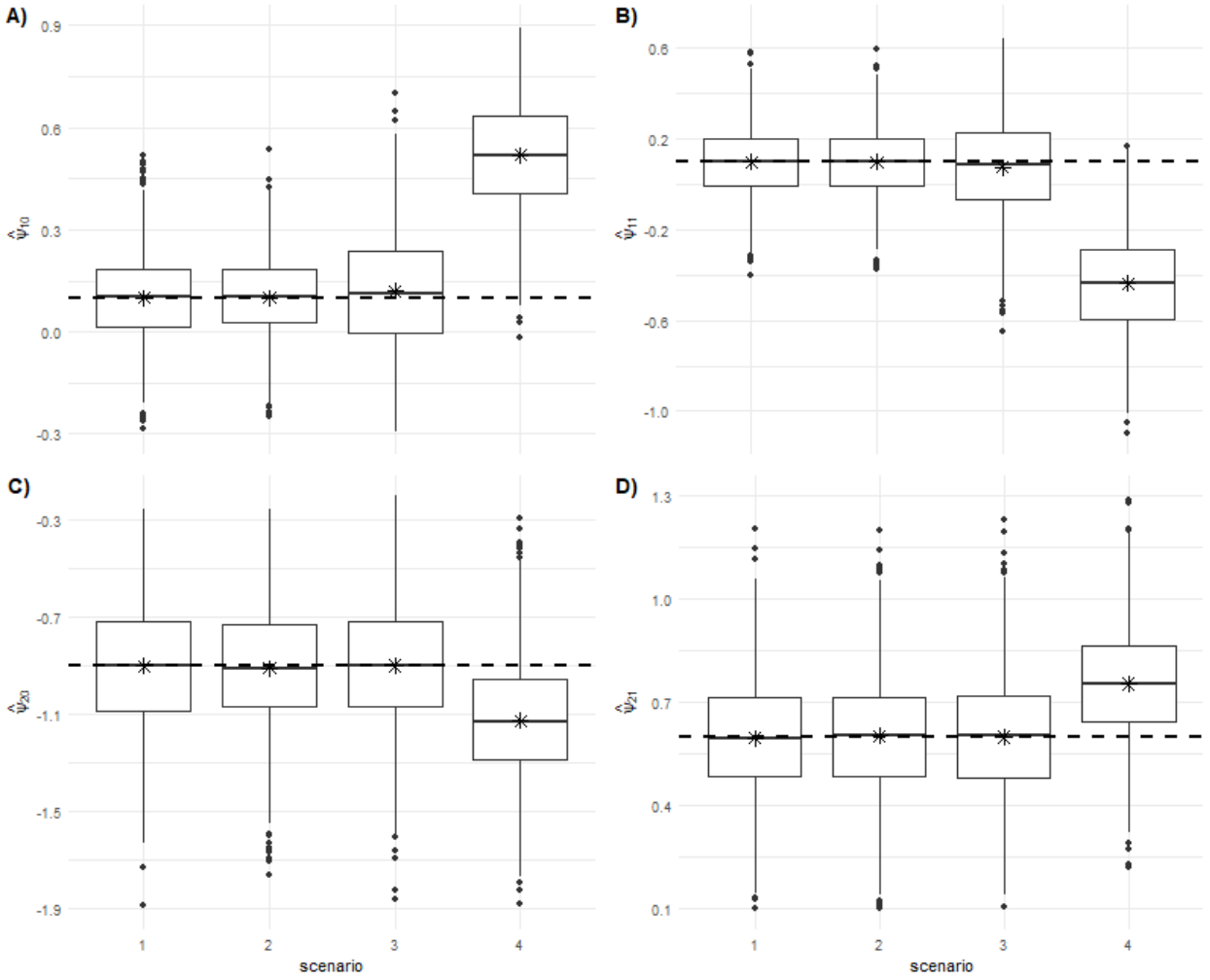

Figure C.15: Distribution of the blip parameter estimates in the first stage (upper row) and second stage (lower row) with DWSurv with sample size n=500 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 60% censoring, with both the probability of censoring dependent on time-varying covariates.

Table C.19: Mean, standard error, bias and root mean squared error of the blip estimators with DWSurv with sample size n=500 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 60% censoring, with the probability of censoring dependent on time-varying covariates.

		dWSurv			
	Sc.	Mean	(SE)	Bias	RMSE
$\hat{\psi}_{10}$	1	0.10	(0.13)	-7.19×10^{-4}	0.13
	2	0.10	(0.12)	-5.55×10^{-4}	0.12
	3	0.12	(0.18)	0.02	0.18
	4	0.53	(0.18)	0.43	0.47
$\hat{\psi}_{11}$	1	0.10	(0.15)	-1.87×10^{-4}	0.15
	2	0.10	(0.15)	-3.83×10^{-4}	0.15
	3	0.08	(0.22)	-0.02	0.22
	4	-0.44	(0.22)	-0.54	0.58
$\hat{\psi}_{20}$	1	-0.90	(0.26)	4.03×10^{-3}	0.26
	2	-0.90	(0.27)	-1.02×10^{-3}	0.27
	3	-0.90	(0.27)	3.04×10^{-3}	0.27
	4	-1.13	(0.26)	-0.23	0.34
$\hat{\psi}_{21}$	1	0.60	(0.17)	-3.06×10^{-3}	0.17
	2	0.60	(0.18)	1.31×10^{-3}	0.18
	3	0.60	(0.18)	-6.95×10^{-4}	0.18
	4	0.75	(0.17)	0.15	0.23

True value of the parameters: $\psi_{10} = 0.1$, $\psi_{11} = 0.1$, $\psi_{20} = -0.9$, $\psi_{21} = 0.6$. Sc. = scenario, SE = standard error, RMSE = root mean squared error.

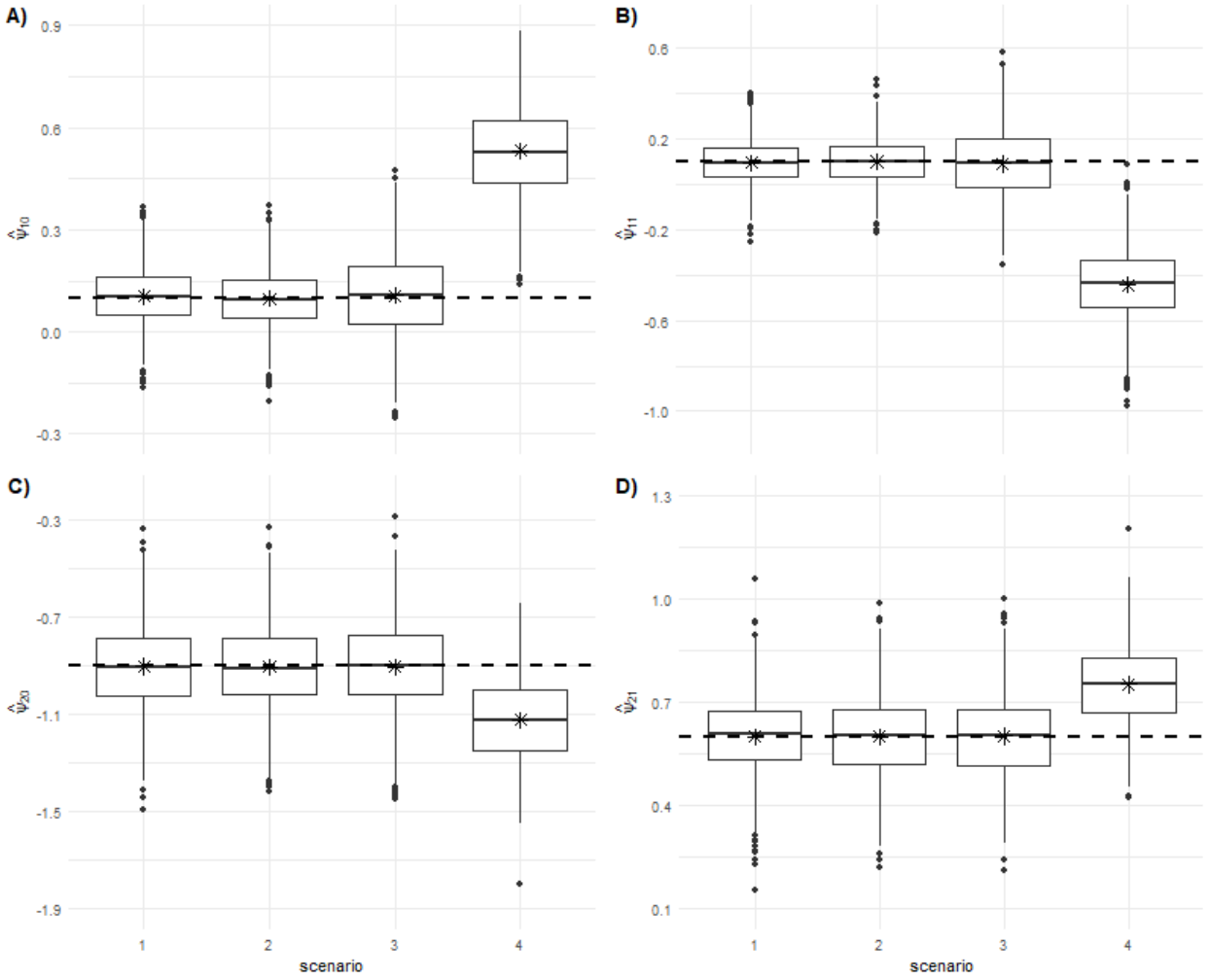

Figure C.16: Distribution of the blip parameter estimates in the first stage (upper row) and second stage (lower row) with DWSurv with sample size n=1000 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 60% censoring, with the probability of censoring dependent on time-varying covariates.

Table C.20: Mean, standard error, bias and root mean squared error of the blip estimators with DWSurv with sample size n=1000 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 60% censoring, with the probability of censoring dependent on time-varying covariates.

		dWSurv			
	Sc.	Mean	(SE)	Bias	RMSE
$\hat{\psi}_{10}$	1	0.10	(0.08)	4.07×10^{-3}	0.08
	2	0.10	(0.09)	-2.44×10^{-3}	0.09
	3	0.11	(0.12)	6.66×10^{-3}	0.12
	4	0.53	(0.13)	0.43	0.45
$\hat{\psi}_{11}$	1	0.09	(0.10)	-5.54×10^{-3}	0.10
	2	0.10	(0.10)	6.85×10^{-4}	0.10
	3	0.09	(0.15)	-9.42×10^{-3}	0.15
	4	-0.44	(0.16)	-0.54	0.57
$\hat{\psi}_{20}$	1	-0.90	(0.17)	-1.47×10^{-3}	0.17
	2	-0.90	(0.17)	-3.48×10^{-3}	0.17
	3	-0.90	(0.18)	-2.74×10^{-3}	0.18
	4	-1.12	(0.18)	-0.22	0.28
$\hat{\psi}_{21}$	1	0.60	(0.12)	1.70×10^{-3}	0.12
	2	0.60	(0.12)	1.37×10^{-3}	0.12
	3	0.60	(0.12)	2.14×10^{-3}	0.12
	4	0.75	(0.12)	0.15	0.19

True value of the parameters: $\psi_{10} = 0.1$, $\psi_{11} = 0.1$, $\psi_{20} = -0.9$, $\psi_{21} = 0.6$. Sc. = scenario, SE = standard error, RMSE = root mean squared error.

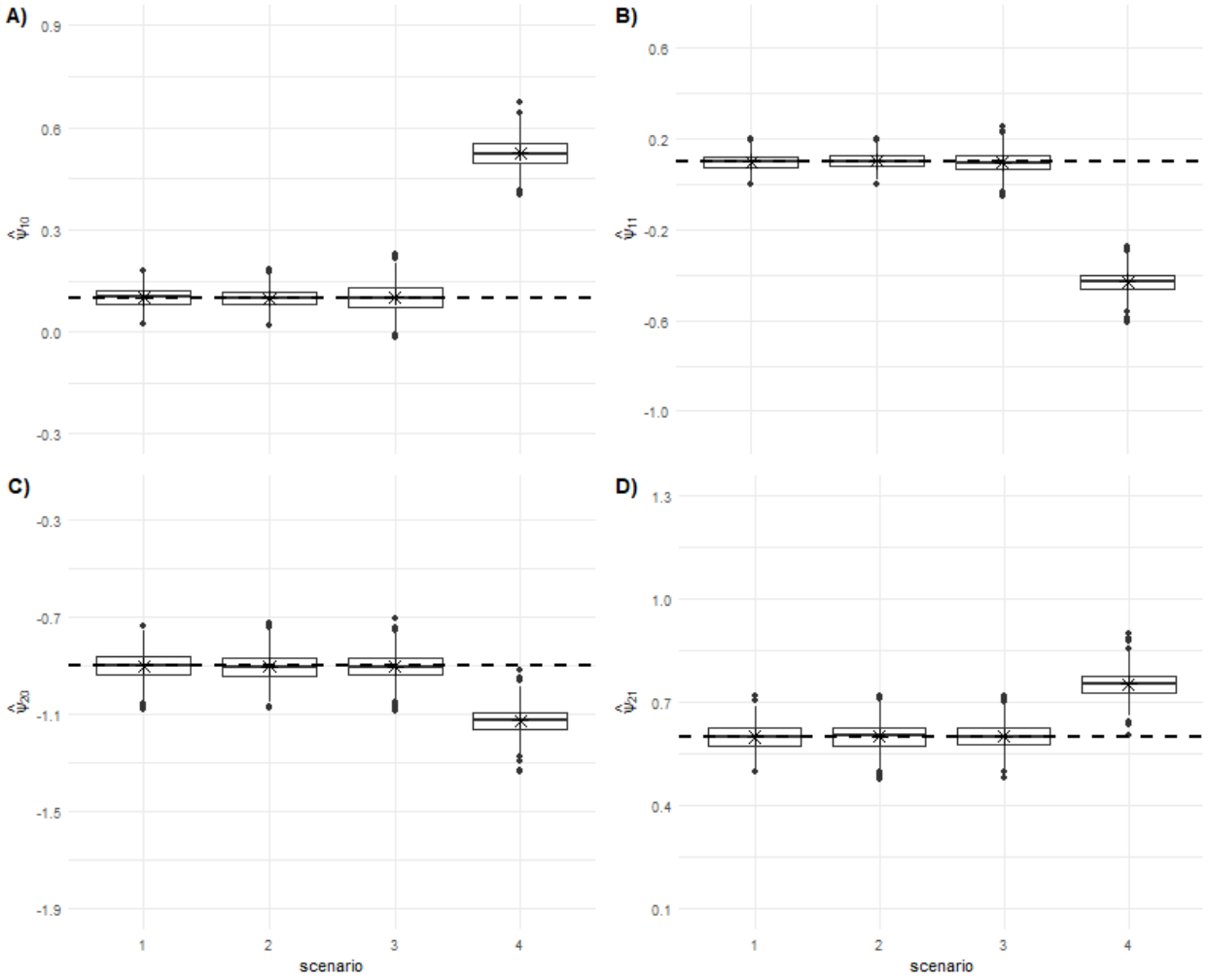

Figure C.17: Distribution of the blip parameter estimates in the first stage (upper row) and second stage (lower row) with DWSurv with sample size n=10,000 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 60% censoring, with the probability of censoring dependent on time-varying covariates.

Table C.21: Mean, standard error, bias and root mean squared error of the blip estimators with DWSurv with sample size n=10,000 across four scenarios: (1) all the models correctly specified, (2) weight models misspecified but treatment-free model correctly specified, (3) treatment-free model misspecified but weight models correctly specified, and (4) all models incorrectly specified. The data were simulated with 60% censoring, with the probability of censoring dependent on time-varying covariates.

		DWSurv			
	Sc.	Mean	(SE)	Bias	RMSE
$\hat{\psi}_{10}$	1	0.10	(0.03)	7.45×10^{-4}	0.03
	2	0.10	(0.03)	-2.11×10^{-3}	0.03
	3	0.10	(0.04)	1.28×10^{-3}	0.04
	4	0.53	(0.04)	0.43	0.43
$\hat{\psi}_{11}$	1	0.10	(0.03)	-1.80×10^{-3}	0.03
	2	0.10	(0.03)	1.85×10^{-3}	0.03
	3	0.10	(0.05)	-3.21×10^{-3}	0.05
	4	-0.43	(0.05)	-0.53	0.53
$\hat{\psi}_{20}$	1	-0.90	(0.05)	2.95×10^{-4}	0.05
	2	-0.90	(0.06)	-1.54×10^{-3}	0.06
	3	-0.90	(0.06)	-2.63×10^{-3}	0.06
	4	-1.13	(0.06)	-0.23	0.23
$\hat{\psi}_{21}$	1	0.60	(0.04)	1.13×10^{-4}	0.04
	2	0.60	(0.04)	1.22×10^{-3}	0.04
	3	0.60	(0.04)	1.81×10^{-3}	0.04
	4	0.75	(0.04)	0.15	0.16

True value of the parameters: $\psi_{10} = 0.1$, $\psi_{11} = 0.1$, $\psi_{20} = -0.9$, $\psi_{21} = 0.6$. Sc. = scenario, SE = standard error, RMSE = root mean squared error.

C.3.3 Ability to Correctly Identify the True Optimal DTR

The ability to correctly identify the true optimal DTR was quantified with the proportion of individuals for whom the optimal DTR was indeed identified. Tables C.22–C.24 show the distribution of the proportion of individuals for whom the stage 1 and/or stage 2 optimal treatment were correctly identified for 1000 simulated data sets with low (30%) or high (60%), independent or dependent censoring.

On average, the proportion of individuals for whom the optimal DTR was correctly identified was high (above 90%) although minimums fell under 50% when censoring was high (60%) and the sample size was small (n=500). In general, a higher proportion of censoring and smaller sample sizes resulted in fewer individuals for whom the optimal DTR was correctly identified. Settings with independent censoring were comparable to settings with dependent (baseline or time-varying) censoring. The method by HNW yielded similar observations.

Table C.22: Ability to identify the optimal DTR with DWSurv and the method by HNW when all models are correctly specified with independent censoring.

		% correctly identified											
		stage 1				stage 2				stages 1 & 2			
		Min	Mean	Med.	Max	Min	Mean	Med.	Max	Min	Mean	Med.	Max
30% independent censoring													
n=500	DWSurv	0.69	0.99	1	1	0.63	0.95	0.96	1	0.63	0.94	0.95	1
	HNW	0.68	0.99	1	1	0.64	0.95	0.96	1	0.64	0.94	0.95	1
n=1000	DWSurv	0.88	1	1	1	0.76	0.96	0.97	1	0.76	0.96	0.97	1
	HNW	0.88	1	1	1	0.78	0.96	0.97	1	0.78	0.96	0.97	1
n=10,000	DWSurv	1	1	1	1	0.95	0.99	0.99	1	0.95	0.99	0.99	1
	HNW	1	1	1	1	0.95	0.99	0.99	1	0.95	0.99	0.99	1
60% independent censoring													
n=500	DWSurv	0.60	0.98	1	1	0.54	0.93	0.95	1	0.45	0.92	0.94	1
	HNW	0.61	0.98	1	1	0.54	0.93	0.95	1	0.41	0.92	0.94	1
n=1000	DWSurv	0.73	1	1	1	0.70	0.95	0.96	1	0.66	0.95	0.96	1
	HNW	0.72	1	1	1	0.68	0.95	0.96	1	0.62	0.95	0.96	1
n=10,000	DWSurv	1	1	1	1	0.95	0.99	0.99	1	0.95	0.99	0.99	1
	HNW	1	1	1	1	0.95	0.99	0.99	1	0.95	0.99	0.99	1

Table C.23: Ability to identify the optimal DTR with DWSurv and the method by HNW when all models are correctly specified with censoring dependent on baseline covariates.

		% correctly identified											
		stage 1				stage 2				stages 1 & 2			
		Min	Mean	Med.	Max	Min	Mean	Med.	Max	Min	Mean	Med.	Max
30% baseline dependent censoring													
n=500	DWSurv	0.70	0.99	1	1	0.76	0.95	0.96	1	0.63	0.94	0.95	1
	HNW	0.68	0.99	1	1	0.75	0.95	0.96	1	0.64	0.94	0.96	1
n=1000	DWSurv	0.79	1	1	1	0.83	0.97	0.97	1	0.79	0.97	0.97	1
	HNW	0.81	1	1	1	0.82	0.97	0.97	1	0.80	0.97	0.97	1
n=10,000	DWSurv	1	1	1	1	0.95	0.99	0.99	1	0.95	0.99	0.99	1
	HNW	1	1	1	1	0.95	0.99	0.99	1	0.95	0.99	0.99	1
60% baseline dependent censoring													
n=500	DWSurv	0.54	0.97	1	1	0.51	0.92	0.94	1	0.48	0.89	0.92	1
	HNW	0.54	0.97	1	1	0.56	0.93	0.94	1	0.51	0.90	0.93	1
n=1000	DWSurv	0.60	0.99	1	1	0.64	0.95	0.96	1	0.57	0.94	0.96	1
	HNW	0.59	0.99	1	1	0.69	0.95	0.96	1	0.56	0.94	0.96	1
n=10,000	DWSurv	1	1	1	1	0.93	0.99	0.99	1	0.93	0.99	0.99	1
	HNW	1	1	1	1	0.94	0.99	0.99	1	0.94	0.99	0.99	1

Table C.24: Ability to identify the optimal DTR with DWSurv when all models are correctly specified with censoring dependent on time-varying covariates.

	% correctly identified											
	stage 1				stage 2				stages 1 & 2			
	Min	Mean	Med.	Max	Min	Mean	Med.	Max	Min	Mean	Med.	Max
30% time-varying dependent censoring												
n=500	0.73	0.99	1	1	0.73	0.95	0.96	1	0.70	0.95	0.96	1
n=1000	0.84	1	1	1	0.84	0.97	0.97	1	0.83	0.97	0.97	1
n=10000	1	1	1	1	0.96	0.99	0.99	1	0.96	0.99	0.99	1
60% time-varying dependent censoring												
n=500	0.58	0.97	1	1	0.53	0.93	0.95	1	0.44	0.92	0.94	1
n=1000	0.71	0.99	1	1	0.74	0.95	0.96	1	0.74	0.95	0.96	1
n=10000	1	1	1	1	0.93	0.99	0.99	1	0.93	0.99	0.99	1

C.3.4 Comparison of Survival Time Distributions Under Different Treatment Assignment Schemes

We aimed to assess whether the optimal DTR estimated by DWSurv would yield optimal (longer) survival times if the treatment assignment followed the estimated optimal DTR. From an initial sample of size 500 or 1000, we estimated the optimal DTR with DWSurv. Then, we generated a larger sample (n=10000) with treatment assignment following the optimal treatment rules estimated by DWSurv. We repeated the same procedure with the method by HNW. We compared the distributions of the resulting survival times with the distributions of the survival times generated with treatment assignment following the true optimal DTR (known from the data generating mechanism) and with treatment assignment following the four possible fixed treatment strategies i.e. always assign treatment 0 ($A_1 = A_2 = 0$), always assign $A_1 = 0$ and $A_2 = 1$, always assign $A_1 = 1$ and $A_2 = 0$ or always assign treatment 1 ($A_1 = A_2 = 1$).

The distribution of the survival times generated according to the optimal DTR estimated by DWSurv was comparable to that of the survival times generated from the true optimal DTR. This conclusion remained true for independent and dependent censoring, for lower or higher proportions of censoring, and for smaller or larger initial sample sizes. The method by HNW yielded similar results.

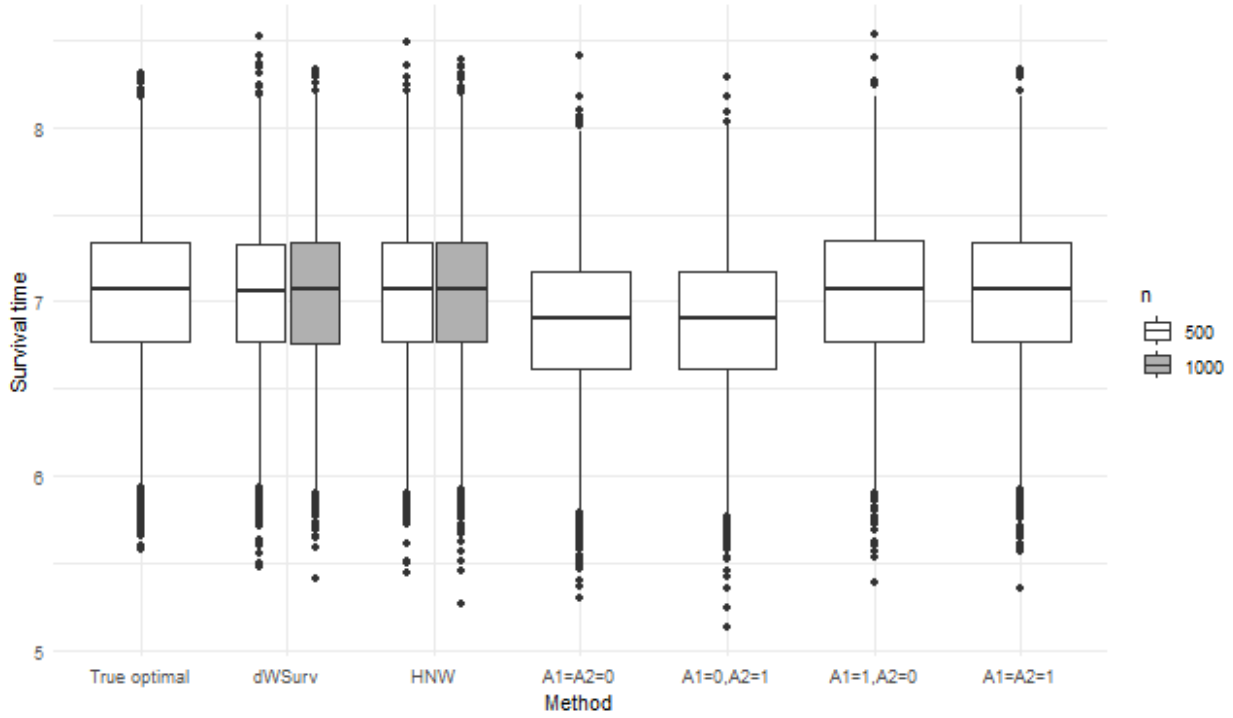

Figure C.18: Distribution of the log-survival times in a large sample (n=10,000) with treatment assignment following the true optimal DTR, the optimal DTR estimated by DWSurv, the optimal DTR estimated by HNW, and the four fixed treatment strategies. For the two schemes relying on the estimation of the optimal DTR by DWSurv or HNW, initial sample sizes of 500 (white) or 1000 (grey) were used. For the other schemes, it is meaningless to specify an initial sample size as the strategies are determined by the data generating mechanism or decided *a priori*. Data were generated with 30% independent censoring.

Table C.25: Summary of the log-survival time distribution under different treatment assignment schemes with 30% independent censoring

		Min	Q1	Mean (SD)	Median	Q3	Max
True optimal DTR		5.58	6.77	7.04 (0.41)	7.07	7.34	8.31
DWSurv	n=500	5.48	6.77	7.04 (0.41)	7.06	7.33	8.52
	n=1000	5.41	6.76	7.05 (0.41)	7.07	7.34	8.33
HNW	n=500	5.44	6.77	7.05 (0.41)	7.07	7.34	8.49
	n=1000	5.26	6.77	7.05 (0.41)	7.07	7.34	8.39
$A_1 = A_2 = 0$		5.30	6.62	6.88 (0.41)	6.90	7.17	8.41
$A_1 = 0, A_2 = 1$		5.13	6.61	6.88 (0.42)	6.90	7.17	8.29
$A_1 = 1, A_2 = 0$		5.38	6.77	7.05 (0.41)	7.07	7.35	8.53
$A_1 = A_2 = 1$		5.36	6.77	7.05 (0.41)	7.07	7.34	8.33

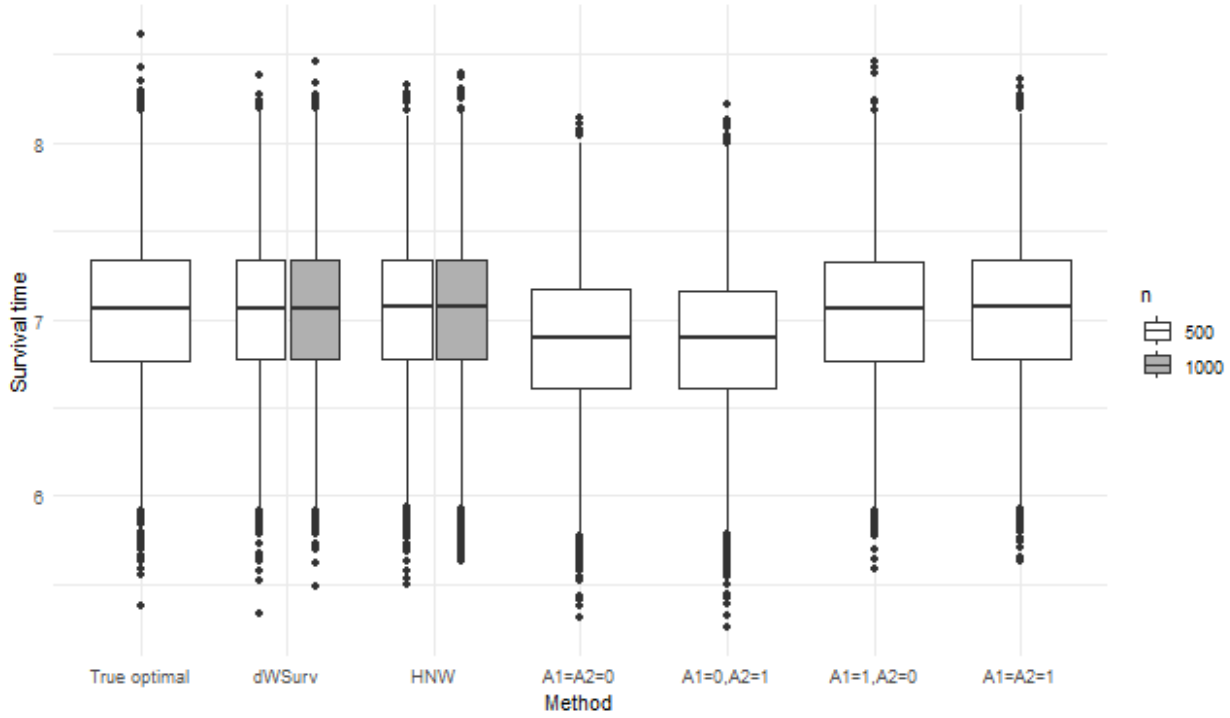

Figure C.19: Distribution of the log-survival times in a large sample (n=10,000) with treatment assignment following the true optimal DTR, the optimal DTR estimated by DWSurv, the optimal DTR estimated by HNW, and the four fixed treatment strategies. For the two schemes relying on the estimation of the optimal DTR by DWSurv or HNW, initial sample sizes of 500 (white) or 1000 (grey) were used. For the other schemes, it is meaningless to specify an initial sample size as the strategies are determined by the data generating mechanism or decided *a priori*. Data were generated with 60% independent censoring.

Table C.26: Summary of the log-survival time distribution under different treatment assignment schemes with 60% independent censoring

		Min	Q1	Mean (SD)	Median	Q3	Max
True optimal DTR		5.37	6.77	7.04 (0.41)	7.07	7.33	8.61
DWSurv	n=500	5.33	6.77	7.04 (0.41)	7.06	7.34	8.38
	n=1000	5.48	6.77	7.05 (0.41)	7.06	7.34	8.46
HNW	n=500	5.50	6.78	7.05 (0.41)	7.07	7.34	8.33
	n=1000	5.64	6.78	7.05 (0.41)	7.07	7.33	8.39
$A_1 = A_2 = 0$		5.31	6.61	6.88 (0.41)	6.90	7.17	8.14
$A_1 = 0, A_2 = 1$		5.26	6.61	6.88 (0.41)	6.89	7.16	8.22
$A_1 = 1, A_2 = 0$		5.58	6.77	7.04 (0.41)	7.07	7.33	8.46
$A_1 = A_2 = 1$		5.63	6.77	7.05 (0.41)	7.07	7.34	8.37

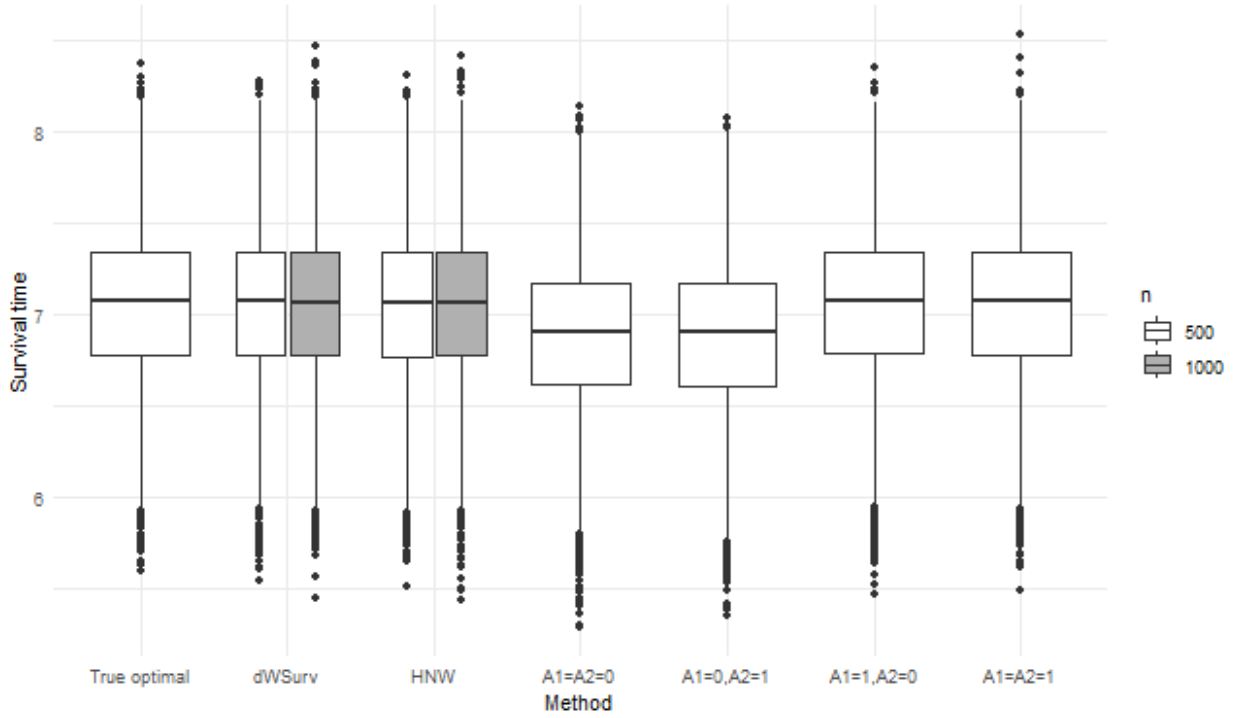

Figure C.20: Distribution of the log-survival times in a large sample (n=10,000) with treatment assignment following the true optimal DTR, the optimal DTR estimated by DWSurv, the optimal DTR estimated by HNW, and the four fixed treatment strategies. For the two schemes relying on the estimation of the optimal DTR by DWSurv or HNW, initial sample sizes of 500 (white) or 1000 (grey) were used. For the other schemes, it is meaningless to specify an initial sample size as the strategies are determined by the data generating mechanism or decided *a priori*. Data were generated with 30% dependent on baseline covariates.

Table C.27: Summary of the log-survival time distribution under different treatment assignment schemes with 30% censoring dependent on baseline covariates

		Min	Q1	Mean (SD)	Median	Q3	Max
True optimal DTR		5.59	6.78	7.05 (0.41)	7.07	7.34	8.38
DWSurv	n=500	5.54	6.78	7.05 (0.41)	7.07	7.34	8.28
	n=1000	5.45	6.77	7.05 (0.41)	7.07	7.34	8.47
HNW	n=500	5.51	6.77	7.05 (0.41)	7.07	7.34	8.32
	n=1000	5.44	6.78	7.05 (0.41)	7.07	7.34	8.42
$A_1 = A_2 = 0$		5.29	6.62	6.88 (0.41)	6.90	7.17	8.14
$A_1 = 0, A_2 = 1$		5.35	6.61	6.88 (0.41)	6.90	7.17	8.08
$A_1 = 1, A_2 = 0$		5.47	6.78	7.05 (0.41)	7.07	7.34	8.36
$A_1 = A_2 = 1$		5.49	6.78	7.05 (0.41)	7.07	7.34	8.53

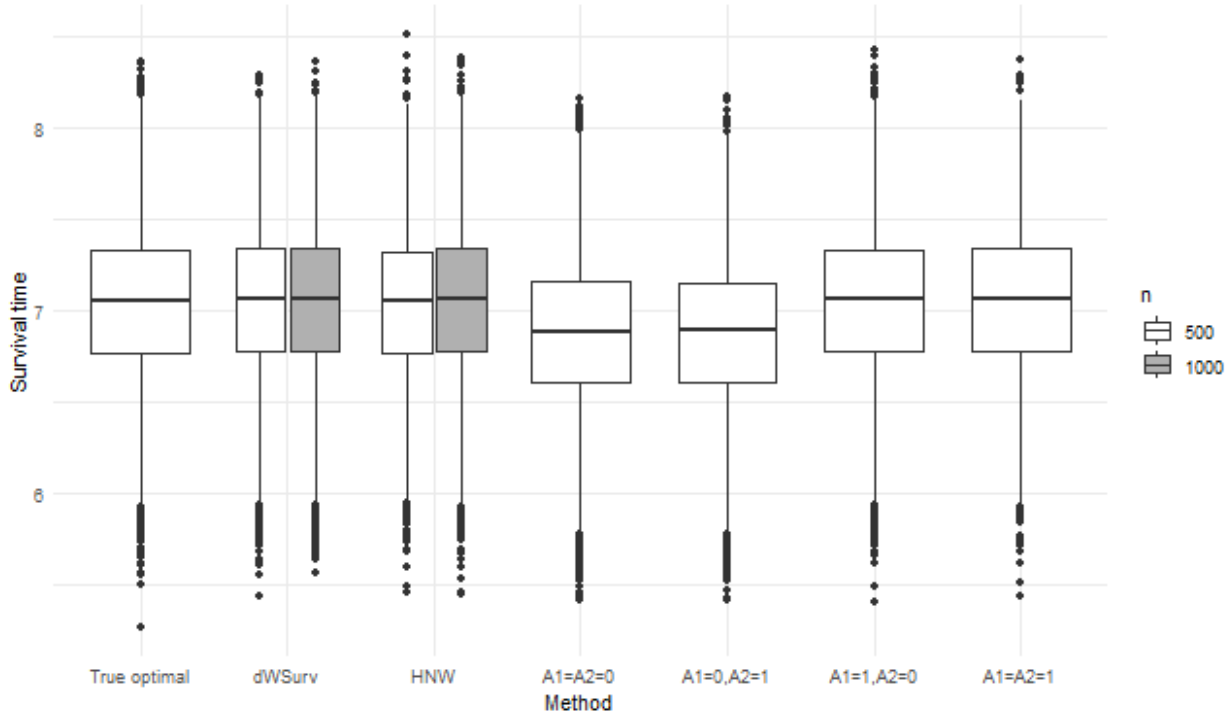

Figure C.21: Distribution of the log-survival times in a large sample (n=10,000) with treatment assignment following the true optimal DTR, the optimal DTR estimated by DWSurv, the optimal DTR estimated by HNW, and the four fixed treatment strategies. For the two schemes relying on the estimation of the optimal DTR by DWSurv or HNW, initial sample sizes of 500 (white) or 1000 (grey) were used. For the other schemes, it is meaningless to specify an initial sample size as the strategies are determined by the data generating mechanism or decided *a priori*. Data were generated with 60% censoring dependent on baseline covariates.

Table C.28: Summary of the log-survival time distribution under different treatment assignment schemes with 60% censoring dependent on baseline covariates

		Min	Q1	Mean (SD)	Median	Q3	Max
True optimal DTR		5.27	6.77	7.04 (0.41)	7.06	7.33	8.37
DWSurv	n=500	5.44	6.78	7.05 (0.41)	7.07	7.34	8.29
	n=1000	5.56	6.78	7.05 (0.41)	7.07	7.34	8.37
HNW	n=500	5.47	6.77	7.04 (0.40)	7.06	7.32	8.51
	n=1000	5.45	6.78	7.05 (0.41)	7.07	7.35	8.39
$A_1 = A_2 = 0$		5.41	6.61	6.87 (0.41)	6.89	7.16	8.16
$A_1 = 0, A_2 = 1$		5.42	6.61	6.88 (0.40)	6.90	7.16	8.17
$A_1 = 1, A_2 = 0$		5.40	6.78	7.05 (0.41)	7.07	7.33	8.42
$A_1 = A_2 = 1$		5.44	6.78	7.05 (0.41)	7.07	7.34	8.37

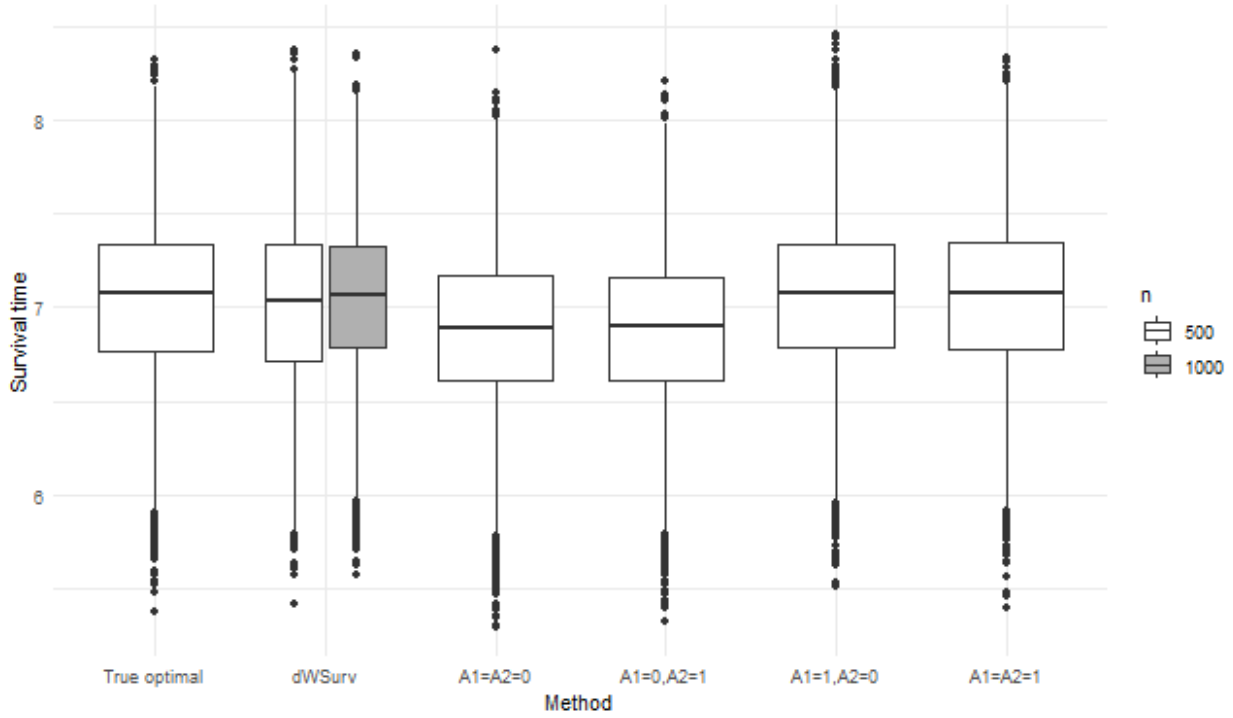

Figure C.22: Distribution of the log-survival times in a large sample (n=10,000) with treatment assignment following the true optimal DTR, the optimal DTR estimated by DWSurv, the optimal DTR estimated by HNW, and the four fixed treatment strategies. For the two schemes relying on the estimation of the optimal DTR by DWSurv or HNW, initial sample sizes of 500 (white) or 1000 (grey) were used. For the other schemes, it is meaningless to specify an initial sample size as the strategies are determined by the data generating mechanism or decided *a priori*. Data were generated with 30% dependent on time-varying covariates.

Table C.29: Summary of the log-survival time distribution under different treatment assignment schemes with 30% censoring dependent on time-varying covariates.

		Min	Q1	Mean (SD)	Median	Q3	Max
True optimal DTR		5.38	6.77	7.05 (0.41)	7.07	7.34	8.32
DWSurv	n=500	5.41	6.72	7.02 (0.43)	7.04	7.34	8.37
	n=1000	5.58	6.78	7.05 (0.41)	7.07	7.33	8.35
$A_1 = A_2 = 0$		5.29	6.61	6.88 (0.41)	6.89	7.17	8.37
$A_1 = 0, A_2 = 1$		5.33	6.61	6.88 (0.41)	6.90	7.16	8.21
$A_1 = 1, A_2 = 0$		5.51	6.78	7.05 (0.41)	7.07	7.34	8.45
$A_1 = A_2 = 1$		5.40	6.77	7.05 (0.41)	7.07	7.34	8.33

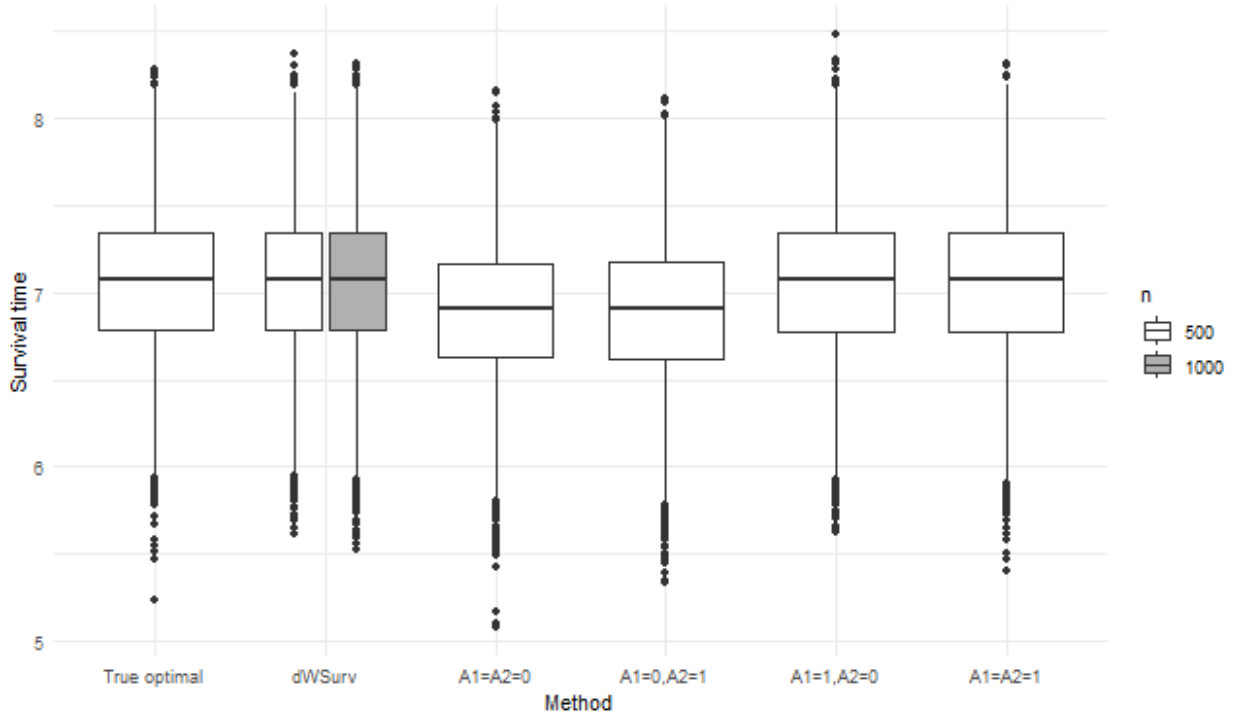

Figure C.23: Distribution of the log-survival times in a large sample (n=10,000) with treatment assignment following the true optimal DTR, the optimal DTR estimated by DWSurv, the optimal DTR estimated by HNW, and the four fixed treatment strategies. For the two schemes relying on the estimation of the optimal DTR by DWSurv or HNW, initial sample sizes of 500 (white) or 1000 (grey) were used. For the other schemes, it is meaningless to specify an initial sample size as the strategies are determined by the data generating mechanism or decided *a priori*. Data were generated with 60% censoring dependent on time-varying covariates.

Table C.30: Summary of the log-survival time distribution under different treatment assignment schemes with 60% censoring dependent on time-varying covariates

		Min	Q1	Mean (SD)	Median	Q3	Max
True optimal DTR		5.23	6.78	7.06 (0.41)	7.07	7.34	8.28
DWSurv	n=500	5.61	6.79	7.06 (0.41)	7.08	7.34	8.37
	n=1000	5.53	6.78	7.05 (0.41)	7.07	7.34	8.31
$A_1 = A_2 = 0$		5.08	6.62	6.89 (0.40)	6.90	7.17	8.16
$A_1 = 0, A_2 = 1$		5.34	6.62	6.88 (0.41)	6.90	7.17	8.11
$A_1 = 1, A_2 = 0$		5.62	6.78	7.05 (0.41)	7.07	7.34	8.48
$A_1 = A_2 = 1$		5.41	6.77	7.05 (0.41)	7.07	7.34	8.31

C.3.5 Expected Survival Time Distribution: Comparison With a Value Search Method

We compare DWSurv to dynamic marginal structural model (dynamic MSM) (Orellana et al., 2010) in terms of expected survival time under optimal DTR. Dynamic MSM is a value search approach which directly maximizes the value of a regime, here the overall survival time, to identify the optimal DTR. The comparison is made in a single-stage setting with the following data generating mechanism:

- Number of simulated data sets: 1,000
- Sample size n=1000
- $X_{11} \sim \text{Uniform}[0.5, 1.5]$
- $X_{12} \sim \text{Bernoulli}(0.6)$
- $A_1 \sim \text{Bernoulli}(\text{expit}(2X_{11} - 1))$
- $\delta \sim \text{Bernoulli}(\text{expit}(X_{12} + 0.1))$
- $\epsilon \sim \text{Normal}(0, 0.09)$
- Log-survival time $\log(T) = 3.7 + 1.5X_{11} - 0.8X_{12} + A_1(\psi_1 + \psi_2 X_{11}) + \epsilon$
- For those with $\delta = 0$, define $C \sim \text{Exp}(1/300)$

The true optimal treatment is given by $\mathbb{I}[\psi_1 + \psi_2 X_{11} > 0]$ which corresponds to $\mathbb{I}[X_{11} > \theta] = \mathbb{I}[X_{11} > -\psi_1/\psi_2]$. We considered two sets of blip parameters, $(\psi_1, \psi_2) = (-0.8, 0.9)$ and $(\psi_1, \psi_2) = (-0.15, 0.2)$.

Dynamic MSM estimates the optimal treatment rule as following:

1. Estimate $P(A_1 = 1|X_1)$ and $P(\Delta = 1|X_{12})$ and calculate the weights $[P(A_1 = a_1|X) \times P(\Delta = \delta|X_{12})]^{-1}$.

2. For each $\theta \in \{0.5, 0.51, \ldots, 1.5\}$:

   2.1 For each observation, determine the optimal treatment from $\mathbb{I}[x_{11} > \theta]$.

   2.2 Keep only the individuals who complied with their optimal treatment i.e. $A_1 = \mathbb{I}[x_{11} > \theta]$, and who experienced an event $\delta = 1$.

   2.3 Estimate $\mathbb{E}[Y^{\text{opt}}]$ in the sample by taking the weighted mean of Y_i with normalized weights $w_i / \sum w_i$.

3. Take $\hat{\theta}^{\text{msm}}$ which maximizes $\mathbb{E}[Y^{\text{opt}}]$.

To compare the performance of dynamic MSM to DWSurv in terms of estimating an optimal treatment that effectively improves the survival time, we use a strategy similar to that presented in Section 4.3 of the article to estimate $\mathbb{E}_\theta[Y^{\text{opt}}]$:

1. For a large sample size (n=10,000), generate X_{11}, X_{12} and ϵ according to the data generating mechanism above. Set $\delta = 1$ (no censoring).

2. Generate $A_1^{\text{msm}} = \mathbb{I}[X_{11} > \hat{\theta}^{\text{msm}}]$ according to the estimated optimal treatment rule by dynamic MSM and $A_1^{\text{DWSurv}} = \mathbb{I}[\hat{\psi}_1 + X_{11}\hat{\psi}_2 > 0]$ according to the estimated optimal treatment rule by DWSurv.

3. Generate two sets of log-survival times as presented above, $\log(T^{\text{msm}})$ with A_1^{msm} and $\log(T^{\text{DWSurv}})$ with A_1^{DWSurv}.

4. Estimate $\mathbb{E}[\log(Y^{\text{opt}})]$ as the mean of Y.

Figures C.24 and C.25 summarize the distribution of $\hat{\theta}$ and $\mathbb{E}[Y^{\text{opt}}]$ with dynamic MSM and DWSurv and the two sets of true blip parameters, $(\psi_1, \psi_2) = (-0.8, 0.9)$ and $(\psi_1, \psi_2) = (-0.15, 0.2)$ across three sample sizes. Summaries of the plotted distributions are presented in Tables C.31 and C.32. DWSurv identifies optimal treatment rule that leads to longer survival time, on average, as compared to dynamic MSM. Moreover, the estimator of θ with DWSurv is more efficient than with dynamic MSM. This was expected as DWSurv focuses on

estimation and inference for the decison rule parameters. When the true blip parameters are $(\psi_1, \psi_2) = (-0.15, 0.2)$, DWSurv sometimes leads to very large estimates for θ, in absolute value (see Table C.32). This is because $\hat{\theta}$ is obtained as $-\hat{\psi}_1/\hat{\psi}_2$ such that $\hat{\psi}_2$ close to zero lead to very large values of $\hat{\theta}$. This is more problematic in small samples in which the estimator of ψ_2 is more variable.

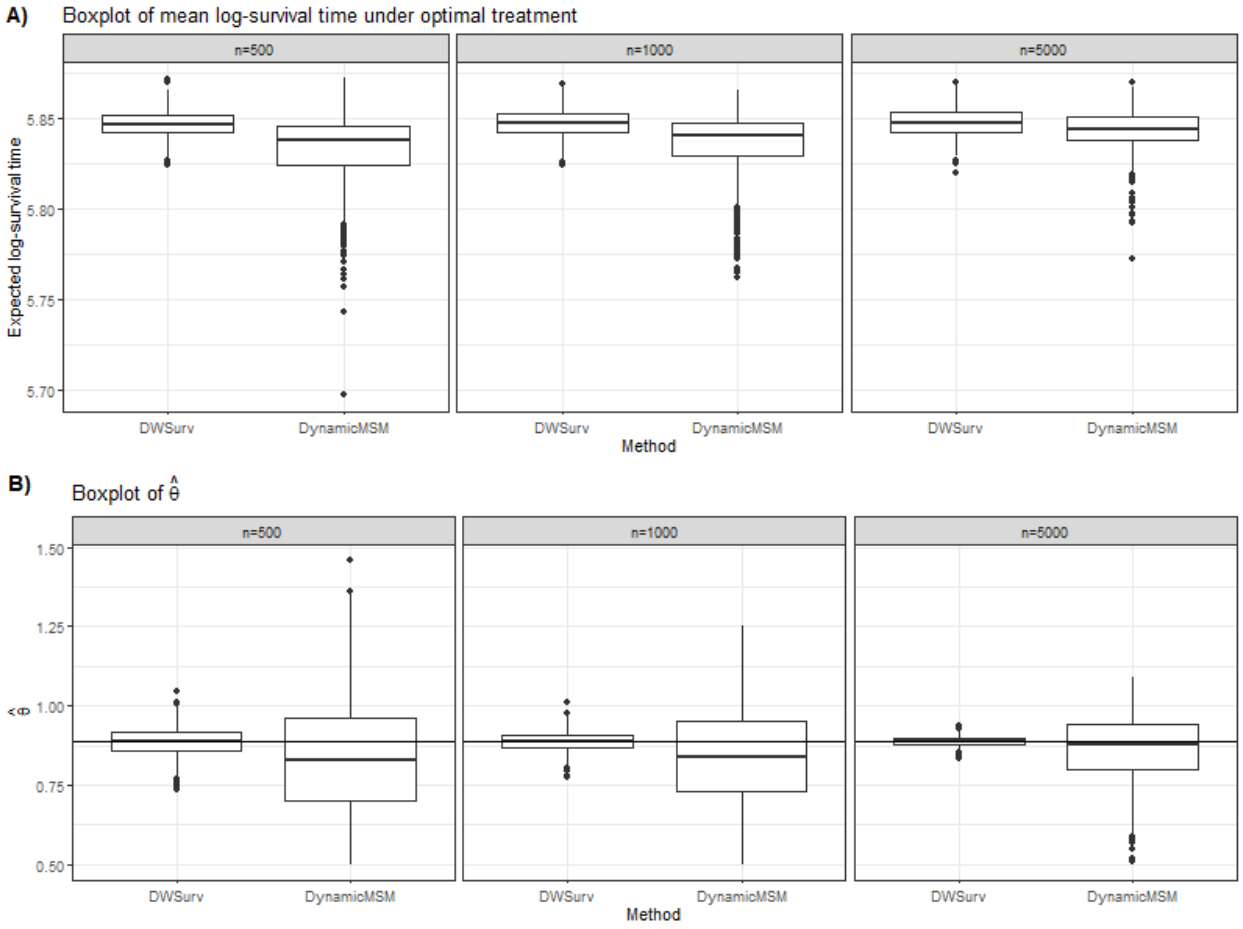

Figure C.24: Distribution of A) the mean log-survival time under optimal treatment and B) the estimate for θ across 1000 simulated data sets with three sample sizes with dynamic MSM and DWSurv with true blip parameters $(\psi_1, \psi_2) = (-0.8, 0.9)$. The horizontal line represents the true value of θ. Data were generated in a single-stage setting with 30% censoring dependent on a baseline covariate.

Table C.31: Distribution of $\hat{\theta}$ with DWSurv and dynamic MSM when $\theta = 0.8/0.9$

Method	Min.	Q1	Median	Mean (SE)	Q3	Max.	Bias	RMSE
n=500								
Dynamic MSM	0.50	0.70	0.83	0.83 (0.18)	0.96	1.46	-0.06	0.18
DWSurv	0.73	0.86	0.89	0.89 (0.05)	0.92	1.05	-9.43×10^{-4}	0.05
n=1000								
Dynamic MSM	0.50	0.73	0.84	0.84 (0.15)	0.95	1.25	-0.05	0.15
DWSurv	0.77	0.87	0.89	0.89 (0.03)	0.91	1.01	10.00^{-3}	0.03
n=5000								
Dynamic MSM	0.51	0.80	0.88	0.86 (0.10)	0.94	1.09	-0.02	0.10
DWSurv	0.83	0.88	0.89	0.89 (0.01)	0.90	0.94	-1.26×10^{-4}	0.01

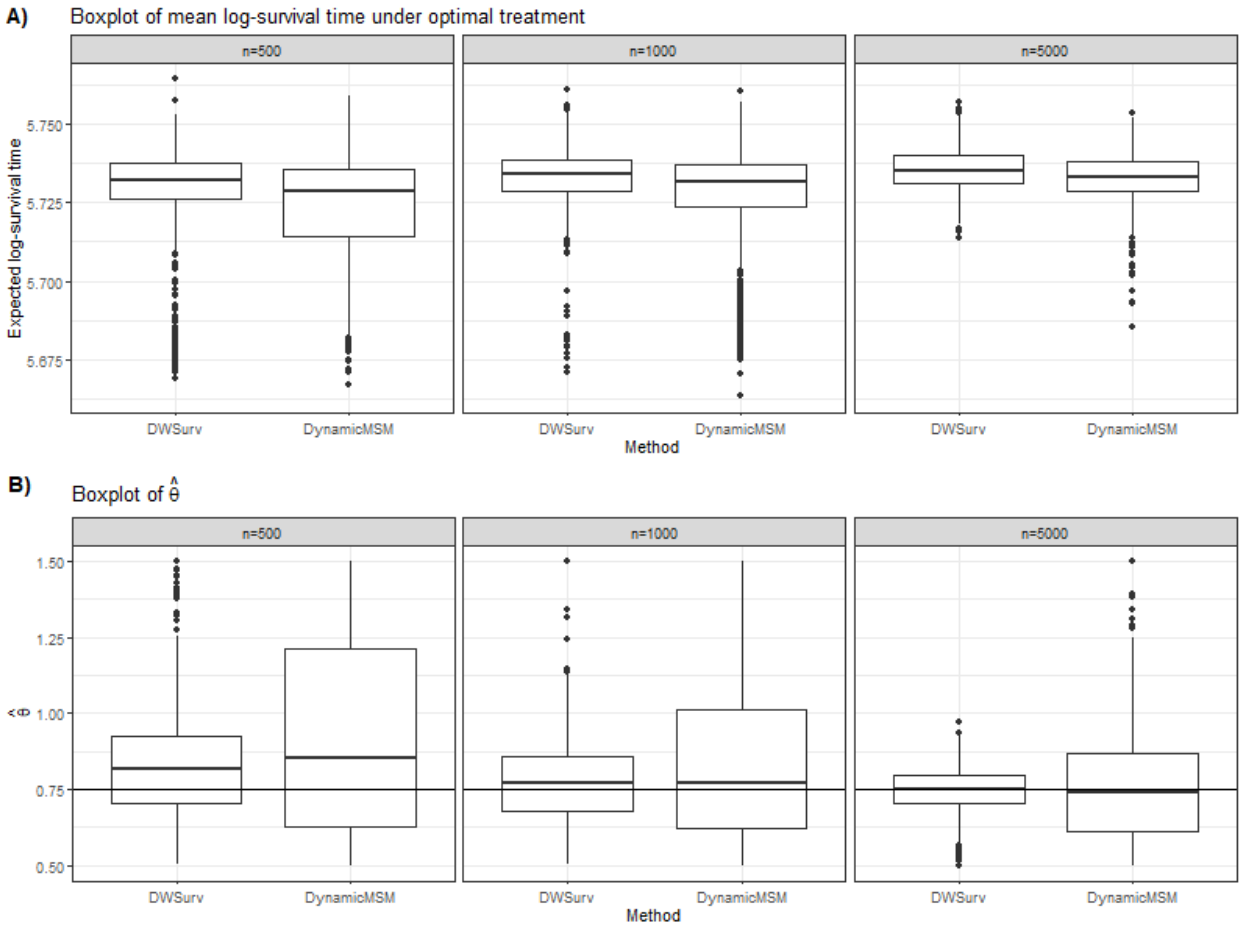

Figure C.25: Distribution of A) the mean log-survival time under optimal treatment and B) the estimate for θ across 1000 simulated data sets with three sample sizes with dynamic MSM and DWSurv with true blip parameters $(\psi_1, \psi_2) = (-0.15, 0.2)$. The boxplots B) are truncated. The horizontal line represents the true value of θ. Data were generated in a single-stage setting with 30% censoring dependent on a baseline covariate.

Table C.32: Distribution of $\hat{\theta}$ with DWSurv and dynamic MSM when $\theta = 0.75$

Method	Min.	Q1	Median	Mean (SE)	Q3	Max.	Bias	RMSE
n=500								
Dynamic MSM	0.50	0.63	0.86	0.92 (0.32)	1.21	1.50	0.17	0.32
DWSurv	-170	0.61	0.79	0.59 (6.08)	0.92	28	-0.16	6.07
n=1000								
Dynamic MSM	0.50	0.62	0.77	0.84 (0.28)	1.01	1.50	0.09	0.28
DWSurv	-8.54	0.61	0.74	0.83 (3.96)	0.85	123	0.08	3.96
n=5000								
Dynamic MSM	0.50	0.61	0.74	0.75 (0.18)	0.87	1.50	9.00×10^{-5}	0.18
DWSurv	0.27	0.70	0.75	0.74 (0.08)	0.80	0.97	-9.72×10^{-3}	0.08

C.4 SERA Data Analysis

C.4.1 Inclusion Criteria

Patients were included in our sample if they met the following inclusion criteria: disease onset less than 1 year prior to baseline with a RA or UA diagnosis at baseline, not already in remission at baseline and on a valid DMARD monotherapy or combination therapy.

C.4.2 Implementation

We estimated the optimal stage 2 treatment decision rule by modeling the log-transformed remission time within stage 2 as a function of the baseline covariates age, gender, smoking status, RA vs. UA diagnosis and treatment received in stage 1 A_1, and of the following time-varying covariates measured at the follow-up visit: disease activity, time since disease onset and pain score. The tailoring variables were disease activity and A_1 such that an interaction term with the stage 2 treatment was included for these two variables. The probability of treatment A_2 was modeled with a logistic regression as a function of age measured at baseline and disease activity, time since disease onset and pain score measured at the follow-up visit.

The probability of being censored within the second stage was also modeled with a logistic regression with the same subset of covariates, plus smoking status. The optimal stage 1 treatment decision rule was estimated by modeling the log-transformed pseudo-remission time as a function of age, gender, smoking status, disease activity, time since disease onset, pain score and RA vs. UA diagnosis, all measured at baseline. Disease activity level at baseline was the only tailoring variable for the stage 1 treatment rule. Logistic regression models were used for the probability of treatment A_1 and the probability of being censored at any point during the study period using the same subset of variables as in the second stage but measured at baseline, plus RA vs. UA diagnosis in the treatment model. In both stages, we used overlap weights.

C.4.3 Definitions and Baseline Characteristics

Table C.33: Key terms and drug categories definitions

Key terms	Definitions
Low disease activity	Defined as 2.6 $<$ DAS28-ESR $\leq$ 3.2 or 2.3 $<$ DAS28-CRP $\leq$ 3.8 if DAS28-ESR is not available
Moderate or high disease activity	Defined as DAS28-ESR $>$ 3.2 or DAS28-CRP $>$ 3.8 if DAS28-ESR is not available
Remission	Defined with DAS28-ESR $\leq$ 2.6. If DAS28-ESR is not available, defined as DAS28-CRP $\leq$ 2.3
DMARD monotherapy	MTX, HCQ, SSZ or LEF
Double DMARD therapy	MTX + SSZ, MTX + HCQ, SSX + HCQ, or any combinations with LEF
Triple DMARD therapy	MTX + SSZ + HCQ
TNFi biologics	Adalimumab, etanercept, infliximab
Non-TNF biologics	Abatacept, rituximab, tocilizumab

DAS28-ESR: Disease Activity Score 28 erythrocyte sedimentation rate, DAS28-CRP: DAS28 C-reactive protein level, MTX: methotrexate, HCQ: hydroxychloroquine, SSZ: sulfasalazine, LEF: leflunomide, TNF: tumor necrosis factor, TNFi: TNF inhibitor

Table C.34: Patients' characteristics at baseline and follow-up visits

Characteristics	Baseline visit (n=488)	Follow-up visit (n=236)
Age in years, mean (SD)	60 (14)	60 (13)
Female, n (%)	308 (63%)	151 (64%)
Smoking status, n (%)		
Current	116 (24%)	59 (25%)
Ex-smoker	180 (37%)	76 (32%)
Never	192 (39%)	101 (43%)
Days since onset, mean (SD)	170 (81)	373 (89)
Disease activity, n (%)		
Low	45 (9%)	90 (38%)
Moderate or high	443 (91%)	146 (62%)
Pain score, mean (SD)	56 (27)	38 (26)
Diagnosis, n (%)		
RA	426 (87%)	205 (87%)
UA	62 (13%)	31 (13%)

SD: standard deviation

C.4.4 Sample R Code

```
library(DTRreg)

model1 <- DWSurv(time = list(~Y1, ~Y2), blip.mod = list(~DA1,

    ~DA2 + A1), treat.mod = list(A1 ~ age + DA1 + onset1 + pain1 +

    RAUA, A2 ~ age + DA2 + onset2 + pain2 + RAUA), cens.mod = list(delta ~

    age + DA1 + onset1 + pain1 + RAUA + smoking, delta ~ age +

    DA2 + onset2 + pain2 + RAUA + smoking), tf.mod = list(~eta2 +

    age + sex + smoking + RAUA + DA1 + onset1 + pain1, ~age +

    sex + smoking + RAUA + DA2 + onset2 + pain2 + A1), var.estim = "asymptotic",

    data = dat, optimization = "min")

summary(model1)

## DTR estimation over 2 stages:

##

## Blip parameter estimates

##             Estimate Std. Error    95% Conf. Int
```

```
## Stage 1 (n = 488)
## (Intercept)   0.0640     0.0754  [-0.0838,0.2118]
##         DA1  -0.0992     0.0795 [-0.2550,0.0566]
##
## Stage 2 (n = 236)
## (Intercept)  -0.0845     0.1238 [-0.3271,0.1581]
##         DA2   0.0813     0.4107  [-0.7235,0.8862]
##          A1   0.2795     0.3186  [-0.3449,0.9039]
##
## Warning: possible non-regularity at stage 1 (prop = 1)
## Warning: possible non-regularity at stage 2 (prop = 1)
##
## Recommended dynamic treatment regimen:
## Stage 1: treat if 0.0640 - 0.0992 DA1 < 0
## Stage 2: treat if -0.0845 + 0.0813 DA2 + 0.2795 A1 < 0
```

Appendix D

Supplemental Materials for Chapter 5

D.1 Computational Times

Table D.1 compares the mean computational time in seconds across the five methods, for one- and two-stage DTRs. Sample size n=100 is not considered in the two-stage DTR to ensure computational stability and sample size n=5,000 is omitted in the two-stage DTR to save computational resources. The naive and adjusted asymptotic variances are the fastest. In the one-stage DTR, they are approximately 200 times faster than the bootstrap methods. In the two-stage DTR, they are 400 times faster. The computational cost of the bootstrap methods compared to the asymptotic variances does not only increase with the number of bootstrap resamples but also with the number of stages in the DTR.

Table D.1: Mean computational time in seconds across 1000 simulated data sets with 1000 bootstrap resamples, if applicable

	One-stage DTR					
Method / Sample size	100	300	500	1000	5000	10,000
Asymptotic (naive)	0.01	0.03	0.02	0.03	0.12	0.25
Asymptotic (adjusted)	0.05	0.03	0.05	0.06	0.15	0.27
Standard bootstrap	9.32	11.11	12.96	17.84	56.34	104.96
Parametric bootstrap 1	9.78	11.87	14.04	19.68	64.51	120.63
Parametric bootstrap 2	9.67	11.79	13.97	19.63	64.16	120.13
	Two-stage DTR					
Method / Sample size	100	300	500	1000	5000	10,000
Asymptotic (naive)	–	0.03	0.04	0.06	–	0.48
Asymptotic (adjusted)	–	0.06	0.07	0.10	–	0.51
Standard bootstrap	–	22.89	26.50	35.87	–	206.79
Parametric bootstrap 1	–	24.64	28.97	40.20	–	243.50
Parametric bootstrap 2	–	24.50	28.83	39.95	–	242.55

D.2 Additional Simulation Results: Unknown Error Distribution

Figure 5.1 in the article shows the coverage of 95% confidence intervals for ψ_{11} when the survival times follow a Log-normal or Weibull distribution. Of interest is to evaluate the performance of the five methods when the assumption of mean zero errors is violated. Recall that the underlying data generating mechanism (see Section 5.4.1 in article) corresponds to a one-stage DTR with a linear treatment-free model specified as $\beta_{10} + \beta_{11}X_{11} + \beta_{12}X_{12}$ with $(\beta_{10}, \beta_{11}, \beta_{12})^T = (4.7, 1.5, -0.8)^T$. Table D.2 shows the corresponding mean/median confidence interval widths across five methods, six sample sizes and two survival time distributions. As expected, larger mean widths are associated with higher coverages. For example, the adjusted asymptotic variance yields relatively larger widths than the other methods for a given sample size, which is in line with its generally high coverage. Figure D.1 shows the coverage of 95% confidence intervals for the main treatment effect ψ_{10}. Again, the violation

of the zero expectation requirement for the errors does not affect the performance of the methods as all methods yield close to nominal coverages when the survival time follows a Weibull distribution. Compared to inferences for ψ_{11}, the adjusted asymptotic variance does not yield higher coverage than the other methods.

Coverages and widths are comparable in the simulations with data generating mechanisms specified with a nonlinear treatment-free model as $\beta_{10} + \beta_{11}X_{11} + \beta_{12}X_{12} + \beta_{13}X_{11}^4$ with $(\beta_{10}, \beta_{11}, \beta_{12}, \beta_{13})^T = (4.7, 3, -0.9, 0.05)^T$.

Table D.2: Mean/median width of 95% confidence intervals for the interaction with treatment $\psi_{11} = 0.1$ with a linear treatment-free model.

Distribution	Method / Sample size	100	300	500	1000	5000	10,000
Log-normal	Asymptotic (naive)	0.91/0.88	0.54/0.54	0.42/0.42	0.30/0.30	0.14/0.14	0.10/0.10
	Asymptotic (adjusted)	1.03/0.99	0.60/0.59	0.47/0.46	0.33/0.33	0.15/0.15	0.10/0.10
	Standard bootstrap	1.00/0.97	0.55/0.55	0.43/0.42	0.30/0.30	0.14/0.14	0.10/0.10
	Parametric bootstrap 1	1.02/1.00	0.58/0.58	0.45/0.44	0.31/0.31	0.14/0.14	0.10/0.10
	Parametric bootstrap 2	0.99/0.98	0.56/0.56	0.43/0.43	0.30/0.30	0.14/0.14	0.10/0.10
Weibull	Asymptotic (naive)	0.97/0.92	0.58/0.56	0.45/0.45	0.32/0.32	0.14/0.14	0.10/0.10
	Asymptotic (adjusted)	1.09/1.04	0.64/0.62	0.50/0.49	0.35/0.35	0.16/0.16	0.11/0.11
	Standard bootstrap	1.07/1.03	0.59/0.58	0.46/0.45	0.32/0.32	0.14/0.14	0.10/0.10
	Parametric bootstrap 1	1.09/1.06	0.62/0.61	0.47/0.47	0.33/0.33	0.15/0.15	0.11/0.11
	Parametric bootstrap 2	1.05/1.03	0.60/0.59	0.46/0.46	0.32/0.32	0.14/0.15	0.10/0.10

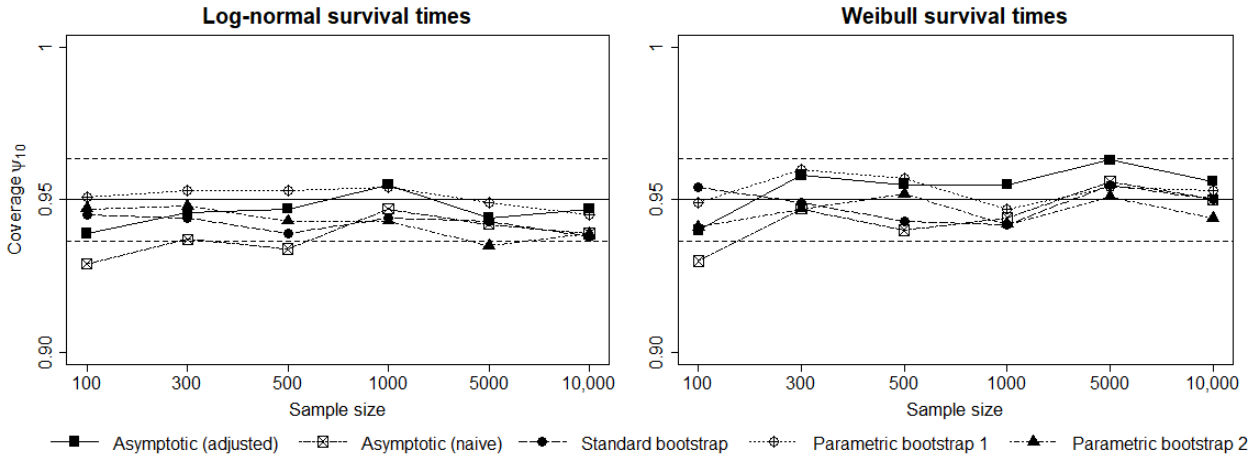

Figure D.1: Coverage of 95% confidence intervals for the main treatment effect ψ_{10} with true Log-normal and Weibull survival times and linear treatment-free model. The horizontal dashed lines represent the bounds for testing the null hypothesis that the coverage equals the nominal rate.

D.3 Additional Simulation Results: Model Misspecification

Figure 5.2 in the article shows the coverage of 95% confidence intervals for ψ_{11} under misspecification of the treatment-free, treatment or censoring model. Recall that the data are generated from a one-stage DTR with a linear treatment-free model. Other ways of misspecifying the models are considered. In the article, the following misspecified models are investigated:

- **Misspecified treatment-free (1)** – data are generated with true treatment-free model $\beta_{10} + \beta_{11}X_{11} + \beta_{12}X_{12}$ but the treatment-free model is misspecified by omitting X_{12};
- **Misspecified treatment (1)** – data are generated with treatment A_1 assigned with probability $P(A_1 = 1) \sim X_{11}$ but the treatment model is misspecified as $P(A_1 = 1) \sim 1$;
- **Misspecified censoring (1)** – data are simulated with censoring indicator δ generated as $P(\delta = 1) \sim X_{12}$ but the censoring model is misspecified as $P(\delta = 1) \sim 1$.

The following misspecifications were also considered:

- **Misspecified treatment-free (2)** – data are generated with true treatment-free model $\beta_{10} + \beta_{11}X_{11} + \beta_{12}X_{12} + \beta_{13}X_{11}^4$ but the treatment-free model is misspecified by omitting the nonlinear term X_{11}^4;
- **Misspecified treatment (2)** – data are generated with treatment A_1 assigned with probability $P(A_1 = 1) \sim X_{11}$ but the treatment model is misspecified as $P(A_1 = 1) \sim X_{12}$, a variable that does not predict treatment;
- **Misspecified censoring (2)** – data are simulated with censoring indicator δ generated as $P(\delta = 1) \sim X_{12}$ but the censoring model is misspecified as $P(\delta = 1) \sim X_{11}$, a

variable that is not associated with censoring.

Figure D.2 shows the coverage of 95% confidence intervals for the two blip parameters ψ_{10} and ψ_{11} under misspecifications of the treatment-free model (1), for which partial results are presented in the article Section 5.4.3, and (2). Table D.3 shows the corresponding mean/median widths. As for the parameter ψ_{11}, misspecification (1), which ignores an important confounder in the treatment-free model, also yields low coverages for ψ_{10} with the adjusted asymptotic variance. The same pattern in mean and median widths is observed for ψ_{10} and ψ_{11} with the adjusted asymptotic variance, that is, the mean widths are large despite the low coverage. Additionally, given that the median widths are relatively similar to that obtained with the other methods, we suspect that the adjusted asymptotic variance estimates exhibit excessive variability (c.f. Section D.5). Conclusions are different under misspecification (2), which ignores the nonlinear component in the treatment-free model. The two asymptotic variances outperform the bootstraps for inferences for ψ_{11} where all bootstraps, and especially the two parametric bootstraps, yield a relatively low coverage across sample sizes.

Figure D.3 shows the coverage of 95% confidence intervals for the two blip parameters ψ_{10} and ψ_{11} under misspecifications of the treatment model (1), for which partial results are presented in the article Section 5.4.3, and (2). All methods perform well under the two investigated misspecifications of the treatment model. Mean and median coverage widths (not shown) exhibit no pattern.

Figure D.4 shows the coverage of 95% confidence intervals for the two blip parameters ψ_{10} and ψ_{11} under misspecifications of the censoring model (1), for which partial results are presented in the article Section 5.4.3, and (2). All methods perform well under the two investigated misspecifications of the censoring model. Mean and median coverage widths (not shown) also exhibit no pattern.

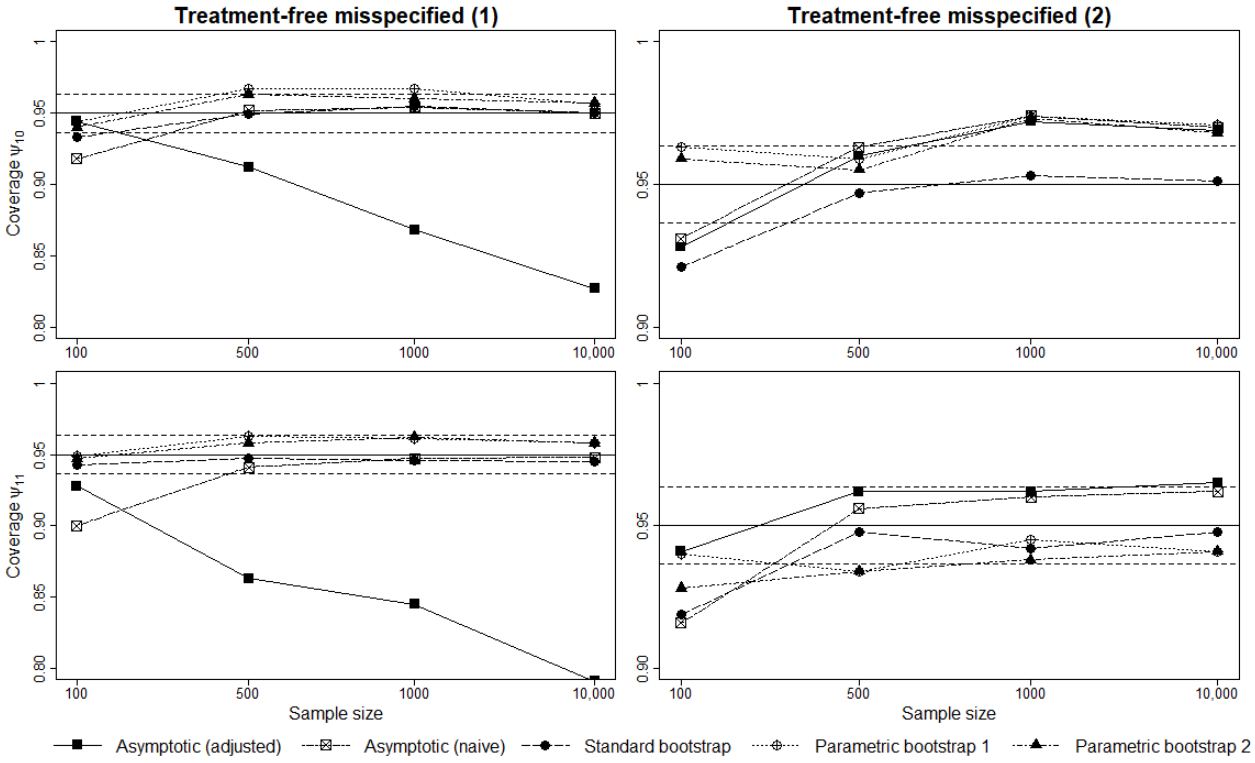

Figure D.2: Coverage of 95% confidence intervals for ψ_{10} and ψ_{11} under two ways of misspecifying the treatment-free model across multiple sample sizes. The horizontal dashed lines represent the bounds for testing the null hypothesis that the coverage equals the nominal rate.

Table D.3: Mean/median width of 95% confidence intervals for ψ_{10} and ψ_{11} when the treatment-free model is misspecified in two different ways.

	$\psi_{10} = 0.1$							
	Treatment-free (1)				Treatment-free (2)			
Method / Sample size	100	500	1000	10,000	100	500	1000	10,000
Asymptotic (naive)	1.09/1.08	0.50/0.50	0.35/0.35	0.11/0.11	0.99/0.98	0.47/0.46	0.33/0.33	0.10/0.10
Asymptotic (adjusted)	2.59/1.25	2.89/0.53	1.08/0.38	0.60/0.11	0.99/0.97	0.46/0.45	0.32/0.32	0.10/0.10
Standard bootstrap	1.18/1.15	0.50/0.50	0.35/0.35	0.11/0.11	1.01/0.99	0.43/0.42	0.30/0.30	0.09/0.09
Parametric bootstrap 1	1.21/1.19	0.53/0.53	0.37/0.37	0.12/0.12	1.08/1.06	0.47/0.47	0.33/0.33	0.10/0.10
Parametric bootstrap 2	1.19/1.18	0.52/0.52	0.37/0.37	0.12/0.12	1.05/1.03	0.46/0.46	0.32/0.32	0.10/0.10
	$\psi_{11} = 0.1$							
	Treatment-free (1)				Treatment-free (2)			
Method / Sample size	100	500	1000	10,000	100	500	1000	10,000
Asymptotic (naive)	1.48/1.45	0.68/0.68	0.48/0.48	0.15/0.15	1.49/1.44	0.72/0.71	0.51/0.51	0.16/0.16
Asymptotic (adjusted)	5.60/1.74	5.99/0.78	2.04/0.55	1.24/0.15	1.57/1.52	0.73/0.72	0.52/0.51	0.16/0.16
Standard bootstrap	1.63/1.59	0.69/0.68	0.48/0.48	0.15/0.15	1.54/1.49	0.66/0.65	0.46/0.46	0.15/0.15
Parametric bootstrap 1	1.67/1.65	0.72/0.72	0.51/0.51	0.16/0.16	1.49/1.45	0.64/0.63	0.45/0.45	0.14/0.14
Parametric bootstrap 2	1.64/1.63	0.71/0.71	0.50/0.50	0.16/0.16	1.44/1.41	0.62/0.62	0.44/0.44	0.14/0.14

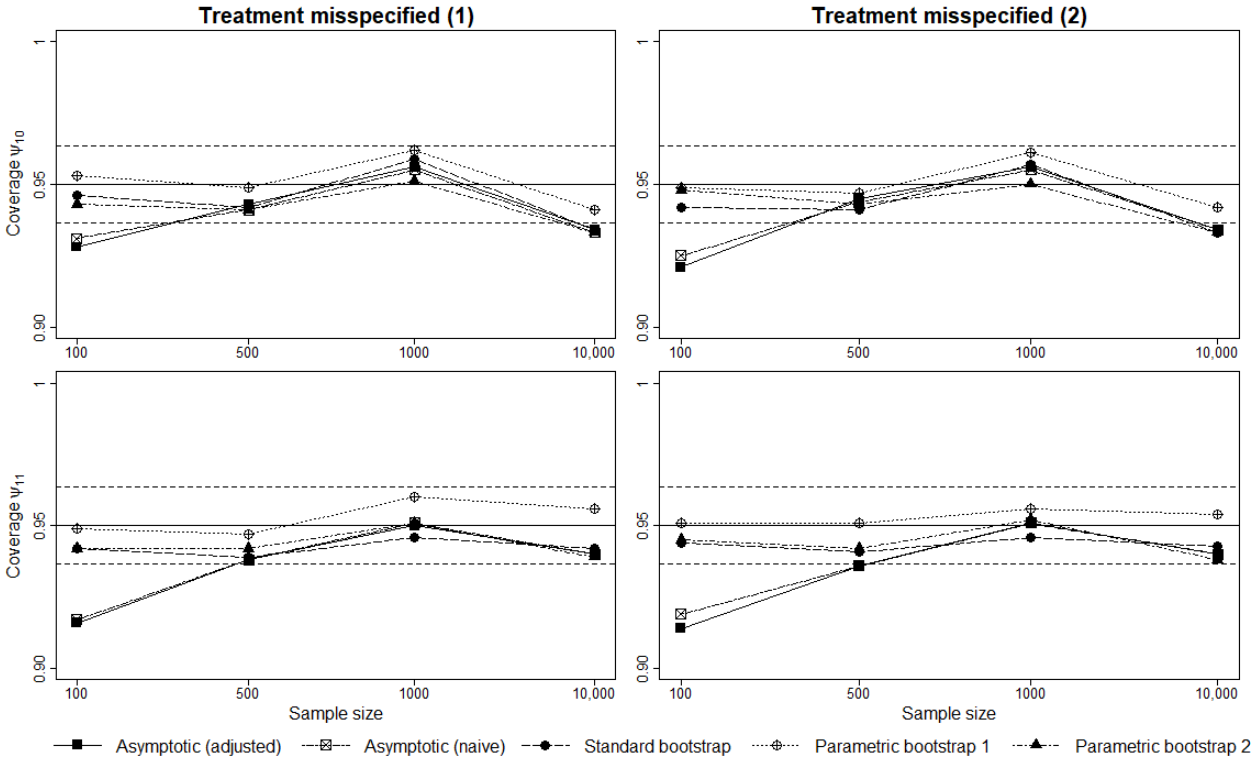

Figure D.3: Coverage of 95% confidence intervals for ψ_{10} and ψ_{11} under two ways of misspecifying the treatment model across multiple sample sizes. The horizontal dashed lines represent the bounds for testing the null hypothesis that the coverage equals the nominal rate.

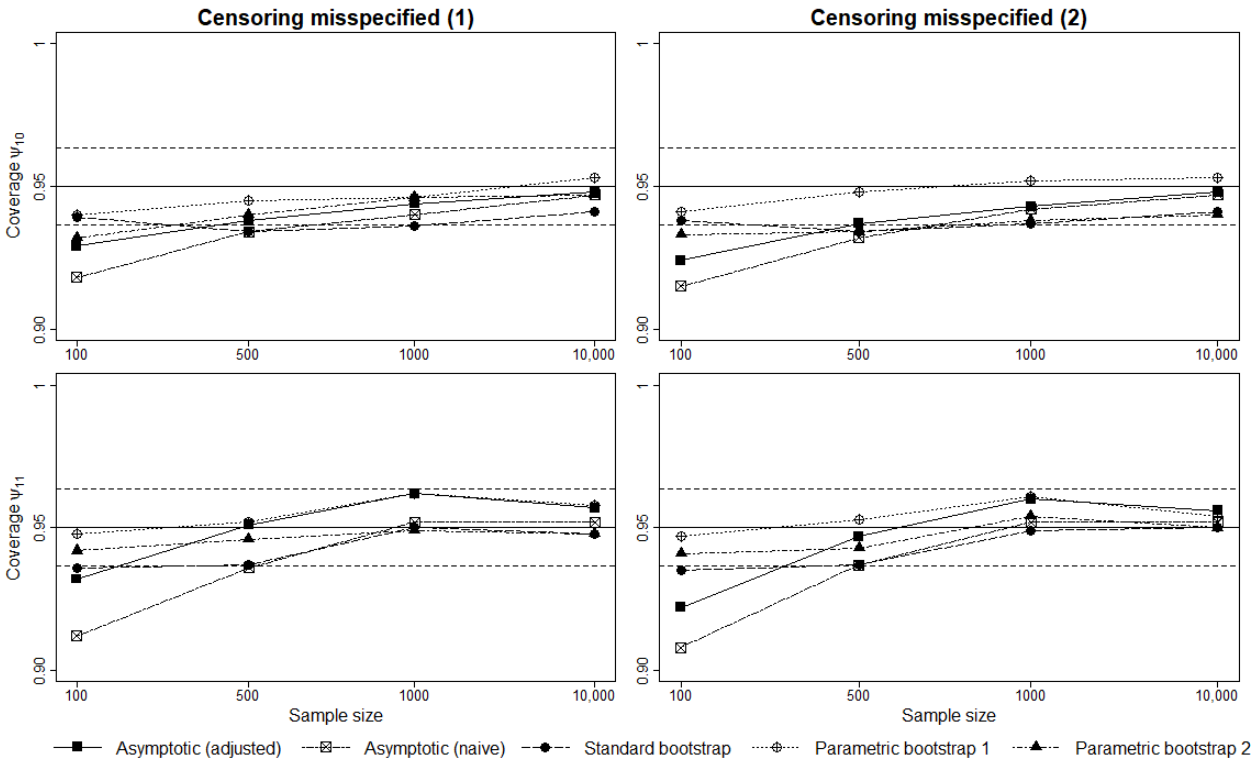

Figure D.4: Coverage of 95% confidence intervals for ψ_{10} and ψ_{11} under two ways of misspecifying the censoring model across multiple sample sizes. The horizontal dashed lines represent the bounds for testing the null hypothesis that the coverage equals the nominal rate.

D.4 Additional Simulation Results: Non-regularity

Figure 5.3 in the article shows the coverage of 95% confidence intervals for ψ_{11} for eight regular and non-regular scenarios across two sample sizes. Recall that the data are generated from a two-stage DTR and that the degree of non-regularity is controlled by the values assigned to the second stage blip parameters $\boldsymbol{\psi_2}$ and by the distribution of X_{22}. Figure D.5 shows the coverages for ψ_{10} across the eight scenarios. All methods perform similarly well across all sample sizes, regardless of the degree of non-regularity. Mean and median confidence interval widths (not shown), as well as results for sample sizes 500 and 10,000 with ψ_{11}, do not bring additional information.

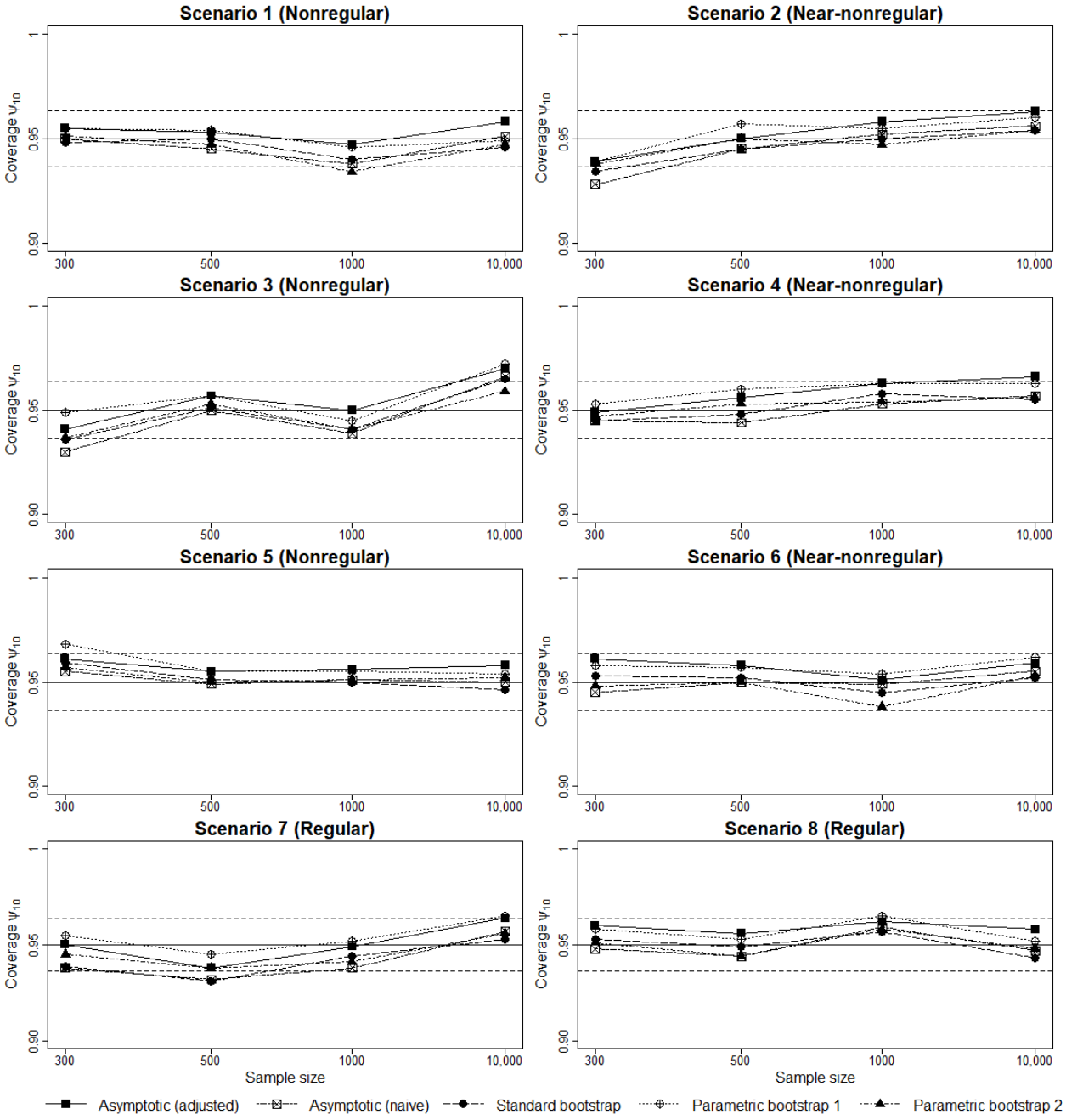

Figure D.5: Coverage of 95% confidence intervals for ψ_{10} in eight regular or nonregular scenarios across multiple sample sizes. The horizontal dashed lines represent the bounds for testing the null hypothesis that the coverage equals the nominal rate.

D.5 Details on the Performance of the Asymptotic Variance

We further investigate the low coverage but average large width of confidence intervals constructed with the adjusted asymptotic variance formula when the treatment-free model is misspecified by omitting an important confounder (c.f. Section D.3). This is not observed when misspecifying the treatment-free model by omitting a nonlinear component. We look at the distributions of the blip estimator $\hat{\psi}_{11}$ and of its standard error estimated with the two asymptotic variance formulae with correctly or incorrectly specified linear or nonlinear treatment-free models. A sample size of n=1000 is used.

Figure D.6 shows the distribution of the estimator $\hat{\psi}_{11}$ under correctly specified or misspecified treatment-free model, when the treatment-free model is truly linear or nonlinear. As expected from the double-robustness property, the estimator is unbiased. Its distribution is more variable under misspecification of the treatment-free model.

Figure D.7 shows the distribution of the standard error for $\hat{\psi}_{11}$ estimated with the adjusted and naive asymptotic variances. Recall that all scenarios yield coverages close to the nominal rate, except the adjusted asymptotic variance which yields a coverage around 0.84 when the linear treatment-free model is misspecified (c.f. Figure 5.2). In this case, Figure D.7 suggests that there are at least one very large standard error estimates. In general, the distribution of the asymptotic standard error estimates is skewed right and is not centered around the Monte Carlo standard error. Table D.4 shows a summary of the distributions presented in Figure D.7. In the scenario where the linear treatment-free model is misspecified, the adjusted asymptotic formula not only estimates a few large standard errors but also smaller ones as compared to the naive standard error estimates, which explains the low coverage despite the large confidence interval widths.

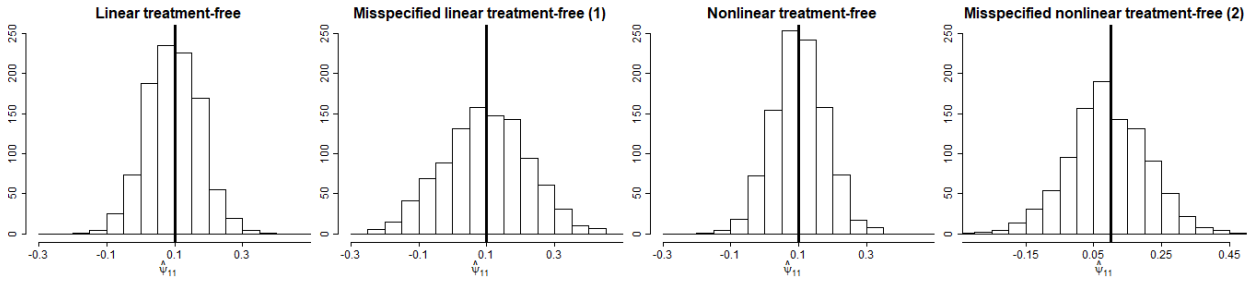

Figure D.6: Distribution of $\hat{\psi}_{11}$ across 1000 simulated data sets when the data are generated with a linear treatment-free model or a nonlinear treatment-free model, and when the treatment-free model is correctly specified or misspecified. The vertical line corresponds to the true parameter value.

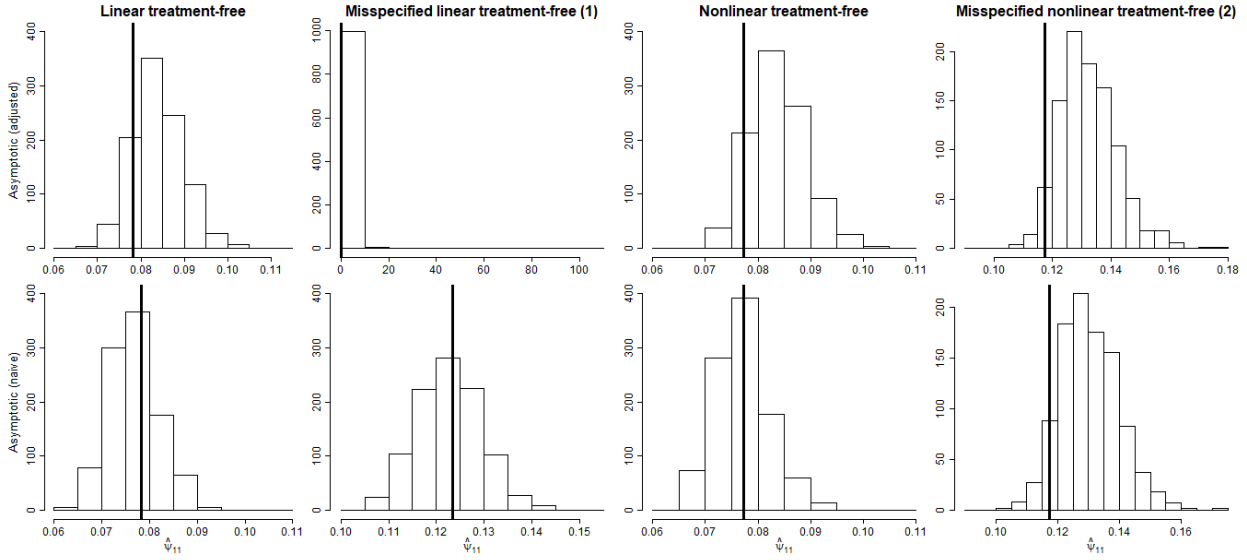

Figure D.7: Distribution of the standard error of $\hat{\psi}_{11}$ across 1000 simulated data sets when the data are generated with a linear treatment-free model or a nonlinear treatment-free model, and when the treatment-free model is correctly specified or misspecified. The vertical line corresponds to Monte Carlo standard errors.

Table D.4: Five-number summary for the distribution of the standard error of $\hat{\psi}_{11}$ when the treatment-free model is misspecified in two different ways.

		Asymptotic (adjusted)	Asymptotic (naive)
Model	Specification	Min Q1 Med Mean Q3 Max	Min Q1 Med Mean Q3 Max
Linear	Correct	0.07 0.08 0.08 0.08 0.09 0.11	0.06 0.07 0.08 0.08 0.08 0.10
	Incorrect	0.02 0.08 0.14 0.52 0.27 104.75	0.10 0.12 0.12 0.12 0.13 0.15
Nonlinear	Correct	0.07 0.08 0.08 0.08 0.09 0.11	0.06 0.07 0.08 0.08 0.08 0.10
	Incorrect	0.11 0.13 0.13 0.13 0.14 0.18	0.10 0.12 0.13 0.13 0.14 0.17

Appendix E

Supplemental Materials for Chapter 6

E.1 CPRD Database, Linkage and Study Cohort Assembling

Data were obtained by linking the CPRD with the Hospital Episodes Statistics (HES). The CPRD records good quality patient-level information on more than 13 million patients, corresponding to a representative sample of the UK population (Herrett et al., 2015, 2010). The

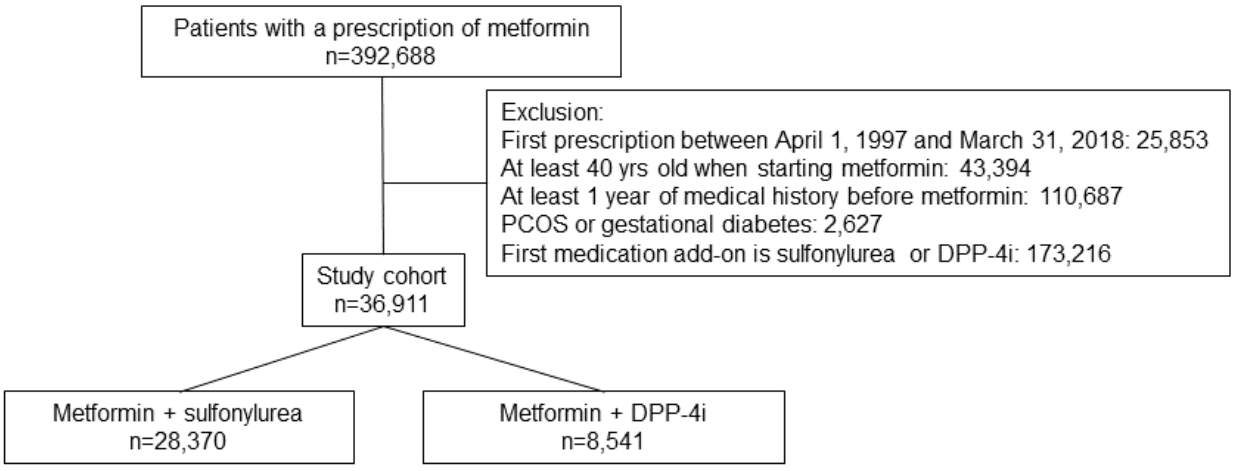

Figure E.1: Patients flowchart for assembling the study cohort. PCOS: polycystic ovarian syndrome, yrs: years.

Table E.1: Outcome and covariates definitions

Outcome	
Cardiovascular event	Defined on the date of a record of one of the following ICD-10 codes in the HES database after the date of study entry: I21.X, I61.X, I63.X, I65.X, I70.0, I70.2, I70.8, I70.9, I73.1, I73.8, I73.9, K55.1, I25.2, I66.3, I.64.
Death	As recorded in the ONS database.
Comorbidities	
History of alcohol misuse	Defined with the presence of a medical code about alcohol misuse at any time before the date of study entry.
History of dyslipidemia	Defined with the presence of a product code (lipid lowering agent) at any time before the date of study entry.
History of hypertension	Defined with the presence of a product code (hypertensive agent) or of a medical code about hypertension at any time before the date of study entry.
History of severe hypoglycemia	Defined with the presence of one of the following ICD-10 codes in the HES database at any time before the date of study entry: E10.0, E11.0, E16.0, E16.1, E16.2.
History of renal disease	Defined with the presence of a medical code about renal disease or failure at any time before the date of study entry.
Other covariates	
BMI	Defined with the most recent record from the date of study entry of (1) BMI or (2) weight and height recorded within 30 days. If BMI was the most recent record, values below 10 or above 90 were set to missing. If weight and height were the most recent records, BMI was calculated accordingly if 10 kg < weight < 500 kg and 1 m < height < 2.25 m. Otherwise, BMI was set to missing.
HbA1c	Defined with the most recent record of HbA1c from the date of study entry. Acceptable measurements were defined as 0 < HbA1c < 20. If HbA1c > 20, then HbA1c was defined as (HbA1c/10.93)+2.15 and was set to missing if this transformation also yielded HbA1c > 20. When more than one acceptable measurements were available on the date of the most recent record before study entry, the mean of the measurements on that date was taken.
SES	Defined with the Index of Multiple Deprivation reported as quintiles.
Smoking status	Medical codes about smoking habits were mapped to three smoking categories: never smoker, ex smoker, current smoker.

data collected in the CPRD gather anthropometric information, lifestyle variables, medical information and prescriptions by general practitioners. The HES database contains hospital admission information, including diagnoses and procedures. Linkage with the CPRD database is possible from April 1, 1997, onward, for approximately 75% of the practices in the CPRD (Herrett et al., 2015). Information on mortality and on deprivation were available from linkages with the Office for National Statistics and Index of Multiple Deprivation and Townsend scores databases, respectively. Figure E.1 shows the patient's flowchart from the

base cohort (patients with at least one prescription of metformin) to the study cohort.

E.2 Covariates and Outcome Definitions

Table E.1 shows definitions of the outcome and covariates. Table E.2 shows the distribution of the time (in years) from the most recent record of a covariate before study entry to study entry. HbA1c appeared to be monitored regularly: 75% of the patients had a record within approximately one month before study entry. BMI was also monitored regularly for most of the patients, with 25% of the patients having BMI recorded exactly on the date of study entry.

Table E.2: Time in years between covariate records and study entry date.

Covariate	Mean (SD)	Min	Q_1	Med	Q_3	Max
Body mass index	0.4 (1.3)	0	0	0.1	0.4	26
HbA1c	0.1 (0.3)	0	0.02	0.04	0.1	11
Smoking status	0.6 (1.4)	0	0.02	0.2	0.6	40
Alcohol misuse	8.3 (8.6)	0	1.6	5.1	13	58
Dyslipidemia	0.2 (0.8)	0	0	0.04	0.1	21
Renal disease	3.3 (4.3)	0	0.9	2.4	4.7	67
Hypertension	0.5 (2.2)	0	0	0.03	0.1	39
Hypoglycemia	4.1 (6.4)	0	0.1	2.0	5.4	42

HbA1c: glycated hemoglobin, Max: maximum, Med: median, Min: minimum, Q_1: first quartile, Q_3: third quartile, SD: standard deviation

E.3 Implementation

E.3.1 Sample `R` Code

The `R` code on the following page shows how to specify DWSurv with all models depending on linear combinations of all covariates and with the following tailoring variables: categories

of HbA1c, BMI and history of severe hypoglycemia. The corresponding R output is shown in Figure E.2.

```
library(DTRreg)
main1 <- dWSurv(time = list(~Y), blip.mod = list(~HBA1C1cat +
    HYPO1 + BMI1), treat.mod = list(TRMT1 ~ AGE1 + GENDER + SES1_imputed +
    SMOKING1 + BMI1 + MET1 + HBA1C1cat + RENAL1 + HYPO1 + HYPER1 +
    ALCOHOL1 + DYSLIP1), tf.mod = list(~AGE1 + GENDER + SES1_imputed +
    SMOKING1 + BMI1 + MET1 + HBA1C1cat + RENAL1 + HYPO1 + HYPER1 +
    ALCOHOL1 + DYSLIP1), cens.mod = list(DELTA ~ TRMT1 + AGE1 +
    GENDER + SES1_imputed + SMOKING1 + BMI1 + MET1 + HBA1C1cat +
    RENAL1 + HYPO1 + HYPER1 + ALCOHOL1 + DYSLIP1), data = mydata,
    var.estim = "asymptotic", asymp.opt = "adjusted")
```

```
> summary(main1)
DTR estimation over 1 stages:

Blip parameter estimates
                 Estimate Std. Error    95% Conf. Int
Stage 1 (n = 35287)
       (Intercept)  -0.9870     0.7628 [-2.4821,0.5080]
HBA1C1catBordeline  -0.8699     0.4469 [-1.7458,0.0060]
      HBA1C1catBad  -1.0948     0.5211 [-2.1161,-0.0734]
          HYPO1Yes   1.6892     0.7060  [0.3054,3.0730]
              BMI1   0.0416     0.0187  [0.0050,0.0782]

Warning: possible non-regularity at stage 1 (prop = 1)

Recommended dynamic treatment regimen:
Stage 1: treat if -0.9870 - 0.8699 HBA1C1catBordeline - 1.0948 HBA1C1catBad + 1.6892 HYPO1Yes + 0.0416 BMI1 > 0
```

Figure E.2: R output for main analysis.

E.3.2 Assessment of the Estimated Rule

One can informally assess the performance of the estimated treatment rule by looking at the Kaplan-Meier curves stratified by whether the estimated optimal treatment was received or not (Figure E.3). The Kaplan-Meier curves include only patients who experienced an event. The dashed lines are Kaplan-Meier curves weighted by $|a - \mathbb{E}(A|\boldsymbol{H})|$, the numerator in the overlap weights, to remove confounding between the treatment assignment and the survival

time. Patients who received their optimal treatment indeed survived longer.

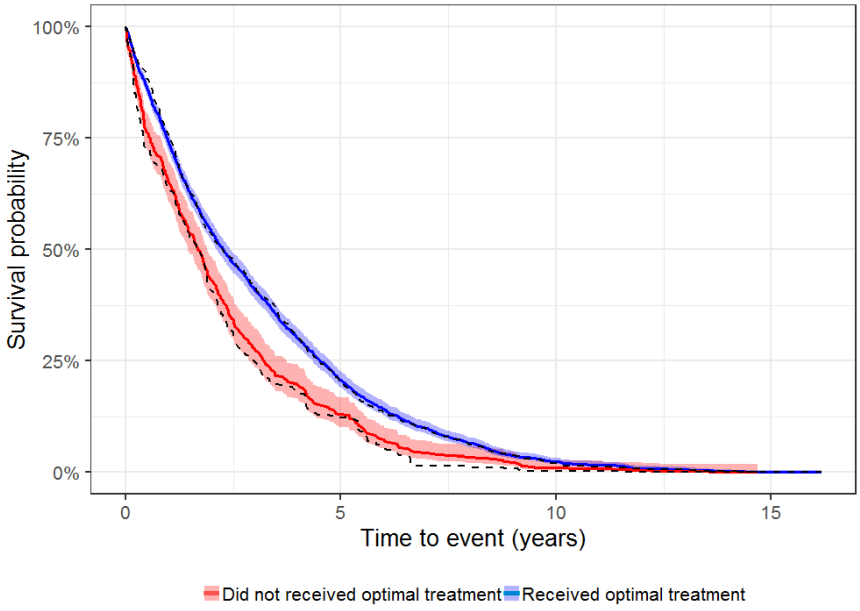

Figure E.3: Kaplan-Meier curves stratified by whether patients followed or not their estimated optimal treatment. The dashed lines are corresponding weighted Kaplan-Meier curves.

E.3.3 Model Checking

Residual plots are shown in Figure E.4 for the fitted values and selected covariates. The plot of fitted values against residuals shows a decreasing trend, which may indicate that the treatment-free model is not well specified. Transformations about age were tried but did not improve the model fit.

The double-robustness of the estimators $\hat{\boldsymbol{\psi}}_{\mathbf{1}}$ is exploited to assess the specification of the treatment-free, treatment and censoring models. If the treatment-free is correctly specified, then the estimators $\hat{\boldsymbol{\psi}}_{\mathbf{1}}$ would be robust to changes in the specification of the treatment and censoring models. Conversely, if the treatment and censoring models are correctly specified, then any specification of the treatment-free model would yield the same estimates of $\boldsymbol{\psi}_{\mathbf{1}}$. We considered 511 different specifications of all three models by including subsets of the

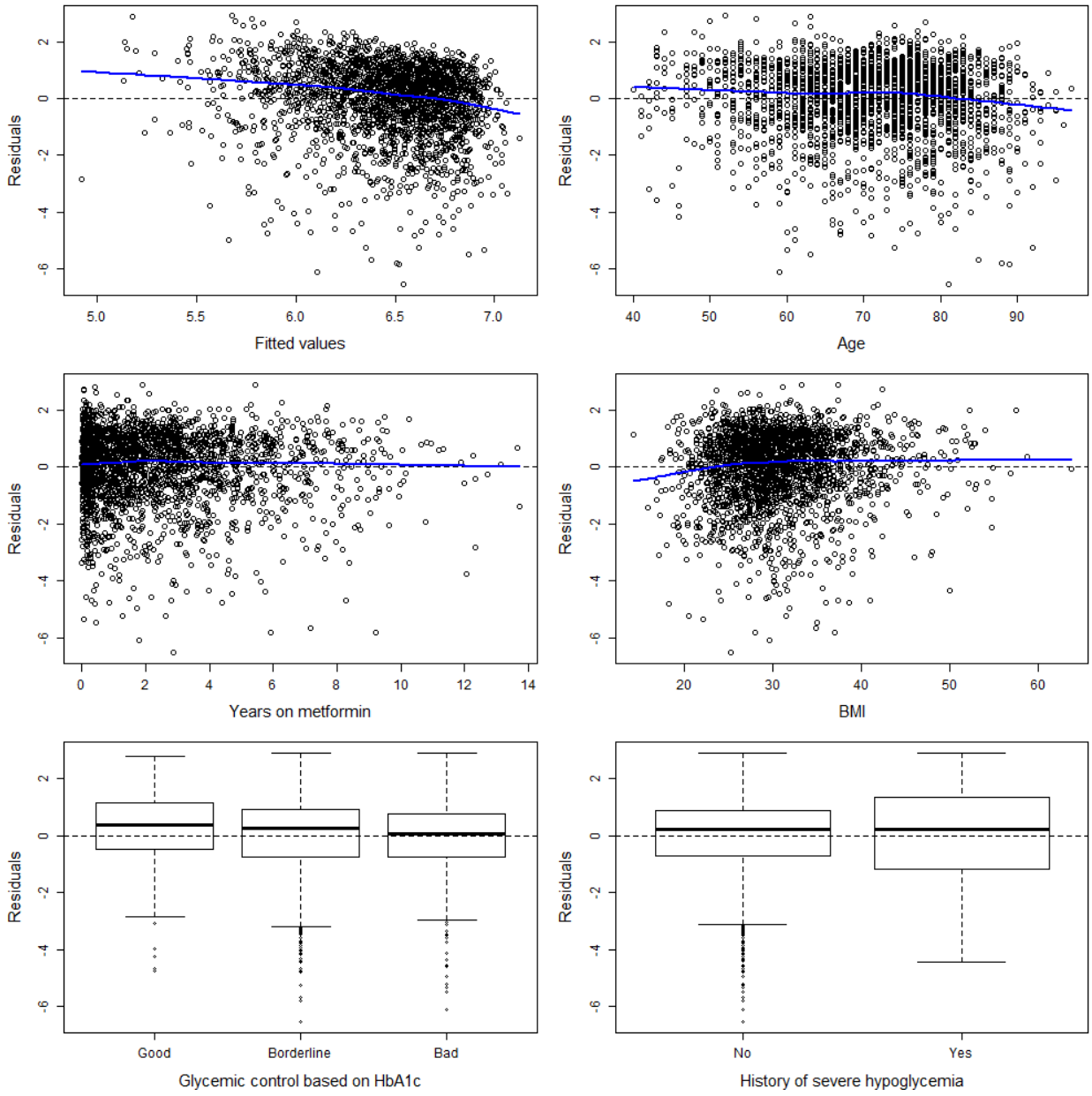

Figure E.4: Residual plots.

covariates originally included in the models. For the treatment-free model, tailoring variables (categories of HbA1c, BMI and severe hypoglycemia) were forced into the model. The left-hand side boxplot in Figure E.5 shows the distribution of the average treatment effect (i.e. the linear combination $\hat{\boldsymbol{\psi}}_1^{\boldsymbol{T}}\boldsymbol{h_1}$) when the censoring and treatment models were held fixed but the treatment-free model was varied. The distribution shows little variability and is centered near the average treatment effect calculated in the main analysis (horizontal dashed line),

suggesting that the treatment and censoring models were reasonably well specified. The right-hand side boxplot in Figure E.5 shows the distribution of the average treatment effect when the censoring and treatment models were varied and the treatment-free model was held fixed. The distribution shows more variability, suggesting that the treatment-free model is not well specified.

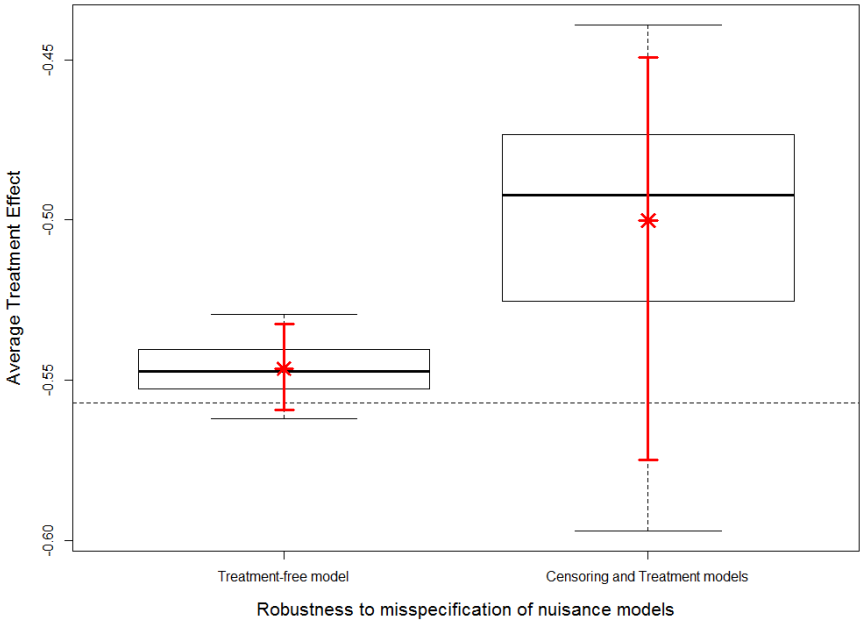

Figure E.5: Double-robustness plot to assess the (mis)specification of the treatment-free, treatment and censoring models. The distribution of the average treatment effect $\hat{\boldsymbol{\psi}}_{\mathbf{1}}^{\boldsymbol{T}}\boldsymbol{h_1}$ is shown across 511 different specifications of the treatment-free model (left-hand side boxplot) and of the censoring and treatment models (right-hand side boxplot). The star shows the mean of the distribution and the thick line points to the 5^{th} and 95^{th} percentiles of the distribution. The horizontal dashed line shows the average treatment effect in the main analysis.

E.4 Sensitivity Analyses

We considered the following sensitivity analyses:

1. We removed "alcohol misuse" from all models because records about alcohol misuse were made a long time before study entry, i.e. on average eight years before study

entry (see Table E.2).

2. We added age as a tailoring variable.

3. We restricted "severe hypoglycemia" to be defined with records made from metformin initiation to study entry.

4. We used the restricted mean survival time (RMST) as the outcome. The RMST truncates the observed censoring and survival times at τ. We chose $\tau = 18.6$ years (6800 days).

5. We used more stringent time-windows to record BMI and HbA1c (3 months, 6 months and 1 year). Only individuals with a record made within the time-window were kept in the analysis.

6. We restricted study entry to after January 1st, 2007, which corresponds to the date when DPP-4i were first approved in the UK.

Results from the sensitivity analyses are presented in Tables E.3–E.5. The sensitivity analyses 1–4 and 6 all led to similar treatment rules as in the main analysis. The observed differences between the parameter estimates in the main analysis and in these five sensitivity analyses were only reflected in the estimated thresholds about BMI in the rule but did not change the conclusions. Restricting the time-window to record BMI and HbA1c significantly changed the point estimates of the treatment rule parameters and changed the estimated rule. Specifically, when restricting BMI and HbA1c measurements within three months before study entry, the treatment rule still recommended DPP-4i for patients with a history of hypoglycemia but recommended sulfonylurea for most patients without a history of hypoglycemia. Only patients with borderline or bad glycemic control and very high BMI (above 63 and above 75, respectively) were recommended to add DPP-4i. Among the 2,645 patients who were recommended to add DPP-4i in the main analysis, 1,190 patients were now recommended to add sulfonylurea. However, even if the resulting rule was different, the

conclusions remained sensibly the same.

Table E.3: Treatment rule parameter estimates in the main analysis and in four sensitivity analyses based on 35,287 patients.

	Main analysis		Without alcohol[a]	+Age[b]	Hypo after metformin[c]	RMST[d]
Tailoring variables	$\hat{\boldsymbol{\psi}}_{\mathbf{1}}$	95% CI	$\hat{\boldsymbol{\psi}}_{\mathbf{1}}$	$\hat{\boldsymbol{\psi}}_{\mathbf{1}}$	$\hat{\boldsymbol{\psi}}_{\mathbf{1}}$	$\hat{\boldsymbol{\psi}}_{\mathbf{1}}$
Intercept	-0.99	(-2.48, 0.51)	-0.98	-0.94	-0.92	-0.99
HbA1c (ref: $\leq 7\%$)						
7% to 10%	-0.87	(-1.75, 0.01)	-0.88	-0.87	-0.85	-0.87
$> 10\%$	-1.09	(-2.12, -0.07)	-1.11	-1.10	-1.06	-1.09
Hypoglycemia	1.69	(0.31, 3.07)	1.71	1.69	1.75	1.69
BMI	0.04	(0.005, 0.08)	0.04	0.04	0.04	0.04
Age				-0.001		

[a]remove history of alcohol misuse from all models. [b]add age as tailoring variable. [c]record hypoglycemia between metformin initiation and study entry. [d]outcome defined as restricted mean survival time (RMST) with $\tau = 18.6$ year. BMI: body mass index, CI: confidence interval, HbA1c: glycated hemoglobin.

Table E.4: Treatment rule parameter estimates in the main analysis based on 35,287 patients and in three sensitivity analyses varying the time-window to record BMI and HbA1c.

	Main analysis		Time-window		
			1 yr (n=32,473)	6 mo (n=28,036)	3 mo (n=22,909)
Tailoring variables	$\hat{\boldsymbol{\psi}}_{\mathbf{1}}$	95% CI	$\hat{\boldsymbol{\psi}}_{\mathbf{1}}$	$\hat{\boldsymbol{\psi}}_{\mathbf{1}}$	$\hat{\boldsymbol{\psi}}_{\mathbf{1}}$
Intercept	-0.99	(-2.48, 0.51)	-0.49	-0.78	-1.21
HbA1c (ref: $\leq 7\%$)					
7% to 10%	-0.87	(-1.75, 0.01)	-0.56	-0.49	0.40
$> 10\%$	-1.09	(-2.12, -0.07)	-0.71	-0.53	0.24
Hypoglycemia	1.69	(0.31, 3.07)	1.44	1.35	2.15
BMI	0.04	(0.005, 0.08)	0.02	0.03	0.01

BMI: body mass index, CI: confidence interval, HbA1c: glycated hemoglobin, yr: year, mo: months.

Table E.5: Treatment rule parameter estimates in the main analysis and in a sensitivity analysis restricting study entry after January 1st, 2007.

	Main analysis (n=35,287)		After 2007 (n=28,463)	
Tailoring variable	$\hat{\boldsymbol{\psi}}_{\mathbf{1}}$	95% CI	$\hat{\boldsymbol{\psi}}_{\mathbf{1}}$	95% CI
Intercept	-0.99	(-2.48, 0.51)	-1.27	(-2.75,0.20)
HbA1c (ref: $\leq 7\%$)				
7% to 10%	-0.87	(-1.75, 0.01)	-0.69	(-1.57, 0.19)
$> 10\%$	-1.09	(-2.12, -0.07)	-0.88	(-1.89, 0.12)
Hypoglycemia	1.69	(0.31, 3.07)	1.91	(0.61, 3.22)
BMI	0.04	(0.005, 0.08)	0.05	(0.02, 0.09)

BMI: body mass index, CI: confidence interval, HbA1c: glycated hemoglobin.

References

Andrews, D. W. K., & Guggenberger, P. (2010). Asymptotic size and a problem with subsampling and with the m-out-of-n bootstrap. *Econometric Theory*, *26*(02), 426–468.
46, 64

Bai, X., Tsiatis, A. A., Lu, W., & Song, R. (2017). Optimal treatment regimes for survival endpoints using a locally-efficient doubly-robust estimator from a classification perspective. *Lifetime Data Analysis*, *23*(4), 585–604.
31, 37, 38, 72, 97

Bender, R., Augustin, T., & Blettner, M. (2005). Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine*, *24*(11), 1713–1723.
178

Bickel, P. J., & Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics*, *9*(6), 1196–1217.
25

Bickel, P. J., & Freedman, D. A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *The Annals of Statistics*, *12*(2), 470–482.
60

Bickel, P. J., Götze, F., & van Zwet, W. R. (1997). Resampling fewer than n observations: Gains, losses, and remedies for losses. *Statistica Sinica*, *7*, 1–31.

24, 46, 64

Bickel, P. J., & Sakov, A. (2008). On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Statistica Sinica*, *18*(3), 967–985.

25, 27

Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics*, *30*(1), 89–99.

29

Bretagnolle, J. (1983). Lois limites du bootstrap de certaines fonctionnelles. In *Annales de l'IHP Probabilités et Statistiques* (Vol. 19, pp. 281–296).

24, 46

Burr, D. (1994). A comparison of certain bootstrap confidence intervals in the Cox model. *Journal of the American Statistical Association*, *89*(428), 1290–1302.

98, 104

Chakraborty, B., Ghosh, P., Moodie, E. E. M., & Rush, A. J. (2016). Estimating optimal shared-parameter dynamic regimens with application to a multistage depression clinical trial. *Biometrics*, *72*(3), 865–876.

151

Chakraborty, B., Laber, E. B., & Zhao, Y. Q. (2013). Inference for optimal dynamic treatment regimes using an adaptive m-out-of-n bootstrap scheme. *Biometrics*, *69*(3), 714–723.

24, 25, 26, 27, 46, 52, 53, 54, 64, 112, 116, 149, 150, 151

Chakraborty, B., Murphy, S. A., & Strecher, V. (2010). Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical Methods in Medical Research*, *19*(3), 317–343.

20, 25, 26, 46, 52, 53, 56, 98, 149, 151, 157, 175

Clayton, H. B., Li, R., Perrine, C. G., & Scanlon, K. S. (2013). Prevalence and reasons for introducing infants early to solid foods: variations by milk feeding type. *Pediatrics*, *131*(4), e1108–e1114.

47

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B*, *34*(2), 187–202.

29

Cui, Y., Zhu, R., & Kosorok, M. R. (2017). Tree based weighted learning for estimating individualized treatment rules with censored data. *Electronic Journal of Statistics*, *11*(2), 3927–3953.

37

Dale, J., Paterson, C., Tierney, A., Ralston, S. H., Reid, D. M., Basu, N., Harvie, J., McKay, N. D., Saunders, S., Wilson, H., & others (2016). The Scottish Early Rheumatoid Arthritis (SERA) Study: an inception cohort and biobank. *BMC Musculoskeletal Disorders*, *17*(1), 461.

85

Daniels, L., Mallan, K. M., Fildes, A., & Wilson, J. (2015). The timing of solid introduction in an obesogenic environment: a narrative review of the evidence and methodological issues. *Australian and New Zealand Journal of Public Health*, *39*(4), 366–373.

47

Efron, B. (1981). Censored data and the bootstrap. *Journal of the American Statistical Association*, *76*(374), 312–319.

98

Efron, B. (1992a). Bootstrap methods: Another look at the jackknife. In *Breakthroughs in Statistics* (pp. 569–593). Springer.

23

Efron, B. (1992b). Six questions raised by the bootstrap. In *Exploring the Limits of Bootstrap* (Vol. 270, pp. 99–126). John Wiley & Sons.
104

Efron, B., & Tibshirani, R. J. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, *1*(1), 54–75.
98, 104, 105

Efron, B., & Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. CRC press.
150

Fan, Y., He, M., Su, L., & Zhou, X. (2019). A smoothed Q-learning algorithm for estimating optimal dynamic treatment regime. *Scandinavian Journal of Statistics*, *46*(2), 446–469.
26, 27

Felson, D. T., Smolen, J. S., Wells, G., Zhang, B., van Tuyl, L. H. D., Funovits, J., Aletaha, D., Allaart, C. F., Bathon, J., Bombardieri, S., & others (2011). American College of Rheumatology/European League Against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. *Arthritis & Rheumatology*, *63*(3), 573–586.
91

Fu, H., Zhou, J., & Faries, D. E. (2016). Estimating optimal treatment regimes via subgroup identification in randomized control trials and observational studies. *Statistics in Medicine*, *35*(19), 3285–3302.
40

Garber, A. J., Abrahamson, M. J., Barzilay, J. I., Blonde, L., Bloomgarden, Z. T., Bush, M. A., Dagogo-Jack, S., DeFronzo, R. A., Einhorn, D., Fonseca, V. A., & others (2019). Consensus statement by the American Association of Clinical Endocrinologists

and American College of Endocrinology on the comprehensive type 2 diabetes management algorithm–2019 executive summary. *Endocrine Practice*, *25*(1), 69–100.
39, 124, 137

Geng, Y., Zhang, H. H., & Lu, W. (2015). On optimal treatment regimes selection for mean survival time. *Statistics in Medicine*, *34*(7), 1169–1184.
37

Gill, R. D., van der Laan, M. J., & Robins, J. M. (1997). Coarsening at random: Characterizations, conjectures, counter-examples. In *Proceedings of the First Seattle Symposium in Biostatistics* (pp. 255–294).
74

Goldberg, Y., & Kosorok, M. R. (2012). Q-learning with censored data. *Annals of Statistics*, *40*(1), 529–560.
30, 31, 33, 71, 97, 124

Growth reference 5-19 years [Computer software manual]. (2007). (Available at http://www.who.int/growthref/en/)
61

Hager, R., Tsiatis, A. A., & Davidian, M. (2018). Optimal two-stage dynamic treatment regimes from a classification perspective with censored survival data. *Biometrics*, *74*(4), 1180–1192.
38, 72, 97, 124

Hernán, M. A., Cole, S. R., Margolick, J., Cohen, M., & Robins, J. M. (2005). Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharmacoepidemiology and Drug Safety*, *14*(7), 477–491.
36, 81

Hernán, M. A., & Robins, J. M. (2010). *Causal Inference*. CRC Boca Raton, FL.

71, 78, 89

Herrett, E., Gallagher, A. M., Bhaskaran, K., Forbes, H., Mathur, R., van Staa, T., & Smeeth, L. (2015). Data resource profile: Clinical Practice Research Datalink (CPRD). *International Journal of Epidemiology*, *44*(3), 827–836.

247, 248

Herrett, E., Thomas, S. L., Schoonen, W. M., Smeeth, L., & Hall, A. J. (2010). Validation and validity of diagnoses in the General Practice Research Database: A systematic review. *British Journal of Clinical Pharmacology*, *69*(1), 4–14.

247

Hirano, K., & Porter, J. R. (2012). Impossibility results for nondifferentiable functionals. *Econometrica*, *80*(4), 1769–1790.

19

Hjort, N. L. (1985). *Bootstrapping Cox's regression model* (Tech. Rep.). Stanford University Lab for Computational Statistics.

98

Hripcsak, G., Ryan, P. B., Duke, J. D., Shah, N. H., Park, R. W., Huser, V., Suchard, M. A., Schuemie M. J., DeFalco, F. J., Perotte, A., & others (2016). Characterizing treatment pathways at scale using the OHDSI network. *Proceedings of the National Academy of Sciences*, *113*(27), 7329–7336.

39, 139

Huang, X., & Ning, J. (2012). Analysis of multi-stage treatments for recurrent diseases. *Statistics in Medicine*, *31*(24), 2805–2821.

32, 35

Huang, X., Ning, J., & Wahed, A. S. (2014). Optimization of individualized dynamic treatment regimes for recurrent diseases. *Statistics in Medicine*, *33*(14), 2363–2378.

31, 32, 33, 34, 35, 36, 71, 72, 81, 90, 97, 124

Infant and young child feeding [Computer software manual]. (2016, January). (Available at http://who.int/mediacentre/factsheets/fs342/en/)
46

Inoue, E., Yamanaka, H., Hara, M., Tomatsu, T., & Kamatani, N. (2007). Comparison of Disease Activity Score (DAS) 28-erythrocyte sedimentation rate and DAS28-C-reactive protein threshold values. *Annals of the Rheumatic Diseases*, *66*(3), 407–409.
91

Inzucchi, S. E., Bergenstal, R. M., Buse, J. B., Diamant, M., Ferrannini, E., Nauck, M., Peters, A. L., Tsapas, A., Wender, R., & Matthews, D. R. (2015). Management of hyperglycemia in type 2 diabetes, 2015: A patient-centered approach: Update to a position statement of the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes Care*, *38*(1), 140–149.
39, 124

Jiang, R., Lu, W., Song, R., & Davidian, M. (2017a). On estimation of optimal treatment regimes for maximizing t-year survival probability. *Journal of the Royal Statistical Society: Series B*, *79*(4), 1165–1185.
32, 37, 38, 72, 97

Jiang, R., Lu, W., Song, R., Hudgens, M. G., & Naprvavnik, S. (2017b). Doubly robust estimation of optimal treatment regimes for survival data – with application to an HIV/AIDS study. *The Annals of Applied Statistics*, *11*(3), 1763–1786.
32, 38, 124

Joffe, M. M. (2001). Administrative and artificial censoring in censored regression models. *Statistics in Medicine*, *20*(15), 2287–2304.
81

Joffe, M. M., Yang, W. P., & Feldman, H. (2012). G-estimation and artificial censoring: Problems, challenges, and applications. *Biometrics*, *68*(1), 275–286.
37, 81

Kalbfleisch, J. D., & Prentice, R. L. (2011). *The Statistical Analysis of Failure Time Data* (Vol. 360). John Wiley & Sons.
27

Karrison, T. G. (1997). Use of Irwin's restricted mean as an index for comparing survival in different treatment groups—interpretation and power considerations. *Controlled Clinical Trials*, *18*(2), 151–167.
31, 90

Kosiborod, M., Lam, C. S. P., Kohsaka, S., Kim, D. J., Karasik, A., Shaw, J., Tangri, N., Goh, S., Thuresson, M., Chen, H., & others (2018). Cardiovascular events associated with SGLT-2 inhibitors versus other glucose-lowering drugs: The CVD-REAL 2 study. *Journal of the American College of Cardiology*, *71*(23), 2628–2639.
39

Krakow, E. F., Hemmer, M., Wang, T., Logan, B., Arora, M., Spellman, S., Couriel, D., Alousi, A., Pidala, J., Last, M., & others (2017). Tools for the precision medicine era: How to develop highly personalized treatment recommendations from cohort and registry data using Q-Learning. *American Journal of Epidemiology*, *186*(2), 160–172.
30

Kramer, M. S., Chalmers, B., Hodnett, E. D., Sevkovskaya, Z., Dzikovich, I., Shapiro, S., Collet, J. P., Vanilovich, I., Mezen, I., Ducruet, T., & others (2001). Promotion of breastfeeding intervention trial (PROBIT): a randomized trial in the Republic of Belarus. *The Journal of the American Medical Association*, *285*(4), 413–420.
47, 59

Kramer, M. S., Guo, T., Platt, R. W., Shapiro, S., Collet, J. P., Chalmers, B., Hodnett, E., Sevkovskaya, Z., Dzikovich, I., Vanilovich, I., & others (2002). Breastfeeding and infant growth: biology or bias? *Pediatrics*, *110*(2), 343–347.

64

Kramer, M. S., Matush, L., Bogdanovich, N., Aboud, F., Mazer, B., Fombonne, E., Collet, J. P., Hodnett, E., Mironova, E., Igumnov, S., & others (2009). Health and development outcomes in 6.5-y-old children breastfed exclusively for 3 or 6 mo. *The American Journal of Clinical Nutrition*, *90*(4), 1070–1074.

59

Kreif, N., Sofrygin, O., Schmittdiel, J., Adams, A., Grant, R., Zhu, Z., van der Laan, M. J., & Neugebauer, R. (2018). Evaluation of adaptive treatment strategies in an observational study where time-varying covariates are not monitored systematically. *arXiv preprint arXiv:1806.11153*.

40

Kuo, S., Yang, C. T., Wu, J. S., & Ou, H. T. (2019). Effects on clinical outcomes of intensifying triple oral antidiabetic drug (OAD) therapy by initiating insulin versus enhancing OAD therapy in patients with type 2 diabetes: A nationwide population-based, propensity-score-matched cohort study. *Diabetes, Obesity and Metabolism*, *21*(2), 312–320.

39

Kuriya, B., Schieir, O., Lin, D., Xiong, J., Pope, J., Boire, G., Haraoui, B., Thorne, J. C., Tin, D., Hitchon, C., & others (2017). Thresholds for the 28-joint disease activity score (DAS28) using C-reactive protein are lower compared to DAS28 using erythrocyte sedimentation rate in early rheumatoid arthritis. *Clinical and Experimental Rheumatology*, *35*(5), 799–803.

91

Laber, E. B., Linn, K. A., & Stefanski, L. A. (2014a). Interactive model building for Q-learning. *Biometrika*, *101*(4), 831–847.

9

Laber, E. B., Lizotte, D. J., Qian, M., Pelham, W. E., & Murphy, S. A. (2014b). Dynamic treatment regimes: Technical challenges and applications. *Electronic Journal of Statistics*, *8*(1), 1225–1272.

26, 27, 151, 157

Laber, E. B., Qian, M., Lizotte, D. J., Pelham, W. E., & Murphy, S. A. (2010). Statistical inference in dynamic treatment regimes. *arXiv preprint arXiv:1006.5831*.

20

Laber, E. B., & Zhao, Y. Q. (2015). Tree-based methods for individualized treatment regimes. *Biometrika*, *102*(3), 501–514.

10

Li, F., Morgan, K. L., & Zaslavsky, A. M. (2018). Balancing covariates via propensity score weighting. *Journal of the American Statistical Association*, *113*(521), 390–400.

18, 72, 80, 126

McGuire, H., Longson, D., Adler, A., Farmer, A., & Lewin, I. (2016). Management of type 2 diabetes in adults: summary of updated NICE guidance. *British Medical Journal*, *353*, i1575.

124

Moodie, E. E. M. (2006). *Inference for Optimal Dynamic Treatment Regimes* (Unpublished doctoral dissertation). University of Washington.

98, 103

Moodie, E. E. M. (2009). A note on the variance of doubly-robust G-estimators. *Biometrika*, *96*(4), 998–1004.

81, 98, 102, 103, 170

Moodie, E. E. M., Dean, N., & Sun, Y. R. (2014). Q-learning: Flexible learning about useful utilities. *Statistics in Biosciences*, *6*(2), 223–243.

9, 13

Moodie, E. E. M., Platt, R. W., & Kramer, M. S. (2009). Estimating response-maximized decision rules with applications to breastfeeding. *Journal of the American Statistical Association*, *104*(485), 155–165.

59

Moodie, E. E. M., & Richardson, T. S. (2010). Estimating optimal dynamic regimes: correcting bias under the null. *Scandinavian Journal of Statistics*, *37*(1), 126–146.

20, 22, 26, 46, 80, 98, 175

Müller, S., & Welsh, A. H. (2009). Robust model selection in generalized linear models. *Statistica Sinica*, *19*(3), 1155–1170.

60

Murphy, S. A. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*(2), 331–355.

11, 12, 45, 50, 71, 97

Murphy, S. A. (2005). An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine*, *24*(10), 1455–1481.

3

Murray, T. A., Yuan, Y., & Thall, P. F. (2018). A Bayesian machine learning approach for optimizing dynamic treatment regimes. *Journal of the American Statistical Association*, *113*(523), 1255–1267.

9

Neugebauer, R., Fireman, B., Roy, J. A., & O'Connor, P. J. (2013). Impact of specific glucose-control strategies on microvascular and macrovascular outcomes in 58,000 adults with type 2 diabetes. *Diabetes Care*, *36*(11), 3510–3516.

40, 138

Neugebauer, R., Schmittdiel, J. A., & van der Laan, M. J. (2016). A case study of the impact of data-adaptive versus model-based estimation of the propensity scores on causal inferences from three inverse probability weighting estimators. *The International Journal of Biostatistics*, *12*(1), 131–155.

40

Newby, R. M., & Davies, P. S. W. (2015). A prospective study of the introduction of complementary foods in contemporary Australian infants: What, when and why? *Journal of Paediatrics and Child Health*, *51*(2), 186–191.

47

Nyström, T., Bodegard, J., Nathanson, D., Thuresson, M., Norhammar, A., & Eriksson, J. W. (2017). Second line initiation of insulin compared with DPP-4 inhibitors after metformin monotherapy is associated with increased risk of all-cause mortality, cardiovascular events, and severe hypoglycemia. *Diabetes Research and Clinical Practice*, *123*, 199–208.

39

Orellana, L., Rotnitzky, A., & Robins, J. M. (2010). Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part I: main content. *The International Journal of Biostatistics*, *6*(2), Article 8.

10, 228

Qian, M., & Murphy, S. A. (2011). Performance guarantees for individualized treatment rules. *Annals of Statistics*, *39*(2), 1180–1210.

9

Rich, B., Moodie, E. E. M., Stephens, D. A., & Platt, R. W. (2010). Model checking with residuals for G-estimation of optimal dynamic treatment regimes. *The International Journal of Biostatistics*, *6*(2), 1–24.

51, 127, 157

Robins, J. M. (1997). Causal inference from complex longitudinal data. In *Latent Variable Modeling and Applications to Causality* (pp. 69–117). Springer.

8, 49

Robins, J. M. (1998). Structural nested failure time models. In *Encyclopedia of Biostatistics* (Vol. 7, pp. 4372–4389). John Wiley and Sons.

36, 37

Robins, J. M. (2000a). Marginal structural models versus structural nested models as tools for causal inference. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials* (pp. 95–133). Springer.

9

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

74

Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium in Biostatistics* (pp. 189–326).

3, 9, 13, 18, 20, 21, 22, 45, 46, 50, 51, 52, 71, 81, 90, 97, 98, 102, 170, 175

Robins, J. M., Blevins, D., Ritter, G., & Wulfsohn, M. (1992). G-estimation of the effect of prophylaxis therapy for Pneumocystis carinii pneumonia on the survival of AIDS patients. *Epidemiology*, *3*(4), 319–336.

36

Robins, J. M., & Greenland, S. (1994). Adjusting for differential rates of prophylaxis therapy for PCP in high-versus low-dose AZT treatment arms in an AIDS randomized trial. *Journal of the American Statistical Association*, *89*(427), 737–749.
71

Robins, J. M., & Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology* (pp. 297–331). Springer.
32, 78

Robins, J. M., Rotnitzky, A., & Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, *90*(429), 106–121.
60

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701.
7

Rubin, D. B. (1980). Randomization analysis of experimental data: The Fisher randomization test comment. *Journal of the American Statistical Association*, *75*(371), 591–593.
8, 49, 74

Rush, A. J., Fava, M., Wisniewski, S. R., Lavori, P. W., Trivedi, M. H., Sackeim, H. A., Thase, M. E., Nierenberg, A. A., Quitkin, F. M., Kashner, T. M., & others (2004). Sequenced treatment alternatives to relieve depression (STAR* D): Rationale and design. *Controlled Clinical Trials*, *25*(1), 119–142.
31

Sengul, I., Akcay-Yalbuzdag, S., Ince, B., Goksel-Karatepe, A., & Kaya, T. (2015). Comparison of the DAS28-CRP and DAS28-ESR in patients with rheumatoid arthritis. *International Journal of Rheumatic Diseases*, *18*(6), 640–645.

91

Shao, J. (1994). Bootstrap sample size in nonregular cases. *Proceedings of the American Mathematical Society*, *122*(4), 1251–1262.

23, 24, 46, 52, 53

Simoneau, G., Moodie, E. E. M., Nijjar, J. S., Platt, R. W., & Scottish Early Rheumatoid Arthritis Inception Cohort Investigators. (2019). Estimating optimal dynamic treatment regimes with survival outcomes. *Journal of the American Statistical Association (in press)*.

97, 100, 101, 115, 123, 126

Simoneau, G., Moodie, E. E. M., Platt, R. W., & Chakraborty, B. (2017). Non-regular inference for dynamic weighted ordinary least squares: understanding the impact of solid food intake in infancy on childhood weight. *Biostatistics*, *19*(2), 233–246.

98, 112, 175

Singh, J. A., Saag, K. G., Bridges, S. L., Akl, E. A., Bannuru, R. R., Sullivan, M. C., Vaysbrot, E., McNaughton, C., Osani, M., Shmerling, R. H., & others (2016). 2015 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. *Arthritis & Rheumatology*, *68*(1), 1–26.

70

Son, K. M., Kim, S. Y., Lee, S. H., Yang, C. M., Seo, Y. I., & Kim, H. A. (2016). Comparison of the disease activity score using the erythrocyte sedimentation rate and C-reactive protein levels in Koreans with rheumatoid arthritis. *International Journal of Rheumatic Diseases*, *19*(12), 1278–1283.

91

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. The MIT press, Cambridge.

45

Vansteelandt, S., & Joffe, M. M. (2014). Structural nested models and G-estimation: The partially realized promise. *Statistical Science*, *29*(4), 707–731.

15

Wahed, A. S., & Tsiatis, A. A. (2004). Optimal estimator for the survival distribution and related quantities for treatment policies in two-stage randomization designs in clinical trials. *Biometrics*, *60*(1), 124–133.

10

Wallace, M. P., & Moodie, E. E. M. (2015). Doubly-robust dynamic treatment regimen estimation via weighted least squares. *Biometrics*, *71*(3), 636–644.

3, 9, 17, 18, 19, 45, 51, 59, 72, 80, 97, 126, 143, 170

Wallace, M. P., Moodie, E. E. M., & Stephens, D. A. (2014). DTRreg: DTR estimation and inference via G-estimation and dynamic WOLS [Computer software manual]. (R package version 1.1)

15, 51

Wallace, M. P., Moodie, E. E. M., & Stephens, D. A. (2016). Model assessment in dynamic treatment regimen estimation via double robustness. *Biometrics*, *72*(3), 855–864.

51, 88, 89, 127

Wallace, M. P., Moodie, E. E. M., & Stephens, D. A. (2017a). Dynamic treatment regimen estimation via regression-based techniques: Introducing R package DTRreg. *Journal of Statistical Software*, *80*(2), 1–20.

72, 97

Wallace, M. P., Moodie, E. E. M., & Stephens, D. A. (2017b). Model validation and selection for personalized medicine using dynamic-weighted ordinary least squares. *Statistical Methods in Medical Research*, *26*(4), 1641–1653.

88

Wallace, M. P., Moodie, E. E. M., & Stephens, D. A. (2017c). An R package for G-estimation of structural nested mean models. *Epidemiology*, *28*(2), e18–e20.

15

Wang, Y., Fu, H., & Zeng, D. (2018). Learning optimal personalized treatment rules in consideration of benefit and risk: with an application to treating type 2 diabetes patients with insulin therapies. *Journal of the American Statistical Association*, *113*(521), 1–13.

40

Watkins, C. (1989). *Learning from delayed rewards* (Unpublished doctoral dissertation). University of Cambridge, England.

45

Watkins, C., & Dayan, P. (1992). Q-learning. *Machine Learning*, *8*(3-4), 279–292.

11

Wei, L. J. (1992). The accelerated failure time model: A useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine*, *11*(14-15), 1871–1879.

29

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: issues and guidance for practice. *Statistics in Medicine*, *30*(4), 377–399.

86

Yu, O. H. Y., Yin, H., & Azoulay, L. (2015). The combination of DPP-4 inhibitors versus sulfonylureas with metformin after failure of first-line treatment in the risk for major cardiovascular events and death. *Canadian Journal of Diabetes*, *39*(5), 383–389.

39, 128

Zhang, B., Tsiatis, A. A., Davidian, M., Zhang, M., & Laber, E. B. (2012a). Estimating optimal treatment regimes from a classification perspective. *Stat*, *1*(1), 103–114.

10

Zhang, B., Tsiatis, A. A., Laber, E. B., & Davidian, M. (2012b). A robust method for estimating optimal treatment regimes. *Biometrics*, *68*(4), 1010–1018.

10

Zhang, B., Tsiatis, A. A., Laber, E. B., & Davidian, M. (2013). Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika*, *100*(3), 681–694.

10, 45, 71, 97

Zhang, Y., Laber, E. B., Tsiatis, A. A., & Davidian, M. (2015). Using decision lists to construct interpretable and parsimonious treatment regimes. *Biometrics*, *71*(4), 895–904.

10, 13

Zhao, Y., Zeng, D., Socinski, M. A., & Kosorok, M. R. (2011). Reinforcement learning strategies for clinical trials in nonsmall cell lung cancer. *Biometrics*, *67*(4), 1422–1433.

37

Zhao, Y. Q., & Laber, E. B. (2014a). Estimation of optimal dynamic treatment regimes. *Clinical Trials*, *11*(4), 400–407.

10

Zhao, Y. Q., Zeng, D., Laber, E. B., & Kosorok, M. R. (2015). New statistical learning methods for estimating optimal dynamic treatment regimes. *Journal of the American Statistical Association*, *110*(510), 583–598.

10, 45, 71, 97

Zhao, Y. Q., Zeng, D., Laber, E. B., Song, R., Yuan, M., & Kosorok, M. R. (2014b). Doubly robust learning for estimating individualized treatment with censored data. *Biometrika*, *102*(1), 151–168.

31, 37, 72

Zhao, Y. Q., Zeng, D., Rush, A. J., & Kosorok, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, *107*(499), 1106–1118.

10

Zhou, X., Mayer-Hamblett, N., Khan, U., & Kosorok, M. R. (2017). Residual weighted learning for estimating individualized treatment rules. *Journal of the American Statistical Association*, *112*(517), 169–187.

10

Zhu, R., Zhao, Y. Q., Chen, G., Ma, S., & Zhao, H. (2017). Greedy outcome weighted tree learning of optimal personalized treatment rules. *Biometrics*, *73*(2), 391–400.

37